

# Testing the Gaussianity and Statistical Isotropy of the Universe

Guest Editors: Dragan Huterer, Eiichiro Komatsu,  
and Sarah Shandera





---

# **Testing the Gaussianity and Statistical Isotropy of the Universe**

Advances in Astronomy

---

# **Testing the Gaussianity and Statistical Isotropy of the Universe**

Guest Editors: Dragan Huterer, Eiichiro Komatsu,  
and Sarah Shandera



---

Copyright © 2010 Hindawi Publishing Corporation. All rights reserved.

This is a special issue published in volume 2010 of "Advances in Astronomy." All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Editorial Board

Cesare Barbieri, Italy  
Joshua S. Bloom, USA  
Elias Brinks, UK  
Michael Brotherton, USA  
Giovanni Carraro, Italy  
Alberto J. Castro-Tirado, Spain  
Michael Disney, UK  
Elmetwally Elabbasy, Egypt  
Nye Evans, UK  
Maurizio Falanga, Switzerland  
Duncan Forbes, Australia  
Andrew Fruchter, USA  
B. T. Gänsicke, UK  
Paul Goldsmith, USA  
Jonathan Grindlay, USA

Martin Hardcastle, UK  
Dean Hines, USA  
Dieter Horns, Germany  
Ivan Hubeny, USA  
John Hughes, USA  
Wing Huen Ip, Taiwan  
Rob Ivison, UK  
Valentina Klochkova, Russia  
Gregory Laughlin, USA  
Myung G. Lee, Republic of Korea  
Karen Leighly, USA  
Jeffrey Linsky, USA  
Mario Mateo, USA  
Ronald Mennickent, Chile  
Zdzislaw E. Musielak, USA

Valery Nakariakov, UK  
Jerome Orosz, USA  
George Pavlov, USA  
Juri Poutanen, Finland  
Somak Raychaudhury, UK  
William Reach, USA  
Jerome Rodriguez, France  
Peter Roming, USA  
Ata Sarajedini, USA  
Regina Schulte-Ladbeck, USA  
Ravi Sheth, USA  
Roberto Turolla, Italy  
Gary Wegner, USA  
Paul J. Wiita, USA  
Edward L. Wright, USA

# Contents

**Testing the Gaussianity and Statistical Isotropy of the Universe**, Dragan Huterer, Eiichiro Komatsu, and Sarah Shandera

Volume 2010, Article ID 697147, 1 page

**Primordial Non-Gaussianity in the Large-Scale Structure of the Universe**, Vincent Desjacques and Uroš Seljak

Volume 2010, Article ID 908640, 23 pages

**Non-Gaussianity from Large-Scale Structure Surveys**, Licia Verde

Volume 2010, Article ID 768675, 15 pages

**Primordial Non-Gaussianity in the Cosmic Microwave Background**, Amit P. S. Yadav and Benjamin D. Wandelt

Volume 2010, Article ID 565248, 27 pages

**Primordial Non-Gaussianity and Bispectrum Measurements in the Cosmic Microwave Background and Large-Scale Structure**, Michele Liguori, Emiliano Sefusatti, James R. Fergusson, and E. P. S. Shellard

Volume 2010, Article ID 980523, 64 pages

**Large-Angle Anomalies in the CMB**, Craig J. Copi, Dragan Huterer, Dominik J. Schwarz, and Glenn D. Starkman

Volume 2010, Article ID 847541, 17 pages

**Testing Gaussianity, Homogeneity, and Isotropy with the Cosmic Microwave Background**,

L. Raul Abramo and Thiago S. Pereira

Volume 2010, Article ID 378203, 25 pages

**Cosmic Strings and Their Induced Non-Gaussianities in the Cosmic Microwave Background**,

Christophe Ringeval

Volume 2010, Article ID 380507, 28 pages

**A Comprehensive Overview of the Cold Spot**, Patricio Vielva

Volume 2010, Article ID 592094, 20 pages

**Primordial Non-Gaussianities from Inflation Models**, Xingang Chen

Volume 2010, Article ID 638979, 43 pages

**Review of Local Non-Gaussianity from Multifield Inflation**, Christian T. Byrnes and Ki-Young Choi

Volume 2010, Article ID 724525, 18 pages

**Non-Gaussianity from Particle Production during Inflation**, Neil Barnaby

Volume 2010, Article ID 156180, 40 pages

**Ekpyrotic Nongaussianity: A Review**, Jean-Luc Lehners

Volume 2010, Article ID 903907, 19 pages

**Second-Order Gauge-Invariant Cosmological Perturbation Theory: Current Status**, Kouji Nakamura

Volume 2010, Article ID 576273, 26 pages



---

**Non-Gaussianity and the Cosmic Microwave Background Anisotropies**, N. Bartolo, S. Matarrese,  
and A. Riotto

Volume 2010, Article ID 157079, 68 pages

**Non-Gaussianity and Statistical Anisotropy from Vector Field Populated Inflationary Models**,  
Emanuela Dimastrogiovanni, Nicola Bartolo, Sabino Matarrese, and Antonio Riotto

Volume 2010, Article ID 752670, 21 pages

## Editorial

# Testing the Gaussianity and Statistical Isotropy of the Universe

**Dragan Huterer,<sup>1</sup> Eiichiro Komatsu,<sup>2</sup> and Sarah Shandera<sup>3</sup>**

<sup>1</sup>Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, MI 48109-1040, USA

<sup>2</sup>Texas Cosmology Center and Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA

<sup>3</sup>Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON, Canada N2L 2Y5

Correspondence should be addressed to Dragan Huterer, [huterer@umich.edu](mailto:huterer@umich.edu)

Received 31 December 2010; Accepted 31 December 2010

Copyright © 2010 Dragan Huterer et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

*Background and Motivation.* The last few years have seen a surge in excitement about measurements of statistics of the primordial fluctuations beyond the power spectrum. New ideas for precision tests of Gaussianity and statistical isotropy in the data are developing simultaneously with proposals for a wide range of new theoretical possibilities. From both the observations and theory, it has become clear that there is a huge discovery potential from upcoming measurements.

The twin principles of statistical isotropy and homogeneity are a crucial ingredient in obtaining most important results in modern cosmology. For example, with these assumptions temperature and density fluctuations in different directions on the sky can be averaged out, leading to accurate constraints on cosmological parameters that we have today. Nevertheless, there is no fundamental reason why these must be obeyed by our universe. Statistical isotropy and homogeneity are starting to be sharply tested using the cosmic microwave background (CMB) and large-scale structure data. Recently, there has been particular activity in these areas, given Wilkinson Microwave Anisotropy Probe remarkable maps, combined with claims of large-angle “anomalies” indicating departures from statistical isotropy as predicted by standard inflationary models.

The statement that primordial curvature fluctuations are nearly Gaussian on scales measured by the CMB is remarkably precise but does not reveal much about their source. Current constraints on the amplitude of the three-point correlation function of fluctuations are nearly four orders of magnitude above predictions from single field slow-roll inflation models and at least an order of magnitude above what is expected just from nonlinearities that develop after the primordial spectrum is laid down. There are a wide spectrum of interesting models that can be ruled out by

tightening this constraint; conversely, a detection of non-Gaussianity would rule out single field slow-roll inflation. While current observations of the CMB fluctuations provide reasonably strong evidence for a primordial source of fluctuations from inflation, only measurements of higher-order statistics can truly shed light on the physics of inflation.

Departures from statistical isotropy and Gaussianity involve a rich set of observable quantities, with diverse signatures that can be measured in the CMB or in large-scale structure using sophisticated statistical methods. These signatures, which carry information about physical processes on cosmological scales, have power to reveal detailed properties of the physics responsible for generating the primordial fluctuations. Even qualitative observational features can identify key properties of the fields involved (e.g., how many fields and which couplings were most relevant) or, alternatively, shed light on the systematic errors in the data. However, because there are so many possibilities from both theory and observation and because many calculations are very technical involving methods such as higher-order perturbation theory, the literature can be daunting.

In this special issue, we have collected articles that summarize the theoretical predictions for departures from Gaussianity or statistical isotropy from a variety of potential sources, together with the observational approaches to test these properties using the CMB or large-scale structure. We hope this collection provides an accessible entry point to these topics as they currently stand, indicating what direction future developments may take place and demonstrating why these questions are so compelling.

*Dragan Huterer  
Eiichiro Komatsu  
Sarah Shandera*

## Review Article

# Primordial Non-Gaussianity in the Large-Scale Structure of the Universe

Vincent Desjacques<sup>1</sup> and Uroš Seljak<sup>1,2</sup>

<sup>1</sup>*Institute for Theoretical Physics, University of Zurich, 8057 Zurich, Switzerland*

<sup>2</sup>*Physics Department, Astronomy Department, and Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720, USA*

Correspondence should be addressed to Vincent Desjacques, [dvince@physik.uzh.ch](mailto:dvince@physik.uzh.ch)

Received 16 January 2010; Revised 3 May 2010; Accepted 19 May 2010

Academic Editor: Dragan Huterer

Copyright © 2010 V. Desjacques and U. Seljak. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Primordial non-Gaussianity is a potentially powerful discriminant of the physical mechanisms that generated the cosmological fluctuations observed today. Any detection of significant non-Gaussianity would thus have profound implications for our understanding of cosmic structure formation. The large-scale mass distribution in the Universe is a sensitive probe of the nature of initial conditions. Recent theoretical progress together with rapid developments in observational techniques will enable us to critically confront predictions of inflationary scenarios and set constraints as competitive as those from the Cosmic Microwave Background. In this paper, we review past and current efforts in the search for primordial non-Gaussianity in the large-scale structure of the Universe.

## 1. Introduction

In generic inflationary models based on the slow roll of a scalar field, primordial curvature perturbations are produced by the inflaton field as it slowly rolls down its potential [1–4]. Most of these scenarios predict a nearly scale-invariant spectrum of adiabatic curvature fluctuations, a relatively small amount of gravity waves, and tiny deviations from Gaussianity in the primeval distribution of curvature perturbations [5–7]. Although the latest measurements of the cosmic microwave background (CMB) anisotropies favor a slightly red power spectrum [8], no significant detection of a  $B$ -mode or some level of primordial non-Gaussianity (NG) has thus far been reported from CMB observations.

While the presence of a  $B$ -mode can only be tested with CMB measurements, primordial deviations from Gaussianity can leave a detectable signature in the distribution of CMB anisotropies *and* in the large-scale structure (LSS) of the Universe. Until recently, it was widely accepted that measurement of the CMB furnished the best probe of primordial non-Gaussianity [9]. However, these conclusions did not take into account the anomalous scale dependence of

the galaxy power spectrum and bispectrum arising from primordial NG of the local  $f_{\text{NL}}^{\text{loc}}$  type [10, 11]. These theoretical results, together with rapid developments in observational techniques, will provide large amount of LSS data to critically confront predictions of non-Gaussian models. In particular, galaxy clustering should provide independent constraints on the magnitude of primordial non-Gaussianity as competitive as those from the CMB and, in the long run, may even give the best constraints.

The purpose of this paper is to provide an overview of the search for a primordial non-Gaussian signal in the large-scale structure. We will begin by briefly summarizing how non-Gaussianity arises in inflationary models (Section 2). Next, we will discuss the impact of primordial non-Gaussianity on the mass distribution in the low-redshift Universe (Section 3). The main body of this paper is Section 4, where we describe in detail a number of methods exploiting the abundance and clustering properties of observed tracers of the LSS to constrain the amount of initial non-Gaussianity. We conclude with a discussion of present and forecasted constraints achievable with LSS surveys (Section 5).

## 2. Models and Observables

Because they assume (i) a single dynamical field (the inflation), (ii) canonical kinetic energy terms (i.e., perturbations propagate at the speed of light), (iii) slow roll (i.e., the timescale over which the inflaton field changes are much larger than the Hubble rate), and (iv) an initial vacuum state, single-field slow-roll models lead to a small level of primordial non-Gaussianity [6, 7, 12]. The lowest-order statistics sensitive to non-Gaussian features in the initial distribution of scalar perturbations  $\Phi(\mathbf{x})$  (we will adopt the standard CMB convention in which  $\Phi(\mathbf{x})$  is Bardeen's curvature perturbation in the matter era) is the 3-point function or bispectrum  $B_\Phi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ , which is a function of any triangle  $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$  (as follows from statistical homogeneity which we assume throughout this paper). It has been recently shown that, in the squeezed limit  $k_3 \ll k_1 \approx k_2$ , the bispectrum of *any* single-field slow-roll inflationary model asymptotes to the local shape (defined in (3)) [13–15]. The corresponding nonlinear parameter predicted by these models is

$$f_{\text{NL}}^{\text{loc}} = \frac{5}{12}(1 - n_s) \approx 0.017 \text{ (single field)}, \quad (1)$$

where  $n_s$  is the tilt or spectral index of the power spectrum  $P_\Phi(k)$ , which is accurately measured to be  $n_s \approx 0.960 \pm 0.013$  [8]. Therefore, any robust measurement of  $f_{\text{NL}}^{\text{loc}}$  well above this level would thus rule out single-field slow-roll inflation.

*2.1. The Shape of the Primordial Bispectrum.* Large, potentially detectable amount of Gaussianity can be produced when at least one of the assumptions listed above is violated, that is, by multiple scalar fields [16, 17], nonlinearities in the relation between the primordial scalar perturbations and the inflaton field [7, 12], interactions of scalar fields [18], a modified dispersion relation, or a departure from the adiabatic Bunch-Davies ground state [19]. Generation of a large non-Gaussian signal is also expected at reheating [20] and in the ekpyrotic scenario [21]. Each of these physical mechanisms leaves a distinct signature in the primordial 3-point function  $B_\Phi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ , a measurement of which would thus provide a wealth of information about the physics of primordial fluctuations. Although the configuration shape of the primordial bispectrum can be extremely complex in some models, there are broadly three classes of shape characterizing the local, equilateral, and folded type of primordial non-Gaussianity [22, 23]. The magnitude of each template “X” is controlled by a dimensionless nonlinear parameter  $f_{\text{NL}}^{\text{X}}$  which we seek to constrain using CMB or LSS observations (instead of attempting a model-independent measurement of  $B_\Phi$ ).

Any non-Gaussianity generated outside the horizon induces a 3-point function that is peaked on squeezed or collapsed triangles for realistic values of the scalar spectral index. The resulting non-Gaussianity depends only on the local value of Bardeen's curvature potential and can thus be conveniently parameterized up to third order by [7, 9, 12, 24]

$$\Phi(\mathbf{x}) = \phi(\mathbf{x}) + f_{\text{NL}}^{\text{loc}} \phi^2(\mathbf{x}) + g_{\text{NL}}^{\text{loc}} \phi^3(\mathbf{x}), \quad (2)$$

where  $\phi(\mathbf{x})$  is an isotropic Gaussian random field and  $f_{\text{NL}}^{\text{loc}}, g_{\text{NL}}^{\text{loc}}$  are dimensionless, phenomenological parameters. Since curvature perturbations are of magnitude  $\mathcal{O}(10^{-5})$ , the cubic-order correction should always be negligibly small compared to the quadratic one when  $\mathcal{O}(f_{\text{NL}}^{\text{loc}}) \sim \mathcal{O}(g_{\text{NL}}^{\text{loc}})$ . However, this condition is not satisfied by some multifield inflationary models such as the curvaton scenario, in which a large  $g_{\text{NL}}^{\text{loc}}$  and a small  $f_{\text{NL}}^{\text{loc}}$  can be simultaneously produced [17]. The quadratic term generates the 3-point function at leading order

$$B_\Phi^{\text{loc}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 2f_{\text{NL}}^{\text{loc}} [P_\phi(k_1)P_\phi(k_2) + (\text{cyc.})], \quad (3)$$

where (cyc.) denotes all cyclic permutations of the indices and  $P_\phi(k)$  is the power spectrum of the Gaussian part  $\phi(\mathbf{x})$  of the Bardeen potential. The cubic-order terms generate a trispectrum  $T_\Phi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$  at leading order.

Equilateral type of non-Gaussianity, which arises in inflationary models with higher-derivative operators such as the DBI model, is well described by the factorizable form [25]

$$B_\Phi^{\text{eq}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 6f_{\text{NL}}^{\text{eq}} \left[ - \left( P_\phi(k_1)P_\phi(k_2) + (\text{cyc.}) \right) - 2 \left( P_\phi(k_1)P_\phi(k_2)P_\phi(k_3) \right)^{2/3} + \left( P_\phi^{1/3}(k_1)P_\phi^{2/3}(k_2)P_\phi(k_3) + (\text{perm.}) \right) \right]. \quad (4)$$

It can be easily checked that the signal is largest in the equilateral configurations  $k_1 \approx k_2 \approx k_3$  and suppressed in the squeezed limit  $k_3 \ll k_1 \approx k_2$ . Note that, in single-field slow-roll inflation, the 3-point function is a linear combination of the local and equilateral shapes [13].

As a third template, we consider the folded or flattened shape which is maximized for  $k_2 \approx k_3 \approx k_1/2$  [26]:

$$B_\Phi^{\text{fol}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = 6f_{\text{NL}}^{\text{fol}} \left[ \left( P_\phi(k_1)P_\phi(k_2) + (\text{cyc.}) \right) + 3 \left( P_\phi(k_1)P_\phi(k_2)P_\phi(k_3) \right)^{2/3} - \left( P_\phi^{1/3}(k_1)P_\phi^{2/3}(k_2)P_\phi(k_3) + (\text{perm.}) \right) \right], \quad (5)$$

and approximate the non-Gaussianity due to modification of the initial Bunch-Davies vacuum in canonical single-field action (the actual 3-point function is not factorizable). As in the previous example,  $B_\Phi^{\text{fol}}$  is suppressed in the squeezed configurations. Unlike  $B_\Phi^{\text{eq}}$ , however, the folded shape induces a scale-dependent bias at large scales (see Section 4.3).

*2.2. Statistics of the Linear Mass Density Field.* Bardeen's curvature potential  $\Phi(\mathbf{x})$  is related to the linear density perturbation  $\delta_0(\mathbf{k}, z)$  at redshift  $z$  through

$$\delta_0(\mathbf{k}, z) = \mathcal{M}(k, z)\Phi(\mathbf{k}), \quad (6)$$

where

$$\mathcal{M}(k, z) = \frac{2k^2 T(k)D(z)}{3\Omega_m H_0^2}. \quad (7)$$

Here,  $T(k)$  is the matter transfer function normalized to unity as  $k \rightarrow 0$ ,  $\Omega_m$  is the present-day matter density, and  $D(z)$  is the linear growth rate normalized to  $1+z$ .  $n$ -point correlators of the linear mass density field can thus be written as

$$\langle \delta_0(\mathbf{k}) \cdots \delta_0(\mathbf{k}_n) \rangle = \left( \prod_{i=1}^n \mathcal{M}(k_i) \right) \langle \Phi(\mathbf{k}_1) \cdots \Phi(\mathbf{k}_n) \rangle. \quad (8)$$

Smoothing inevitably arises when comparing observations of the large-scale structure with theoretical predictions from, for example, perturbation theory (PT), which are valid only in the weakly nonlinear regime [27], or from the spherical collapse model which ignores the strongly nonlinear internal dynamics of the collapsing regions [28, 29]. For this reason, we introduce the *smoothed* linear density field

$$\delta_R(\mathbf{k}, z) = \mathcal{M}(k, z) W_R(k) \Phi(\mathbf{k}) \equiv \mathcal{M}_R(k, z) \Phi(\mathbf{k}), \quad (9)$$

where  $W_R(k)$  is a (spherically symmetric) window function of characteristic radius  $R$  or mass scale  $M$  that smooths out the small-scale nonlinear fluctuations. We will assume a top-hat filter throughout. Furthermore, since  $M$  and  $R$  are equivalent variables, we shall indistinctly use the notations  $\delta_R$  and  $\delta_M$  in what follows.

**2.3. Topological Defects Models.** In addition to inflationary scenarios, there is a whole class of models, known as topological defect models, in which cosmological fluctuations are sourced by an inhomogeneously distributed component which contributes a small fraction of the total energy momentum tensor [30, 31]. The density field is obtained as the convolution of a discrete set of points with a specific density profile. Defects are deeply rooted in particle physics as they are expected to form at a phase transition. Since the early Universe may have plausibly undergone several phase transitions, it is rather unlikely that no defects at all were formed. Furthermore, high-redshift tracers of the LSS may be superior to CMB at finding non-Gaussianity sourced by topological defects [32]. However, CMB observations already provide stringent limits on the energy density of a defect component [8], so we shall only minimally discuss the imprint of these scenarios in the large-scale structure. Phenomenological defect models are, for instance,

$$\delta(\mathbf{x}) = \phi(\mathbf{x}) + \alpha_{\text{NL}}(\phi^2(\mathbf{x}) - \langle \phi^2 \rangle), \quad (10)$$

in which the initial matter density  $\delta(\mathbf{x})$  (rather than the curvature perturbation  $\Phi(\mathbf{x})$ ) contains a term proportional to the square of a Gaussian scalar field  $\phi(\mathbf{x})$  [9], or the  $\chi^2$  model (also known as isocurvature CDM model) in which  $\delta(\mathbf{x}) \propto \phi^2(\mathbf{x})$  [33].

### 3. Evolution of the Matter Density Field with Primordial NG

In this section, we summarize a number of results relative to the effect of primordial NG on the mass density field. These will be useful to understand the complexity that arises when considering biased tracers of the density field (see Section 4).

**3.1. Setting Up Non-Gaussian Initial Conditions.** Investigating the impact of non-Gaussian initial conditions (ICs) on the large-scale structure traced by galaxies and so forth requires simulations large enough so that many long-wavelength modes are sampled. At the same time, the simulations should resolve dark-matter halos hosting luminous red galaxies (LRGs) or quasars (QSOs), so that one can construct halo samples whose statistical properties mimic as closely as possible those of the real data. This favors the utilization of pure N-body simulations for which a large dynamical range can be achieved.

The evolution of the matter density field with primordial non-Gaussianity has been studied in series of large cosmological N-body simulations seeded with Gaussian and non-Gaussian initial conditions; see, for example, [11, 34–43]. For generic non-Gaussian (scalar) random fields, we face the problem of setting up numerical simulations with a prescribed correlation structure [44]. While an implementation of the equilateral and folded bispectrum shape requires the calculation of several computationally demanding convolutions, the operation is straightforward for primordial NG described by a local mapping such as the  $\chi^2$  or the  $f_{\text{NL}}^{\text{loc}}$  model. In the latter case, the local transformation  $\Phi = \phi + f_{\text{NL}}^{\text{loc}} \phi^2$  is performed before multiplication by the matter transfer function  $T(k)$  (computed with publicly available Boltzmann codes [45, 46]). The (dimensionless) power spectrum of the Gaussian part  $\phi(\mathbf{x})$  of the Bardeen potential is the usual power-law  $\Delta_\phi^2(k) \equiv k^3 P_\phi(k)/(2\pi^2) = A_\phi (k/k_0)^{n_s-1}$ . Unless otherwise stated, we shall assume a normalization  $A_\phi = 7.96 \times 10^{-10}$  at the pivot point  $k_0 = 0.02 \text{ Mpc}^{-1}$ . To date, essentially all numerical studies of structure formation with inflationary non-Gaussianity have implemented the local shape solely, so we will focus on this model in what follows.

Non-Gaussian corrections to the primordial curvature perturbations can renormalize the input (unrenormalized) power spectrum of fluctuations used to seed the simulations [47]. For the local  $f_{\text{NL}}^{\text{loc}}$  model with  $|f_{\text{NL}}^{\text{loc}}| \lesssim 100$ , renormalization effects are unlikely to be noticeable due to the limited dynamical range of current cosmological simulations. However, they can be significant, for example, in simulations of a local cubic coupling  $g_{\text{NL}}^{\text{loc}} \phi^3$  with a large primordial trispectrum [48]. The cubic-order term  $g_{\text{NL}}^{\text{loc}} \phi^3$  renormalizes the amplitude  $A_\phi$  of the power spectrum of initial curvature perturbations to  $A_\phi \rightarrow A_\phi + 6g_{\text{NL}}^{\text{loc}} \langle \phi^2 \rangle$ , where

$$\langle \phi^2 \rangle = \int \frac{d^3k}{(2\pi)^3} P_\phi(k). \quad (11)$$

For scale-invariant initial conditions,  $\langle \phi^2 \rangle$  has a logarithmic divergence at large and small scales. In practice, however, the finite box size and the resolution of the simulations naturally furnish a low- and high- $k$  cutoff. The effective correction to the amplitude of density fluctuations  $\delta\sigma_8$  in the  $g_{\text{NL}}^{\text{loc}} \phi^3$  model thus is

$$\delta\sigma_8 = 3g_{\text{NL}}^{\text{loc}} \left( \frac{Lk_0}{2\pi} \right)^{1-n_s} [1 - N^{n_s-1}] \frac{A_\phi}{1-n_s}, \quad (12)$$

where  $N$  is the number of mesh points along one dimension and  $L$  is the simulation box length. For  $g_{\text{NL}}^{\text{loc}} = 10^6$ ,  $L = 1.5 h^{-1} \text{Gpc}$ , and  $N = 1024$ , for instance, we obtain  $\delta\sigma_8 \approx 0.015$ .

To generate the initial particle distribution, the Zelovich approximation is commonly used instead of the exact gravitational dynamics. This effectively corresponds to starting from non-Gaussian initial conditions [49]. Since the transition to the true gravitational dynamics proceeds rather gradually [50], one should ensure that the initial expansion factor is much smaller than that of the outputs analyzed. Alternatively, it is possible to generate more accurate ICs based on second-order Lagrangian perturbation theory (2LPT) [51]. At fixed initial expansion factor, they reduce transients such that the true dynamics is recovered more rapidly [52].

**3.2. Mass Density Probability Distribution.** In the absence of primordial NG, the probability distribution function (PDF) of the initial smoothed density field (the probability that a randomly placed cell of volume  $V$  has some specific density) is Gaussian. Namely, all normalized or reduced *smoothed* cumulants  $S_J$  of order  $J \geq 3$  are zero. An obvious signature of primordial NG would thus be an initially nonvanishing skewness  $S_3 = \langle \delta_R^3 \rangle_c / \langle \delta_R^2 \rangle^2$  or kurtosis  $S_4 = \langle \delta_R^4 \rangle_c / \langle \delta_R^2 \rangle^3 - 3 / \langle \delta_R^2 \rangle$  [36, 53, 54]. Here, the subscript  $c$  denotes the connected piece of the  $n$ -point moment that cannot be simplified into a sum over products of lower-order moments. At third order, for instance, the cumulant of the smoothed density field is an integral of the 3-point function:

$$\langle \delta_R^3 \rangle_c = \int \frac{d^3 k_1}{(2\pi)^3} \int \frac{d^3 k_2}{(2\pi)^3} \int \frac{d^3 k_3}{(2\pi)^3} B_R(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z), \quad (13)$$

where

$$B_R(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, z) = \mathcal{M}_R(k_1, z) \mathcal{M}_R(k_2, z) \mathcal{M}_R(k_3, z) \times B_\Phi(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \quad (14)$$

is the bispectrum of the smoothed linear density fluctuations at redshift  $z$ . Note that, owing to  $S_3(R, z) \propto D(z)^{-1}$ , the product  $\sigma S_3(R)$  does not depend on redshift. Over the range of scale accessible to LSS observations,  $\sigma S_3(R)$  is a monotonic decreasing function of  $R$  that is of magnitude  $\sim 10^{-4}$  for the local, equilateral, and folded templates discussed above (Figure 1). Strictly speaking, all reduced moments should be specified to fully characterize the density PDF, but a reasonable description of the density distribution can be achieved with moments up to the fourth order.

Numerical and analytic studies generally find that a density PDF initially skewed towards positive values produces more overdense regions while a negatively skewed distribution produces larger voids. Gravitational instabilities also generate a positive skewness in the density field, reflecting the fact that the evolved density distribution exhibits an extended tail towards large overdensities [49, 55–59]. This gravitationally induced signal eventually dominates the primordial contribution such that, at fixed normalization amplitude, non-Gaussian models deviate more strongly

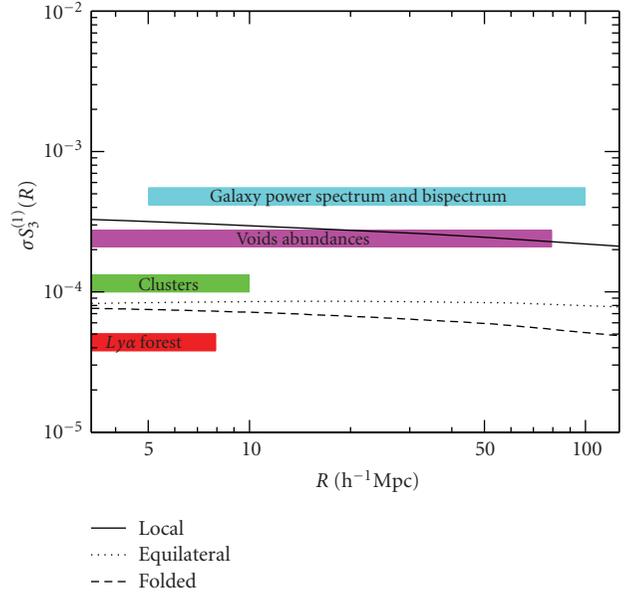


FIGURE 1: Skewness  $\sigma S_3^{(1)}(R)$  of the smoothed density field in unit of  $f_{\text{NL}}^{\text{X}}$  for the local, equilateral, and folded bispectrum shapes. The skewness for the equilateral and folded templates is a factor of  $\sim 3$  smaller than in the local model. In any case, this implies that  $|\sigma_R S_3(R)| \ll 1$  on the scales probed by the large-scale structure for realistic values of the nonlinear coupling parameter,  $|f_{\text{NL}}^{\text{X}}| \lesssim 100$ . The shaded regions approximately indicate the range of scales probed by various LSS tracers. For the galaxy power spectrum and bispectrum, the upper limit sensitively depends upon the surveyed volume.

from the Gaussian paradigm at high redshift. The time evolution of the normalized cumulants  $S_J$  can be worked out for generic Gaussian and non-Gaussian ICs using, for example, PT or the spherical collapse approximation. For Gaussian ICs, PT predicts the normalized cumulants to be time independent to the lowest nonvanishing order, with a skewness  $S_3 \approx 34/7$ , whereas, for non-Gaussian ICs, the linear contribution to the cumulants decays as  $S_J(R, z) = S_J(R, \infty) / D^{J-2}(z)$  [60, 61].

The persistence of the primordial hierarchical amplitude  $S_J(R, \infty)$  sensitively depends upon the magnitude of  $S_N$ ,  $N \geq J$ , relative to unity. For example, an initially large nonvanishing kurtosis could source skewness with a time-dependence and amplitude similar to those induced by nonlinear gravitational evolution [60]. Although there is an infinite class of non-Gaussian models, we can broadly divide them into weakly and strongly non-Gaussian. In weak NG models, the primeval signal in the normalized cumulants is rapidly obliterated by gravity-induced non-Gaussianity. This is the case of hierarchical scaling models where  $n$ -point correlation functions satisfy  $\xi_n \propto \xi_2^{n-1}$  with  $\xi_2 \ll 1$  at large-scales. By contrast, strongly NG initial conditions dominate the evolution of the normalized cumulants. This occurs when the hierarchy of correlation functions obeys the dimensional scaling  $\xi_n \propto \xi_2^{n/2}$ , which arises in the particular case of  $\chi^2$  initial conditions [62] or in defect models such as texture [37, 63, 64]. These scaling laws have been successfully

confronted with numerical investigations of the evolution of cumulants [37, 38]. We note that the scaling of the contribution induced by gravity is, however, different for the kurtosis [65], suggesting that the latter is a better probe of the nature of initial conditions.

Although gravitational clustering tends to erase the memory of initial conditions, numerical simulations of non-Gaussian initial conditions show that the occurrence of highly underdense and overdense regions is very sensitive to the presence of primordial NG. In fact, the imprint of primordial NG is best preserved in the negative tail of the PDF  $P(\rho_R)$  of the evolved (and smoothed) density field  $\rho_R$  [40]. A satisfactory description of this effect can be obtained from an Edgeworth expansion of the initial smoothed overdensity field [66]. At high densities  $\rho_R \gg 1$ , the non-Gaussian modification approximately scales as  $\rho_R^{3/5}$  whereas, at low densities  $\rho_R \simeq 0$ , the deviation is steeper,  $\rho_R^{6/5}$ . Taking into account the weak-scale dependence of  $\sigma_{S_3}(R)$  further enhances this asymmetry.

**3.3. Power Spectrum and Bispectrum.** Primordial-non-Gaussianity also imprints a signature in Fourier-space statistics of the matter density field. Positive values of  $f_{\text{NL}}^{\text{loc}}$  tend to increase the small-scale matter power spectrum  $P_\delta(k)$  [10, 40, 67] and the large-scale matter bispectrum  $B_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$  [10, 68]. In the weakly nonlinear regime where 1-loop PT applies, the Fourier mode of the density field for growing-mode initial conditions reads [56, 69]

$$\begin{aligned} \delta(\mathbf{k}, z) &= \delta_0(\mathbf{k}, z) + \frac{1}{(2\pi)^3} \int d^3q_1 d^3q_2 \delta_{\text{D}}(\mathbf{k} - \mathbf{q}_1 - \mathbf{q}_2) \\ &\quad \times F_2(\mathbf{q}_1, \mathbf{q}_2) \delta_0(\mathbf{q}_1, z) \delta_0(\mathbf{q}_2, z). \end{aligned} \quad (15)$$

The kernel  $F_2(\mathbf{k}_1, \mathbf{k}_2) = 5/7 + \mu(k_1/k_2 + k_2/k_1)/2 + 2\mu^2/7$ , where  $\mu$  is the cosine of the angle between  $\mathbf{k}_1$  and  $\mathbf{k}_2$ , describes the nonlinear 2nd-order evolution of the density field. It is nearly independent of  $\Omega_m$  and  $\Omega_\Lambda$  and vanishes in the (squeezed) limit  $\mathbf{k}_1 = -\mathbf{k}_2$  as a consequence of the causality of gravitational instability. At 1-loop PT, (15) generates the mass power spectrum

$$\begin{aligned} P_\delta(k, z) &= P_\delta^{\text{G}}(k, z) + \Delta P_\delta^{\text{NG}}(k, z) = P_0(k, z) \\ &\quad + [P_{(22)}(k, z) + P_{(13)}(k, z)] + \Delta P_\delta^{\text{NG}}(k, z). \end{aligned} \quad (16)$$

Here,  $P_0(k, z)$  is the linear matter power spectrum at redshift  $z$ ,  $P_{(22)}$  and  $P_{(13)}$  are the standard one-loop contributions in the case of Gaussian ICs [69, 70], and

$$\Delta P_\delta^{\text{NG}}(k, z) = 2 \int \frac{d^3q}{(2\pi)^3} F_2(\mathbf{q}, \mathbf{k} - \mathbf{q}) B_0(-\mathbf{k}, \mathbf{q}, \mathbf{k} - \mathbf{q}, z) \quad (17)$$

is the correction due to primordial NG [67]. This term scales as  $\propto D^3(z)$ , so the effect of non-Gaussianity is largest at low redshift. Moreover, because  $F_2(\mathbf{k}_1, \mathbf{k}_2)$  vanishes in the squeezed limit, (17) is strongly suppressed at small wavenumbers, even in the local  $f_{\text{NL}}^{\text{loc}}$  model for which

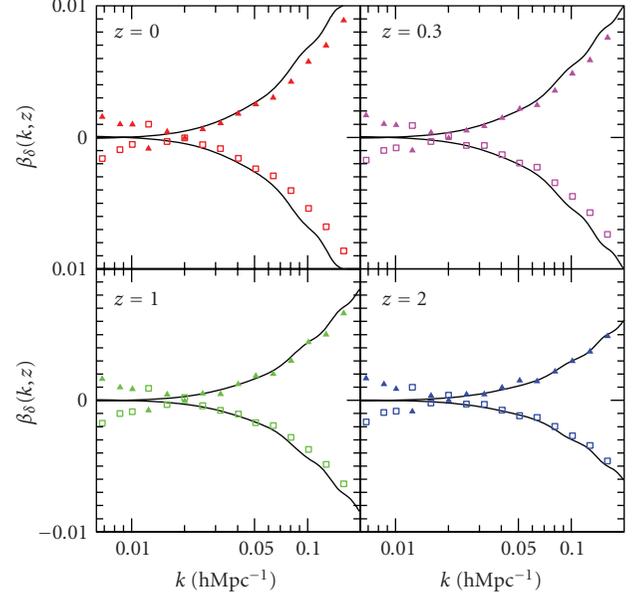


FIGURE 2: Non-Gaussian fractional correction  $\beta_\delta(k, z) = \Delta P_\delta^{\text{NG}}(k, z)/P_\delta^{\text{G}}(k, z)$  to the matter power spectrum that originates from primordial non-Gaussianity of the local type. Results are shown at redshift  $z = 0, 0.5, 1$ , and  $2$  for  $f_{\text{NL}}^{\text{loc}} = +100$  (filled symbols) and  $f_{\text{NL}}^{\text{loc}} = -100$  (empty symbols). The solid curves indicate the prediction from a 1-loop perturbative expansion.

$B_0(-\mathbf{k}, \mathbf{q}, \mathbf{k} - \mathbf{q}, z)$  is maximized in the same limit (i.e.,  $|\mathbf{k}| \rightarrow 0$ ). For  $f_{\text{NL}}^{\text{loc}} \sim \mathcal{O}(10^2)$ , the magnitude of this correction is at a percent level in the weakly nonlinear regime  $k \lesssim 0.1 \text{ hMpc}^{-1}$  [41, 42, 71], in good agreement with the measurements (see Figure 2). Extensions of the renormalization group description of dark-matter clustering [72] to non-Gaussian initial density, and velocity perturbations can improve the agreement up to wavenumbers  $k \lesssim 0.25 \text{ hMpc}^{-1}$  [73, 74].

It is also instructive to compare measurements of the matter bispectrum  $B_\delta(k_1, k_2, k_3)$  with perturbative predictions. To second order in PT, the matter bispectrum is the sum of a primordial contribution and two terms induced by gravitational instability [56, 75] (we will henceforth omit the explicit  $z$ -dependence for brevity):

$$\begin{aligned} B_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= B_0(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \\ &\quad + [2F_2(\mathbf{k}_1, \mathbf{k}_2)P_0(k_1)P_0(k_2) + (\text{cyc.})] \\ &\quad + \int \frac{d^3q}{(2\pi)^3} [F_2(\mathbf{q}, \mathbf{k}_3 - \mathbf{q}) \\ &\quad \times T_0(\mathbf{q}, \mathbf{k}_3 - \mathbf{q}, \mathbf{k}_1, \mathbf{k}_2) + (\text{cyc.})], \end{aligned} \quad (18)$$

where  $T_0(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$  is the primordial trispectrum of the density field. A similar expression can also be derived for the matter trispectrum, which turns out to be less sensitive to gravitationally induced nonlinearities [76]. The reduced bispectrum  $Q_3$  is conveniently defined as

$$Q_3(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = \frac{B_\delta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)}{[P_\delta(k_1)P_\delta(k_2) + \text{cyclic}]}. \quad (19)$$

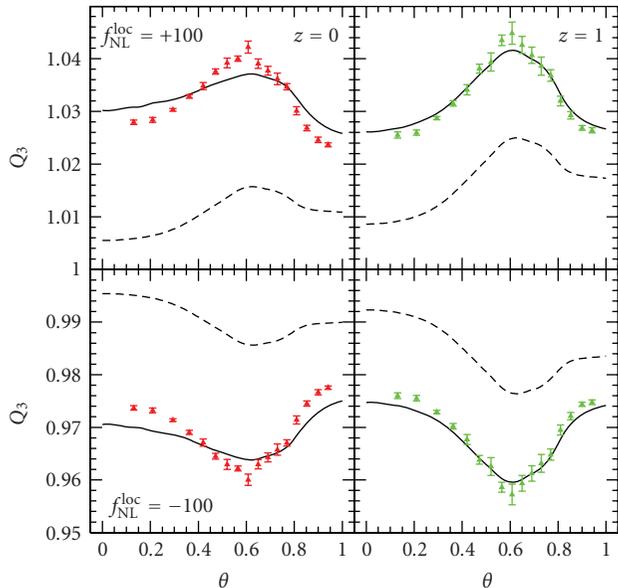


FIGURE 3: Reduced matter bispectrum  $Q_3$  as a function of the angle  $\theta$  between  $\mathbf{k}_1$  and  $\mathbf{k}_2$  for a fixed  $k_1 = 0.094 \text{ hMpc}^{-1}$  and  $k_2 = 1.5k_1$ . The panels show ratios between the non-Gaussian and Gaussian  $Q_3$  for  $f_{\text{NL}}^{\text{loc}} = +100$  (top) and  $-100$  (bottom). Dashed lines correspond to tree-level PT while continuous line indicates the 1-loop PT prediction.

For Gaussian initial conditions,  $Q_3$  is independent of time and, at tree-level PT, is constant and equal to  $Q_3(k, k, k) = 4/7$  for equilateral configurations [56]. For general triangles, moreover, it approximately retains this simple behavior, with a dependence on triangle shape through  $F_2(\mathbf{k}_1, \mathbf{k}_2)$  [10]. Figure 3 illustrates the effect of primordial NG of the local  $f_{\text{NL}}^{\text{loc}}$  type on the shape dependence of  $Q_3$  for a particular set of triangle configurations. As can be seen, the inclusion of 1-loop corrections greatly improves the agreement with the numerical data [77]. An important feature that is not apparent in Figure 3 is the fact that the primordial part to the reduced-matter bispectrum scales as  $Q_3 \propto 1/\mathcal{M}_R(k)$  for approximately equilateral triangles (and under the assumption that  $f_{\text{NL}}^{\text{loc}}$  is scale independent) [10]. This anomalous scaling considerably raises the ability of the matter bispectrum to constrain primordial NG of the local  $f_{\text{NL}}^{\text{loc}}$  type. Unfortunately, neither the matter bispectrum nor the power spectrum is directly observable with the large-scale structure of the Universe. Temperature anisotropies in the redshifted 21 cm background from the preionization epoch could in principle furnish a direct measurement of these quantities [78–80], but foreground contamination may severely hamper any detection. Weak lensing is another direct probe of the dark matter, although we can only observe it in projection along the line of sight [81].

As we will see shortly, however, this large-scale anomalous scaling is also present in the bispectrum and power spectrum of observable tracers of the large-scale structure such as galaxies. It is this unique signature that will make future all-sky LSS surveys competitive with CMB experiments.

**3.4. Velocities.** Primordial non-Gaussianity also leaves a signature in the large-scale coherent bulk motions which, in the linear regime, are directly related to the linear density field [55]. The various non-Gaussian models considered by [82] tend to have lower-velocity dispersion and bulk flow than fiducial Gaussian model, regardless of the sign of the skewness. However, while the probability distribution of velocity components is sensitive to primordial NG of the local type, in defect models it can be very close to Gaussian, even when the density field is strongly non-Gaussian, as a consequence of the central limit theorem [83, 84]. In this regard, correlation of velocity differences could provide a better test of non-Gaussian initial conditions [85].

To measure peculiar velocities, one must subtract the Hubble flow from the observed redshift. This requires an estimate of the distance which is only available for nearby galaxies and clusters (although, e.g., the kinetic Sunyaev-Zel’dovich (kSZ) effect could be used to measure the bulk motions of distant galaxy clusters [86]). So far, measurements of the local galaxy density and velocity fields [87] as well as reconstruction of the initial density PDF from galaxy density and velocity data [88] are consistent with Gaussian initial conditions.

## 4. LSS Probe of Primordial Non-Gaussianity

Discrete and continuous tracers of the large-scale structure, such as galaxies, the Ly $\alpha$  forest, the 21 cm hydrogen line, and so forth, provide a distorted image of the matter density field. In CDM cosmologies, galaxies form inside overdense regions [89] and this introduces a bias between the mass and the galaxy distribution [90]. As a result, distinct samples of galaxies trace the matter distribution differently, the most luminous galaxies preferentially residing in the most massive DM halos. This biasing effect, which concerns most tracers of the large-scale structure, remains to be fully understood. Models of galaxy clustering assume, for instance, that the galaxy biasing relation only depends on the local mass density, but the actual mapping could be more complex [91, 92]. Because of biasing, tracers of the large-scale structure will be affected by primordial non-Gaussianity in a different way than the mass density field. In this section, we describe a number of methods exploiting the abundance and clustering properties of biased tracers to constrain the level of primordial NG. We focus on galaxy clustering as it provides the tightest limits on primordial NG (see Section 5).

**4.1. Halo Finding Algorithm.** Locating groups of bound particles, or DM halos, in simulations is central to the methods described below. In practice, we aim at extracting halo catalogs with statistical properties similar to those of observed galaxies or quasars. This, however, proves to be quite difficult because the relation between observed galaxies and DM halos is somewhat uncertain. Furthermore, there is freedom at defining a halo mass.

An important ingredient is the choice of the halo identification algorithm. Halo finders can be broadly divided

into two categories: friends-of-friends (FOF) finders [93] and spherical overdensity (SO) finders [94]. While the mass of an SO halo is defined by the radius at which the inner overdensity exceeds  $\Delta_{\text{vir}}(z)$  (typically  $\sim$  a few hundred times the background density  $\bar{\rho}(z)$ ), the mass of an FOF halo is given by the number of particles within a linking length  $b$  from each other ( $b \sim 0.15 - 0.2$  in unit of mean interparticle distance). These definitions are somewhat arbitrary and may suit specific purposes only. In what follows, we shall present mainly results for SO halos as their mass estimate is more closely connected to the predictions of the spherical collapse model, on which most of the analytic formulae presented in this section are based. The question of how the spherical overdensity masses can be mapped onto friends-of-friends masses remains a matter of debate (e.g., [95]). Clearly, however, since the peak height  $\nu(M, z)$  depends on the halo mass  $M$  through the variance  $\sigma_M$  (see below), any systematic difference will be reflected in the value of  $\nu$  associated to a specific halo sample. As we will see shortly, this affects the size of the fractional deviation from the Gaussian mass function.

Catalogs of mock galaxies with luminosities comparable to those of the targeted survey provide an additional layer of complication that can be used, among others, to assess the impact of observational errors on the non-Gaussian signal. However, most numerical studies of cosmic structure formation with primordial NG have not yet addressed this level of sophistication, so we will discuss results based on statistics of dark matter halos only.

**4.2. Abundances of Voids and Bound Objects.** It has long been recognized that departure from Gaussianity can significantly affect the abundance of highly biased tracers of the mass density field, as their frequency sensitively depends upon the tails of the initial density PDF [96–98]. The (extended) Press-Schechter approach has been extensively applied to ascertain the effect of primordial NG on the high mass tail of the mass function.

**4.2.1. Press-Schechter Approach.** The Press-Schechter theory [99] and its extensions based on excursion sets [100–102] predict that the number density  $n(M, z)$  of halos of mass  $M$  at redshift  $z$  is entirely specified by a multiplicity function  $f(\nu)$ ,

$$n(M, z) = \frac{\bar{\rho}}{M^2} \nu f(\nu) \frac{d \ln \nu}{d \ln M}, \quad (20)$$

where the peak height or significance  $\nu(M, z) = \delta_c(z)/\sigma_M$  is the typical amplitude of fluctuations that produce those halos. Here and henceforth,  $\sigma_M$  denotes the variance of the initial density field  $\delta_M$  smoothed on mass scale  $M \propto R^3$  and linearly extrapolated to present epoch, whereas  $\delta_c(z) \approx 1.68D(0)/D(z)$  is the critical linear overdensity for (spherical) collapse at redshift  $z$ . In the standard Press-Schechter approach,  $n(M, z)$  is related to the level excursion probability  $P(> \delta_c, M)$  that the linear density contrast of a region of mass  $M$  exceeds  $\delta_c(z)$ ,

$$\nu f(\nu) = -2 \frac{\bar{\rho}}{M} \frac{dP}{dM} = \sqrt{\frac{2}{\pi}} \nu e^{-\nu^2/2}, \quad (21)$$

where the last equality assumes Gaussian initial conditions. The factor of 2 is introduced to account for the contribution of low density regions embedded in overdensities at scale  $> M$ . In the extended Press-Schechter theory,  $\delta_M$  evolves with  $M$  and  $\nu f(\nu)$  is the probability that a trajectory is absorbed by the constant barrier  $\delta = \delta_c$  (as is appropriate in the spherical collapse approximation) on mass scale  $M$ . In general, the exact form of  $f(\nu)$  depends on the barrier shape [103] and the filter shape [104]. Note also that  $\int d\nu f(\nu) = 1$ , which ensures that all the mass is contained in halos.

Despite the fact that the Press-Schechter mass function overpredicts (underpredicts) the abundance of low (high) mass objects, it can be used to estimate the fractional deviation from Gaussianity. In this formalism, the non-Gaussian fractional correction to the multiplicity function is  $R(\nu, f_{\text{NL}}^X) \equiv f(\nu, f_{\text{NL}}^X)/f(\nu, 0) = (dP/dM)(> \delta_c, M, f_{\text{NL}}^X)/(dP/dM)(> \delta_c, M, 0)$ , which is readily computed once the non-Gaussian density PDF  $P(\delta_M)$  is known. In the simple extensions proposed by [105, 106],  $P(\delta_M)$  is expressed as the inverse transform of a cumulant-generating function. In [106], the saddle-point technique is applied directly to  $P(\delta_M)$ . The resulting Edgeworth expansion is then used to obtain  $P(> \delta_c, M)$ . Neglecting cumulants beyond the skewness, one obtain (we momentarily drop the subscript  $M$  for convenience)

$$R_{\text{LV}}(\nu, f_{\text{NL}}^X) \approx 1 + \frac{1}{6} \sigma S_3 (\nu^3 - 3\nu) - \frac{1}{6} \frac{d(\sigma S_3)}{d \ln \nu} \left( \nu - \frac{1}{\nu} \right), \quad (22)$$

after integration over regions above  $\delta_c(z)$ . In [105], it is the level excursion probability  $P(> \delta_c, M)$  that is calculated within the saddle-point approximation. This approximation better asymptotes to the exact large mass tail, which exponentially deviates from the Gaussian tail. To enforce the normalization of the resulting mass function, one may define  $\nu_\star = \delta_\star/\sigma$  with  $\delta_\star = \delta_c \sqrt{1 - S_3 \delta_c/3}$ , and use [105, 107]

$$\nu_\star f(\nu_\star) = M^2 \frac{n_{\text{NG}}(M, z)}{\bar{\rho}} \frac{d \ln M}{d \ln \nu_\star}. \quad (23)$$

The fractional deviation from the Gaussian mass function then becomes

$$R_{\text{MVJ}}(\nu, f_{\text{NL}}^X) \approx \exp\left(\frac{\nu^3}{6} \sigma S_3\right) \left[ -\frac{\sigma \nu^2}{6 \nu_\star} \frac{d S_3}{d \ln \nu} + \frac{\nu_\star}{\nu} \right]. \quad (24)$$

Both formulae have been shown to give reasonable agreement with numerical simulations of non-Gaussian cosmologies [41, 108, 109] (but note that [11, 110] have reached somewhat different conclusions). Expanding  $\delta_\star$  at the first order in  $f_{\text{NL}}^X$  shows that these two theoretical expectations differ in the coefficient of the  $\nu \sigma S_3$  term. Therefore, it is interesting to consider also the approximation

$$R(\nu, f_{\text{NL}}^X) \approx \exp\left(\frac{\nu^3}{6} \sigma S_3\right) \left[ 1 - \frac{\nu}{2} \sigma S_3 - \frac{\nu}{6} \frac{d(\sigma S_3)}{d \ln \nu} \right], \quad (25)$$

which is designed to match better the Edgeworth expansion of [106] when the peak height is  $\nu \sim 1$  [48]. When the

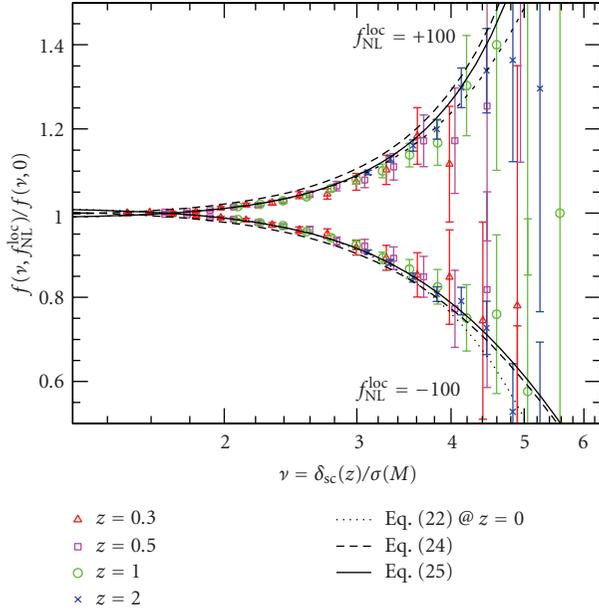


FIGURE 4: Fractional deviation from the Gaussian mass function as a function of the peak height  $\nu = \delta_c/\sigma$ . Different symbols refer to different redshifts as indicated. The various curves are theoretical prediction at  $z = 0$  (see text). Halos were identified using a spherical overdensity (SO) finder with a redshift-dependent overdensity threshold  $\Delta_{\text{vir}}(z)$  (with  $\Delta_{\text{vir}}(z)$  increasing from  $\sim 200$  at high redshift to attain  $\sim 350$  at  $z = 0$ ). Error bars denote Poisson errors. For illustration,  $M = 10^{15} M_\odot/h$  corresponds to a significance  $\nu = 3.2, 5.2, 7.7$  at redshift  $z = 0, 1$  and  $2$ , respectively. Similarly,  $M = 10^{14} M_\odot/h$  and  $10^{13} M_\odot/h$  correspond to  $\nu = 1.9, 3, 4.5$  and  $1.2, 1.9, 2.9$ , respectively.

primordial trispectrum is large (i.e., when  $g_{\text{NL}}^{\text{loc}} \sim 10^6$ ), terms involving the kurtosis must be included [48, 105, 106, 111]. In this case, it is also important to take into account a possible renormalization of the fluctuation amplitude,  $\sigma_8 \rightarrow \sigma_8 + \delta\sigma_8$  (12), to which the high mass tail of the mass function is exponentially sensitive [48].

Figure 4 shows the effect of primordial NG of the local  $f_{\text{NL}}^{\text{loc}}$  type on the mass function of SO halos identified with a redshift-dependent overdensity threshold  $\Delta_{\text{vir}}(z)$  (motivated by spherical collapse in a  $\Lambda$ CDM cosmology [112]). Overall, the approximation (25) agrees better with the measurements than (24), which somewhat overestimates the data for  $f_{\text{NL}}^{\text{loc}} = 100$ , and than (22), which is not always positive definite for  $f_{\text{NL}}^{\text{loc}} = -100$ . However, as can be seen in Figure 5, while the agreement with the data is reasonable for the SO halos, the theory strongly overestimates the effect in the mass function of FOF halos. Reference [109], who use a FOF algorithm with  $b = 0.2$ , makes the replacement  $\delta_c \rightarrow \delta_c \sqrt{q}$  with  $q \simeq 0.75$  to match the Gaussian and non-Gaussian mass functions. A physical motivation of this replacement is provided by [113, 114], who demonstrate that the diffusive nature of the collapse barrier introduces a similar factor  $q = (1 + D_B)^{-1}$ , regardless of the initial conditions. However, the value of the diffusion constant  $D_B$  was actually measured from simulations that use a SO

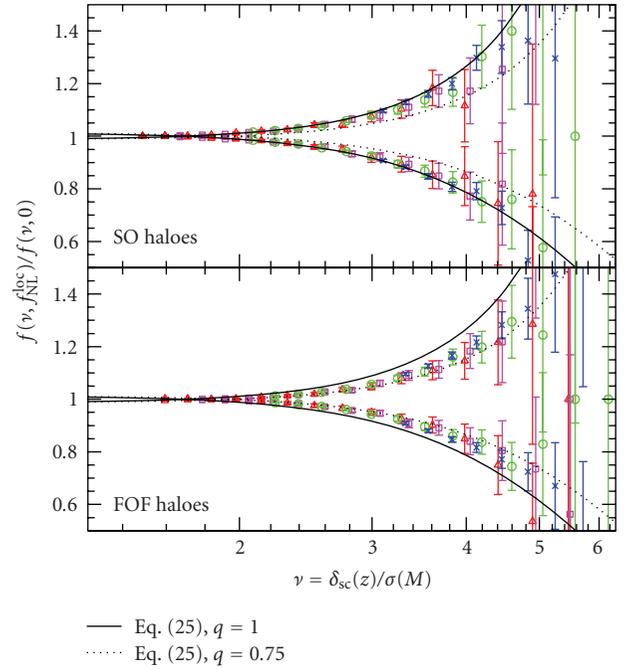


FIGURE 5: Fractional deviation from the Gaussian mass function as a function of the peak height  $\nu = \delta_c/\sigma$ . Different symbols refer to different redshifts as in Figure 4. The curves are the theoretical prediction (25) at  $z = 0$  with  $q = 1$  (solid) and  $q = 0.75$  (dotted). In the top panel, halos were identified using a spherical overdensity (SO) finder with a redshift-dependent overdensity threshold  $\Delta_{\text{vir}}(z)$  whereas, in the bottom panel, a Friends-of-Friends (FOF) finding algorithm with linking length  $b = 0.2$  was used.

finder with constant  $\Delta_{\text{vir}} = 200$  [115]. In the excursion set approach of [116], the value of  $q$  is obtained by normalizing the Gaussian mass function to simulation (i.e., it has nothing to do with the collapse dynamics) and, therefore, depends on the halo finder. Figure 5 demonstrates that this is also the case for the non-Gaussian correction  $R(\nu, f_{\text{NL}}^{\text{loc}})$ : choosing  $q \simeq 0.75$  as advocated in [109] gives good agreement for FOF halos, but strongly underestimates the effect for SO halos, for which the best-fit  $q$  is close to unity. As we will see below, the strength of the non-Gaussian bias may also be sensitive to the choice of halo finder.

More sophisticated formulations based on extended Press-Schechter (EPS) theory and/or modifications of the collapse criterion look promising since they can reasonably reproduce both the Gaussian halo counts and the dependence on  $f_{\text{NL}}^{\text{X}}$  [114, 117, 118]. The probability of first upcrossing can, in principle, be derived for any non-Gaussian density field and any choice of smoothing filter [119, 120]. For a general filter, the non-Markovian dynamics generates additional terms in the non-Gaussian correction to the mass function that arise from 3-point correlators of the smoothed density  $\delta_M$  at different mass scales [114]. However, large error bars still make difficult to test for the presence of such subleading terms. For generic moving barriers  $B(\sigma)$  such as those appearing in models of triaxial

collapse [121, 122], the leading contribution to the non-Gaussian correction approximately is [117]

$$\begin{aligned} R(\nu, f_{\text{NL}}^{\text{X}}) &\approx 1 + \frac{1}{6} \sigma \mathcal{S}_3 H_3 \left( \frac{B(\sigma)}{\sigma} \right) \\ &\approx 1 + \frac{1}{6} \sigma \mathcal{S}_3 \sqrt{q} (q\nu^3 - 3\nu), \end{aligned} \quad (26)$$

where  $H_3(\nu) \equiv \nu^3 - 3\nu$  and the last equality assumes  $\nu \gg 1$ . For SO halos, (26) with  $q \sim 0.7$  does not fit to the measured correction  $R(\nu, f_{\text{NL}}^{\text{loc}})$  better than (25). However, the ellipsoidal collapse barrier with  $q \sim 0.7$  matches better the Gaussian mass function for moderate peak height  $\nu \lesssim 2$  [118].

Parameterizations of the fractional correction based on N-body simulations have also been considered. While [43] considers a fourth-order polynomial fit to account for values of  $f_{\text{NL}}^{\text{loc}}$  as large as 750, [11] exploits the fact that, for sufficiently small  $f_{\text{NL}}^{\text{loc}}$ , there is a one-to-one mapping between halos in Gaussian and non-Gaussian cosmologies. In both cases, the fitting functions are consistent with the simulations at the few percent level.

**4.2.2. Clusters Abundance.** Rich clusters of galaxies trace the rare, high-density peaks in the initial conditions and thus offer the best probe of the high-mass tail of the multiplicity function. To infer the cluster mass function, the X-ray and millimeter windows are better suited than the optical-wave range because selection effects can be understood better (see, however, [123]).

Following early theoretical [96, 97, 124–126] and numerical [35, 127–129] work on the effect of non-Gaussian initial conditions on the multiplicity function of cosmic structures, the abundance of clusters and X-ray counts in non-Gaussian cosmologies has received much attention in the literature. At fixed normalization of the observed abundance of local clusters, the protoclusters associated with high redshift ( $2 < z < 4$ ) Ly $\alpha$  emitters are much more likely to develop in strongly non-Gaussian models than in the Gaussian paradigm [39, 110, 130]. Considering the redshift evolution of cluster abundances thus can break the degeneracy between the initial density PDF and the background cosmology. Simple extensions of the Press-Schechter formalism similar to those considered above have been shown to capture reasonably well the cluster mass function over a wide range of redshift for various non-Gaussian scenarios [131]. Scaling relations between the cluster mass, X-ray temperature and Compton  $y$ -parameter calibrated using theory, observations and detailed simulations of cluster formation [132, 133], can be exploited to predict the observed distribution functions of X-ray and SZ signals and assess the capability of cluster surveys to test the nature of the initial conditions [134–140].

An important limitation of this method is that the impact of realistic models of primordial non-Gaussianity on cluster abundances is small compared to systematics in current and upcoming surveys [141–143]. Given the current uncertainties in the redshift evolution of clusters (one commonly assumes that clusters are observed at the epoch they collapse [142]), the selection effects in the calibration

of X-ray and SZ fluxes with halo mass, the freedom in the definition of the halo mass, the degeneracy with the normalization amplitude  $\sigma_8$  (for positive  $f_{\text{NL}}^{\text{X}}$ , the mass function is more enhanced at the high mass end, and this is similar to an increase in the amplitude of fluctuations  $\sigma_8$  [144]) and the low number statistics, the prospects of using the cluster mass function only to place competitive limits on  $f_{\text{NL}}^{\text{X}}$  with the current data are small. A two-fold improvement in cluster mass calibration is required to provide constraints comparable to CMB measurements [143].

**4.2.3. Voids Abundance.** The frequency of cosmic voids offers a probe of the low density tail of the initial PDF [145]. The Press-Schechter formalism can also be applied to ascertain the sensitivity of the voids abundance to non-Gaussian initial conditions. Voids are defined as regions of mass  $M$  whose density is less than some critical value  $\delta_v \leq 0$  or, alternatively, whose three eigenvalues of the tidal tensor [146] lie below some critical value  $\lambda_v \leq 0$  [66, 118, 145, 147]. An important aspect in the calculation of the mass function of voids is the overcounting of voids located inside collapsing regions. This voids-in-clouds problem, as identified by [148]), can be solved within the excursion set theory by studying a two barriers problem:  $\delta_c$  for halos and  $\delta_v$  for voids. Including this effect reduces the frequency of the smallest voids [118]. Neglecting this complication notwithstanding, the differential number density of voids of radius  $R$  is [145, 147]

$$\frac{dn}{dR} \approx \frac{9}{2\pi^2} \sqrt{\frac{\pi}{2}} \frac{|\nu_v|}{R^4} e^{-\nu_v^2/2} \frac{d \ln |\nu_v|}{d \ln M} \left[ 1 - \frac{1}{6} \sigma \mathcal{S}_3 H_3(|\nu_v|) \right], \quad (27)$$

where  $\nu_v = \delta_v/\sigma_M$ . While a positive  $f_{\text{NL}}^{\text{X}}$  produces more massive halos, it generates fewer large voids [118, 145]. Hence, the effect is qualitatively different from a simple rescaling of the normalization amplitude  $\sigma_8$ . A joint analysis of both abundances of clusters and cosmic voids might thus provide interesting constraints on the shape of the primordial 3-point function. There are, however, several caveats to this method, including the fact that there is no unique way to define voids [145]. Clearly, voids identification algorithms will have to be tested on numerical simulations [149] before a robust method can be applied to real data.

**4.3. Galaxy 2-Point Correlation.** Before [90] showed that, in Gaussian cosmologies, correlations of galaxies and clusters can be amplified relative to the mass distribution, it was argued that primeval fluctuations have a non-Gaussian spectrum [150, 151] to explain the observed strong correlation of Abell clusters [152, 153]. Along these lines, [154] pointed out that primordial non-Gaussianity could significantly increase the amplitude of the two-point correlation of galaxies and clusters on large-scales. However, except from [155] who showed that correlations of high density peaks in non-Gaussian models are significantly stronger than in the Gaussian model with identical mass power spectrum, subsequent work focused mostly on abundances (Section 4.2) or higher order statistics such as the bispectrum (Section 4.4). It is

only recently that [11] have demonstrated the strong scale-dependent bias arising in non-Gaussian models of the local  $f_{\text{NL}}^{\text{loc}}$  type.

**4.3.1. The Non-Gaussian Bias.** In the original derivation of [11], the Laplacian is applied to the local mapping  $\Phi = \phi + f_{\text{NL}}^{\text{loc}} \phi^2$  in order to show that, upon substitution of the Poisson equation, the overdensity in the neighborhood of density peaks is spatially modulated by a factor proportional to the local value of  $\phi$ . Taking into account the coherent motions induced by gravitational instabilities, the scale-dependent bias correction reads

$$\Delta b_\kappa(k, f_{\text{NL}}^{\text{loc}}) = 3 f_{\text{NL}}^{\text{loc}} [b_1(M) - 1] \delta_c(0) \frac{\Omega_m H_0^2}{k^2 T(k) D(z)}, \quad (28)$$

where  $b_1(M)$  is the linear, Eulerian bias of halos of mass  $M$ . The original result missed out a multiplicative factor of  $T(k)^{-1}$  which was introduced subsequently by [157] upon a derivation of (28) in the limit of high density peaks. The peak-background split approach [103, 158, 159] promoted by [156] shows that the scale-dependent bias applies to any tracer of the matter density field whose (Gaussian) multiplicity function depends on the local mass density only. In this approach, the Gaussian piece of the potential is decomposed into short- and long-wavelength modes,  $\phi = \phi_l + \phi_s$ . The short-wavelength piece of the density field is then given by the convolution

$$\delta_s = \mathcal{M} \star \Phi_s = \mathcal{M} \star \phi_s (1 + 2 f_{\text{NL}}^{\text{loc}} \phi_l) + f_{\text{NL}}^{\text{loc}} \mathcal{M} \star \phi_s^2, \quad (29)$$

where  $\mathcal{M}$  is the transfer function (7). Ignoring the white-noise term, this provides an intuitive explanation of the effect in terms of a local rescaling of the small-scale amplitude of matter fluctuations or, equivalently, a local rescaling of the critical density threshold,

$$\sigma_s \longrightarrow \sigma_s (1 + 2 f_{\text{NL}}^{\text{loc}} \phi_l(x)). \quad (30)$$

Assuming that the mass function depends only on the peak height  $\nu = \delta_c / \sigma_s$ , the long-wavelength part of the halo overdensity becomes [156] (see also [11, 71, 160])

$$\begin{aligned} \delta_l^{\text{h}}(\mathbf{x}) &= \frac{1}{n(\nu)} n \left( \frac{\delta_c - \delta_l(\mathbf{x})}{\sigma_s (1 + 2 f_{\text{NL}}^{\text{loc}} \phi_l(\mathbf{x}))} \right) - 1 \\ &\approx -\frac{1}{\sigma_s} (\delta_l(\mathbf{x}) + 2 f_{\text{NL}}^{\text{loc}} \phi_l(\mathbf{x})) \frac{d \ln n}{d\nu}. \end{aligned} \quad (31)$$

Upon a Fourier transformation and using the fact that, in the Gaussian case,  $\delta_l^{\text{h}}(\mathbf{k}) = b_L \delta_l(\mathbf{k})$  with the Lagrangian bias  $b_L = -\sigma_s^{-1} d \ln n / d\nu$ , we recover the non-Gaussian bias correction (28) provided that the tracers move coherently with the dark matter, that is,  $b_L = b_1(M) - 1$  [161]. As emphasized in [11], the scale-dependence arises from the fact that the non-Gaussian curvature perturbation  $\Phi(\mathbf{x})$  is related to the density through the Poisson equation (6) (so that  $\delta_l(\mathbf{k}) = \mathcal{M}(k) \phi_l(\mathbf{k})$ ). There is no such effect in the (local)  $\chi^2$  model, (10), nor in texture-seeded cosmologies [162] for instance.

The derivation of [157], based on the clustering of regions of the smoothed density field  $\delta_M$  above threshold  $\delta_c(z)$ , is formally valid for high density peaks only. However, it is general enough to apply to any shape of primordial bispectrum. The 2-point correlation function of that level excursion set, which was first derived by [124], can be expressed in the high threshold limit ( $\nu \gg 1$ ) as

$$\begin{aligned} \xi_{>\nu}(\mathbf{r}) &= -1 + \exp \left\{ \sum_{n=2}^{\infty} \sum_{j=1}^{n-1} \frac{\gamma^n \sigma^{-n}}{j!(n-j)!} \right. \\ &\quad \left. \times \xi_R^{(n)} \left( \begin{array}{cc} \mathbf{x}_1, \dots, \mathbf{x}_1, & \mathbf{x}_2, \dots, \mathbf{x}_2 \\ j \text{ times} & (n-j) \text{ times} \end{array} \right), z=0 \right\}, \end{aligned} \quad (32)$$

where  $\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2$ . For most non-Gaussian models in which the primordial 3-point function is the dominant correction, this expansion can be truncated at the third order and Fourier transformed to yield the non-Gaussian correction  $\Delta P_{>\nu}(k)$  to the power spectrum. Assuming a small level of primordial NG, we can also write  $\Delta P_{>\nu}(k) \approx 2 b_L \Delta b_\kappa P_R(k)$  where  $b_L \approx \nu^2 / \delta_c$ , and eventually obtain

$$\Delta b_\kappa(k, f_{\text{NL}}^{\text{X}}) \equiv b_\phi(k) \mathcal{F}(k, f_{\text{NL}}^{\text{X}}) = \left( \frac{2 b_L \delta_c(z)}{\mathcal{M}_R(k, 0)} \right) \mathcal{F}(k, f_{\text{NL}}^{\text{X}}). \quad (33)$$

The dependence on the shape of the 3-point function is encoded in the function  $\mathcal{F}(k, f_{\text{NL}}^{\text{X}})$  [157, 163]

$$\begin{aligned} \mathcal{F}(k, f_{\text{NL}}^{\text{X}}) &= \frac{1}{16\pi^2 \sigma^2} \int_0^\infty dk_1 k_1^2 \mathcal{M}_R(k_1, 0) \\ &\quad \times \int_{-1}^{+1} d\mu \mathcal{M}_R(\sqrt{\alpha}, 0) \frac{B_\Phi(k_1, \sqrt{\alpha}, k)}{P_\Phi(k)}, \end{aligned} \quad (34)$$

where  $\alpha^2 = k^2 + k_1^2 + 2 \mu k k_1$ . Note that, for  $f_{\text{NL}}^{\text{loc}} < 0$ , this first order approximation always breaks down at sufficiently small  $k$  because  $\Delta P_{>\nu}(k) < 0$ .

Figure 6 shows the non-Gaussian halo bias (33) induced by the local, equilateral and folded bispectrum [163]. In the local and folded non-Gaussianity, the deviation is negligible at  $k = 0.1 \text{ hMpc}^{-1}$ , but increases rapidly with decreasing wavenumber. Still, the large-scale correction is much smaller for the folded template, and nearly absent for the equilateral type, which make them much more difficult to detect with galaxy surveys [163]. To get insights into the behavior of  $\Delta b_\kappa(k, f_{\text{NL}}^{\text{X}})$  at large-scales, let us identify the dominant contribution to  $\mathcal{F}(k, f_{\text{NL}}^{\text{X}})$  in the limit  $k \ll 1$ . Setting  $\mathcal{M}_R(\sqrt{\alpha}, 0) \approx \mathcal{M}_R(k_1, 0)$  and expanding  $P_\Phi(\sqrt{\alpha})$  at second order in  $k/k_1$ , we find after some algebra

$$\begin{aligned} \mathcal{F}(k, f_{\text{NL}}^{\text{loc}}) &\approx f_{\text{NL}}^{\text{loc}}, \\ \mathcal{F}(k, f_{\text{NL}}^{\text{eq}}) &\approx f_{\text{NL}}^{\text{eq}} \left[ 3 \Sigma_R \left( \frac{2(n_s - 4)}{3} \right) k^{2(4-n_s)/3} \right. \\ &\quad \left. + \frac{1}{2} (n_s - 4) \Sigma_R(-2) k^2 \right] \sigma_R^{-2}, \\ \mathcal{F}(k, f_{\text{NL}}^{\text{fol}}) &\approx \frac{3}{2} f_{\text{NL}}^{\text{fol}} \Sigma_R \left( \frac{n_s - 4}{3} \right) k^{(4-n_s)/3} \sigma_R^{-2}, \end{aligned} \quad (35)$$

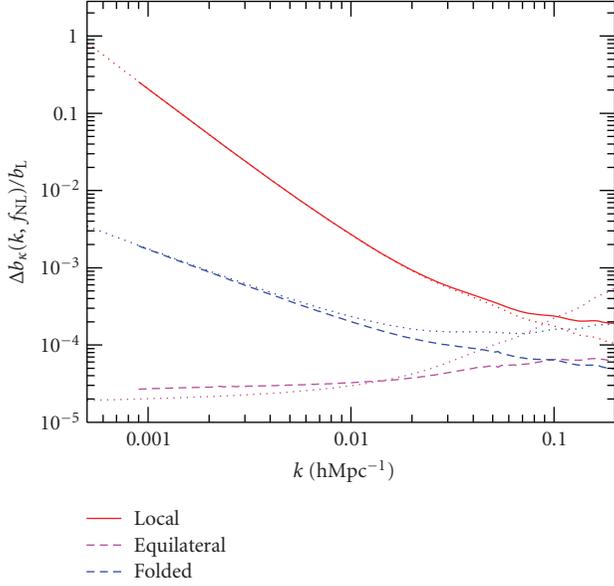


FIGURE 6: Non-Gaussian bias correction (33) for the local, equilateral and folded primordial bispectrum as a function of wavenumber. Results are shown at  $z = 0$  for a smoothing radius  $R = 5 \text{ h}^{-1} \text{ Mpc}$  and a nonlinear parameter  $f_{\text{NL}}^{\text{X}} = 1$ . The dotted line represent an analytic approximation, (35), which is valid at large-scales  $k \ll 1$ . Note that while the magnitude of  $\Delta b_{\kappa}(k, f_{\text{NL}}^{\text{loc}})$  does not change with  $R$ ,  $\Delta b_{\kappa}(k, f_{\text{NL}}^{\text{eq}})$  and  $\Delta b_{\kappa}(k, f_{\text{NL}}^{\text{fol}})$  strongly depends on the smoothing radius (see text).

assuming no running scalar index, that is,  $dn_s/d \ln k = 0$ . The auxiliary function  $\Sigma_R(n)$  is defined as

$$\Sigma_R(n) = \frac{1}{2\pi^2} \int_0^{\infty} dk k^{(2+n)} \mathcal{M}_R(k, 0)^2 P_{\phi}(k). \quad (36)$$

Hence, we have  $\Sigma_R(0) \equiv \sigma_R^2$ . As can be seen in Figure 6, these approximations capture relatively well the large scale non-Gaussian bias correction induced by the equilateral and folded type of non-Gaussianity. For a nearly scale-invariant spectrum  $n_s \approx 1$ , the effect scales as  $\Delta b_{\kappa} \propto k$  and  $\Delta b_{\kappa} \propto \text{const.}$ , respectively. Another important feature of the equilateral and folded non-Gaussian bias is the dependence on the mass scale  $M$  through the multiplicative factor  $\sigma_R^{-2}$ . Indeed, choosing  $R = 1 \text{ h}^{-1} \text{ Mpc}$  instead of  $R = 5 \text{ h}^{-1} \text{ Mpc}$  as done in Figure 6 would suppress the effect by a factor of  $\sim 3$ . In the high peak limit,  $\sigma_R^{-2} \approx b_L/\delta_c(z)$  which cancels out the dependence on redshift but enhances the sensitivity to the halo bias, that is,  $\Delta b_{\kappa} \propto b_L^2$  for the equilateral and folded shapes whereas  $\Delta b_{\kappa} \propto b_L$  in the local model.

At this point, it is appropriate to mention a few caveats to these calculations. Firstly, (28) assumes that the tracers form after a spherical collapse, which may be a good approximation for the massive halos only. If one instead considers the ellipsoidal collapse dynamics, in which the evolution of a perturbation depends upon the three eigenvalues of the initial tidal shear,  $\delta_c(0)$  should be replaced by its ellipsoidal counterparts  $\delta_{\text{ec}}(0)$  which is always larger than the spherical value [121]. In this model, the scale-dependent bias  $\Delta b_{\kappa}$  is thus enhanced by a factor  $\delta_{\text{ec}}(0)/\delta_c(0)$  [11, 160]. Secondly,

(28) assumes that the biasing of the surveyed objects is described by the peak height  $\nu$  only or, equivalently, the hosting halo mass  $M$ . However, this may not be true for quasars whose activity may be triggered by merger of halos [164, 165]. Reference [156] used the EPS formalism to estimate the bias correction  $\Delta b_{\text{merger}}$  induced by mergers

$$\Delta b_{\text{merger}} = \delta_c^{-1}, \quad (37)$$

so the factor  $b_1(M) - 1$  should be replaced by  $b_1(M) - 1 - \delta_c^{-1} \approx b_1(M) - 1.6$ . The validity of this result should be evaluated with cosmological simulations of quasars formation. In this respect, semianalytic models of galaxy formation suggest that merger-triggered objects such as quasars do not cluster much differently than other tracers of the same mass [166]. However, this does not mean that the same should hold for the non-Gaussian scale-dependent bias. Still, since the recent merger model is an extreme case it seems likely that the actual bias correction is  $0 < \Delta b_{\text{merger}} < \delta_c^{-1}$ . Thirdly, the scale-dependent bias has been derived using the Newtonian approximation to the Poisson equation, so one may wonder whether general relativistic (GR) corrections to  $\mathcal{M}_R(k)^{-1}$  may suppress the effect on scales comparable to the Hubble radius. Reference [167] showed how large-scale primordial NG induced by GR corrections propagates onto small scales once cosmological perturbations reenter the Hubble radius in the matter dominated era. This effect generates a scale-dependent bias comparable, albeit of opposite sign to that induced by local NG [163]. More recently, [168] argues that there are no GR corrections to the non-Gaussian bias and that the scaling  $\Delta b_{\kappa} \propto k^{-2}$  applies down to smallish wavenumbers.

We can also ask ourselves whether higher-order terms in the series expansion (32) furnish corrections to the non-Gaussian bias similar to (28). The quadratic coupling  $f_{\text{NL}}^{\text{loc}} \phi^2$  induces a second order correction to the halo power spectrum which reads [48]

$$\begin{aligned} \Delta P_{\text{h}}(k) &= \frac{4}{3} (f_{\text{NL}}^{\text{loc}})^2 [b_1(M) - 1]^2 \delta_c^2(z) S_3^{(1)}(M) \\ &\times \mathcal{M}_R(k, 0) P_{\phi}(k). \end{aligned} \quad (38)$$

Its magnitude relative to the term linear in  $f_{\text{NL}}^{\text{loc}}$ , (28), is approximately 0.03 at redshift  $z = 1.8$  and for a halo mass  $M = 10^{13} M_{\odot}/h$ . Although its contribution becomes increasingly important at higher redshift, it is fairly small for realistic values of  $f_{\text{NL}}^{\text{loc}}$ . In local NG model, the power spectrum of biased tracers of the density field can also be obtained from a local Taylor series in the evolved (Eulerian) density contrast  $\delta$  and the Gaussian part  $\phi$  of the initial (Lagrangian) curvature perturbation [47, 71]. Using this approach, it can be shown that the halo power spectrum arising from the first order terms of the local bias expansion can be cast into the form [47]

$$P_{\text{h}}(k) = \left[ b_1(M) + f_{\text{NL}}^{\text{loc}} b_{\phi}(k) \right]^2 P_R(k). \quad (39)$$

Hence, we also obtain a second order term proportional to  $(f_{\text{NL}}^{\text{loc}})^2 \mathcal{M}_R^{-2} P_R(k) = (f_{\text{NL}}^{\text{loc}})^2 P_{\phi}(k)$  which, however, contributes only at very small wavenumber  $k \lesssim 0.001 \text{ h}^{-1} \text{ Mpc}$ .

All this suggests that (28) is the dominant contribution to the non-Gaussian bias in the wavenumber range  $0.001 \lesssim k \lesssim 0.1 \text{ hMpc}^{-1}$ .

Finally, a non-Gaussian, scale-dependent bias correction can also arise in the local, deterministic bias ansatz  $\delta_h(\mathbf{x}) = b_1\delta(\mathbf{x}) + b_2\delta(\mathbf{x})^2/2 + \dots$  [169] if the initial density field is non-Gaussian. Here,  $b_N$  is the  $N$ th-order bias parameters (here again, the first-order bias is  $b_1 \equiv 1 + b_L$ ). In this approach, the correction is induced by the correlation  $b_1 b_2 \langle \delta(\mathbf{x}_1)\delta^2(\mathbf{x}_2) \rangle$  between the linear and quadratic term in the galaxy biasing relation (which is in fact a collapsed or squeezed 3-point function) and thus reads [67, 68]

$$\Delta b_\kappa(k, f_{\text{NL}}^{\text{loc}}) = 2f_{\text{NL}}^{\text{loc}} b_2 \sigma_R^2 \mathcal{M}_R(k, 0)^{-1}. \quad (40)$$

Even though  $b_2 \sigma_R^2 \approx b_L \delta_c$  in the high-threshold limit  $\nu \gg 1$ ,  $b_2 \sigma_R^2$  behaves very differently than  $b_L \delta_c$  for moderate peak height because  $b_2$  is proportional to the second derivative of the mass function  $n(\nu)$ . So far however, (28) appears to describe reasonably well the numerical results for a wide range of halo bias.

**4.3.2. Comparison with Simulations.** In order to fully exploit the potential of forthcoming large-scale surveys, a number of studies have tested the theoretical prediction against the outcome of large numerical simulations [11, 41–43, 71, 109].

At the lowest order, there are two additional albeit relatively smaller corrections to the Gaussian bias which arise from the dependence of both the halo number density  $n(M, z)$  and the matter power spectrum  $P_\delta(k, z)$  on primordial NG [41]. Firstly, assuming the peak-background split holds, the change in the mean number density of halos induces a scale-independent offset which we denote  $\Delta b_1(f_{\text{NL}}^{\text{loc}})$ . In terms of the non-Gaussian fractional correction  $R(\nu, f_{\text{NL}}^{\text{loc}})$  to the mass function, this contribution is

$$\Delta b_1(f_{\text{NL}}^{\text{loc}}) = -\frac{1}{\sigma} \frac{\partial}{\partial \nu} \ln[R(\nu, f_{\text{NL}}^{\text{loc}})]. \quad (41)$$

It is worth noticing that  $\Delta b_1(f_{\text{NL}}^{\text{loc}})$  has a sign opposite to that of  $f_{\text{NL}}^{\text{loc}}$ , because the bias decreases when the mass function goes up. In practice, the approximation (25), which matches well the SO data for  $\nu \lesssim 4$ , can be used for moderate values of the peak height. For FOF halos with linking length  $b = 0.2$ , one should make the replacement  $\delta_c \rightarrow \delta_c \sqrt{q}$  with  $q \approx 0.75$  in the calculation of the scale-independent offset. It is sensible to evaluate  $\Delta b_1(f_{\text{NL}}^{\text{loc}})$  at a mass scale  $\langle M \rangle$  equal to the average halo mass of the sample. Secondly, we also need to account for the change in the matter power spectrum (see Figure 2 in Section 3). Summarizing, local non-Gaussianity adds a correction  $\Delta b(k, f_{\text{NL}}^{\text{loc}})$  to the bias  $b(k)$  of dark matter halos that reads [41]

$$\Delta b(k, f_{\text{NL}}^{\text{loc}}) = \Delta b_\kappa(k, f_{\text{NL}}^{\text{loc}}) + \Delta b_1(f_{\text{NL}}^{\text{loc}}) + b_1(M)\beta_\delta(k, f_{\text{NL}}^{\text{loc}}) \quad (42)$$

at first order in  $f_{\text{NL}}^{\text{loc}}$ . As can be seen in Figure 7, the inclusion of these extra terms substantially improves the comparison between the theory and the simulations. Considering only

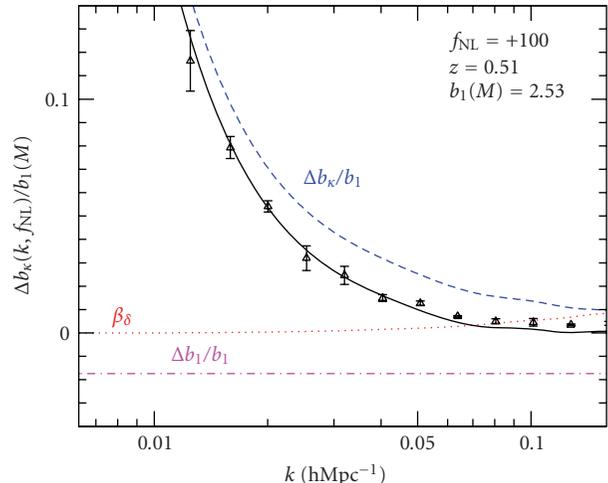


FIGURE 7: Non-Gaussian bias correction (filled symbols) for halos of mass  $M > 2 \times 10^{13} M_\odot/h$  extracted at  $z = 0.5$  from simulations of the local  $f_{\text{NL}}^{\text{loc}}$  model. The solid curve represents the theoretical model (42). The dashed, dotted-dashed-dotted, and dotted curves show the scale-dependent bias  $\Delta b_\kappa$ , the scale-independent offset  $\Delta b_1$  and the contribution from the matter power spectrum  $b_1(M)\beta_\delta$  that arise at first order in  $f_{\text{NL}}^{\text{loc}}$  (see Figure 2). The error bars indicate the scatter among 8 realizations of  $1024^3$  simulations with box size  $L = 1600 \text{ h}^{-1} \text{ Mpc}$ .

the scale-dependent shift  $\Delta b_\kappa$  leads to an apparent suppression of the effect in simulations relative to the theory. Including the scale-independent offset  $\Delta b_1$  considerably improves the agreement at wavenumbers  $k \lesssim 0.05 \text{ hMpc}^{-1}$ . Finally, adding the scale-dependent term  $b_1(M)\beta_\delta$  further adjusts the match at small scale  $k \gtrsim 0.05 \text{ hMpc}^{-1}$  by making the non-Gaussian bias shift less negative. Along these lines, [71] find that the inclusion of  $\Delta b_1$  to the bias also improves the agreement with measurements of  $\Delta b(k, f_{\text{NL}}^{\text{loc}})$  obtained for FOF halos, and show that taking into account second- and higher-order corrections could extend the validity of the theory up to scales  $k \sim 0.1 - 0.3 \text{ hMpc}^{-1}$ .

The non-Gaussian bias correction can be measured in the cross- and autopower spectrum of dark matter halos,  $P_{h\delta}(k)$  and  $P_h(k)$ . To compute these quantities, dark matter particles and halo centers are interpolated onto a regular cubical mesh. The resulting dark matter and halo fluctuation fields,  $\delta_m(\mathbf{k})$  and  $\delta_h(\mathbf{k})$ , are then Fourier transformed to yield the matter-matter, halo-matter and halo-halo power spectra  $P_\delta(k)$ ,  $P_{h\delta}(k)$  and  $P_h(k)$ , respectively.  $P_h(k)$  is then corrected for the shot noise, which is assumed to be  $1/\bar{n}_h$  if dark matter halos are a Poisson sampling of some continuous field. This discreteness correction is negligible for  $P_\delta(k)$  due to the large number of dark matter particles. On linear scales ( $k \lesssim 0.01 \text{ hMpc}^{-1}$ ), the halo bias  $b(\mathbf{k}) = \delta_h(\mathbf{k})/\delta_m(\mathbf{k})$  approaches the constant value  $b_1(M)$  which needs to be measured accurately as it controls the strength of the scale-dependent bias correction  $\Delta b_\kappa$ . In this respect, the ratio  $P_{h\delta}(k)/P_\delta(k)$  is a better proxy for the halo bias since it is less sensitive to shot-noise.

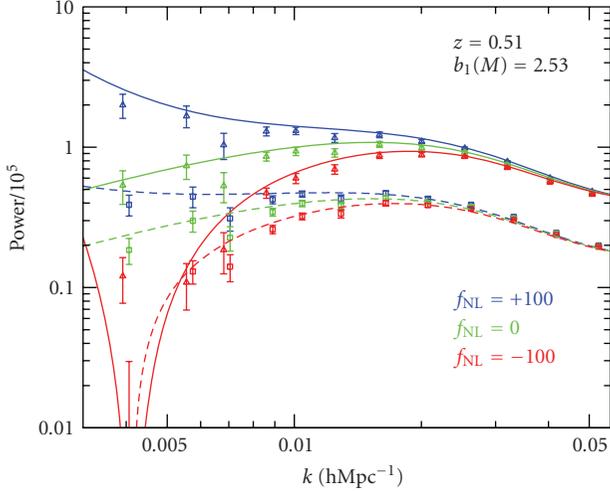


FIGURE 8: Halo-halo (solid curve) and halo-matter (dashed curve) power spectra  $P_h(k)$  and  $P_{h\delta}(k)$  measured in simulations of the Gaussian and  $f_{\text{NL}}^{\text{loc}} = \pm 100$  models for halos of mass  $M > 2 \times 10^{13} M_\odot/h$  at redshift  $z = 1$ . The error bars represent the scatter among 8 realizations. For  $f_{\text{NL}}^{\text{loc}} = -100$ , the crosspower spectrum is negative on scales  $k \lesssim 0.005 \text{ hMpc}^{-1}$ , in good agreement with the theoretical prediction.

Auto- and crosspower analyses may not agree with each other if the halos and dark matter do not trace each other on scale  $k \lesssim 0.01 \text{ hMpc}^{-1}$  where the non-Gaussian bias is large, that is, if there is stochasticity. Figure 8 shows  $P_{h\delta}(k)$  and  $P_h(k)$  averaged over 8 realizations of the models with  $f_{\text{NL}}^{\text{loc}} = 0, \pm 100$  [41]. The same Gaussian random seed field  $\phi$  was used in each set of runs so as to minimize the sampling variance. Measurements of the non-Gaussian bias correction obtained with the halo-halo or the halo-matter power spectrum are in a good agreement with each other, indicating that non-Gaussianity does not induce stochasticity and the predicted scaling (28) applies equally well for the auto- and cross-power spectrum. However, while a number of numerical studies of the  $f_{\text{NL}}^{\text{loc}}$  model have confirmed the scaling  $\Delta b_\kappa(k, f_{\text{NL}}^{\text{loc}}) \propto \mathcal{M}_R(k)^{-1}$  and the redshift dependence  $\propto D(z)^{-1}$  [11, 41, 43, 109], the exact amplitude of the non-Gaussian bias correction remains somewhat debatable. Reference [41] who use SO halos and [71] who use FOF halos find satisfactory agreement with the theory once the scale-independent offset  $\Delta b_1$  is included. By contrast, see [43], who use the same FOF halos as [71], argue that the scale-dependent piece  $\Delta b_\kappa$  requires, among others, a multiplicative correction of the form  $(1 - \beta_1 f_{\text{NL}}^{\text{loc}})$ , with  $\beta_1 \sim 4 \times 10^{-4} > 0$ . Similarly, [109] who also use FOF halos find that the theory is a good fit to the simulations only upon replacing  $b_L$  by  $qb_L$  in (33), with  $q \simeq 0.75$ . Part of the discrepancy may be probably due to the fact that the last two references do not include  $\Delta b_1$ , which leads to an apparent suppression of the effect (see Figure 7). Another possible source of discrepancy may be choice of the halo finder which, as seen in Figure 5, has an impact on the strength of the non-Gaussian correction to the mass function. This possibility is investigated in Figure 9, which shows the non-Gaussian bias correction obtained

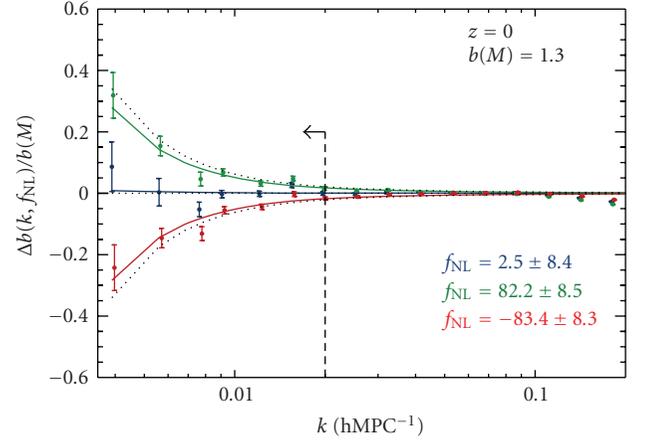


FIGURE 9: Fractional correction to the Gaussian halo bias in the  $f_{\text{NL}}^{\text{loc}} = \pm 100$  and Gaussian models. In contrast to Figure 8, halos were identified with a FOF finder of linking length  $b = 0.2$ . Only the wavemodes to the left of the vertical line were used to fit  $\Delta b_\kappa(k, f_{\text{NL}}^{\text{loc}})$ . For this low biased sample, the scale-independent correction is  $|\Delta b_1| \lesssim 0.003$  and can thus be ignored. The best-fit value of  $f_{\text{NL}}^{\text{loc}}$  and the corresponding  $1\sigma$  error is quoted for each model (Figure taken from [170]).

with FOF halos. For this low biased sample, the scale-independent correction is  $|\Delta b_1| \lesssim 0.003$  and can thus be neglected. The best-fit values of  $f_{\text{NL}}^{\text{loc}}$  are significantly below the input values of  $\pm 100$ , in agreement with the findings of [43, 109] (note, however, that this suppression is more consistent with  $\delta_c$  being rescaled by  $\sqrt{q}\delta_c \approx 0.86\delta_c$  and  $b_L$  being unchanged). This indicates that the choice of halo finder may also affect the magnitude of the scale-dependent non-Gaussian bias. Discrepancies have also been observed between the theoretical and measured non-Gaussian bias corrections in non-Gaussian models of the local cubic-order coupling  $g_{\text{NL}}^{\text{loc}}\phi^3$  [48]. Understanding all these results clearly requires a better modeling of halo clustering.

**4.3.3. Redshift-Space Distortions.** Peculiar velocities generate systematic differences between the spatial distribution of data in real and redshift space. These redshift-space distortions must be properly taken into account in order to extract  $f_{\text{NL}}^{\text{X}}$  from redshift surveys. On the linear scales of interest, the redshift-space power spectrum of biased tracers reads as [171, 172]

$$P_h^s(k, \mu) = [b_1^2 P_\delta(k) + 2b_1 f \mu^2 P_{\delta\theta}(k) + f^2 \mu^4 P_\theta(k)], \quad (43)$$

where  $P_{\delta\theta}$  and  $P_\theta$  are the density-velocity and velocity-divergence power spectra,  $\mu$  is the cosine of the angle between the wavemode  $\mathbf{k}$  and the line of sight and  $f$  is the logarithmic derivative of the growth factor. For  $P_\theta$ , the 1-loop correction due to primordial NG is identical to (17) provided  $F_2(\mathbf{k}_1, \mathbf{k}_2)$

is replaced by the kernel  $G_2(\mathbf{k}_1, \mathbf{k}_2) = 3/7 + \mu(k_1/k_2 + k_2/k_1)/2 + 4\mu^2/7$  describing the 2nd order evolution of the velocity divergence [62]. For  $P_{\delta\theta}$ , this correction is

$$\Delta P_{\delta\theta}^{\text{NG}}(k) = \int \frac{d^3\mathbf{q}}{(2\pi)^3} [F_2(\mathbf{q}, \mathbf{k} - \mathbf{q}) + G_2(\mathbf{q}, \mathbf{k} - \mathbf{q})] \times B_0(-\mathbf{k}, \mathbf{q}, \mathbf{k} - \mathbf{q}). \quad (44)$$

Again, causality implies that  $G_2(\mathbf{k}_1, \mathbf{k}_2)$  vanishes in the limit  $\mathbf{k}_1 = -\mathbf{k}_2$ . For unbiased tracers with  $b_1 = 1$ , the linear Kaiser relation is thus recovered at large scales  $k \lesssim 0.01 \text{ hMpc}^{-1}$  (this is consistent with the analysis of [173]). For biased tracers, we still expect the Kaiser formula to be valid, but the distortion parameter  $\beta$  should now be equal to  $\beta = f/(b_1 + \Delta b_\kappa)$ , where  $\Delta b_\kappa(k, f_{\text{NL}}^X)$  is the scale-dependent bias induced by the primordial non-Gaussianity.

**4.3.4. Mitigating Cosmic Variance and Shot-Noise.** Because of the finite number of large scale wavemodes accessible to a survey, any large-scale measurement of the power spectrum is limited by the cosmic (or sampling) variance caused by the random nature of the wavemodes. For discrete tracers such as galaxies, the shot noise is another source of error. Restricting ourselves to weak primordial NG, the relative error on the power spectrum  $P$  is  $\sigma_P/P \approx 1/\sqrt{N}(1 + \sigma_n^2/P)$ , where  $N$  is the number of independent modes measured and  $\sigma_n^2$  is the shot-noise [174]. Under the standard assumption of Poisson sampling,  $\sigma_n^2$  equals the inverse of the number density  $1/\bar{n}$  and causes a scale-independent enhancement of the power spectrum. The extent to which one can improve the observational limits on the nonlinear parameters  $f_{\text{NL}}^X$  will strongly depend on our ability to minimize the impact of these two sources of errors. By comparing differently biased tracers of the same surveyed volume [175, 176] and suitably weighting galaxies (by the mass of their host halo for instance) [177, 178], it should be possible to circumvent these problems and considerably improve the detection level.

Figure 9 illustrates how the impact of sampling variance on the measurement of  $f_{\text{NL}}^{\text{loc}}$  can be mitigated. Namely, the data points show the result of taking the ratio  $P_h(k, f_{\text{NL}}^{\text{loc}})/P_\delta(k, f_{\text{NL}}^{\text{loc}})$  for each set of runs with same Gaussian random seed field  $\phi$  before averaging over the realizations. This procedure is equivalent to the multitracers method advocated by [175]. Here,  $P_\delta$  can be thought as mimicking the power spectrum of a nearly unbiased tracer of the mass density field with high number density. Although, in practical applications, using the dark matter field works better [170], in real data  $P_\delta$  should be replaced by a tracer of the same surveyed volume different than the one used to compute  $P_h$ . Figure 9 also shows that upon taking out most of the cosmic variance, there is some residual noise caused by the discrete nature of the dark matter halos. As shown recently [178] however, weighting the halos according to their mass can dramatically reduce the shot noise relative to the Poisson expectation, at least when compared against the dark matter. Applying such a weighting may thus significantly improve the error on the nonlinear parameter  $f_{\text{NL}}^{\text{loc}}$ , but this should be explored in realistic simulations

of galaxies, especially because the halo mass  $M$  may not be easily measurable from real data [170]. This approach undoubtedly deserves further attention as it has the potential to substantially improve the extraction of the primordial non-Gaussian signal from galaxy surveys.

To conclude this section, it is worth noting that, while the PDF of power values  $P(\mathbf{k})$  has little discriminatory power (for large surveyed volume, it converges towards the Rayleigh distribution as a consequence of the central limit theorem) [179], the covariance of power spectrum measurements (which is sensitive to the selection function, but also to correlations among the phase of the Fourier modes) may provide quantitative limits on certain type of non-Gaussian models [174, 180].

**4.4. Galaxy Bispectrum and Higher-Order Statistics.** Higher statistics of biased tracers, such as the galaxy bispectrum, are of great interest as they are much more sensitive to the shape of the primordial 3-point function than the power spectrum [10, 42, 68, 181, 182]. Therefore, they could break some of the degeneracies affecting the non-Gaussian halo bias (e. g., the leading order scale-dependent correction to the Gaussian bias induced by the local quadratic and cubic coupling are fully degenerated [48]).

**4.4.1. Normalized Cumulants of the Galaxy Distribution.** The skewness of the galaxy count probability distribution function could provide constraints on the amount of non-Gaussianity in the initial conditions. As discussed in Section 3, however, it is difficult to disentangle the primordial and gravitational causes of skewness in low redshift data unless the initial density field is strongly non-Gaussian. The first analyzes of galaxy catalogs in terms of count-in-cells densities all reached the conclusion that the skewness (and higher-order moments) of the observed galaxy count PDF is consistent with the value predicted by gravitational instability of initially Gaussian fluctuations [50, 57, 60, 183–185]. Back then however, most of the galaxy samples available were not large enough to accurately determine the  $S_j$  at large-scales [186]. Despite the two orders of magnitude increase in surveyed volume, these measurements are still sensitive to cosmic variance, that is, to the presence of massive superclusters or large voids. Nevertheless, the best estimates of the first normalized cumulants  $S_j$  of the galaxy PDF strongly suggest that high order galaxy correlation functions indeed follow the hierarchical scaling predicted by the gravitational clustering of Gaussian ICs [187]. There is no evidence for strong non-Gaussianity in the initial density field as might be seeded by cosmic strings or textures [188].

The genus statistics of constant density surfaces through the galaxy distribution measures the relative abundance of low and high density regions as a function of the smoothing scale  $R$  and, therefore, could also be used as a diagnostic tool for primordial non-Gaussianity. For a Gaussian random field, the genus curve (i.e., the genus number as a function of the density contrast) is symmetric about  $\delta_R = 0$  regardless the value of  $R$ . Primordial NG and nonlinear gravitational

evolution can disrupt this symmetry [189]. The effect of non-Gaussian ICs on the topology of the galaxy distribution has been explored in a number of papers [35, 190–193]. For large values of  $R$  and realistic amount of primordial NG, the genus statistics can also be expanded in a series whose coefficients are the normalized cumulants  $S_J$  of the smoothed galaxy density field. In other words, the genus statistics essentially provides another measure of the (large-scale) cumulants. So far, measurements from galaxy data are broadly consistent with Gaussian initial conditions [194, 195].

*4.4.2. Galaxy Bispectrum.* Most of the scale-dependence of the primordial  $n$ -point functions is integrated out in the normalized cumulants, which makes them weakly sensitive to primordial NG. However, while the effect of non-Gaussian initial conditions, galaxy bias, gravitational instabilities and so forth, are strongly degenerated in the  $S_J$ , they imprint distinct signatures in the galaxy bispectrum  $B_g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ , an accurate measurement of which could thus constrain the shape of the primordial 3-point function.

In the original derivation of [181], the large-scale (unfiltered) galaxy bispectrum in the  $f_{\text{NL}}^{\text{loc}}$  model is given by

$$\begin{aligned} B_g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) &= b_1^3 B_0(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \\ &+ b_1^2 b_2 [P_0(k_1)P_0(k_2) + (\text{cyc.})] \\ &+ 2b_1^3 [F_2(\mathbf{k}_1, \mathbf{k}_2)P_0(k_1)P_0(k_2) + (\text{cyc.})]. \end{aligned} \quad (45)$$

Again,  $b_1$  and  $b_2$  are the first- and second-order bias parameters that describe the galaxy biasing relation assumed local and deterministic [169]. The first term in the right-hand side is the primordial contribution which, for equilateral configurations and in the  $f_{\text{NL}}^{\text{loc}}$  model, scales as  $\mathcal{M}_R(k, z)^{-1}$  like in the matter bispectrum, (18). The two last terms are the contribution from nonlinear bias and the tree-level correction from gravitational instabilities, respectively. They have the smallest signal in squeezed configurations.

As recognized by [68, 182], (45) misses an important term that may significantly enhance the sensitivity of the galaxy bispectrum to non-Gaussian initial conditions. This contribution is sourced by the trispectrum  $T_R(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$  of the smoothed mass density field

$$\frac{1}{2} b_1^2 b_2 \int \frac{d^3 q}{(2\pi)^3} T_R(\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}, \mathbf{k}_3 - \mathbf{q}) + (2 \text{ perms.}). \quad (46)$$

At large-scale, this simplifies to the sum of the linearly evolved primordial trispectrum  $T_0(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$  and a coupling between the primordial bispectrum  $B_0(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$  (linear in  $f_{\text{NL}}^{\text{X}}$ ) and the second order PT corrections (through the kernel  $F_2(\mathbf{k}_1, \mathbf{k}_2)$ ). In the case of local non-Gaussianity and for equilateral configurations, the first piece proportional to  $T_0$  scales as  $(f_{\text{NL}}^{\text{loc}})^2 k^{-4}$  times the Gaussian tree-level prediction, with the same redshift dependence. Hence, it is similar to the second order correction  $(f_{\text{NL}}^{\text{loc}})^2 \mathcal{M}_R^{-2} P_R(k)$  that appears in the halo power spectrum (see (39)). The second piece linear in  $f_{\text{NL}}^{\text{X}}$  generates a signal at large-scales

for essentially all triangle shapes in the local model as well as in the case of equilateral NG. This second contribution is maximized in the squeezed limit (where it is one order of magnitude larger than the result obtained by [181]) which helps disentangling it from the Gaussian terms. Note that a strong dependence on triangle shape is also present in other NG scenarios such as the  $\chi^2$  model [62].

This newly derived contributions are claimed to lead to more than one order of magnitude improvement in certain limits [182], but it is not yet clear whether these gains can be fully realized with upcoming galaxy surveys. To accurately predict the constraints that could be achieved with future measurements of the galaxy bispectrum, a comparison of these predictions with the halo bispectrum extracted from numerical simulations is highly desirable. To date, the only numerical study [42] has measured the halo bispectrum for some isosceles triangles ( $k_1 = k_2$ ). While the shape dependence is in reasonable agreement with the theory, the observed  $k$ -dependence appears to depart from the predicted scaling.

*4.5. Intergalactic Medium and the Ly $\alpha$  Forest.* Primordial non-Gaussianity also affects the intergalactic medium (IGM) as a positive  $f_{\text{NL}}^{\text{X}}$  enhances the formation of high-mass halos at early times and, therefore, accelerates reionization [196–198]. At lower redshift, small box hydrodynamical simulations of the Ly $\alpha$  forest indicate that non-Gaussian initial conditions could leave a detectable signature in the Ly $\alpha$  flux PDF, power spectrum and bispectrum [199]. However, while differences appear quite pronounced in the high transmissivity tail of the flux PDF (i.e., in underdense regions), the Ly $\alpha$  1D flux power spectrum seems little affected. Given the small box size of these hydrodynamical simulations, it is worth exploring the effect in large N-body cosmological simulations using a semianalytic modeling of the Ly $\alpha$  forest [200], even though such an approach only provides a very crude approximation to the temperature-density diagram of the IGM in hydrodynamical simulations. Figure 10 shows the imprint of local type NG on the Ly $\alpha$  3D flux power spectrum (which is not affected by projection effects) extracted at  $z = 2$  from a series of large simulations. The Ly $\alpha$  transmitted flux is calculated in the Gunn-Peterson approximation [201]. A clear signature similar to the non-Gaussian halo bias can be seen. As expected, it is of opposite sign since the Ly $\alpha$  forest is anti-biased relative to the mass density field (overdensities are mapped onto relatively low flux transmission).

To estimate the strength of the signal (see [200] for the details), one can assume that the (real space) optical depth  $\tau(\mathbf{x})$  to Ly $\alpha$  absorption at comoving position  $\mathbf{x}$  is approximately [202]

$$\tau(\mathbf{x}) = \bar{\tau} [1 + \delta_g(\mathbf{x})]^\alpha, \quad (47)$$

where  $\delta_g$  is the gas density,  $\bar{\tau} \sim 1$  is the optical depth at mean gas density and  $\alpha \sim 1 - 2$  is some parameter that depends on the exact thermal history of the low density IGM. The above relation holds for the moderate overdensities  $\delta_g \lesssim 10$  that are responsible for most of the Ly $\alpha$  absorption features. To

relate the gas density to the smoothed linear density field, we could make the simple ansatz  $\delta_g \equiv \delta_R$  [203]. In this linear approximation, however, the large-scale bias  $b_F$  of the Ly $\alpha$  flux density field is much larger than that measured in detailed numerical simulations (e.g.,  $b_F^2 \simeq 0.017$  at  $z = 3$  [204]). Therefore, one may want to consider the lognormal mapping [205, 206]

$$1 + \delta_g = \exp\left(\delta_R - \frac{\sigma_R^2}{2}\right), \quad (48)$$

to better capture nonlinearities in the gas density field. Expanding  $\exp(\delta_R)$  at second order and noticing that in the presence of weak non-Gaussianity, the joint PDF  $P(\delta_R(\mathbf{x}_1), \delta_R(\mathbf{x}_2))$  can generically be expanded into an Edgeworth series where the primordial 3-point function is the dominant correction, it is straightforward to compute the Ly $\alpha$  3D flux power spectrum for nonzero  $f_{\text{NL}}^{\text{X}}$ . Upon a Fourier transformation, we arrive at

$$\frac{P_{\text{F}}(k, f_{\text{NL}}^{\text{X}})}{P_{\text{F}}(k, 0)} = 1 - 4g_{\text{F}}\sigma_{\text{R}}^2\mathcal{M}_{\text{R}}(k, z)^{-1}\mathcal{F}(k, f_{\text{NL}}^{\text{X}}). \quad (49)$$

where  $g_{\text{F}}$  is some auxiliary function of  $(\bar{\tau}, \alpha, \sigma_{\text{R}})$ . This result is valid for any model of primordial NG characterized by an initial bispectrum. In the  $f_{\text{NL}}^{\text{loc}}$  model, the large-scale non-Gaussian Ly $\alpha$  bias scales as  $\Delta b_{\text{F}}(k, f_{\text{NL}}^{\text{X}}) \approx -2g_{\text{F}}\sigma_{\text{R}}^2\mathcal{M}_{\text{R}}(k, z)^{-1} \propto k^{-2}T(k)^{-1}$  like the non-Gaussian halo bias. Assuming that  $\bar{\tau} = 0.7$ ,  $\sigma_{\text{R}} = 1.8$  and  $\alpha = 1.65$  yields a mean flux  $\bar{F} \approx 0.8$  and a ratio  $P_{\text{F}}(k = 0.01, f_{\text{NL}}^{\text{X}})/P_{\text{F}}(k = 0.01, 0) \approx 1 \mp 0.13$  for  $f_{\text{NL}}^{\text{loc}} = \pm 100$  comparable in magnitude to that seen in Figure 10. A detection of this effect, although challenging in particular because of continuum uncertainties, could be feasible with future data sets. Summarizing, the Ly $\alpha$  should provide interesting information on the non-Gaussian signal over a range of scale and redshift not easily accessible to galaxy and CMB observations [199, 200].

## 5. Current Limits and Prospects

As the importance of primordial non-Gaussianity relative to the non-Gaussianity induced by gravitational clustering and galaxy bias increases towards high redshift, the optimal strategy to constrain the nonlinear coupling parameter(s) with LSS is to use large-scale, high-redshift observations [32].

*5.1. Existing Constraints on Primordial NG.* The non-Gaussian halo bias presently is the only LSS method that provides a robust limit on the magnitude of a primordial 3-point function of the local shape. It is a broadband effect that can be easily measured with photometric redshifts. The authors of [156] have applied (28) to constrain the value of  $f_{\text{NL}}^{\text{loc}}$  using a compilation of large-scale clustering data. Their constraint arise mostly from the QSO sample at median redshift  $z = 1.8$ , which covers a large comoving volume and is highly biased,  $b_1 = 2.7$ . They obtain

$$-29 < f_{\text{NL}}^{\text{loc}} < +69, \quad (50)$$

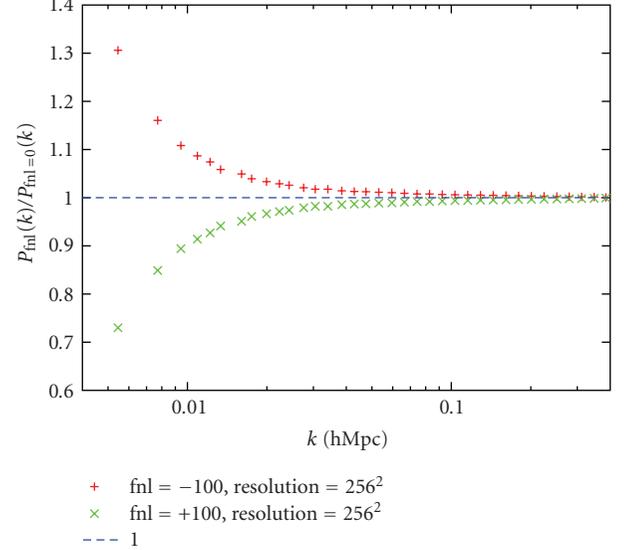


FIGURE 10: Ratio between the  $z = 2$  Ly $\alpha$  flux power spectrum extracted from simulations of Gaussian and non-Gaussian initial conditions. The mean transmission is set to  $\bar{F} = 0.8$  and the power-law exponent  $\alpha = 1.65$  (see text).

at 95% confidence level. These limits are competitive with those from CMB measurements,  $-10 < f_{\text{NL}}^{\text{loc}} < +74$  [207]. It is straightforward to translate this 2- $\sigma$  limit into a constraint on the cubic order coupling  $g_{\text{NL}}^{\text{loc}}$  since the non-Gaussian scale-dependent bias  $\Delta b_{\text{F}}(k, g_{\text{NL}}^{\text{loc}})$  has the same functional form as  $\Delta b_{\text{F}}(k, f_{\text{NL}}^{\text{loc}})$  [48]. Assuming  $f_{\text{NL}}^{\text{loc}} = 0$ , one obtains

$$-3.5 \times 10^5 < g_{\text{NL}}^{\text{loc}} < +8.2 \times 10^5. \quad (51)$$

These limits are comparable with those inferred from the analysis of CMB data.

Measurements of the galaxy bispectrum in several redshift catalogs have shown evidence for a configuration shape dependence in agreement with that predicted from gravitational instability, ruling out  $\chi^2$  initial conditions at the 95% C.L. [208, 209]. Recent analyses of the SDSS LRGs catalogue indicate that the shape dependence of the reduced 3-point correlation  $Q_3 \sim \xi_3/(\xi_2)^2$  is also consistent with Gaussian ICs [210], although a primordial (hierarchical) non-Gaussian contribution in the range  $Q_3 \sim 0.5 - 3$  cannot be ruled out [211]. Other LSS probes of primordial non-Gaussianity, such as the abundance of massive clusters, are still too affected by systematics to furnish tight constraints on the shape and magnitude of a primordial 3-point function. Still, the observation of a handful of unexpectedly massive high-redshift clusters has been interpreted as evidence of a substantial degree of primordial NG [212–214].

*5.2. Future Prospects.* Improving the current limits will further constrain the physical mechanisms for the generation of cosmological perturbations.

The non-Gaussian halo bias also leaves a signature in cross-correlation statistics of weak cosmic shear (galaxy-galaxy and galaxy-CMB) [215, 216] and in the integrated Sachs-Wolfe (ISW) effect [156, 160, 217]. Measurements of the lensing bispectrum could also constrain a number of non-Gaussian models [218]. However, galaxy clustering will undoubtedly offer the most promising LSS diagnostic of primordial non-Gaussianity. The detectability of a local primordial bispectrum has been assessed in a series of papers. It is expected that future all-sky galaxy surveys will achieve constraints of the order of  $\Delta f_{\text{NL}}^{\text{loc}} \sim 1$  assuming that all systematics are reasonably under control [47, 106, 156, 160, 217, 219–221]. Realistic models of cubic type non-Gaussianity [48], modifications of the initial vacuum state or horizon-scale GR corrections [163] should also be tested with future measurement of the galaxy power spectrum.

Upcoming observations of high redshift clusters will provide increased leverage on measurement of primordial non-Gaussianity with abundances and possibly put limits on any nonlinear parameter  $f_{\text{NL}}^{\text{X}}$  at the level of a few tens [139]. Combining the information provided by the evolution of the mass function and power spectrum of galaxy clusters should yield constraints with a precision  $\Delta f_{\text{NL}}^{\text{loc}} \sim 10$  for a wide field survey covering half of the sky [214]. Alternatively, using the full covariance of cluster counts (which is sensitive to the non-Gaussian halo bias) can furnish constraints of  $\Delta f_{\text{NL}}^{\text{loc}} \sim 1 - 5$  for a Dark Energy Survey-type experiment [222, 223].

As emphasized in “Section 4” however, the exact magnitude of the non-Gaussian bias is still uncertain partly due to the freedom at the definition of the halo mass and the uncertainty in the correspondence between simulated quantities and observables. Understanding this type of systematics will be crucial to set reliable constraints on a primordial non-Gaussian component. To fully exploit the potential of future galaxy surveys, it will also be essential to extend the theoretical and numerical analyses to other bispectrum shapes than the local template used so far. Ultimately, the gain that can be achieved will critically depend on our ability to minimize the impact of sampling variance and shot-noise. In this regards, multi-tracers methods combined with optimal weighting schemes should deserve further attention as they hold the promise to become the most accurate method to extract the primordial non-Gaussian signal from galaxy surveys [175–178].

## Acknowledgments

The authors give their special thanks to Nico Hamaus and Shirley Ho for sharing with them material prior to publication, and Emiliano Sefusatti for providing us with the data shown in Figure 3. The authors would also like to thank Martin Crocce, Christopher Hirata, Ilian Iliev, Tsz Yan Lam, Patrick McDonald, Nikhil Padmanabhan, and Anze Slosar for collaboration on these issues, and Tobias Baldauf for comments and for a careful reading of this paper. This paper was supported by the Swiss National Foundation (Contract

no. 200021-116696/1) and made extensive use of the NASA Astrophysics Data System and arXiv.org preprint server.

## References

- [1] V. F. Mukhanov and G. V. Chibisov, “Quantum fluctuations and a nonsingular universe,” *Soviet Journal of Experimental and Theoretical Physics Letters*, vol. 33, pp. 532–535, 1981.
- [2] A. A. Starobinsky, “Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations,” *Physics Letters B*, vol. 117, pp. 175–178, 1982.
- [3] S. W. Hawking, “The development of irregularities in a single bubble inflationary universe,” *Physics Letters B*, vol. 115, no. 4, pp. 295–297, 1982.
- [4] A. H. Guth and S.-Y. Pi, “Fluctuations in the new inflationary universe,” *Physical Review Letters*, vol. 49, no. 15, pp. 1110–1113, 1982.
- [5] T. J. Allen, B. Grinstein, and M. B. Wise, “Non-Gaussian density perturbations in inflationary cosmologies,” *Physics Letters B*, vol. 197, no. 1-2, pp. 66–70, 1987.
- [6] T. Falk, R. Rangarajan, and M. Srednicki, “Dependence of density perturbations on the coupling constant in a simple model of inflation,” *Physical Review D*, vol. 46, no. 10, pp. 4232–4234, 1992.
- [7] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, “The three-point correlation function of the cosmic microwave background in inflationary models,” *Astrophysical Journal*, vol. 430, no. 2, pp. 447–457, 1994.
- [8] E. Komatsu, J. Dunkley, M. R. Nolte et al., “Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal*, vol. 180, no. 2, pp. 330–376, 2009.
- [9] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, “Large-scale structure, the cosmic microwave background and primordial non-Gaussianity,” *Monthly Notices of the Royal Astronomical Society*, vol. 313, no. 1, pp. 141–147, 2000.
- [10] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, “Probing primordial non-Gaussianity with large-scale structure,” *Physical Review D*, vol. 69, no. 10, Article ID 103513, 2004.
- [11] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, “Imprints of primordial non-Gaussianities on large-scale structure: scale-dependent bias and abundance of virialized objects,” *Physical Review D*, vol. 77, no. 12, article 123514, 2008.
- [12] D. S. Salopek and J. R. Bond, “Nonlinear evolution of long-wavelength metric fluctuations in inflationary models,” *Physical Review D*, vol. 42, no. 12, pp. 3936–3962, 1990.
- [13] J. Maldacena, “Non-Gaussian features of primordial fluctuations in single field inflationary models,” *Journal of High Energy Physics*, vol. 2003, no. 5, article 013, 2003.
- [14] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant second-order perturbations and non-Gaussianity from inflation,” *Nuclear Physics B*, vol. 667, no. 1-2, pp. 119–148, 2003.
- [15] P. Creminelli and M. Zaldarriaga, “A single-field consistency relation for the three-point function,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2004, no. 10, article 006, 2004.
- [16] A. Linde and V. Mukhanov, “Non-Gaussian isocurvature perturbations from inflation,” *Physical Review D*, vol. 56, no. 2, pp. R535–R539, 1997.
- [17] D. H. Lyth, C. Ungarelli, and D. Wands, “Primordial density perturbation in the curvaton scenario,” *Physical Review D*, vol. 67, no. 2, Article ID 023503, 2003.

- [18] T. Falk, R. Rangarajan, and M. Srednicki, “The angular dependence of the three-point correlation function of the cosmic microwave background radiation as predicted by inflationary cosmologies,” *Astrophysical Journal Letters*, vol. 403, pp. L1–L3, 1993.
- [19] J. Lesgourgues, D. Polarski, and A. A. Starobinsky, “Quantum-to-classical transition of cosmological perturbations for non-vacuum initial states,” *Nuclear Physics B*, vol. 497, no. 1-2, pp. 479–508, 1997.
- [20] G. Dvali, A. Gruzinov, and M. Zaldarriaga, “Cosmological perturbations from inhomogeneous reheating, freeze-out, and mass domination,” *Physical Review D*, vol. 69, no. 8, Article ID 083505, 2004.
- [21] J.-L. Lehners and P. J. Steinhardt, “Non-Gaussian density fluctuations from entropically generated curvature perturbations in ekpyrotic models,” *Physical Review D*, vol. 77, no. 6, Article ID 063533, 2008.
- [22] D. Babich, P. Creminelli, and M. Zaldarriaga, “The shape of non-Gaussianities,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2004, no. 8, article 009, 2004.
- [23] J. R. Fergusson and E. P. S. Shellard, “Shape of primordial non-Gaussianity and the CMB bispectrum,” *Physical Review D*, vol. 80, no. 4, Article ID 043510, 2009.
- [24] E. Komatsu and D. N. Spergel, “Acoustic signatures in the primary microwave background bispectrum,” *Physical Review D*, vol. 63, no. 6, Article ID 063002, 2001.
- [25] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, “Limits on non-Gaussianities from WMAP data,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2006, no. 5, article 004, 2006.
- [26] P. D. Meerburg, J. P. van der Schaar, and P. S. Corasaniti, “Signatures of initial state modifications on bispectrum statistics,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2009, no. 5, article 018, 2009.
- [27] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, “Large-scale structure of the Universe and cosmological perturbation theory,” *Physics Report*, vol. 367, no. 1–3, pp. 1–248, 2002.
- [28] J. E. Gunn and J. R. Gott III, “On the infall of matter into clusters of galaxies and some effects on their evolution,” *Astrophysical Journal*, vol. 176, article 1, 1972.
- [29] J. R. Bond and S. T. Myers, “The peak-patch picture of cosmic catalogs. I. Algorithms,” *Astrophysical Journal*, vol. 103, no. 1, pp. 1–39, 1996.
- [30] R. Durrer, “Topological defects in cosmology,” *New Astronomy Reviews*, vol. 43, no. 2–4, pp. 111–156, 1999.
- [31] A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects*, Cambridge University Press, Cambridge, UK, 2000.
- [32] L. Verde, R. Jimenez, M. Kamionkowski, and S. Matarrese, “Tests for primordial non-Gaussianity,” *Monthly Notices of the Royal Astronomical Society*, vol. 325, no. 1, pp. 412–418, 2001.
- [33] P. J. E. Peebles, “An isocurvature cold dark matter cosmogony. II. Observational tests,” *Astrophysical Journal*, vol. 510, no. 2, pp. 531–540, 1999.
- [34] L. Moscardini, S. Matarrese, F. Lucchin, and A. Messina, “Non-Gaussian initial conditions in cosmological N-body simulations. II—cold dark matter models,” *Monthly Notices of the Royal Astronomical Society*, vol. 248, pp. 424–438, 1991.
- [35] D. H. Weinberg and S. Cole, “Non-Gaussian fluctuations and the statistics of galaxy clustering,” *Monthly Notices of the Royal Astronomical Society*, vol. 259, pp. 652–694, 1992.
- [36] P. Coles, L. Moscardini, F. Lucchin, S. Matarrese, and A. Messina, “Skewness as a test of non-Gaussian primordial density fluctuations,” *Monthly Notices of the Royal Astronomical Society*, vol. 264, p. 749, 1993.
- [37] E. Gaztanaga and P. Maehoenen, “Large-scale clustering from non-Gaussian texture models,” *Astrophysical Journal Letters*, vol. 462, no. 1, article L1, 1996.
- [38] M. White, “Higher order moments of the density field in a parametrized sequence of non-Gaussian theories,” *Monthly Notices of the Royal Astronomical Society*, vol. 310, no. 2, pp. 511–516, 1999.
- [39] H. Mathis, J. Silk, L. M. Griffiths, and M. Kunz, “Constraining cosmic microwave background consistent primordial voids with cluster evolution,” *Monthly Notices of the Royal Astronomical Society*, vol. 350, no. 1, pp. 287–297, 2004.
- [40] M. Grossi, E. Branchini, K. Dolag, S. Matarrese, and L. Moscardini, “The mass density field in simulated non-Gaussian scenarios,” *Monthly Notices of the Royal Astronomical Society*, vol. 390, no. 1, pp. 438–446, 2008.
- [41] V. Desjacques, U. Seljak, and I. T. Iliev, “Scale-dependent bias induced by local non-Gaussianity: a comparison to N-body simulations,” *Monthly Notices of the Royal Astronomical Society*, vol. 396, no. 1, pp. 85–96, 2009.
- [42] T. Nishimichi, A. Taruya, K. Koyama, and C. Sabiu, “Scale dependence of halo bispectrum from non-Gaussian initial conditions in cosmological N-body simulations,” *JCAP*, no. 07, article no. 002, 2010.
- [43] A. Pillepich, C. Porciani, and O. Hahn, “Halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions,” *Monthly Notices of the Royal Astronomical Society*, vol. 402, pp. 191–206, 2010.
- [44] R. Vio, P. Andreani, and W. Wamstecker, “Numerical simulation of non-Gaussian random fields with prescribed correlation structure,” *Publications of the Astronomical Society of the Pacific*, vol. 113, no. 786, pp. 1009–1020, 2001.
- [45] U. Seljak and M. Zaldarriaga, “A line-of-sight integration approach to cosmic microwave background anisotropies,” *Astrophysical Journal*, vol. 469, no. 2, pp. 437–444, 1996.
- [46] A. Lewis, A. Challinor, and A. Lasenby, “Efficient computation of cosmic microwave background anisotropies in closed Friedmann-Robertson-Walker models,” *Astrophysical Journal*, vol. 538, no. 2, pp. 473–476, 2000.
- [47] P. McDonald, “Primordial non-Gaussianity: large-scale structure signature in the perturbative bias model,” *Physical Review D*, vol. 78, no. 12, Article ID 123519, 2008.
- [48] V. Desjacques and U. Seljak, “Signature of primordial non-Gaussianity of  $\phi^3$  type in the mass function and bias of dark matter haloes,” *Physical Review D*, vol. 81, no. 2, Article ID 123519, 2010.
- [49] R. Juszkiewicz, F. R. Bouchet, and S. Colombi, “Skewness induced by gravity,” *Astrophysical Journal*, vol. 412, no. 1, pp. L9–L12, 1993.
- [50] C. M. Baugh, E. Gaztanaga, and G. Efstathiou, “A comparison of the evolution of density fields in perturbation theory and numerical simulations—II. Counts-in-cells analysis,” *Monthly Notices of the Royal Astronomical Society*, vol. 274, pp. 1049–1070, 1995.
- [51] R. Scoccimarro, “Transients from initial conditions: a perturbative analysis,” *Monthly Notices of the Royal Astronomical Society*, vol. 299, no. 4, pp. 1097–1118, 1998.
- [52] M. Crocce, S. Pueblas, and R. Scoccimarro, “Transients from initial conditions in cosmological simulations,” *Monthly Notices of the Royal Astronomical Society*, vol. 373, no. 1, pp. 369–381, 2006.

- [53] X. Luo and D. N. Schramm, “Kurtosis, skewness, and non-Gaussian cosmological density perturbations,” *Astrophysical Journal*, vol. 408, no. 1, pp. 33–42, 1993.
- [54] E. L. Lokas, R. Juszkiewicz, D. H. Weinberg, and F. R. Bouchet, “Kurtosis of large-scale cosmic fields,” *Monthly Notices of the Royal Astronomical Society*, vol. 274, pp. 730–744, 1995.
- [55] P. J. E. Peebles, *The Large-Scale Structure of the Universe*, Princeton University Press, Princeton, NJ, USA, 1980.
- [56] J. N. Fry, “The Galaxy correlation hierarchy in perturbation-theory,” *Astrophys Journal*, vol. 279, pp. 499–510, 1984.
- [57] P. Coles and C. S. Frenk, “Skewness and large-scale structure,” *Monthly Notices of the Royal Astronomical Society*, vol. 253, pp. 727–737, 1991.
- [58] F. R. Bouchet, R. Juszkiewicz, S. Colombi, and R. Pellat, “Weakly nonlinear gravitational instability for arbitrary  $\Omega$ ,” *Astrophysical Journal*, vol. 394, no. 1, pp. L5–L8, 1992.
- [59] O. Lahav, M. Itoh, S. Inagaki, and Y. Suto, “Non-Gaussian signatures from Gaussian initial fluctuations: evolution of skewness and kurtosis from cosmological simulations in the highly nonlinear regime,” *Astrophysical Journal*, vol. 402, no. 2, pp. 387–397, 1993.
- [60] J. N. Fry and R. J. Scherrer, “Skewness in large-scale structure and non-Gaussian initial conditions,” *Astrophysical Journal*, vol. 429, no. 1, pp. 36–42, 1994.
- [61] E. Gaztañaga and P. Fosalba, “Cosmological perturbation theory and the spherical collapse model—II. Non-Gaussian initial conditions,” *Monthly Notices of the Royal Astronomical Society*, vol. 301, no. 2, pp. 524–534, 1998.
- [62] R. Scoccimarro, “Gravitational clustering from  $\chi^2$  initial conditions,” *Astrophysical Journal*, vol. 542, no. 1, pp. 1–8, 2000.
- [63] N. Turok and D. N. Spergel, “Scaling solution for cosmological models at large  $N$ ,” *Physical Review Letters*, vol. 66, no. 24, pp. 3093–3096, 1991.
- [64] R. Durrer, R. Juszkiewicz, M. Kunz, and J.-P. Uzan, “Skewness as a probe of non-Gaussian initial conditions,” *Physical Review D*, vol. 62, no. 2, Article ID 021301, 5 pages, 2000.
- [65] M. J. Chodorowski and F. R. Bouchet, “Kurtosis in large-scale structure as a constraint on non-Gaussian initial conditions,” *Monthly Notices of the Royal Astronomical Society*, vol. 279, no. 2, pp. 557–563, 1996.
- [66] T. Y. Lam and R. K. Sheth, “The non-linear probability distribution function in models with local primordial non-Gaussianity,” *Monthly Notices of the Royal Astronomical Society*, vol. 395, no. 3, pp. 1743–1748, 2009.
- [67] A. Taruya, K. Koyama, and T. Matsubara, “Signature of primordial non-Gaussianity on the matter power spectrum,” *Physical Review D*, vol. 78, no. 12, Article ID 123534, 2008.
- [68] E. Sefusatti, “One-loop perturbative corrections to the matter and galaxy bispectrum with non-Gaussian initial conditions,” *Physical Review D*, vol. 80, no. 12, Article ID 123002, 24 pages, 2009.
- [69] M. H. Goroff, B. Grinstein, S.-J. Rey, and M. B. Wise, “Coupling of modes of cosmological mass density fluctuations,” *Astrophysical Journal*, vol. 311, pp. 6–14, 1986.
- [70] N. Makino, M. Sasaki, and Y. Suto, “Analytic approach to the perturbative expansion of nonlinear gravitational fluctuations in cosmological density and velocity fields,” *Physical Review D*, vol. 46, no. 2, pp. 585–602, 1992.
- [71] T. Giannantonio and C. Porciani, “Structure formation from non-Gaussian initial conditions: multivariate biasing, statistics, and comparison with N-body simulations,” *Physical Review D*, vol. 81, no. 6, article 063530, 2010.
- [72] S. Matarrese and M. Pietroni, “Resumming cosmic perturbations,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2007, no. 6, article 026, 2007.
- [73] K. Izumi and J. Soda, “Renormalized Newtonian cosmic evolution with primordial non-Gaussianity,” *Physical Review D*, vol. 76, no. 8, Article ID 083517, 2007.
- [74] N. Bartolo, J. P. Beltrn Almeida, S. Matarrese, M. Pietroni, and A. Riotto, “Signatures of primordial non-Gaussianities in the matter power-spectrum and bispectrum: the time-RG approach,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2010, no. 3, article 011, 2010.
- [75] P. Catelan and L. Moscardini, “Kurtosis and large-scale structure,” *Astrophysical Journal*, vol. 426, no. 1, pp. 14–18, 1994.
- [76] L. Verde and A. F. Heavens, “On the trispectrum as a Gaussian test for cosmology,” *Astrophysical Journal*, vol. 553, no. 1, pp. 14–24, 2001.
- [77] E. Sefusatti, M. Crocce, and V. Desjacques, “The Matter bispectrum in N-body simulations with non-Gaussian initial conditions,” *MNRAS*, vol. 406, no. 2, pp. 1014–1028, 2010.
- [78] A. Loeb and M. Zaldarriaga, “Measuring the small-scale power spectrum of cosmic density fluctuations through 21 cm tomography prior to the epoch of structure formation,” *Physical Review Letters*, vol. 92, no. 21, Article ID 211301, 4 pages, 2004.
- [79] A. Cooray, “Large-scale non-Gaussianities in the 21 cm background anisotropies from the era of reionization,” *Monthly Notices of the Royal Astronomical Society*, vol. 363, no. 3, pp. 1049–1056, 2005.
- [80] A. Pillepich, C. Porciani, and S. Matarrese, “The bispectrum of redshifted 21 centimeter fluctuations from the dark ages,” *Astrophysical Journal*, vol. 662, no. 1, pp. 1–14, 2007.
- [81] M. Bartelmann and P. Schneider, “Weak gravitational lensing,” *Physics Report*, vol. 340, no. 4-5, pp. 291–472, 2001.
- [82] F. Lucchini, S. Matarrese, A. Messina, L. Moscardini, and G. Tormen, “Velocity fields in non-Gaussian cold dark matter models,” *Monthly Notices of the Royal Astronomical Society*, vol. 272, pp. 859–868, 1995.
- [83] R. J. Scherrer, “Linear velocity fields in non-Gaussian models for large-scale structure,” *Astrophysical Journal*, vol. 390, no. 2, pp. 330–337, 1992.
- [84] R. Moessner, “Statistics of peculiar velocities from cosmic strings,” *Monthly Notices of the Royal Astronomical Society*, vol. 277, pp. 927–932, 1995.
- [85] P. Catelan and R. J. Scherrer, “Velocity differences as a probe of non-Gaussian density fields,” *Astrophysical Journal*, vol. 445, no. 1, pp. 1–6, 1995.
- [86] G. P. Holder, “Measuring cluster peculiar velocities and temperatures at centimeter and millimeter wavelengths,” *Astrophysical Journal*, vol. 602, no. 1, pp. 18–25, 2004.
- [87] L. Kofman, E. Bertschinger, J. M. Gelb, A. Nusser, and A. Dekel, “Evolution of one-point distributions from Gaussian initial fluctuations,” *Astrophysical Journal*, vol. 420, no. 1, pp. 44–57, 1994.
- [88] A. Nusser, A. Dekel, and A. Yahil, “Evidence for Gaussian initial fluctuations from the 1.2 Jansky IRAS survey,” *Astrophysical Journal*, vol. 449, no. 2, pp. 439–445, 1995.
- [89] S. D. M. White and M. J. Rees, “Core condensation in heavy halos—a two-stage theory for galaxy formation and clustering,” *Monthly Notices of the Royal Astronomical Society*, vol. 183, pp. 341–358, 1978.
- [90] N. Kaiser, “On the spatial correlations of Abell clusters,” *Astrophysical Journal Letters*, vol. 284, pp. L9–L12, 1984.

- [91] V. Desjacques, “Baryon acoustic signature in the clustering of density maxima,” *Physical Review D*, vol. 78, no. 10, Article ID 103503, 2008.
- [92] P. McDonald and A. Roy, “Clustering of dark matter tracers: generalizing bias for the coming era of precision LSS,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2009, no. 8, article 020, 2009.
- [93] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, “The evolution of large-scale structure in a universe dominated by cold dark matter,” *Astrophysical Journal*, vol. 292, pp. 371–394, 1985.
- [94] C. Lacey and S. Cole, “Merger rates in hierarchical models of galaxy formation—part two—comparison with N-body simulations,” *Monthly Notices of the Royal Astronomical Society*, vol. 271, article 676, 1998.
- [95] Z. Lukić, D. Reed, S. Habib, and K. Heitmann, “The structure of halos: implications for group and cluster cosmology,” *Astrophysical Journal*, vol. 692, pp. 217–228, 2009.
- [96] F. Lucchin and S. Matarrese, “The effect of non-Gaussian statistics on the mass multiplicity of cosmic structures,” *Astrophysical Journal*, vol. 330, pp. 535–544, 1988.
- [97] S. Colafrancesco, F. Lucchin, and S. Matarrese, “The mass function from local density maxima—groups and clusters of galaxies,” *Astrophysical Journal*, vol. 345, pp. 3–11, 1989.
- [98] P. Coles and J. D. Barrow, “Non-Gaussian statistics and the microwave background radiation,” *Monthly Notices of the Royal Astronomical Society*, vol. 228, pp. 407–426, 1987.
- [99] W. H. Press and P. Schechter, “Formation of galaxies and clusters of galaxies by self-similar gravitational condensation,” *Astrophysical Journal*, vol. 187, pp. 425–438, 1974.
- [100] J. A. Peacock and A. F. Heavens, “Alternatives to the Press-Schechter cosmological mass function,” *Monthly Notices of the Royal Astronomical Society*, vol. 243, pp. 133–143, 1990.
- [101] S. Cole, “Modeling galaxy formation in evolving dark matter halos,” *Astrophysical Journal*, vol. 367, no. 1, pp. 45–53, 1991.
- [102] J. R. Bond, S. Cole, G. Efstathiou, and N. Kaiser, “Excursion set mass functions for hierarchical Gaussian fluctuations,” *Astrophysical Journal*, vol. 379, no. 2, pp. 440–460, 1991.
- [103] R. K. Sheth and G. Tormen, “Large-scale bias and the peak background split,” *Monthly Notices of the Royal Astronomical Society*, vol. 308, pp. 119–126, 1999.
- [104] M. Maggiore and A. Riotto, “The halo mass function from excursion set theory. I. Gaussian fluctuations with non-Markovian dependence on the smoothing scale,” *The Astrophysical Journal*, vol. 711, no. 2, article 907, 2010.
- [105] S. Matarrese, L. Verde, and R. Jimenez, “The abundance of high-redshift objects as a probe of non-Gaussian initial conditions,” *Astrophysical Journal*, vol. 541, pp. 10–24, 2000.
- [106] M. Lo Verde, A. Miller, S. Shandera, and L. Verde, “Effects of scale-dependent non-Gaussianity on cosmological structures,” *Journal of Cosmology and Astro-Particle Physics*, vol. 4, article 14, 2008.
- [107] P. Valageas, “Mass function and bias of dark matter halos for non-Gaussian initial conditions,” *Astronomy & Astrophysics*, vol. 514, 14 pages, 2010.
- [108] M. Grossi, K. Dolag, E. Branchini, S. Matarrese, and L. Moscardini, “Evolution of massive haloes in non-Gaussian scenarios,” *Monthly Notices of the Royal Astronomical Society*, vol. 382, no. 3, pp. 1261–1267, 2007.
- [109] M. Grossi, L. Verde, C. Carbone et al., “Large-scale non-Gaussian mass function and halo bias: tests on N-body simulations,” *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 1, pp. 321–332, 2009.
- [110] X. Kang, P. Norberg, and J. Silk, “Can a large-scale structure probe cosmic microwave background-constrained non-Gaussianity?” *Monthly Notices of the Royal Astronomical Society*, vol. 376, pp. 343–347, 2007.
- [111] M. Maggiore and A. Riotto, “The halo mass function from excursion set theory with a non-Gaussian trispectrum,” *Monthly Notices of the Royal Astronomical Society*, vol. 405, no. 2, pp. 1244–1252, 2010.
- [112] V. R. Eke, S. Cole, and C. S. Frenk, “Cluster evolution as a diagnostic for Omega,” *Monthly Notices of the Royal Astronomical Society*, vol. 282, pp. 263–280, 1996.
- [113] M. Maggiore and A. Riotto, “The halo mass function from excursion set theory. III. Non-Gaussian fluctuations,” *The Astrophysical Journal*, vol. 717, no. 1, article 526, 2010.
- [114] M. Maggiore and A. Riotto, “The halo mass function from excursion set theory. II. the diffusing barrier,” *The Astrophysical Journal*, vol. 717, no. 1, article 515, 2010.
- [115] B. E. Robertson, A. V. Kravtsov, J. Tinker, and A. Zentner, “Collapse barriers and halo abundance: testing the excursion set ansatz,” *Astrophysical Journal*, vol. 696, pp. 636–652, 2009.
- [116] R. K. Sheth and G. Tormen, “An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier,” *Monthly Notices of the Royal Astronomical Society*, vol. 329, pp. 61–75, 2002.
- [117] T. Y. Lam and R. K. Sheth, “Halo abundances in the  $f_{\text{nl}}$  model,” *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 4, pp. 2143–2151, 2009.
- [118] T. Y. Lam, R. K. Sheth, and V. Desjacques, “The initial shear field in models with primordial local non-Gaussianity and implications for halo and void abundances,” *Monthly Notices of the Royal Astronomical Society*, vol. 399, no. 3, pp. 1482–1494, 2009.
- [119] P. P. Avelino and P. T. P. Viana, “The cloud-in-cloud problem for non-Gaussian density fields,” *Monthly Notices of the Royal Astronomical Society*, vol. 314, no. 2, pp. 354–358, 2000.
- [120] K. T. Inoue and M. Nagashima, “Analytic approach to the cloud-in-cloud problem for non-Gaussian density fluctuation,” *Astrophysical Journal*, vol. 574, pp. 9–18, 2002.
- [121] R. K. Sheth, H. J. Mo, and G. Tormen, “Ellipsoidal collapse and an improved model for the number and spatial distribution of dark matter haloes,” *Monthly Notices of the Royal Astronomical Society*, vol. 323, pp. 1–12, 2001.
- [122] V. Desjacques, “Environmental dependence in the ellipsoidal collapse model,” *Monthly Notices of the Royal Astronomical Society*, vol. 388, no. 2, pp. 638–658, 2008.
- [123] E. Rozo, R. H. Wechsler, B. P. Koester, A. E. Evrard, and T. A. McKay, “Optically-selected cluster catalogs as a precision cosmology tool,” preprint.
- [124] S. Matarrese, F. Lucchin, and S. A. Bonometto, “A path-integral approach to large-scale matter distribution originated by non-Gaussian fluctuations,” *Astrophysical Journal Letters*, vol. 310, pp. L21–L26, 1986.
- [125] P. Catelan, F. Lucchin, and S. Matarrese, “Peak number density of non-Gaussian random fields,” *Physical Review Letters*, vol. 61, pp. 267–270, 1988.
- [126] S. Borgani and S. A. Bonometto, “Galaxy density in biased theories of galaxy origin,” *Astronomy & Astrophysics*, vol. 215, pp. 17–20, 1989.
- [127] S. Matarrese, F. Lucchin, A. Messina, and L. Moscardini, “Non-Gaussian initial conditions in cosmological N-body simulations. III—groups in cold dark matter,” *Monthly Notices of the Royal Astronomical Society*, vol. 253, pp. 35–46, 1991.

- [128] C. Park, D. N. Spergel, and N. Turok, "Large-scale structure in a texture-seeded cold dark matter cosmogony," *Astrophysical Journal*, vol. 372, no. 2, pp. L53–L57, 1991.
- [129] S. Borgani, P. Coles, L. Moscardini, and M. Plionis, "The angular distribution of clusters in skewed CDM models / cold dark matter," *Monthly Notices of the Royal Astronomical Society*, vol. 266, p. 524, 1994.
- [130] H. Mathis, J. M. Diego, and J. Silk, "The case for non-Gaussianity on cluster scales," *Monthly Notices of the Royal Astronomical Society*, vol. 353, no. 2, pp. 681–688, 2004.
- [131] J. Robinson and J. E. Baker, "Evolution of the cluster abundance in non-Gaussian models," *Monthly Notices of the Royal Astronomical Society*, vol. 311, no. 4, pp. 781–792, 2000.
- [132] R. Stanek, A. E. Evrard, H. Böhringer, P. Schuecker, and B. Nord, "The X-ray luminosity-mass relation for local clusters of galaxies," *Astrophysical Journal*, vol. 648, no. 2, pp. 956–968, 2006.
- [133] D. Nagai, "The impact of galaxy formation on the Sunyaev-Zel'dovich effect of galaxy clusters," *Astrophysical Journal*, vol. 650, no. 2, pp. 538–549, 2006.
- [134] W. A. Chiu, J. P. Ostriker, and M. A. Strauss, "Using cluster abundances and peculiar velocities to test the Gaussianity of the cosmological density field," *Astrophysical Journal*, vol. 494, article 479, 1998.
- [135] A. J. Benson, C. Reichardt, and M. Kamionkowski, "Statistics of Sunyaev-Zel'dovich cluster surveys," *Monthly Notices of the Royal Astronomical Society*, vol. 331, no. 1, pp. 71–84, 2002.
- [136] S. Sadeh, Y. Rephaeli, and J. Silk, "Impact of a non-Gaussian density field on Sunyaev-Zeldovich observables," *Monthly Notices of the Royal Astronomical Society*, vol. 368, no. 4, pp. 1583–1598, 2006.
- [137] S. Sadeh, Y. Rephaeli, and J. Silk, "Cluster abundances and Sunyaev-Zel'dovich power spectra: effects of non-Gaussianity and early dark energy," *Monthly Notices of the Royal Astronomical Society*, vol. 380, pp. 637–645, 2007.
- [138] E. Sefusatti, C. Vale, K. Kadota, and J. Frieman, "Primordial non-Gaussianity and dark energy constraints from cluster surveys," *Astrophysical Journal*, vol. 658, pp. 669–679, 2007.
- [139] C. Fedeli, L. Moscardini, and S. Matarrese, "The clustering of galaxy clusters in cosmological models with non-Gaussian initial conditions: predictions for future surveys," *Monthly Notices of the Royal Astronomical Society*, vol. 397, pp. 1125–1137, 2009.
- [140] M. Roncarelli, L. Moscardini, E. Branchini, et al., "Imprints of primordial non-Gaussianities in X-ray and SZ signals from galaxy clusters," *Monthly Notices of the Royal Astronomical Society*, vol. 42, no. 2, pp. 923–933, 2010.
- [141] J. Oukbir, J. G. Bartlett, and A. Blanchard, "X-ray galaxy clusters: constraints on models of galaxy formation," *Astronomy and Astrophysics*, vol. 320, no. 2, pp. 365–377, 1997.
- [142] J. P. Henry, "Measuring cosmological parameters from the evolution of cluster X-ray temperatures," *Astrophysical Journal*, vol. 534, no. 2, pp. 565–580, 2000.
- [143] A. Amara and A. Refregier, "Power spectrum normalization and the non-Gaussian halo model," *Monthly Notices of the Royal Astronomical Society*, vol. 351, no. 1, pp. 375–383, 2004.
- [144] R. Mandelbaum and U. Seljak, "A robust lower limit on the amplitude of matter fluctuations in the universe from cluster abundance and weak lensing," *Journal of Cosmology and Astro-Particle Physics*, vol. 2007, no. 6, article 24, 2007.
- [145] M. Kamionkowski, L. Verde, and R. Jimenez, "The void abundance with non-Gaussian primordial perturbations," *Journal of Cosmology and Astro-Particle Physics*, vol. 2009, no. 1, article 010, 2009.
- [146] A. G. Doroshkevich, "The space structure of perturbations and the origin of rotation of galaxies in the theory of fluctuation," *Astrofizika*, vol. 6, pp. 591–600, 1970.
- [147] H. Song and J. Lee, "The mass function of void groups as a probe of primordial non-Gaussianity," *Astrophysical Journal Letters*, vol. 701, pp. L25–L28, 2009.
- [148] R. K. Sheth and R. van de Weygaert, "A hierarchy of voids: much ado about nothing," *Monthly Notices of the Royal Astronomical Society*, vol. 350, pp. 517–538, 2004.
- [149] J. M. Colberg, F. Pearce, C. Foster et al., "The Aspen-Amsterdam void finder comparison project," *Monthly Notices of the Royal Astronomical Society*, vol. 387, no. 2, pp. 933–944, 2008.
- [150] A. Dressler, "Galaxy morphology in rich clusters—implications for the formation and evolution of galaxies," *Astrophysical Journal*, vol. 236, pp. 351–365, 1980.
- [151] P. J. E. Peebles, "The sequence of cosmogony and the nature of primeval departures from homogeneity," *Astrophysical Journal*, vol. 274, pp. 1–13, 1983.
- [152] N. A. Bahcall and R. M. Soneira, "The spatial correlation function of rich clusters of galaxies," *Astrophysical Journal*, vol. 270, pp. 20–38, 1983.
- [153] A. A. Klypin and A. I. Kopylov, "The spatial covariance function for rich clusters of galaxies," *Soviet Astronomy Letters*, vol. 9, article 41, 1983.
- [154] B. Grinstein and M. B. Wise, "Non-Gaussian fluctuations and the correlations of galaxies or rich clusters of galaxies," *Astrophysical Journal*, vol. 310, pp. 19–22, 1986.
- [155] A. P. A. Andrade, A. L. B. Ribeiro, and C. A. Wuensche, "High order correction terms for the peak-peak correlation function in nearly-Gaussian models," *Astronomy and Astrophysics*, vol. 457, no. 2, pp. 385–391, 2006.
- [156] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, "Constraints on local primordial non-Gaussianity from large scale structure," *Journal of Cosmology and Astro-Particle Physics*, vol. 2008, no. 8, article 031, 2008.
- [157] S. Matarrese and L. Verde, "The effect of primordial non-Gaussianity on halo bias," *Astrophysical Journal Letters*, vol. 677, no. 2, pp. L77–L80, 2008.
- [158] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, "The statistics of peaks of Gaussian random fields," *Astrophysical Journal*, vol. 304, pp. 15–61, 1986.
- [159] S. Cole and N. Kaiser, "Biased clustering in the cold dark matter cosmogony," *Monthly Notices of the Royal Astronomical Society*, vol. 237, pp. 1127–2013, 1989.
- [160] N. Afshordi and A. J. Tolley, "Primordial non-Gaussianity, statistics of collapsed objects, and the integrated Sachs-Wolfe effect," *Physical Review D*, vol. 78, no. 12, Article ID 123507, 2008.
- [161] H. J. Mo and S. D. M. White, "An analytic model for the spatial clustering of dark matter haloes," *Monthly Notices of the Royal Astronomical Society*, vol. 282, pp. 347–361, 1996.
- [162] R. Y. Cen, J. P. Ostriker, D. N. Spergel, and N. Turok, "A hydrodynamic approach to cosmology: texture-seeded cold dark matter and hot dark matter cosmogonies," *Astrophysical Journal*, vol. 383, no. 1, pp. 1–18, 1991.
- [163] L. Verde and S. Matarrese, "Detectability of the effect of Inflationary non-Gaussianity on halo bias," *Astrophysical Journal Letters*, vol. 706, pp. L91–L95, 2009.
- [164] E. Hatziminaoglou, G. Mathez, J.-M. Solanes, A. Manrique, and E. Salvador-Solé, "Major mergers of haloes, the growth of massive black holes and the evolving luminosity function of quasars," *Monthly Notices of the Royal Astronomical Society*, vol. 343, no. 2, pp. 692–704, 2003.

- [165] P. F. Hopkins, L. Hernquist, T. J. Cox, et al., “Black holes in galaxy mergers: evolution of quasars,” *Astrophysical Journal*, vol. 630, pp. 705–715, 2005.
- [166] S. Bonoli, F. Shankar, S. White, V. Springel, and S. Wyithe, “On merger bias and the clustering of quasars,” *Monthly Notices of the Royal Astronomical Society*, vol. 404, no. 1, pp. 399–408, 2010.
- [167] N. Bartolo, S. Matarrese, and A. Riotto, “Signatures of primordial non-Gaussianity in the large-scale structure of the universe,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2005, no. 10, article 010, 2005.
- [168] D. Wands and A. Slosar, “Scale-dependent bias from primordial non-Gaussianity in general relativity,” *Physical Review D*, vol. 79, no. 12, Article ID 123507, 2009.
- [169] J. N. Fry and E. Gaztañaga, “Biasing and hierarchical statistics in large-scale structure,” *Astrophysical Journal*, vol. 413, no. 2, pp. 447–452, 1993.
- [170] N. Hamaus, et al., in preparation, 2010.
- [171] N. Kaiser, “Clustering in real space and in redshift space,” *Monthly Notices of the Royal Astronomical Society*, vol. 227, pp. 1–21, 1987.
- [172] A. J. S. Hamilton, “Linear redshift distortions: a review,” in *The Evolving Universe*, D. Hamilton, Ed., vol. 231 of *Astrophysics and Space Science Library*, p. 185, 1998.
- [173] T. Y. Lam, V. Desjacques, and R. K. Sheth, “The non-linear redshift space probability distribution function in models with local primordial non-Gaussianity,” *Monthly Notices of the Royal Astronomical Society*, vol. 402, no. 4, pp. 2397–2402, 2010.
- [174] H. A. Feldman, N. Kaiser, and J. A. Peacock, “Power-spectrum analysis of three-dimensional redshift surveys,” *Astrophysical Journal*, vol. 426, no. 1, pp. 23–37, 1994.
- [175] U. Seljak, “Extracting primordial non-Gaussianity without cosmic variance,” *Physical Review Letters*, vol. 102, no. 2, Article ID 021302, 2009.
- [176] P. McDonald and U. Seljak, “How to evade the sample variance limit on measurements of redshift-space distortions,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2009, no. 10, article 007, 2009.
- [177] A. Slosar, “Optimal weighting in  $f_{NL}$  constraints from large scale structure in an idealised case,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2009, no. 3, article 004, 2009.
- [178] U. Seljak, N. Hamaus, and V. Desjacques, “How to suppress the shot noise in galaxy surveys,” *Physical Review Letters*, vol. 103, no. 9, article 091303, 2009.
- [179] Z. Fan and J. M. Bardeen, “Distributions of Fourier modes of cosmological density fields,” *Physical Review D*, vol. 51, no. 12, pp. 6714–6721, 1995.
- [180] A. J. Stirling and J. A. Peacock, “Power correlations in cosmology: limits on primordial non-Gaussian density fields,” *Monthly Notices of the Royal Astronomical Society*, vol. 283, p. L99, 1996.
- [181] E. Sefusatti and E. Komatsu, “Bispectrum of galaxies from high-redshift galaxy surveys: primordial non-Gaussianity and nonlinear galaxy bias,” *Physical Review D*, vol. 76, no. 8, Article ID 083004, 2007.
- [182] D. Jeong and E. Komatsu, “Primordial non-Gaussianity, scale-dependent bias, and the bispectrum of galaxies,” *Astrophysical Journal*, vol. 703, no. 2, pp. 1230–1248, 2009.
- [183] E. Gaztañaga, “N-point correlation functions in the CfA and SSRS redshift distribution of galaxies,” *Astrophysical Journal*, vol. 398, no. 1, pp. L17–L20, 1992.
- [184] E. Gaztanaga and J. Yokoyama, “Probing the statistics of primordial fluctuations and their evolution,” *Astrophysical Journal*, vol. 403, pp. 450–465, 1993.
- [185] F. R. Bouchet, M. A. Strauss, M. Davis, K. B. Fisher, A. Yahil, and J. P. Huchra, “Moments of the counts distribution in the 1.2 Jansky IRAS galaxy redshift survey,” *Astrophysical Journal*, vol. 417, p. 36, 1993.
- [186] S. Luo and E. Vishniac, “Bayesian approaches to testing the nature of primordial density fluctuation,” *Astrophysical Journal*, vol. 443, no. 2, pp. 469–478, 1995.
- [187] D. J. Croton, E. Gaztañaga, C. M. Baugh, et al., “The 2dF galaxy redshift survey: higher order galaxy correlation functions,” *Monthly Notices of the Royal Astronomical Society*, vol. 352, pp. 1232–1244, 2004.
- [188] W. J. Frith, P. J. Outram, and T. Shanks, “High order 2MASS galaxy correlation functions: probing the Gaussianity of the primordial density field,” *Monthly Notices of the Royal Astronomical Society*, vol. 373, pp. 759–768, 2006.
- [189] T. Matsubara and Y. Suto, “Nonlinear evolution of genus in a primordial random Gaussian density field,” *Astrophysical Journal*, vol. 460, no. 1, pp. 51–58, 1996.
- [190] P. Coles, L. Moscardini, M. Plionis, F. Lucchin, S. Matarrese, and A. Messina, “Topology in two dimensions. IV—CDM models with non-Gaussian initial conditions,” *Monthly Notices of the Royal Astronomical Society*, vol. 260, pp. 572–588, 1993.
- [191] T. Matsubara and J. Yokoyama, “Genus statistics of the large-scale structure with non-Gaussian density fields,” *Astrophysical Journal*, vol. 463, p. 409, 1996.
- [192] C. Hikage, A. Taruya, and Y. Suto, “Genus statistics for galaxy clusters and nonlinear biasing of dark matter halos,” *Astrophysical Journal*, vol. 556, pp. 641–652, 2001.
- [193] C. Hikage, P. Coles, M. Grossi et al., “The effect of primordial non-Gaussianity on the topology of large-scale structure,” *Monthly Notices of the Royal Astronomical Society*, vol. 385, no. 3, pp. 1613–1620, 2008.
- [194] J. R. Gott III, D. C. Hambrick, M. S. Vogeley et al., “Genus topology of structure in the sloan digital sky survey: model testing,” *Astrophysical Journal*, vol. 675, no. 1, pp. 16–28, 2008.
- [195] J. B. James, M. Colless, G. F. Lewis, and J. A. Peacock, “Topology of non-linear structure in the 2dF Galaxy Redshift Survey,” *Monthly Notices of the Royal Astronomical Society*, vol. 394, no. 1, pp. 454–466, 2009.
- [196] X. Chen, A. Cooray, N. Yoshida, and N. Sugiyama, “Can non-Gaussian cosmological models explain the WMAP high optical depth for reionization?” *Monthly Notices of the Royal Astronomical Society*, vol. 346, pp. L31–L35, 2003.
- [197] P. P. Avelino and A. R. Liddle, “Cosmic reionization constraints on the nature of cosmological perturbations,” *Monthly Notices of the Royal Astronomical Society*, vol. 371, pp. 1755–1759, 2006.
- [198] D. Crociani, L. Moscardini, M. Viel, and S. Matarrese, “The effects of primordial non-Gaussianity on the cosmological reionization,” *Monthly Notices of the Royal Astronomical Society*, vol. 394, no. 1, pp. 133–141, 2009.
- [199] M. Viel, E. Branchini, K. Dolag, M. Grossi, S. Matarrese, and L. Moscardini, “Primordial non-Gaussianities in the intergalactic medium,” *Monthly Notices of the Royal Astronomical Society*, vol. 393, no. 3, pp. 774–782, 2009.
- [200] S. Ho, et al., in preparation, 2010.
- [201] J. E. Gunn and B. A. Peterson, “On the density of neutral hydrogen in intergalactic space,” *Astrophysical Journal*, vol. 142, pp. 1633–1641, 1965.

- [202] R. A. C. Croft, D. H. Weinberg, N. Katz, and L. Hernquist, “Recovery of the power spectrum of mass fluctuations from observations of the Ly-alpha forest,” *Astrophysical Journal*, vol. 495, article 44, 1998.
- [203] M. Zaldarriaga, R. Scoccimarro, and L. Hui, “Inferring the linear power spectrum from the Ly $\alpha$  forest,” *Astrophysical Journal*, vol. 590, no. 1 I, pp. 1–7, 2003.
- [204] P. McDonald, “Toward a measurement of the cosmological geometry at  $z \sim 2$ : predicting Ly $\alpha$  forest correlation in three dimensions and the potential of future data sets,” *Wireless Networks*, vol. 585, pp. 34–51, 2003.
- [205] P. Coles and B. Jones, “A lognormal model for the cosmological mass distribution,” *Monthly Notices of the Royal Astronomical Society*, vol. 248, pp. 1–13, 1991.
- [206] H. Bi and A. F. Davidsen, “Evolution of structure in the intergalactic medium and the nature of the Ly-alpha forest,” *Astrophysical Journal*, vol. 479, p. 523, 1997.
- [207] E. Komatsu, K. M. Smith, J. Dunkley, et al., “Seven-year wilkinson microwave anisotropy probe (WMAP) observations: cosmological interpretation,” preprint.
- [208] J. A. Frieman and E. Gaztañaga, “The projected three-point correlation function: theory and observations,” *Astrophysical Journal Letters*, vol. 521, no. 2, pp. L83–L86, 1999.
- [209] R. Scoccimarro, H. A. Feldman, J. N. Fry, and J. A. Frieman, “The bispectrum of IRAS redshift catalogs,” *Astrophysical Journal*, vol. 546, no. 2, pp. 652–664, 2001.
- [210] G. V. Kulkarni, R. C. Nichol, R. K. Sheth, H.-J. Seo, D. J. Eisenstein, and A. Gray, “The three-point correlation function of luminous red galaxies in the Sloan Digital Sky Survey,” *Monthly Notices of the Royal Astronomical Society*, vol. 378, pp. 1196–1206, 2007.
- [211] E. Gaztañaga, A. Cabré, F. Castander, M. Crocce, and P. Fosalba, “Clustering of luminous red galaxies—III. Baryon acoustic peak in the three-point correlation,” *Monthly Notices of the Royal Astronomical Society*, vol. 399, no. 2, pp. 801–811, 2009.
- [212] J. A. Willick, “Constraints on primordial non-Gaussianity from the high-redshift cluster MS 1054-03,” *Astrophysical Journal*, vol. 530, no. 1, pp. 80–95, 2000.
- [213] R. Jimenez and L. Verde, “Implications for primordial non-Gaussianity ( $f_{NL}$ ) from weak lensing masses of high- $z$  galaxy clusters,” *Physical Review D*, vol. 80, no. 12, Article ID 127302, 2009.
- [214] B. Sartoris, S. Borgani, C. Fedeli, et al., “The potential of X-ray cluster surveys to constrain primordial non-Gaussianity,” preprint.
- [215] D. Jeong, E. Komatsu, and B. Jain, “Galaxy-CMB and galaxy-galaxy lensing on large scales: sensitivity to primordial non-Gaussianity,” *Physical Review D*, vol. 80, no. 12, Article ID 123527, 25 pages, 2009.
- [216] C. Fedeli and L. Moscardini, “Cosmic shear statistics in cosmologies with non-Gaussian initial conditions,” *Monthly Notices of the Royal Astronomical Society*, vol. 405, no. 1, pp. 681–694, 2010.
- [217] C. Carbone, L. Verde, and S. Matarrese, “Non-Gaussian halo bias and future galaxy surveys,” *Astrophysical Journal Letters*, vol. 684, pp. L1–L4, 2008.
- [218] M. Takada and B. Jain, “Cosmological parameters from lensing power spectrum and bispectrum tomography,” *Monthly Notices of the Royal Astronomical Society*, vol. 348, pp. 897–915, 2004.
- [219] Y. Gong, X. Wang, Z. Zheng, and X.-L. Chen, “Primordial non-Gaussianity from LAMOST surveys,” *Research in Astronomy and Astrophysics*, vol. 10, no. 2, article 107, 2010.
- [220] E. Sefusatti, M. Liguori, A. P. S. Yadav, M. G. Jackson, and E. Pajer, “Constraining running non-Gaussianity,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2009, no. 12, article 022, 2009.
- [221] C. Carbone, O. Mena, and L. Verde, “Cosmological parameters degeneracies and non-Gaussian halo bias,” *JCAP*, no. 07, article no. 020, 2010.
- [222] M. Oguri, “Self-calibrated cluster counts as a probe of primordial non-Gaussianity,” *Physical Review Letters*, vol. 102, no. 21, Article ID 211301, 2009.
- [223] C. Cunha, D. Huterer, and O. Dore, “Primordial non-Gaussianity from the covariance of galaxy cluster counts,” *Phys. Rev. D*, no. 82, article no. 023004, 2010.

## Review Article

# Non-Gaussianity from Large-Scale Structure Surveys

**Licia Verde**

*Institució Catalana de Recerca i Estudis Avançat (ICREA) & Instituto de Ciencias del Cosmos,  
Universitat de Barcelona-IEEC (ICC-UB), Martí i Franques 1, 08028 Barcelona, Spain*

Correspondence should be addressed to Licia Verde, [liciaverde@icc.ub.edu](mailto:liciaverde@icc.ub.edu)

Received 27 January 2010; Accepted 18 March 2010

Academic Editor: Dragan Huterer

Copyright © 2010 Licia Verde. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the advent of galaxy surveys which provide large samples of galaxies or galaxy clusters over a volume comparable to the horizon size (SDSS-III, HETDEX, Euclid, JDEM, LSST, Pan-STARRS, CIP, etc.) or mass-selected large cluster samples over a large fraction of the extra-galactic sky (Planck, SPT, ACT, CMBPol, B-Pol), it is timely to investigate what constraints these surveys can impose on primordial non-Gaussianity. I illustrate here three different approaches: higher-order correlations of the three dimensional galaxy distribution, abundance of rare objects (extrema of the density distribution), and the large-scale clustering of halos (peaks of the density distribution). Each of these avenues has its own advantages, but, more importantly, these approaches are highly complementary under many respects.

## 1. Introduction

The recent advances in the understanding of the origin and evolution of the Universe have been driven by the advent of high-quality data, in unprecedented amount (just think of WMAP and SDSS, e.g.). Despite this, most of the information about cosmological parameters come from the analysis of a massive compression of the data: the power spectrum of their statistical fluctuations over the mean. The power spectrum is a complete statistical description of a random field only if it is Gaussian.

Even the simplest inflationary models predict deviations from Gaussian initial conditions. These deviations are expected to be small, although “small” in some models may be “detectable.” For a thorough review of inflationary non-Gaussianity see [1]; for our purpose it will be sufficient to say that to describe inflation-motivated departures from Gaussian initial conditions many write [2–5]

$$\Phi = \phi + f_{\text{NL}}(\phi^2 - \langle \phi^2 \rangle). \quad (1)$$

Here  $\phi$  denotes a gaussian field and  $\Phi$  denotes Bardeen’s gauge-invariant potential, which, on sub-Hubble scales reduces to the usual Newtonian peculiar gravitational potential, up to a minus sign. In the literature, there are two conventions for (1): the large-scale structure (LSS) and the

Cosmic Microwave Background (CMB) one. In the LSS convention  $\Phi$  is linearly extrapolated at  $z = 0$ ; in the CMB convention  $\Phi$  is instead primordial: thus  $f_{\text{NL}}^{\text{LSS}} = g(z = \infty)/g(0)f_{\text{NL}}^{\text{CMB}} \sim 1.3f_{\text{NL}}^{\text{CMB}}$ , where  $g(z)$  denotes the linear growth suppression factor relative to an Einstein-de-Sitter Universes. In the past few years it has become customary to always report  $f_{\text{NL}}^{\text{CMB}}$  values even if, for simplicity as it will be clear below, one carries out the calculations with  $f_{\text{NL}}^{\text{LSS}}$ .

While for simplicity one may just assume  $f_{\text{NL}}$  in (1) to be a constant (yielding the so-called *local* model or *local-type*) in reality the expression is more complicated and  $f_{\text{NL}}$  is scale and configuration dependent. Even if the bispectrum does not completely specify non-Gaussianity, in most practical applications, the non-Gaussianity is set by writing down the bispectrum of  $\Phi$ . For example, one can see that for the local model the bispectrum is

$$B_{\Phi}(k_1, k_2, k_3) = 2f_{\text{NL}}P_{\phi}(k_1)P_{\phi}(k_2) + 2 \text{cyc}, \quad (2)$$

where  $P$  denotes the power spectrum and it is often assumed that  $P_{\phi} = P_{\Phi}$ ; “cyc.” denotes two cyclic terms over  $k_1, k_2, k_3$ .

It has been shown [6] that for non-Gaussianity of the local type the bispectrum is dominated by the so-called *squeezed* configurations, triangles where one wave-vector length is much smaller than the other two. Models such as the curvaton for example, have a non-Gaussianity of

the local type. Standard, single-field slow roll inflation also yields a local non-Gaussianity but with an unmeasurably small  $f_{\text{NL}}$  (see [1] and references therein). On the other hand, many inflationary models have an *equilateral*-type non-Gaussianity, that is, the bispectrum is dominated by equilateral triangles [6]. References [7, 8] have proposed a functional form which closely approximates the behavior of the inflationary bispectrum and which is useful for efficient data-analysis:

$$B(k_1, k_2, k_3) = 6f_{\text{NL}} \left\{ -P(k_1)P(k_2) + 2\text{cyc.} \right. \\ \left. - 2[P(k_1)P(k_2)P(k_3)]^{2/3} \right. \\ \left. + P^{1/3}(k_1)P^{2/3}(k_2)P(k_3) + 5\text{cyc.} \right\}. \quad (3)$$

Note that the same numerical value for  $f_{\text{NL}}$  gives rise to a larger skewness in the local case than in the equilateral case (e.g., see [9]), explaining why the CMB constraints on  $f_{\text{NL}}$  are weaker for the equilateral case [7, 8, 10].

Specific deviations from a single field, slow roll, canonical kinetic energy, Bunch-Davies vacuum, leave their specific signature on the bispectrum “shape” (i.e., the dependence of  $B$  on the shape of the triangle made by the three  $\vec{k}$  vectors), see discussion in [11] and references therein.

Non-Gaussianity therefore offers a probe of aspects of inflation (namely, the interactions of the inflaton) that are difficult to probe by other means (i.e., measuring the shape of the primordial power spectrum and properties of the stochastic background of gravity waves). So, how could primordial non-Gaussianity be tested?

One could look at the early Universe: by looking at CMB anisotropies we can probe cosmic fluctuations at a time when their statistical distribution should have been close to their original form but the signal is small. On the other hand, one could analyze the statistics of the large-scale structures, close to the present-day, when the overdensities are larger, but this is a more complicated approach, since gravitational instability (for the dark matter distribution) and bias (for galaxies or clusters of galaxies) introduce non-Gaussian features in an initially Gaussian field and they mask the signal one is after. Finally, the abundance of rare events (such as galaxy clusters and high-redshift galaxies) probes the tails of the PDF of the density field, which are extremely sensitive to deviations from Gaussianity. Here, I will concentrate on the signature of non-Gaussianity on large-scale structure (i.e., at redshift  $z \lesssim 1$ ) as they can be traced by galaxy surveys (i.e., I will not consider wide field weak gravitational lensing surveys); other contributions to this review will focus on non-Gaussian signatures on the CMB, thus here it will be sufficient to give only a very brief and succinct introduction to the subject.

At recombination fluctuations are small (recall that  $\delta_\Phi \sim 10^{-5}$ ), and the CMB temperature fluctuations are directly related to  $\Phi$  making this a very clean probe. However, effectively only one redshift can be tested giving us only a 2-dimensional information.

The most widespread technique for testing Gaussianity in the CMB is to use the CMB bispectrum:

$$\langle a_{\ell_1}^{m_1} a_{\ell_2}^{m_2} a_{\ell_3}^{m_3} \rangle = B_{\ell_1 \ell_2 \ell_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \quad (4)$$

where the  $a_\ell^m$  are the coefficients of the spherical harmonic expansion of the CMB temperature fluctuation:  $\Delta T/T = \sum_{\ell m} a_\ell^m Y_\ell^m$  and the presence of the 3-J symbol ensures that the bispectrum is defined if  $l_1 + l_2 + l_3 = \text{even}$ ,  $\ell_j + \ell_k \geq l_i \geq |\ell_j - \ell_k|$  (triangle rule), and that  $m_1 + m_2 + m_3 = 0$ . It should be however clear that secondary CMB anisotropies and foregrounds also induce a CMB bispectrum which can mask or partially mimic the signal see, for example [5, 12–17] and references therein.

In the last few years, this area of research has received an impulse, motivated by the recent full sky CMB data from WMAP. In particular it has been shown that the constraints can be greatly improved by effectively “reconstructing” the potential  $\Phi$  from CMB temperature and polarization data rather than simply using the temperature bispectrum alone [18, 19]. This technique would yield constraints on non-Gaussianity of the local type of  $\Delta f_{\text{NL}} \sim 1$  for an ideal experiment and  $\Delta f_{\text{NL}} \sim 3$  for the Planck satellite. This is particularly promising as  $f_{\text{NL}}$  of order unity or larger is produced by broad classes of inflationary models (see e.g., [1] and references therein).

Currently, the most stringent constraints for the local type are  $27 < f_{\text{NL}} < 147$  at the 95% confidence (central value 87) from WMAP 3 years data [20]; and from the WMAP 5 years data,  $-9 < f_{\text{NL}} < 111$  at the 95% confidence level (central value 55) [10] and  $-4 < f_{\text{NL}} < 80$  [21]. Despite the heated debate on whether  $f_{\text{NL}} = 0$  is ruled out or not, the two measurements are not necessarily in conflict: the two central values differ by only about  $1\sigma$ ; different, although not independent, data sets were used with different galactic cuts, and the maximum multipole considered in the analyses is also different. What makes the subject very interesting, is that, if the central value for  $f_{\text{NL}}$  is truly around 60, forthcoming data will yield a highly-significant detection.

## 2. Higher-Order Correlations

Theoretical considerations (see discussion in e.g., [11] and references therein) lead us to define primordial non-Gaussianity by its bispectrum. While in principle there may be types of non-Gaussianity which would be more directly related to higher-order correlations (e.g., [22] and references therein), and while a full description of a non-Gaussian distribution would require the specification of all the higher-order correlations, it is clear that quantities such as the bispectrum enclose information about the phase correlation between  $k$ -modes. In the Gaussian case, different Fourier modes are uncorrelated (by definition of Gaussian random phases) and a statistic like the power spectrum does not carry information about phases. The bispectrum is the lowest-order correlation with zero expectation value in a Gaussian random field. But, even if the initial conditions were Gaussian, nonlinear evolution due to gravitational instability generates a nonzero bispectrum. In particular, gravitational

instability has its own “signature” bispectrum, at least in the next-to-leading order in cosmological perturbation theory [23]:

$$B(\vec{k}_1, \vec{k}_2, \vec{k}_3) = 2P(k_1)P(k_2)J(\vec{k}_1, \vec{k}_2) + 2\text{cyc}, \quad (5)$$

where  $J(\vec{k}_1, \vec{k}_2)$  is the gravitational instability “kernel” which depends very weakly on cosmology and for an Einstein-de-Sitter Universe is

$$J(\vec{k}_1, \vec{k}_2) = \frac{5}{7} + \frac{\vec{k}_1 \cdot \vec{k}_2}{2k_1k_2} \left( \frac{k_1}{k_2} + \frac{k_2}{k_1} \right) + \frac{2}{7} \left( \frac{\vec{k}_1 \cdot \vec{k}_2}{k_1k_2} \right)^2. \quad (6)$$

In the highly nonlinear regime the detailed form of the kernel changes, but it is something that could be computed and calibrated by extending perturbation theory beyond the next-to-leading order and by comparing with numerical N-body simulations (see other contributions in this issue). It was recognized a decade ago [4] that this signal is quite large compared to any expected primordial non-Gaussianity and that the primordial signal “redshifts away” compared to the gravitational signal. In fact, a primordial signal given by a local type of non-Gaussianity parameterized by a given  $f_{\text{NL}}$ , would affect the late-time dark matter density bispectrum with a contribution of the form

$$B^{f_{\text{NL local}}}(\vec{k}_1, \vec{k}_2, \vec{k}_3, z) = 2f_{\text{NL}}P(k_1)P(k_2) \frac{\mathcal{F}(\vec{k}_1, \vec{k}_2)}{D(z)/D(z=0)} + 2\text{cyc}, \quad (7)$$

where  $D(z)$  is the linear growth function which in an Einstein-de-Sitter universe goes like  $(1+z)^{-1}$  and

$$\mathcal{F} = \frac{\mathcal{M}(k_3)}{\mathcal{M}(k_1)\mathcal{M}(k_2)}; \quad \mathcal{M}(k) = \frac{2}{3} \frac{k^2 T(k)}{H_0^2 \Omega_{m,0}}, \quad (8)$$

$T(k)$  denoting the transfer function,  $H_0$  the Hubble parameter, and  $\Omega_{m,0}$  the matter density parameter. Clearly the two contributions have different scale and redshift dependence and the two kernel shapes in configuration space are different, thus, making the two components, at least in principle and for high signal-to-noise, separable.

Unfortunately, with galaxy surveys, one does not observe the dark matter distribution directly. Dark matter halos are believed to be hosts for galaxy formation, and different galaxies at different redshifts populate halos following different prescriptions. In large-scale structure studies, often the assumption of linear scale independent bias is made. A linear bias will not introduce a nonzero bispectrum in a Gaussian field and its effect on a field with a nonzero bispectrum is only to rescale its bispectrum amplitude. This is, however, an approximation, possibly roughly valid at large scales for dark matter halos, and when looking at the power spectrum, but unlikely to be true in detail. To go beyond the linear bias assumption, often the assumption of quadratic bias is made, where the relation between dark matter overdensity field and galaxy field is specified by two parameters:  $b_1$  and  $b_2$ ,  $\delta_g(x) = b_1 \delta_{\text{DM}}(x) + b_2 (\delta_{\text{DM}}^2 - \langle \delta_{\text{DM}}^2 \rangle)$ ;  $b_1$  and  $b_2$  are assumed

to be scale-independent (although this assumption must break down at some point) but they can vary with redshift. Clearly, a quadratic bias will introduce non-Gaussianity even on an initially Gaussian field. In summary, for local non-Gaussianity and scale-independent quadratic bias we have [4, 24]

$$B(\vec{k}_1, \vec{k}_2, \vec{k}_3, z) = 2P(k_1)P(k_2)b_1(z)^3 \times \left[ f_{\text{NL}} \frac{\mathcal{F}(\vec{k}_1, \vec{k}_2)}{D(z)} + J(\vec{k}_1, \vec{k}_2) + \frac{b_2(z)}{2b_1(z)} \right] + \text{cyc}. \quad (9)$$

Before the above expression can be compared to observations it needs to be further complicated by redshift space distortions (and shot noise). Realistic surveys use the redshift as a proxy for distance, but gravitationally induced peculiar velocities distort the redshift-space galaxy distribution. We will not go into these details here as including redshift space distortions (and shot noise) will not change the gist of the message.

From a practical point of view, it is important to note that photometric surveys, although in general can cover larger volumes than spectroscopic ones, are not suited for this analysis: the projection effects due to the photo- $z$  smearing along the line-of-sight is expected to suppress significantly the sensitivity of the measured bispectrum to the shape of the primordial one (see e.g., [25, 26]).

Reference [4] concluded that “CMB is likely to provide a better probe of such (local) non-Gaussianity.” Much more recently, reference [27] revisited the issue and found that, assuming a given known redshift dependence of the  $(b_1, b_2)$  bias parameters and an all sky survey from  $z = 0$  to  $z = 5$  with a galaxy number density of at least  $5 \times 10^{-4} \text{ h}^3/\text{Mpc}^3$ , the galaxy bispectrum can provide constraints on the  $f_{\text{NL}}$  parameter competitive with CMB. However, for all planned surveys, the forecasted errors are much larger than Planck forecasted errors. This holds qualitatively also for the equilateral case.

While the gravitationally induced non-Gaussian signal in the bispectrum has been detected to high statistical significance (see [28] and references therein, see also other contributions to this issue), the nonlinear bias signature is not uncontroversial, and there have been so far no detection of any extra (primordial) bispectrum contributions.

Of course one could also consider higher-order correlations. One of the advantages of considering, for example, the trispectrum is that, contrary to the bispectrum, it has very weak nonlinear growth [29], but has the disadvantage that the signal is delocalized: the number of possible configurations grows fast with the dimensionality  $n$  of the  $n$ -point function!

In summary, higher-order correlations as observed in the CMB or in the evolved Universe, can be used to determine the bispectrum *shape*. The two approaches should be seen as complementary as they are affected by different systematic effects and probe different scales. The next two probes we

consider have a less-rich sensitivity to the bispectrum shape, but their own peculiar advantages.

### 3. The Mass Function

The abundance of collapsed objects (dark matter halos as traced, e.g., by galaxies and galaxy clusters) contains important information about the properties of initial conditions on galaxy and clusters scales. The Gaussian assumption plays a central role in analytical predictions for the abundance and statistical properties of the first objects to collapse in the Universe. In this context, the formalism proposed by Press and Schechter [30], with its later extensions and improvements, has become the standard lore for predicting the number of collapsed dark matter halos as a function of redshift. However, even a small deviation from Gaussianity can have a deep impact on those statistics which probe the tails of the distribution. This is indeed the case for the abundance of high-redshift objects like galaxies and clusters at  $z \gtrsim 1$  which correspond to high peaks, that is, rare events, in the underlying dark matter density field. Therefore, even small deviations from Gaussianity might be potentially detectable by looking at the statistics of high-redshift systems. Before proceeding let us introduce some definitions.

We are interested in predictions for rare objects, that is the collapsed objects that form in extreme peaks of the density field  $\delta(x) = \delta\rho/\rho$ . The statistics of collapsed objects can be described by the statistics of the density perturbation smoothed on some length scale  $R$  (or equivalently a mass scale  $M = 4/3\pi R^3\rho$ ),  $\delta_R$ .

To incorporate non-Gaussian initial conditions into predictions for the smoothed density field, we need an expression for the probability distribution function (PDF) for  $\delta_R$ . For a particular real-space expansion like (1), one may make a formal change of variable in the Gaussian PDF to generate a normalized distribution [31]. However, this may not be possible in general and the change of variables does not work for the smoothed cumulants of the density field. In general, the PDF for a generic non-Gaussian distribution can be written exactly as a function of the cumulant's generating function  $\mathcal{W}_R$  for the smoothed density field:

$$\mathcal{P}(\delta_R)d\delta_R = \int \frac{d\lambda}{2\pi} \exp[-i\lambda\delta_R + \mathcal{W}_R(\lambda)]d\delta_R \quad (10)$$

with

$$\mathcal{W}_R(\lambda) = \sum_{n=2}^{\infty} \frac{(i\lambda)^n}{n!} \mu_{n,R}, \quad (11)$$

where  $\mu_{n,R}$  denote the cumulants and, for example,  $\mu_{2,R} = \sigma_R^2 = \langle \delta_R^2 \rangle$  and the skewness  $\mu_{3,R}$  is related to the normalized skewness of the smoothed density field  $S_{3,R} = \mu_{3,R}/\mu_{2,R}^2$ . It is useful to define a ‘‘skewness per  $f_{\text{NL}}$  unit’’  $S_{3,R}^{f_{\text{NL}}=1}$  so that  $S_{3,R} = f_{\text{NL}} S_{3,R}^{f_{\text{NL}}=1}$ . The skewness  $\mu_{3,R}$  is related to the underlying bispectrum by

$$\mu_{3,R} = \int \frac{d^3k_1 d^3k_2 d^3k_3}{(2\pi)^6} B_{\delta,R}(\vec{k}_1, \vec{k}_2, \vec{k}_3) \delta_{k_1+k_2+k_3}^D, \quad (12)$$

where  $\delta^D$  denotes the Dirac delta function and  $B_{\delta,R}$  denotes the bispectrum of the  $\delta$  overdensity field smoothed on scale  $R$ . It is related to the potential one trivially by remembering the Poisson equation:  $\delta_R(\vec{k}) = \mathcal{M}(k)W_R(k)\Phi(\vec{k})$ . Here,  $W_R(k)$  denotes the smoothing kernel, usually taken to be the Fourier transform of the top hat window. In any practical application the dimensionality of the integration can be reduced by collapsing the expressing  $\vec{k}_3$  as a function of  $\vec{k}_1$  and  $\vec{k}_2$ .

It is important at this point to make a small digression to specify definitions of key quantities. Even in linear theory, the normalized skewness of the density field depends on redshift; however in the Press and Schechter framework one should always use linearly extrapolated quantities at  $z = 0$ . In this context therefore, when writing  $S_{3,R} = f_{\text{NL}} S_{3,R}^{f_{\text{NL}}=1}$ , if  $S_{3,R}^{f_{\text{NL}}=1}$  is that if the density field extrapolated linearly at  $z = 0$  then  $f_{\text{NL}}$  must be the LSS one and not the CMB one.

To compute the abundance of collapsed objects from the PDF one will then follow the Press and Schechter swindle: first compute

$$\mathcal{P}(> \delta_c | z, R) = \int_{\delta_c(z)}^{\infty} d\delta_R \mathcal{P}(\delta_R), \quad (13)$$

where  $\delta_c$  denotes the critical threshold for collapse; then the number of collapsed objects is

$$n(M, z)dM = 2 \frac{3H_0^2 \Omega_m}{8\pi GM} \left| \frac{d\mathcal{P}(> \delta_c | z, R)}{dM} \right| M. \quad (14)$$

Note that the redshift dependence is usually enclosed only in  $\delta_c$ :  $\sigma_M$  is computed on the field linearly extrapolated at  $z = 0$ , and  $\delta_c(z) = \Delta_c(z)D(z=0)/D(z)$  and  $\Delta_c(z)$  depends very weakly on redshift and  $\Delta_c(z=0) \sim 1.68$ .

Equation (14) however cannot be computed analytically and exactly starting from (10): some approximations need to be done in order to obtain an analytically manageable expression. Two approaches have been taken so far in the literature which we will briefly review below.

**3.1. MVJ Approach.** The authors of [31] proceed by first performing the integration over  $\delta_R$  to obtain an exact expression for  $\mathcal{P}(> \delta_c | z, M)$ . At this point they expand the generating functional to the desired order, for example, keeping only terms up to the skewness, then perform a Wick rotation to change variables and finally a saddle-point approximation to evaluate the remaining integral. The saddle point approximation is very good for large thresholds  $\delta_c/\sigma_M \gg 1$ , thus for rare and massive peaks. For the final expression for the mass function they obtain

$$\hat{n}(M, z) = 2 \frac{3H_0^2 \Omega_{m,0}}{8\pi GM^2} \frac{1}{\sqrt{2\pi}\sigma_M} \exp\left[-\frac{\delta_*^2}{2\sigma_M^2}\right] \times \left| \frac{1}{6} \frac{\delta_c^2}{\sqrt{1 - S_{3,M}\delta_c/3}} \frac{dS_{3,M}}{d \ln M} + \frac{\delta_*}{\sigma_M} \frac{d\sigma_M}{d \ln M} \right|, \quad (15)$$

where  $\sigma_M$  denotes the *rms* value of the density field, the subscript  $M$  denotes that the density field has been smoothed

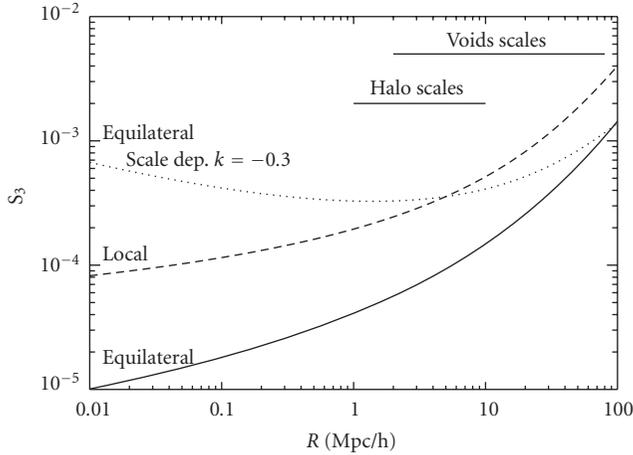


FIGURE 1: Skewness  $S_{3,R}$  of the density field at  $z = 0$  as a function of the smoothing scale  $R$  for different types of non-Gaussianity. Figure reproduced from [32].

on a scale  $R(M)$  corresponding to  $R(M) = [M3/(4\rho)]^{1/3}$ , and  $\delta_* = \delta_c \sqrt{1 - \delta_c S_{3,M}/3}$ .

This derivation shows that the mass function in principle depends on all cumulants, but that if non-Gaussianity is described by a bispectrum (and all higher order connected correlations are assumed to be zero or at least negligible), it depends only on the skewness. The mass function does not carry explicit information about the shape of non-Gaussianity. Nevertheless for a given numerical value of  $f_{\text{NL}}$  the skewness can have different amplitude and scale dependence for different models of non-Gaussianity, as illustrated in Figure 1.

**3.2. LMSV Approach.** The authors of [9] (LMSV) instead proceed by using the saddle point approximation in the expression for  $\mathcal{P}(\delta_R)$  and then using the Edgeworth expansion truncated at the desired order. The resulting simplified expression for the PDF can then be integrated to obtain  $\mathcal{P}( > \delta_c | z, R)$  and derived to obtain the mass function. The final mass function in this approximation is given by

$$\begin{aligned} \hat{n}(M, z) &= 2 \frac{3H_0^2 \Omega_{m,0}}{8\pi GM} \frac{1}{\sqrt{2\pi x s}} \exp\left[-\frac{\delta_c^2}{2\sigma_M^2}\right] \\ &\times \left| \frac{d \ln \sigma_M}{dM} \left( \frac{\delta_c}{\sigma_M} + \frac{S_{3,M} \sigma_M}{6} \left( \frac{\delta_c^4}{\sigma_M^4} - 2 \frac{\delta_c^2}{\sigma_M^2} - 1 \right) \right) \right. \\ &\quad \left. + \frac{1}{6} \frac{d S_{3,M}}{dM} \sigma_M \left( \frac{\delta_c^2}{\sigma_M^2} - 1 \right) \right|. \end{aligned} \quad (16)$$

Without knowledge of all higher cumulants one is forced to use approximate expressions for the mass function. By truncating the Edgeworth expansion at the linear order in the skewness, the resulting PDF is no longer positive-definite. The number of terms that should be kept in the expansion depends on  $\delta_c(z)/\sigma_M$ . The truncation used is a good approximation of the true PDF if  $\delta_c(z)/\sigma_M$  is small

(and non-Gaussianity is small) but for rare events (the tails of the distribution)  $\delta_c(z)/\sigma_M$  is large. One thus expect this approximation to break down for large masses, high redshift, and high  $f_{\text{NL}}$ .

Reference [9] quantified the range of validity of their approximation by assuming that when terms proportional to  $S_3^2$  become important is no longer valid to neglect terms proportional to higher-order cumulants. Then they define the validity regime of their mass function to be where corrections from the  $S_3^2$  are unimportant. They find, as expected, that for very massive objects the approximation breaks down and that the upper mass limit for applicability of the mass function decreases with redshift and  $f_{\text{NL}}$ . But for low masses, redshifts and  $f_{\text{NL}}$  their formula is better than the MVJ. On the other hand the MVJ range of validity extends to higher masses, redshifts and  $f_{\text{NL}}$  values, as expected, as MVJ applied the saddle point approximation to  $\mathcal{P}( > \delta_c | M, z)$  which is an increasingly good approximation for rare objects.

Of course, the natural observable to apply this method to are not only galaxy surveys (and the clusters found there), but, especially suited, are the mass-selected large clusters surveys offered by on-going Sunyaev-Zeldovich experiments (e.g., Planck, ACT, SPT).

A detailed comparison with N-body simulations is the next logical step to pursue.

**3.3. Comparison with N-Body Simulations.** Before we proceed we should consider that the Press and Schechter formulation of the mass function even in the Gaussian initial conditions case, can be significantly improved see for example, [34–36]. Much improved expressions have been extensively calibrated on Gaussian initial conditions N-body simulations. The major limitations in both the MVJ and LMSV derivations (since they follow the classic Press and Schechter formulation) are the assumption of spherical collapse and the sharp  $k$ -space filtering. In addition, the excursion set improvement on the original Press and Schechter swindle relies on the random-phase hypothesis, which is not satisfied for non-Gaussian initial conditions. Since these improvements of the mass function have not yet been generalized to generic non-Gaussian initial conditions (but work is on-going, see other contributions in this issue) the analytical results above should be used to model fractional *corrections* to the Gaussian case.

Thus the non-Gaussian mass function,  $n_{\text{NG}}(M, z)$  can be written as a function of a Gaussian one,  $n_{\text{G}}(M, z)$  (accurately calibrated on N-body simulations) with a non-Gaussian *correction factor*  $\mathcal{R}$  (see e.g., [9, 37]):

$$n_{\text{NG}}(M, z) = n_{\text{G}}(M, z) \mathcal{R}(S_3, M, z), \quad (17)$$

where

$$\mathcal{R}(S_3, M, z) = \frac{\hat{n}(M, z, f_{\text{NL}})}{\hat{n}(M, z, f_{\text{NL}} = 0)} \quad (18)$$

and  $\hat{n}$  is given by the MVJ or LMSV approximation. The correction  $\mathcal{R}$  can then be calibrated on N-body simulations.

Reference [33] argues that the same correction that is in the Gaussian case modifies the collapse threshold,  $\delta_c$ , to

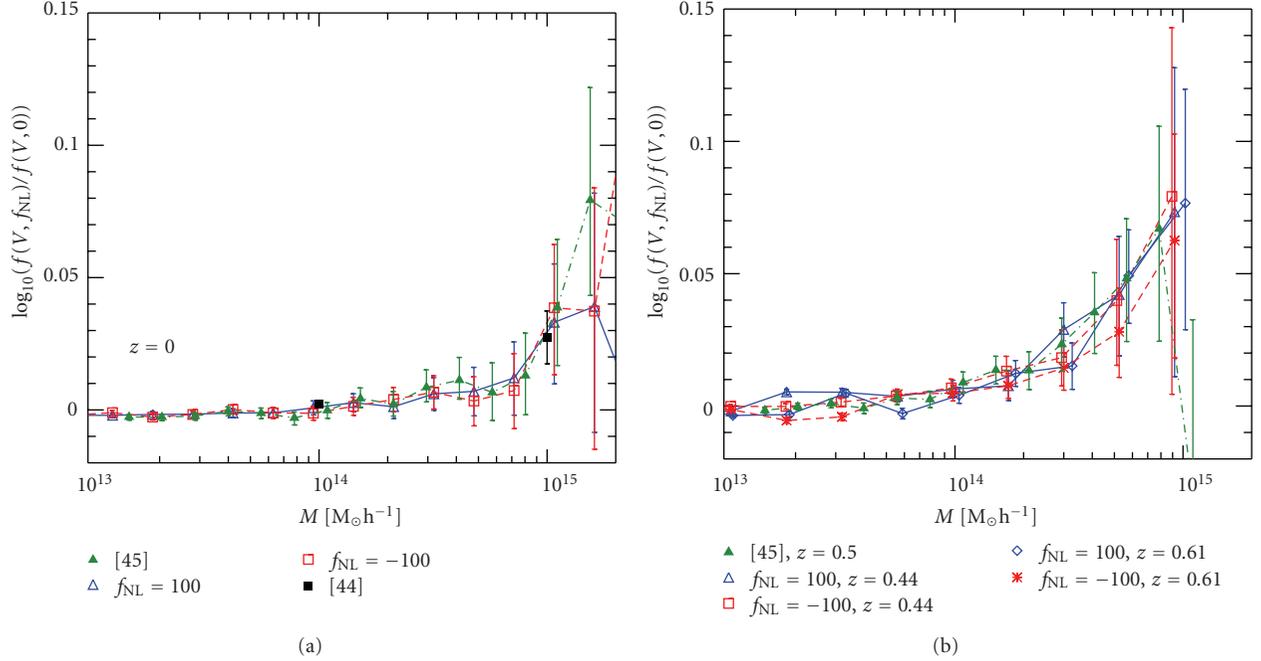


FIGURE 2: Correction to the Gaussian mass function as measured in different non-Gaussian simulations. There is now agreement between different simulations. The y axis should be interpreted as  $\text{Log}_{10}\mathcal{R}(M, z, f_{\text{NL}})$ . For negative values of  $f_{\text{NL}}$  the absolute value of the non-Gaussian correction is considered. Reproduced from [33, Figures 4 and 5].

improve over the original Press and Schechter formulation, may apply to the non-Gaussian correction. The detailed physical interpretation of this is still matter of debate in the literature [38–40]. In summary, reference [33] proposes to write the non-Gaussian correction factor for the MVJ [31] case as

$$\begin{aligned} \mathcal{R}_{\text{NG}}(M, z, f_{\text{NL}}) &= \exp\left[\delta_{ec}^3 \frac{S_{3,M}}{6\sigma_M^2}\right] \\ &\times \left| \frac{1}{6} \frac{\delta_{ec}}{\sqrt{1 - \frac{\delta_{ec} S_{3,M}}{3}}} \frac{dS_{3,M}}{d \ln \sigma_M} + \sqrt{1 - \frac{\delta_{ec} S_{3,M}}{3}} \right|, \end{aligned} \quad (19)$$

and for the LMSV [9] case:

$$\begin{aligned} \mathcal{R}_{\text{NG}}(M, z, f_{\text{NL}}) &= 1 + \frac{1}{6} \frac{\sigma_M^2}{\delta_{ec}} \\ &\times \left[ S_{3,M} \left( \frac{\delta_{ec}^4}{\sigma_M^4} - 2 \frac{\delta_{ec}^2}{\sigma_M^2} - 1 \right) + \frac{dS_{3,M}}{d \ln \sigma_M} \left( \frac{\delta_{ec}^2}{\sigma_M^2} - 1 \right) \right], \end{aligned} \quad (20)$$

where  $\delta_{ec}$  denotes the modified critical density for collapse, which for high peaks is  $\delta_{ec} \sim \delta_c \sqrt{q}$ . Reference [33] calibrated these expressions on N-body simulations to find  $q = 0.75$ . We anticipate here that the validity of this extrapolation (i.e.,

in terms of a correction to the critical collapse threshold) can be tested independently on the large-scale non-Gaussian halo bias as described in Section 4. Note that, in both cases, in the limit of small non-Gaussianity the correction factors reduce to

$$\mathcal{R} = 1 + S_{3,M} \frac{\delta_{ec}^3}{6\sigma_M^2}. \quad (21)$$

Non-Gaussian mass functions have been computed from simulations and compared with different theoretical predictions in several works [41–45]. In the past, conflicting results were reported, but the issue seems to have been settled, there is agreement among mass function measured from different non-Gaussian simulations performed by three different groups as shown for example in Figure 2. As expected both MVJ and LMSV prescriptions for the non-Gaussian correction to the mass function agree with the simulation results, provided one makes the substitution  $\delta_c \rightarrow \delta_{ec}$ , with some tentative indication that MVJ may be better for very massive objects while LMSV performs better for less rare events. This is shown in Figure 3 where the points represent measurements from N-body simulations presented in [33].

**3.4. Voids.** While galaxy clusters form at the highest overdensities of the primordial density field and probe the high-density tail of the PDF, voids form in the low-density regions and thus probe the low-density tail of the PDF. Most of the volume of the evolved universe is underdense, so it seems interesting to pay attention to the distribution of underdense regions. A void distribution function can be derived in an

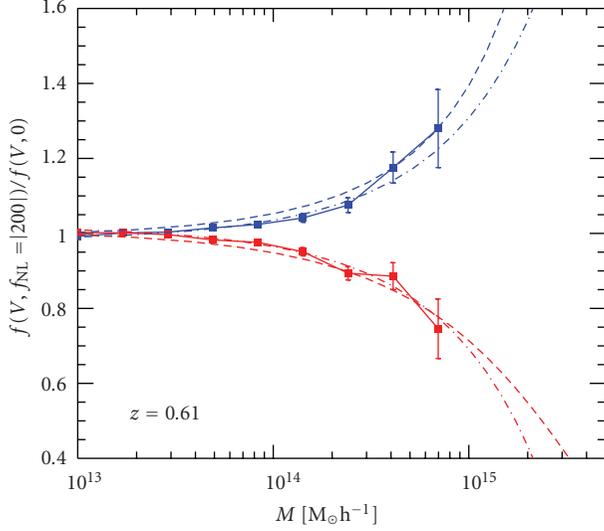


FIGURE 3: The points show the non-Gaussian correction to the mass function as measured in the N-body simulations of [33]. Blue corresponds to  $f_{\text{NL}} = 200$  and red to  $f_{\text{NL}} = -200$ . The dashed lines correspond to the MVJ formulation and the dot-dashed lined to the LMSV formulation. In both cases the substitution  $\delta_c \rightarrow \delta_{ec}$  has been performed. The y axis should be interpreted as  $\mathcal{R}(M, z, f_{\text{NL}} = 200)$ . Reproduced from Figure 7 [33].

analogous way to the Press Schechter mass function by realizing that negative density fluctuations grow into voids [32], a critical underdensity  $\delta_v$  is necessary for producing a void and this plays the role of the critical overdensity  $\delta_c$  for producing bound objects (halos). The more underdense a void is the more negative  $\delta_v$  becomes. The precise value of  $\delta_v(z)$  depends on the precise definition of a void (and may depend on the the observables used to find voids); realistic values of  $\delta_v(z = 0)$  are expected to be  $\gtrsim -1$ . In the absence of a better prescription, here, following [32],  $\delta_v$  is treated as a phenomenological parameter and results are shown for a range of  $\delta_v$  values. To derive the non-Gaussian void probability function one proceeds as above with the only subtlety that  $\delta_v$  is negative and that  $\mathcal{P}(< \delta) = 1 - \mathcal{P}(> \delta)$  thus  $|d\mathcal{P}(< \delta)/dM| = |d\mathcal{P}(> \delta)/dM|$ . Thus the void PDF as a function of  $|\delta_v|$  can be obtained from the PDF of MVJ [31] or LMSV [9], provided one keeps track of the sign of each term. For example in the LMSV approximation the void distribution function becomes [32]

$$\begin{aligned} \hat{n}(R, z, f_{\text{NL}}) &= \frac{9}{2\pi^2} \sqrt{\frac{\pi}{2}} \frac{1}{R^4} e^{-\delta_c^2/2\sigma_M^2} \\ &\times \left\{ \left| \frac{d \ln \sigma_M}{d \ln M} \right| \right. \\ &\times \left[ -\frac{|\delta_v|}{\sigma_M} \frac{S_{3,M} \sigma_M}{6} \left( \frac{\delta_v^4}{\sigma_M^4} - 2 \frac{\delta_v^2}{\sigma_M^2} - 1 \right) \right] \\ &\left. + \frac{1}{6} \frac{dS_3}{dM} \sigma_M \left( \frac{\delta_v^2}{\sigma_M^2} - 1 \right) \right\}, \end{aligned} \quad (22)$$

where the expression is reported as a function of the smoothing radius rather than the mass, since a void Lagrangian radius is probably easier to determine than its mass.

Note that while a positive skewness ( $f_{\text{NL}} > 0$ ) boosts the number of halos at the high-mass end (and slightly suppress the number of low-mass halos), it is a negative skewness that will increase the voids size distribution at the largest voids end (and slightly decrease it for small void sizes). Reference [32] concluded that the abundance of voids is sensitive to non-Gaussianity,  $|\delta_v|$  is expected to be smaller than  $\delta_c$  by a factor 2 to 3. If voids probe the same scales as halos then they should provide constraints on  $f_{\text{NL}}$  2 to 3 times worse. However voids may probe slightly larger scales than halos, in many non-Gaussian models,  $S_3^{f_{\text{NL}}=1}$  increases with scales (see e.g., Figure 1), compensating for the threshold.

The approach reviewed here provides a rough estimate of the fractional change in abundance due to primordial non-Gaussianity but will not provide reliably the abundance itself. It is important to stress here that rigorously quantitative results will need to be calibrated on cosmological simulations and mock survey catalogs.

#### 4. Effects on the Halo Power Spectrum

Recently, reference [43, 46] showed that primordial non-Gaussianity affects the clustering of dark matter halos (i.e., density extrema) inducing a scale-dependent bias for halos on large scales. This can be seen for example by considering halos as regions where the (smoothed) linear density field exceeds a suitable threshold. All correlations and peaks considered in the section are those of the *initial* density field (linearly extrapolated to the present time). Thus for example in the Gaussian case [47–49] for high peaks we would have the following relation between the correlation function of halos of mass  $M$ ,  $\xi_{h,M}(r)$  and that of the dark matter distribution smoothed on scale  $R$ , corresponding to mass  $M$ ,  $\xi_R(r)$ :

$$\xi_{h,M}(r) \simeq \left( b_{h,L}^G \right)^2 \xi_R(r), \quad (23)$$

where  $b_L^G \simeq \delta_c/\sigma_R^2$  denotes the Lagrangian halo bias (in the Gaussian case), although more refined expressions can be found in for example, [50, 51].

The Lagrangian bias appears here because correlations and peaks are those of the initial density field (linearly extrapolated). Making the standard assumptions that halos move coherently with the underlying dark matter, one can obtain the final Eulerian bias as  $b_E = 1 + b_L$ , using the techniques outlined in [50–53]. Below we will omit the subscript  $E$  for Eulerian bias.

The two-point correlation function of regions above a high threshold has been obtained, for the general non-Gaussian case, in [54–56]:

$$\xi_{h,M}(|\mathbf{x}_1 - \mathbf{x}_2|) = -1 + \exp[X], \quad (24)$$

where

$$X = \sum_{N=2}^{\infty} \sum_{j=1}^{N-1} \frac{\nu^N \sigma_R^{-N}}{j!(N-j)!} \xi^{(N)} \left[ \begin{matrix} \mathbf{x}_1, \dots, \mathbf{x}_1, & \mathbf{x}_2, \dots, \mathbf{x}_2 \\ j \text{ times} & (N-j) \text{ times} \end{matrix} \right], \quad (25)$$

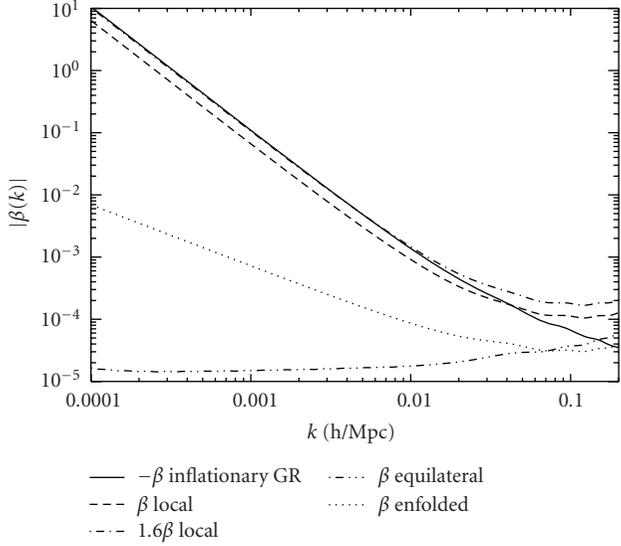


FIGURE 4: The scale-dependence of the large-scale halo bias induced by a nonzero bispectrum for different types of non-Gaussianity. The dashed line corresponds to the local type and the dot-dot-dotted to equilateral type. Figure reproduced from [57].

where  $\nu = \delta_c \sigma_R$ . For large separations the exponential can be expanded to first order. This is what we will do in what follows but we will comment on this choice below.

For small non-Gaussianities (Effectively that is for values of  $f_{\text{NL}}$  consistent with observations.), we can keep terms up to the three-point correlation function  $\xi^{(3)}$ , obtaining that the correction to the halo correlation function,  $\Delta\xi_h$  due to a non-zero three-point function is given by:

$$\begin{aligned} \Delta\xi_h &= \frac{\nu_R^3}{2\sigma_R^3} \left[ \xi_R^{(3)}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_2) + \xi_R^{(3)}(\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_2) \right] \\ &= \frac{\nu_R^3}{\sigma_R^3} \xi_R^{(3)}(\mathbf{x}_1, \mathbf{x}_1, \mathbf{x}_2). \end{aligned} \quad (26)$$

For a general bispectrum  $B(k_1, k_2, k_3)$  this yields a correction to the power spectrum (see [46] for steps in the derivation):

$$\begin{aligned} \frac{\Delta P}{P} &= \frac{\delta_c(z)}{\mathcal{M}_R(k)} \frac{1}{4\pi^2 \sigma_R^2} \int dk_1 k_1^2 \mathcal{M}_R(k_1) \\ &\times \int_{-1}^1 d\mu \mathcal{M}_R(\sqrt{\alpha}) \frac{\beta_\phi(k_1, \sqrt{\alpha}, k)}{P_\phi(k)}, \end{aligned} \quad (27)$$

where we have made the substitution  $\alpha = k_1^2 + k^2 + 2k_1 k \mu$ . Here  $\mathcal{M}_R = W_R \mathcal{M}$  and  $W_R$  and  $\mathcal{M}$  were introduced in Section 3. The effect on the halo bias is  $\Delta b_h^L/b_h^L = (1/2)(\Delta P/P)$  and thus

$$b_h^{\text{fNL}} = 1 + \frac{\Delta_c(z)}{\sigma_R^2 D^2(z)} [1 + \delta_c(z) \beta_R(k)], \quad (28)$$

where the expression for  $\beta$  can be obtained by comparing to (27). The term  $\Delta_c(z)/[\sigma_R^2 D^2(z)] \simeq b^G - 1$  can be recognized as the Gaussian Lagrangian halo bias.

So far the derivation is generic for all types of non-Gaussianity specified by a given bispectrum. We can then

consider specific cases. In particular for local non-Gaussianity we obtain

$$\begin{aligned} \beta_R(k) &= \frac{2f_{\text{NL}}}{8\pi^2 \sigma_R^2 \mathcal{M}_R(k)} \int dk_1 k_1^2 \mathcal{M}_R(k_1) P_\phi(k_1) \\ &\times \int_{-1}^1 d\mu \mathcal{M}_R(\sqrt{\alpha}) \left[ \frac{P_\phi(\sqrt{\alpha})}{P_\phi(k)} + 2 \right]. \end{aligned} \quad (29)$$

Thus  $\Delta b_h/b_h$  is  $2f_{\text{NL}}$  times a redshift-dependent factor  $\Delta_c(z)/D(z) = \delta_c(z)$ , times a  $k$ - and mass-dependent factor. The function  $\beta_R(k)$  is shown as the dashed line in Figure 4. This result for the local non-Gaussianity has been derived in at least three other ways. Reference [43] generalizes the Kaiser [47] argument of high peak bias for the local non-Gaussianity. Starting from  $\nabla^2 \Phi = \nabla^2 \phi + 2f_{\text{NL}}[\phi \nabla^2 \phi + |\nabla \phi|^2]$  where near peaks  $|\nabla \phi|^2$  is negligible they obtain  $\delta = \delta^{\text{fNL}=0} [1 + 2f_{\text{NL}} \phi]$ . The Poisson equation to convert  $\phi$  in  $\delta$  then gives the scale-dependence. More details are presented elsewhere in this issue.

Reference [58] works in the peak-background split. This approach is especially useful to understand that it is the coupling between very large and small scales introduced by local (squeezed-configurations) non-Gaussianity to boost (or suppress) the peaks clustering. In this approach, the density field can be written as  $\rho(\vec{x}) = \bar{\rho}(1 + \delta_l + \delta_s)$  where  $\delta_l$  denotes long wavelength fluctuations and  $\delta_s$ , short wavelength fluctuations.  $\delta_l$  is the one responsible for modulating halo formation (i.e., to boost peaks above the threshold for collapse), so the halo number density is  $n = \bar{n}(1 + b_{h,L} \delta_l)$  and  $b_{h,L} = \bar{n}^{-1} \partial n / \partial \delta_l$ .

In the local non-Gaussian case they decompose the Gaussian field  $\phi$  as a combination of long and short wavelength fluctuations  $\phi = \phi_l + \phi_s$  thus  $\Phi = \phi_l + f_{\text{NL}} \phi_l^2 + (1 + 2f_{\text{NL}} \phi_l) \phi_s + f_{\text{NL}} \phi_s^2 + \text{const}$ . Also in this non-Gaussian case one can split the density field  $\delta$  in  $\delta_l$  and  $\delta_s$  and relate this to  $f_{\text{NL}}$  (it is easier to work in Fourier space):  $\delta_l(k) = \alpha(k) \phi_l(k)$  and  $\delta_s = \alpha(k) [(1 + 2f_{\text{NL}} \phi_l) \phi_s + f_{\text{NL}} \phi_s^2] \equiv \alpha(k) [X_1 \phi_s + X_2 \phi_s^2]$ , the last equality giving the definition of  $X_1$  and  $X_2$ . Note that  $\delta_s$  cannot be ignored here because  $\phi_l$  enters in  $X_1$ , in other words, local non-Gaussianity couples long and short wavelength modes. The local halo number density is now function of  $\delta_l$ ,  $X_1$ , and  $X_2$  yielding the following result for the Lagrangian halo bias:

$$\begin{aligned} b_{h,L} &= \bar{n}^{-1} \left[ \frac{\partial n}{\partial \delta_l} + 2f_{\text{NL}} \frac{d\phi_l}{d\delta_l} \frac{\partial n}{\partial X_1} \right] \\ &= b_{h,L}^G + 2f_{\text{NL}} \frac{d\phi_l}{d\delta_l} \frac{\partial \ln n}{\partial \ln \sigma_8} \\ &\equiv b_{h,L}^G \left( 1 + 2f_{\text{NL}} \alpha(k)^{-1} \delta_c \right), \end{aligned} \quad (30)$$

where  $\alpha(k)$  encloses the scale-dependence of the effect. Reference [59] rederives the ellipsoidal collapse for small deviations from Gaussianity of the local type. They find that a non-zero  $f_{\text{NL}}$  modifies the threshold for collapse, the modification is proportional to  $f_{\text{NL}}$ . This should sound familiar from Section 3. They then use the definition  $b_{h,L} = \bar{n}^{-1} \partial n / \partial \delta_c$  keeping track of the fact that  $\delta$  is “modulated” by  $f_{\text{NL}}$ .

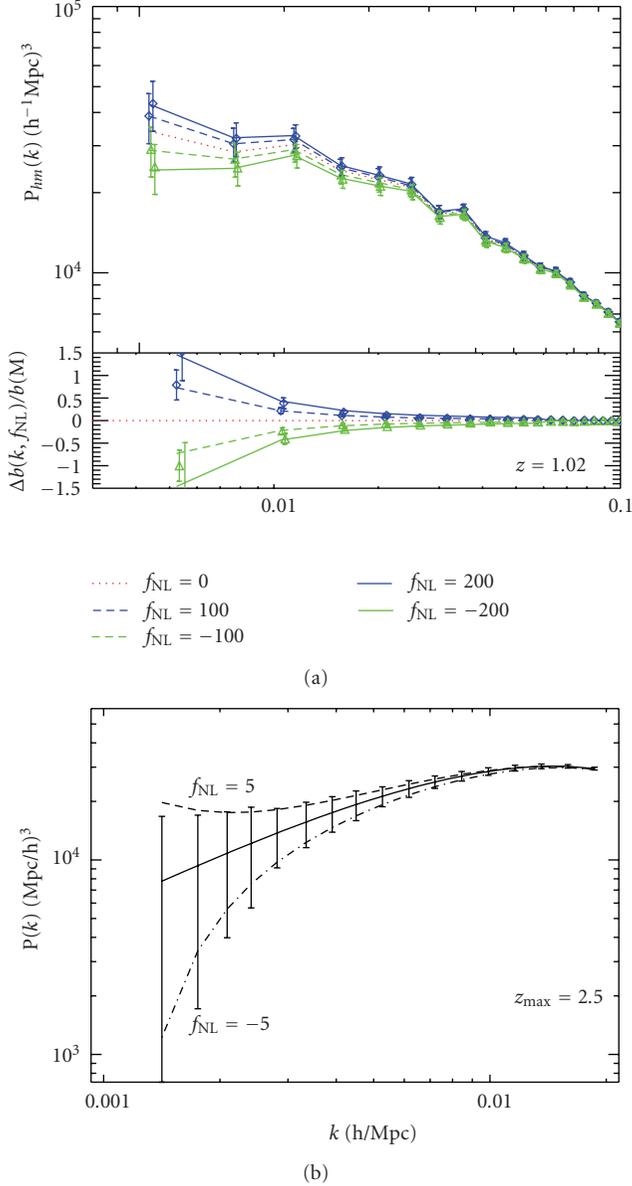


FIGURE 5: Effect of the non-Gaussian halo bias on the power spectrum. In the left-top of (a) panel we show the halo-matter cross-power spectrum for masses above  $10^{13} M_{\odot}$  at  $z = 1.02$ . The left-bottom panel of (a) shows the ratio of the non-Gaussian to Gaussian bias. Figure reproduced from [33]. The  $f_{NL}$  values reported in the figure legend should be interpreted as  $f_{NL}^{LSS}$ . On the right panel we show the expected effect and error-bars for the large-scale power spectrum for a survey like LSST. Figure reproduced from [26].

The effect of the non-Gaussian halo bias on the power spectrum is shown in Figure 5 where the points are measurements from an N-body simulations of [33] (see figure caption for more details).

The above result

$$\Delta b = f_{NL} \delta_c (b_h^G - 1) \beta_R^{f_{NL}=1}(k), \quad (31)$$

where  $\beta_R(k) = f_{NL} \beta_R^{f_{NL}=1}(k)$ , can be improved in several ways.

First of all, we have not made any distinction between the redshift at which the object is being observed ( $z_o$ ) and that at which is being formed ( $z_f$ ). Except for the rarest events this should be accounted for. The Gaussian Lagrangian bias expression used so far is an approximation, a more accurate expression is [50, 52, 53]

$$b_{h,L}^G(z_o, M, z_f) = \frac{1}{D(z_o)} \left[ \frac{\delta_c(z_f)}{\sigma_M^2} - \frac{1}{\delta_c(z_f)} \right]. \quad (32)$$

Then, the halo bias expressions are derived within the ‘‘classical’’ Press and Schechter theory, as we have seen in Section 3, subsequent improvements on the mass function can be seen as a correction to the collapse threshold. In the expression for the Gaussian halo bias  $b_L^G = \bar{n}^{-1} \partial n^G / \partial \delta_c$  one can consider mass functions that are better fit to simulations than the standard Press and Schechter one obtaining:

$$b_{h,L}^G(z_o, M, z_f) = \frac{1}{D(z_o)} \left[ \frac{q \delta_c(z_f)}{\sigma_M^2} - \frac{1}{\delta_c(z_f)} \right] + \frac{2p}{\delta_c(z_f) D(z_o)} \left[ 1 + \left( \frac{q \delta_c^2(z_f)}{\sigma_M^2} \right)^p \right]^{-1}. \quad (33)$$

The parameters  $q$  and  $p$  account for nonspherical collapse and fit to numerical simulations yield  $q \sim 0.75$ ,  $p = 0.3$  for example, [34]. In this expression the term in the second line is usually subdominant. The term ‘‘ $-1/\delta_c$ ’’ in the first line is known as ‘‘antibias’’, and it becomes negligible for old halos  $z_f \gg z_o$ . Note that by including the antibias correction in  $b_g$  of (31) one recovers the ‘‘recent mergers’’ approximation of [58].

The same correction should also apply to the non-Gaussian correction to the halo bias:

$$\Delta b = f_{NL} q' \delta_c (b_h^G - 1) \beta_R^{f_{NL}=1}(k), \quad (34)$$

where  $q'$  should coincide with  $q$  above; it can be calibrated to N-body simulations and is found indeed to be  $q' = 0.75$  [33]. The non-Gaussian halo bias prediction and results from N-body simulations with local non-Gaussianity are shown in Figure 6.

Finally one may note that for  $f_{NL}$  large and negative, (27) and (28) would formally yield  $b_h^{f_{NL}}$  and  $P_h(k)$  negative on large enough scales. This is a manifestation of the breakdown of the approximations made: (a) all correlations of higher order than the bispectrum were neglected, for large NG this truncation may not hold; (b) The exponential in (24) was expanded to linear order. This however could be easily corrected for, remembering that the  $P(k)$  obtained from (27) is in reality the Fourier transform of  $X$ , the argument of the exponential. One would then compute the halo correlation function using (24) and Fourier transforming back to obtain the halo power-spectrum.

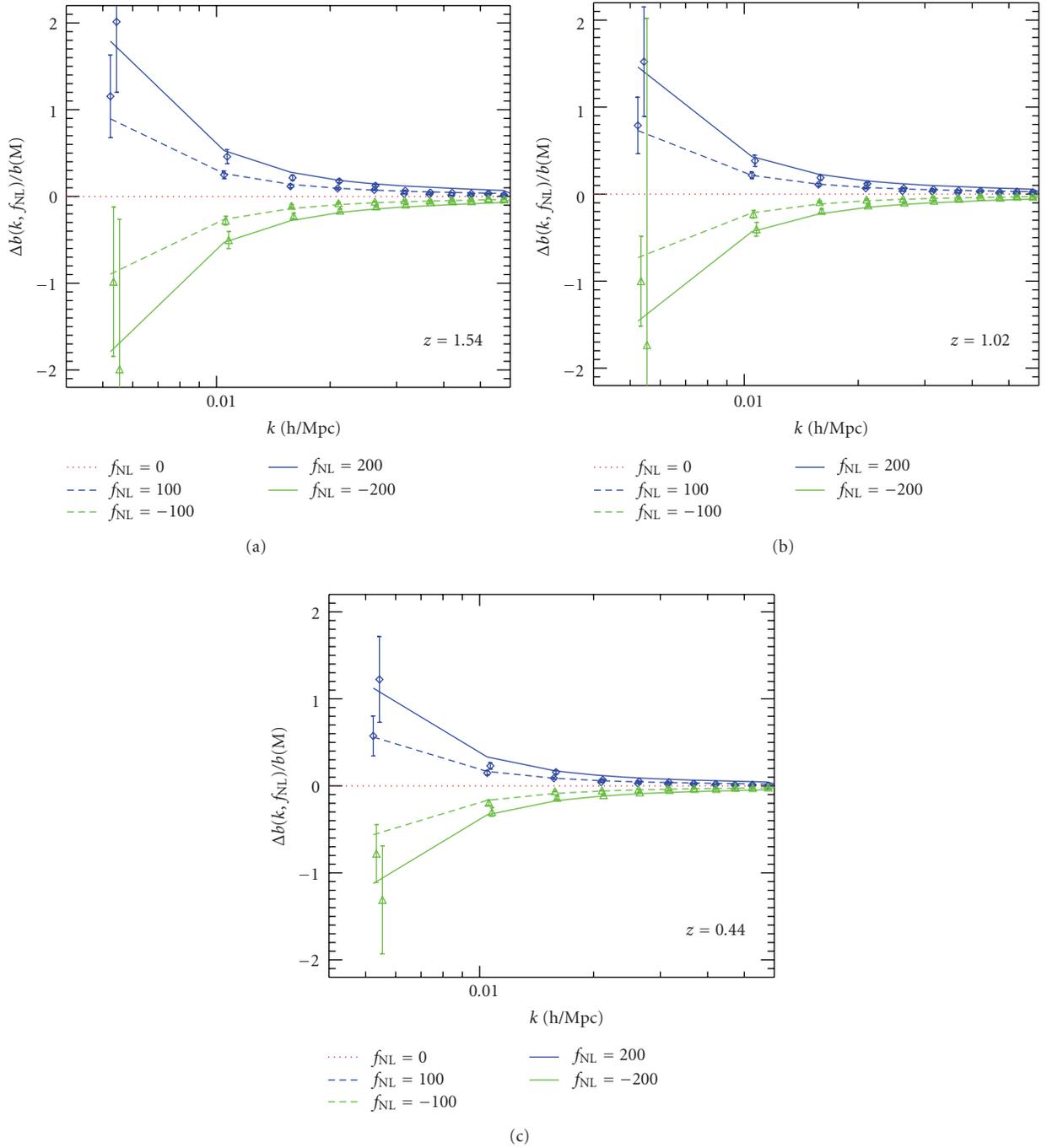


FIGURE 6: The quantity  $\Delta b/b$  as function of  $k$ , for simulation snapshots at  $z = 0.44, 1.02,$  and  $1.54$ . Simulation outputs and theory lines are shown for  $f_{\text{NL}} = \pm 100$  and  $f_{\text{NL}} = \pm 200$ . Figure reproduced from [33]. The  $f_{\text{NL}}$  values reported in this figure legend should be interpreted as  $f_{\text{NL}}^{\text{LSS}}$ .

So far we have concentrated on local non-Gaussianity, but the expression of (27) and (28) is more general. Using this formulation, reference [57] computed the quantity  $\beta_R^{f_{\text{NL}}=1}(k)$  for several types of non-Gaussianity (equilateral, local, and enfolded); this is shown in Figure 4. It is clear that the non-Gaussian halo-bias effect has some sensitivity to the bispectrum shape, for example the effect for the equilateral type of non-Gaussianity is suppressed by orders

of magnitude compared to the local-type and the flattened case is somewhere in the middle. Figure 4 also shows a type of non-Gaussianity arising from General-relativistic (GR) corrections on scales comparable to the Hubble radius. Note that perturbations on super-Hubble scales are initially needed in order to “feed” the GR correction terms. In this respect the significance of this contribution is analogous to the well-known large-scale anticorrelation between CMB

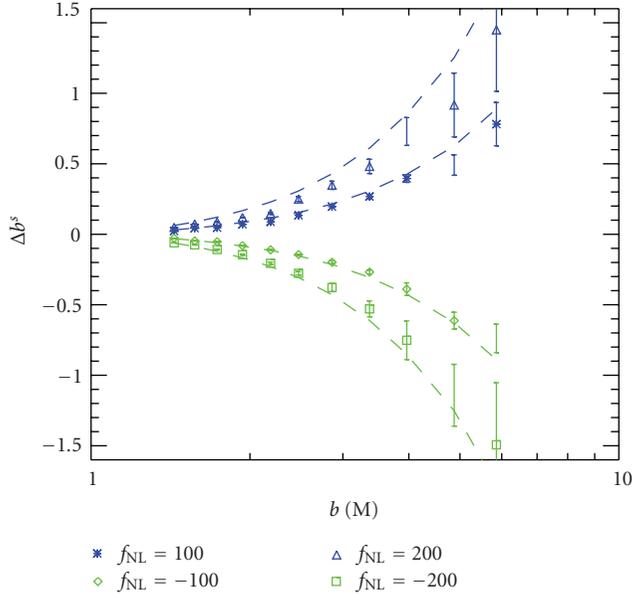


FIGURE 7: Non-Gaussian halo bias correction as function of the Gaussian halo bias. Figure reproduced from [33]. The  $f_{\text{NL}}$  values reported in the figure legend should be interpreted as  $f_{\text{NL}}^{\text{LSS}}$ .

temperature and E-mode polarization, it is a consequence of the properties of the inflationary mechanism to lay down the primordial perturbations. This effect has the same magnitude as a local non-Gaussianity with  $f_{\text{NL}} \gtrsim 1$ .

The next logical step is then to ask how well present or forthcoming data could constrain non-Gaussianity using the halo-bias effect. It is interesting to note that surveys that aim at measuring Baryon Acoustic Oscillations (BAO) in the galaxy distribution to constrain dark energy are well suited to also probe non-Gaussianity; they cover large volumes and their galaxy number density is well suited so that on the scale of interest (both for BAO and non-Gaussianity) shot noise does not dominate the signal. Photometric surveys are also well suited: as the non-Gaussian signal is localized at very large scales and is a smooth function of  $k$ , the photo- $z$  smearing effects are unimportant.

The theory developed so far describes the clustering of halos while we observe galaxies. Different galaxy populations occupy dark matter halos following different prescriptions. If we think in the halo-model framework (e.g., [60] and references therein) at very large scales only the “two-halo” contribution matters and the details of the halo occupation distribution (the “one-halo” term) become unimportant.

What is important to keep in mind is that the effect of the non-Gaussianity parameter one wants to measure,  $f_{\text{NL}}$ , is fully degenerate with the value of the Gaussian (small scales) halo bias. Figure 7 shows the dependence of the non-Gaussian correction on the Gaussian bias.

Thus highly biased tracers will show a larger non-Gaussian effect for the same  $f_{\text{NL}}$  value. Of course for a given cosmological model the Gaussian bias can be measured accurately by comparing the predicted dark matter power spectrum with the observed one. Alternatively, two

TABLE 1: Current recent  $2 - \sigma$  constraints on local  $f_{\text{NL}}$ .

Data/method	$f_{\text{NL}}$	Reference
Photometric LRG-bias	$63_{-331}^{+101}$	[58]
Spectroscopic LRG-bias	$70_{-191}^{+139}$	[58]
QSO-bias	$8_{-77}^{+47}$	[58]
Combined	$28_{-57}^{+42}$	[58]
NVSS-ISW	$105_{-1157}^{+755}$	[58]
NVSS-ISW	$236 \pm 127(1 - \sigma)$	[59]
WMAP3-Bispectrum	$30 \pm 84$	[10]
WMAP3-Bispectrum	$32 \pm 68$	[8]
WMAP3-Bispectrum	$87 \pm 60$	[20]
WMAP-Bispectrum	$38 \pm 42$	[11, 17, 21]
WMAP5-Bispectrum	$51 \pm 60$	[10]
WMAP5-Minkowski	$-57 \pm 121$	[10]

TABLE 2: Forecasts  $1 - \sigma$  constraints on local  $f_{\text{NL}}$ .

Data/method	$\Delta f_{\text{NL}}(1 - \sigma)$	Reference
BOSS-bias	18	[63]
ADEPT/Euclid-bias	1.5	[63]
PANNStarrs-bias	3.5	[63]
LSST-bias	0.7	[63]
LSST-ISW	7	[59]
BOSS-bispectrum	35	[27]
ADEPT/Euclid-bispectrum	3.6	[27]
Planck-Bispectrum	3	[19]
BPOL-Bispectrum	2	[19]

differently biased tracers can be used in tandem to disentangle the two effects [61, 62].

Since clustering amplitude may depend on the entire halo history, it becomes then interesting to model in details the dependence of the effect on the halo merger tree (Reid et al., in preparation).

*4.1. Outlook for the Future.* How well can this method do to constrain primordial non-Gaussianity compared with the other techniques presented here? The Integrated Sachs Wolfe (ISW) effect offers a window to probe clustering on the largest scales (where the signal is large); on the other hand, a measurement of clustering of tracers of dark matter halos is a very direct window into this effect. A Fisher matrix approach [58, 59, 63] shows that the ISW signal is weighted at relatively low redshift (where dark energy starts dominating) while the non-Gaussian signal grows with redshift, thus making the shape of the halo power spectrum a more promising tool. An overview of current constraints from different approaches can be found in Table 1 and future forecasts in Table 2, for non-Gaussianity of the local type. Large, mass-selected cluster samples as produced by SZ-based experiments will provide a optimally suited data-set for this technique (see e.g., [64]).

While for a given  $f_{\text{NL}}$  model such as the local one, methods that exploit the non-Gaussian bias seem to yield the smallest error-bars for large-scale structure, it should be

kept in mind that the bispectrum can be used to investigate the full configuration dependence of  $f_{\text{NL}}$  and thus is a very powerful tool to discriminate between different type of non-Gaussianity. In addition CMB-bispectrum and halo bias test non-Gaussianity on very large scales while the large scale structure bispectrum mostly probes mildly nonlinear scales. As primordial non-Gaussianity may be scale-dependent, all these techniques are highly complementary.

The above estimates assume that the underlying cosmological model is known. The large-scale shape of the power spectrum can be affected by cosmology. Carbone et al. (in prep.) explore possible degeneracies between  $f_{\text{NL}}$  and cosmological parameters. They find that the parameters that are most strongly correlated with  $f_{\text{NL}}$  are parameters describing dark energy clustering, neutrino mass and running of the primordial power spectrum spectral slope. For surveys that cover a broad redshift range the error on  $f_{\text{NL}}$  degrade little when marginalizing over these extra parameters; the peculiar redshift dependence of the non-Gaussian signal lifts the degeneracy.

## 5. Conclusions

A natural question to ask at this point may be “what observable will have better chances to constrain primordial non-Gaussianity?”

In principle the abundance of rare events is a very powerful probe of non-Gaussianity; however, in practice, it is limited by the practical difficulty of determining the mass of the observed objects and its corresponding large uncertainty in the determination. This point is stressed for example, in [9]. With the advent of high-precision measurements of gravitational lensing by massive clusters, the mass uncertainty, at least for small to moderate size clusters samples can be greatly reduced. Forthcoming Sunyaev-Zeldovich experiments will provide large samples of mass-selected clusters which could then be followed up by lensing mass measurements (see e.g., [65, 66]). So far there is only one very high redshift ( $z = 1.4$ ) very massive  $M \simeq 8 \times 10^{14} M_{\odot}$  with high-precision mass determination via gravitational lensing [67]. Reference [68] pointed out that this object is extremely rare, for Gaussian initial conditions there should be 0.002 such objects or less in the surveyed area, which is uncomfortably low probability. But the cluster mass is very well determined; a non-Gaussianity still compatible with CMB constraints could bring the probability of observing of the object to more comfortable values. This result should be interpreted as a “proof of principle” showing that this a potentially powerful avenue to pursue.

The measurement of the three-point correlation function allows one to map directly the shape-dependence of the bispectrum. For large-scale structures the limiting factors are the large non-Gaussian contribution induced by gravitational evolution and the uncertainty of the nonlinear behavior of galaxy bias.

The halo-bias approach can yield highly competitive constraints, but it is less sensitive to the bispectrum shape. Still, the big difference in the magnitude and shape of

TABLE 3: Forecasted non-Gaussianity constraints: (A) [20] (B) [63] (C) [69, 70] (E) [57]) (F) e.g., [15].

type NG	CMB Bispectrum		Halo bias	
	Planck	BPol	Euclid	LSST
$1 - \sigma$ errors				
Local	3 <sup>(A)</sup>	2 <sup>(A)</sup>	1.5 <sup>(B)</sup>	0.7 <sup>(B)</sup>
Equilateral	25 <sup>(C)</sup>	14 <sup>(C)</sup>	—	—
Enfolded	$\mathcal{O}10$	$\mathcal{O}10$	39 <sup>(E)</sup>	18 <sup>(E)</sup>
$\# \sigma$ detection				
GR	N/A	N/A	1 <sup>(E)</sup>	2 <sup>(E)</sup>
secondaries	3 <sup>(F)</sup>	5 <sup>(F)</sup>	N/A	N/A

the scale-dependent biasing factor between different non-Gaussian models implies that the halo bias can become a useful tool to study shapes when combined with for example, measurements of the CMB bispectrum. Table 3 highlights this complementarity. For example, one could envision different scenarios.

If non-Gaussianity is local with negative  $f_{\text{NL}}$  and CMB obtains a detection, then the halo bias approach should also give a high-significance detection (GR correction and primordial contributions add up), while if it is local but with positive  $f_{\text{NL}}$ , the halo-bias approach could give a lower statistical significance for small  $f_{\text{NL}}$  as the GR correction contribution has the opposite sign.

If CMB detects  $f_{\text{NL}}$  at the level of  $\sim 10$  and of a form that is close to local, but halo bias does not detect it, then the CMB bispectrum is given by secondary effects.

If CMB detects non-Gaussianity but is not of the local type, then halo bias can help discriminate between equilateral and enfolded shapes; if halo bias sees a signal it indicates enfolded type; if halo bias does not see a signal it indicates equilateral type. Thus even a nondetection of the halo-bias effect, in combination with CMB constraints can have an important discriminatory power.

In any case, if the simplest inflationary scenario holds, for surveys like Euclid and LSST, the halo-bias approach is expected to detect a non-Gaussian signal very similar to the local type signal with an amplitude of  $f_{\text{NL}} \sim -1.5$  which is due to large-scales GR corrections to the Poisson equation. This effect should leave no imprint in the CMB; once again the combination of the two observable can help enormously to discriminate among models for the origin of cosmological structures.

In addition we should bear in mind that non-Gaussianity may be scale-dependent. In fact for models like DBI inflation it is expected to be scale-dependent. A proposed parameterization of the scale-dependence of non-Gaussianity is given by

$$B_{\phi}(\vec{k}_1, \vec{k}_2, \vec{k}_3) = f_{\text{NL}} \left( \frac{K}{k_p} \right)^{n_{\text{ng}}} F(\vec{k}_1, \vec{k}_2, \vec{k}_3), \quad (35)$$

where  $K = (k_1 k_2 k_3)^{1/3}$  [70],  $k_p$  denotes the pivot and  $n_{\text{ng}}$  the slope or running of non-Gaussianity, although other authors prefer to use  $K = (k_1 + k_2 + k_3)/3$  [9, 59] as for

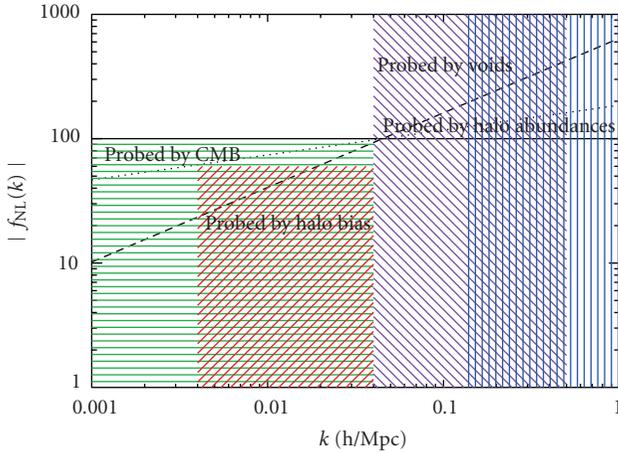


FIGURE 8: Scale-dependent  $f_{\text{NL}}$  and scales probed by different approaches mentioned here. The solid line has  $n_g = 0$ , the dotted line has  $n_g = 0.2$ , and the dashed one has  $n_g = 0.6$ . Hashed areas for CMB and halo-bias show allowed regions.

squeezed configurations  $K \neq 0$ . It is still an open issue which parameterization is better in practice.

In any case different observables probe different scales (see Figure 8) and their complementary means that “the combination is more than the sum of the parts.”

What is clear, however, is that the thorny systematic effects that enter in all these approaches will require that a variety of complementary avenues be taken to establish a robust detection of primordial non-Gaussianity.

## Acknowledgments

This work is supported by MICCIN Grant AYA2008-03531 and FP7-IDEAS-Phys.LSS 240117. The author would like to thank her closest collaborators in many of the articles reviewed here: Carmelita Carbone, Klaus Dolag, Margherita Grossi, Alan Heavens, Raul Jimenez, Marc Kamionkowski, Marilena LoVerde, Sabino Matarrese, Lauro Moscardini, Sarah Shandera, and her collaborators for the reported work-in-progress: Carmelita Carbone, Olga Mena, Beth Reid. She would also like to thank the referee for a careful review of the manuscript.

## References

- [1] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, “Non-Gaussianity from inflation: theory and observations,” *Physics Reports*, vol. 402, no. 3-4, pp. 103–266, 2004.
- [2] D. S. Salopek and J. R. Bond, “Nonlinear evolution of long-wavelength metric fluctuations in inflationary models,” *Physical Review D*, vol. 42, no. 12, pp. 3936–3962, 1990.
- [3] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, “The three-point correlation function of the cosmic microwave background in inflationary models,” *The Astrophysical Journal*, vol. 430, no. 2, pp. 447–457, 1994.
- [4] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, “Large-scale structure, the cosmic microwave background and primordial non-Gaussianity,” *Monthly Notices of the Royal Astronomical Society*, vol. 313, no. 1, pp. 141–147, 2000.
- [5] E. Komatsu and D. N. Spergel, “Acoustic signatures in the primary microwave background bispectrum,” *Physical Review D*, vol. 63, no. 6, Article ID 063002, 2001.
- [6] D. Babich, P. Creminelli, and M. Zaldarriaga, “The shape of non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, no. 8, pp. 199–217, 2004.
- [7] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, “Limits on non-Gaussianities from WMAP data,” *Journal of Cosmology and Astroparticle Physics*, no. 5, article 004, 2006.
- [8] P. Creminelli, L. Senatore, M. Zaldarriaga, and M. Tegmark, “Limits on  $f_{\text{NL}}$  parameters from Wilkinson Microwave Anisotropy Probe three-year data,” *Journal of Cosmology and Astroparticle Physics*, no. 3, article 005, 2007.
- [9] M. LoVerde, A. Miller, S. Shandera, and L. Verde, “Effects of scale-dependent non-Gaussianity on cosmological structures,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article 014, 2008.
- [10] E. Komatsu, J. Dunkley, M. R. Nolta, et al., “Five-year Wilkinson microwave anisotropy probe (WMAP) observations: cosmological interpretation,” *The Astrophysical Journal Supplement Series*, vol. 180, no. 2, article 330, 2008.
- [11] E. Komatsu, N. Afshordi, N. Bartolo, et al., “Non-Gaussianity as a probe of the physics of the primordial universe and the astrophysics of the low redshift universe,” in *AGB Stars and Related Phenomena: The Astronomy and Astrophysics Decadal Survey*, vol. 2010 of *Astronomy*, p. 158, 2009.
- [12] D. M. Goldberg and D. N. Spergel, “Microwave background bispectrum. II. A probe of the low redshift universe,” *Physical Review D*, vol. 59, no. 10, Article ID 103002, 6 pages, 1999.
- [13] L. Verde and D. N. Spergel, “Dark energy and cosmic microwave background bispectrum,” *Physical Review D*, vol. 65, no. 4, Article ID 043007, 2002.
- [14] A. Cooray and W. Hu, “Weak gravitational lensing bispectrum,” *The Astrophysical Journal*, vol. 548, pp. 7–18, 2001.
- [15] A. Mangilli and L. Verde, “Non-Gaussianity and the CMB bispectrum: confusion between primordial and lensing-Rees-Sciama contribution?” *Physical Review D*, vol. 80, no. 12, Article ID 123007, 2009.
- [16] P. Serra and A. Cooray, “Impact of secondary non-Gaussianities on the search for primordial non-Gaussianity with CMB maps,” *Physical Review D*, vol. 77, no. 10, Article ID 107305, 2008.
- [17] D. Hanson, K. M. Smith, A. Challinor, and M. Liguori, “CMB lensing and primordial non-Gaussianity,” *Physical Review D*, vol. 80, no. 8, Article ID 083004, 2009.
- [18] A. P. S. Yadav and B. D. Wandelt, “CMB tomography: reconstruction of adiabatic primordial scalar potential using temperature and polarization maps,” *Physical Review D*, vol. 71, no. 12, Article ID 123004, 2005.
- [19] A. P. S. Yadav, E. Komatsu, and B. D. Wandelt, “Fast estimator of primordial non-Gaussianity from temperature and polarization anisotropies in the cosmic microwave background,” *The Astrophysical Journal*, vol. 664, no. 2, pp. 680–686, 2007.
- [20] A. P. S. Yadav and B. D. Wandelt, “Evidence of primordial non-Gaussianity ( $f_{\text{NL}}$ ) in the wilkinson microwave anisotropy probe 3-year data at  $2.8\sigma$ ,” *Physical Review Letters*, vol. 100, no. 18, Article ID 181301, 2008.
- [21] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $f_{\text{NL}}^{\text{local}}$  from WMAP 5-year data,” *Journal of Cosmology and Astro-Particle Physics*, vol. 9, article 006, 2009.

- [22] P. Chingangbam and C. Park, "Statistical nature of non-Gaussianity from cubic order primordial perturbations: CMB map simulations and genus statistic," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article 019, 2009.
- [23] P. Catelan, F. Lucchin, S. Matarrese, and L. Moscardini, "Eulerian perturbation theory in non-flat universes: second-order approximation," *Monthly Notices of the Royal Astronomical Society*, vol. 276, pp. 39–56, 1995.
- [24] S. Matarrese, L. Verde, and A. F. Heavens, "Large-scale bias in the universe: bispectrum method," *Monthly Notices of the Royal Astronomical Society*, vol. 290, no. 4, pp. 651–662, 1997.
- [25] L. Verde, A. F. Heavens, and S. Matarrese, "Projected bispectrum in spherical harmonics and its application to angular galaxy catalogues," *Monthly Notices of the Royal Astronomical Society*, vol. 318, no. 2, pp. 584–595, 2000.
- [26] LSST Science Collaborations, "LSST Science Book, Version 2.0," <http://arxiv.org/abs/0912.0201>.
- [27] E. Sefusatti and E. Komatsu, "Bispectrum of galaxies from high-redshift galaxy surveys: primordial non-Gaussianity and nonlinear galaxy bias," *Physical Review D*, vol. 76, no. 8, Article ID 083004, 2007.
- [28] L. Verde, A. F. Heavens, W. J. Percival, et al., "The 2dF galaxy redshift survey: the bias of galaxies and the density of the universe," *Monthly Notices of the Royal Astronomical Society*, vol. 335, no. 2, pp. 432–440, 2002.
- [29] L. Verde and A. F. Heavens, "On the trispectrum as a Gaussian test for cosmology," *The Astrophysical Journal*, vol. 553, no. 1, pp. 14–24, 2001.
- [30] W. H. Press and P. Schechter, "Formation of galaxies and clusters of galaxies by self-similar gravitational condensation," *The Astrophysical Journal*, vol. 187, pp. 425–438, 1974.
- [31] S. Matarrese, L. Verde, and R. Jimenez, "The abundance of high-redshift objects as a probe of non-Gaussian initial conditions," *The Astrophysical Journal*, vol. 541, no. 1, pp. 10–24, 2000.
- [32] M. Kamionkowski, L. Verde, and R. Jimenez, "The void abundance with non-gaussian primordial perturbations," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 1, article 010, 2009.
- [33] M. Grossi, L. Verde, C. Carbone, et al., "Large-scale non-Gaussian mass function and halo bias: tests on N-body simulations," *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 1, pp. 321–332, 2009.
- [34] R. K. Sheth and G. Tormen, "Large-scale bias and the peak background split," *Monthly Notices of the Royal Astronomical Society*, vol. 308, no. 1, pp. 119–126, 1999.
- [35] A. Jenkins, C. S. Frenk, S. D. M. White, et al., "The mass function of dark matter haloes," *Monthly Notices of the Royal Astronomical Society*, vol. 321, no. 2, pp. 372–384, 2001.
- [36] D. S. Reed, R. Bower, C. S. Frenk, A. Jenkins, and T. Theuns, "The halo mass function from the dark ages through the present day," *Monthly Notices of the Royal Astronomical Society*, vol. 374, no. 1, pp. 2–15, 2007.
- [37] L. Verde, R. Jimenez, M. Kamionkowski, and S. Matarrese, "Tests for primordial non-Gaussianity," *Monthly Notices of the Royal Astronomical Society*, vol. 325, no. 1, pp. 412–418, 2001.
- [38] J. Lee and S. F. Shandarin, "Comparison of analytical mass functions with numerical simulations," *The Astrophysical Journal*, vol. 517, no. 1, pp. L5–L8, 1999.
- [39] B. E. Robertson, A. V. Kravtsov, J. Tinker, and A. R. Zentner, "Collapse barriers and halo abundance: testing the excursion set ansatz," *The Astrophysical Journal*, vol. 696, pp. 636–652, 2009.
- [40] M. Maggiore and A. Riotto, "The halo mass function from excursion set theory. III. Non-Gaussian fluctuations," <http://arxiv.org/abs/0903.1251>.
- [41] X. Kang, P. Norberg, and J. Silk, "Can a large-scale structure probe cosmic microwave background-constrained non-Gaussianity?" *Monthly Notices of the Royal Astronomical Society*, vol. 376, pp. 343–347, 2007.
- [42] M. Grossi, K. Dolag, E. Branchini, S. Matarrese, and L. Moscardini, "Evolution of massive haloes in non-Gaussian scenarios," *Monthly Notices of the Royal Astronomical Society*, vol. 382, no. 3, pp. 1261–1267, 2007.
- [43] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, "Imprints of primordial non-Gaussianities on large-scale structure: scale-dependent bias and abundance of virialized objects," *Physical Review D*, vol. 77, no. 12, Article ID 123514, 2008.
- [44] V. Desjacques, U. Seljak, and I. T. Iliev, "Scale-dependent bias induced by local non-Gaussianity: a comparison to N-body simulations," *Monthly Notices of the Royal Astronomical Society*, vol. 396, no. 1, pp. 85–96, 2009.
- [45] A. Pillepich, C. Porciani, and O. Hahn, "Table of contents," *Monthly Notices of the Royal Astronomical Society*, p. 1959, 2009.
- [46] S. Matarrese and L. Verde, "The effect of primordial non-Gaussianity on halo bias," *The Astrophysical Journal*, vol. 677, no. 2, pp. L77–L80, 2008.
- [47] N. Kaiser, "On the spatial correlations of Abell clusters," *The Astrophysical Journal Letters*, vol. 284, pp. L9–L12, 1984.
- [48] H. D. Politzer and M. B. Wise, "Relations between spatial correlations of rich clusters of galaxies," *The Astrophysical Journal*, vol. 285, pp. L1–L3, 1984.
- [49] L. G. Jensen and A. S. Szalay, "N-point correlations for biased galaxy formation," *The Astrophysical Journal*, vol. 305, pp. L5–L9, 1986.
- [50] H. J. Mo and S. D. M. White, "An analytic model for the spatial clustering of dark matter haloes," *Monthly Notices of the Royal Astronomical Society*, vol. 282, no. 2, pp. 347–361, 1996.
- [51] P. Catelan, F. Lucchin, S. Matarrese, and C. Porciani, "The bias field of dark matter haloes," *Monthly Notices of the Royal Astronomical Society*, vol. 297, no. 3, pp. 692–712, 1998.
- [52] G. Efstathiou, C. S. Frenk, S. D. M. White, and M. Davis, "Gravitational clustering from scale-free initial conditions," *Monthly Notices of the Royal Astronomical Society*, vol. 235, pp. 715–748, 1988.
- [53] S. Cole and N. Kaiser, "Biased clustering in the cold dark matter cosmogony," *Monthly Notices of the Royal Astronomical Society*, vol. 237, pp. 1127–1146, 1989.
- [54] B. Grinstein and M. B. Wise, "Non-Gaussian fluctuations and the correlations of galaxies or rich clusters of galaxies," *The Astrophysical Journal*, vol. 310, pp. 19–22, 1986.
- [55] S. Matarrese, F. Lucchin, and S. A. Bonometto, "A path-integral approach to large-scale matter distribution originated by non-Gaussian fluctuations," *The Astrophysical Journal Letters*, vol. 310, no. 1, pp. L21–L26, 1986.
- [56] F. Lucchin, S. Matarrese, and N. Vittorio, "Scale-invariant clustering and primordial biasing," *The Astrophysical Journal Letters*, vol. 330, pp. L21–L23, 1988.
- [57] L. Verde and S. Matarrese, "Detectability of the effect of inflationary non-Gaussianity on halo bias," *The Astrophysical Journal*, vol. 706, no. 1, pp. L91–L95, 2009.
- [58] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, "Constraints on local primordial non-Gaussianity from large scale structure," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 031, 2008.

- [59] N. Afshordi and A. J. Tolley, “Primordial non-Gaussianity, statistics of collapsed objects, and the integrated Sachs-Wolfe effect,” *Physical Review D*, vol. 78, no. 12, Article ID 123507, 2008.
- [60] A. Cooray and R. Sheth, “Halo models of large scale structure,” *Physics Report*, vol. 372, no. 1, pp. 1–129, 2002.
- [61] U. Seljak, “Extracting primordial non-Gaussianity without cosmic variance,” *Physical Review Letters*, vol. 102, no. 2, Article ID 021302, 2009.
- [62] A. Slosar, “Optimal weighting in  $f_{NL}$  constraints from large scale structure in an idealised case,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 3, article 004, 2009.
- [63] C. Carbone, L. Verde, and S. Matarrese, “Non-Gaussian halo bias and future galaxy surveys,” *The Astrophysical Journal*, vol. 684, pp. L1–L4, 2008.
- [64] C. Fedeli, L. Moscardini, and S. Matarrese, “The clustering of galaxy clusters in cosmological models with non-Gaussian initial conditions: predictions for future surveys,” *Monthly Notices of the Royal Astronomical Society*, vol. 397, no. 2, pp. 1125–1137, 2009.
- [65] C. Sealfon, L. Verde, and R. Jimenez, “Stacking weak-lensing signals of Sunyaev-Zel’dovich clusters to constrain cluster physics,” *The Astrophysical Journal*, vol. 649, no. 1, pp. 118–128, 2006.
- [66] E. Rozo, E. S. Rykoff, A. Evrard, et al., “Constraining the scatter in the mass-richness relation of maxbcg clusters with weak lensing and X-ray data,” *The Astrophysical Journal*, vol. 699, no. 1, pp. 768–781, 2009.
- [67] M. J. Jee, P. Rosati, H. C. Ford, et al., “Hubble space telescope weak-lensing study of the galaxy cluster xmmu j2235.3–2557 at  $z \sim 1.4$ : a surprisingly massive galaxy cluster when the universe is one-third of its current age,” *The Astrophysical Journal*, vol. 704, no. 1, pp. 672–686, 2009.
- [68] R. Jimenez and L. Verde, “Implications for primordial non-Gaussianity ( $f_{NL}$ ) from weak lensing masses of high- $z$  galaxy clusters,” *Physical Review D*, vol. 80, no. 12, Article ID 127302, 2009.
- [69] D. Baumann, M. G. Jackson, P. Adshead, et al., “Probing inflation with CMB polarization,” in *CMB Polarization Workshop: Theory and Foregrounds: Cmbpol Mission Concept Study*, S. Dodelson, D. Baumann, A. Cooray, J. Dunkley, and A. Fraisse, Eds., vol. 1141 of *AIP Conference Series*, pp. 10–120, 2009.
- [70] E. Sefusatti, M. Liguori, A. P. S. Yadav, M. G. Jackson, and E. Pajer, “Constraining running non-Gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article 022, 2009.

## Review Article

# Primordial Non-Gaussianity in the Cosmic Microwave Background

Amit P. S. Yadav<sup>1</sup> and Benjamin D. Wandelt<sup>2,3</sup>

<sup>1</sup>*Institute for Advanced Study, School of Natural Sciences, Einstein Drive, Princeton, NJ 08540, USA*

<sup>2</sup>*Department of Astronomy and Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

<sup>3</sup>*Institut d'Astrophysique de Paris, 98 bis, Boulevard Arago, 75014 Paris, France*

Correspondence should be addressed to Amit P. S. Yadav, ayadav@ias.edu

Received 17 March 2010; Accepted 3 June 2010

Academic Editor: Eiichiro Komatsu

Copyright © 2010 A. P. S. Yadav and B. D. Wandelt. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the last few decades, advances in observational cosmology have given us a standard model of cosmology. We know the content of the universe to within a few percent. With more ambitious experiments on the way, we hope to move beyond the knowledge of what the universe is made of, to why the universe is the way it is. In this paper we focus on primordial non-Gaussianity as a probe of the physics of the dynamics of the universe at the very earliest moments. We discuss (1) theoretical predictions from inflationary models and their observational consequences in the cosmic microwave background (CMB) anisotropies; (2) CMB-based estimators for constraining primordial non-Gaussianity with an emphasis on bispectrum templates (3) current constraints on non-Gaussianity and what we can hope to achieve in the near future and (4) nonprimordial sources of non-Gaussianities in the CMB such as bispectrum due to second order effects, three way crosscorrelation between primary-lensing-secondary CMB, and possible instrumental effects.

## 1. Motivation

In the last few decades the advances in observational cosmology have led the field to its “golden age.” Cosmologists are beginning to nail down the basic cosmological parameters. We now know that we live in a Universe which is  $13.7 \pm 0.1$  Gyr old and is spatially flat to about 1% and is made of  $4.6 \pm 0.1\%$  baryons,  $22.8 \pm 1.3\%$  dark matter, and remaining  $72.6 \pm 1.5\%$  in the form of dark energy. Although we know the constituents to high accuracy, we still do not completely understand the physics of the beginning, the nature of dark energy and dark matter. Many upcoming CMB experiments complimented with observational campaign to map 3D structure of the Universe and new particle physics constraints from the Large Hadron Collider will enable us to move beyond the knowledge of what the universe is made of, to why the universe is the way it is. In this paper we focus on learning about the physics responsible for the initial conditions for the universe.

Inflation [1–4] is perhaps one of the most promising paradigms for the early universe, which, apart from solving

some of the problems of the Big Bang model like the flatness and horizon problem, also gives a mechanism for producing the seed perturbations for structure formation [5–9] and other testable predictions.

Most observational probes based on 2-point statistics like CMB power spectrum still allow vast number of inflationary models. Moreover, the alternatives to inflation such as cyclic models are also compatible with the data. Characterizing the non-Gaussianity in the primordial perturbations has emerged as powerful probe of the early universe. The amplitude of non-Gaussianity is described in terms of dimensionless nonlinearity parameter  $f_{\text{NL}}$  (defined in Section 3). Different models of inflation predict different amounts of  $f_{\text{NL}}$ , starting from  $O(1)$  to  $f_{\text{NL}} \sim 100$ , above which values have been excluded by the WMAP data already. Non-Gaussianity from the simplest inflation models that are based on a slowly rolling scalar field is very small [10–15]; however, a very large class of more general models with, for example, multiple scalar fields, features in inflaton potential, nonadiabatic fluctuations, noncanonical kinetic terms, deviations from Bunch-Davies vacuum, among others [16], for

a review and references therein generates substantially higher amounts of non-Gaussianity.

The measurement of the bispectrum of the CMB anisotropies is one of the most promising and “clean” way of constraining  $f_{\text{NL}}$ . Many efficient methods for evaluating bispectrum of CMB temperature anisotropies exist [17–21]. So far, the bispectrum tests of non-Gaussianity have not detected any significant  $f_{\text{NL}}$  in temperature fluctuations mapped by COBE [22] and WMAP [18, 23–28]. On the other hand, some authors have claimed non-Gaussian signatures in the WMAP temperature data [29–33]. These signatures cannot be characterized by  $f_{\text{NL}}$  and are consistent with nondetection of  $f_{\text{NL}}$ .

Currently the constraints on the  $f_{\text{NL}}$  come from temperature anisotropy data alone. By also having the polarization information in the cosmic microwave background, one can improve sensitivity to primordial fluctuations [34, 35]. Although the experiments have already started characterizing polarization anisotropies [36–39], the errors are large in comparison to temperature anisotropy. The upcoming experiments such as Planck will characterize polarization anisotropy to high accuracy.

The organization of the paper is as follow. In Section 2 we review the inflationary cosmology focusing on how the microscopic quantum fluctuations during inflation get converted into macroscopic sees perturbations for structure formation, and as CMB anisotropies. In Section 3 we discuss theoretical predictions for non-Gaussianity from the inflationary cosmology. In Section 4 we show how the primordial non-Gaussianity is connected to the CMB bispectrum, and describe/review CMB bispectrum-based estimators to constrain primordial non-Gaussianity ( $f_{\text{NL}}$ ). In Section 5 we discuss the current constraints on  $f_{\text{NL}}$  by CMB bispectrum and what we can hope to achieve in near future. We also discuss non-primordial sources of non-Gaussianity which contaminate primordial bispectrum signal. In Section 6 we discuss other methods for constraining  $f_{\text{NL}}$  besides CMB bispectrum. Finally in Section 7 we summarize with concluding remarks.

## 2. Introduction: The Early Universe

One of the most promising paradigms of the early universe is inflation [1, 2, 4], which, apart from solving the flatness, homogeneity, and isotropy problem, also gives a mechanism for producing the seed perturbations for structure formation and other testable predictions<sup>1</sup> (for a recent review of inflationary cosmology see [40]). During inflation, the universe goes through an exponentially expanding phase. From the Friedman equation, the condition for the accelerated expansion is

$$\rho + 3p < 0. \quad (1)$$

For both matter and radiation this condition is not satisfied. But it turns out that for a scalar field, the above condition

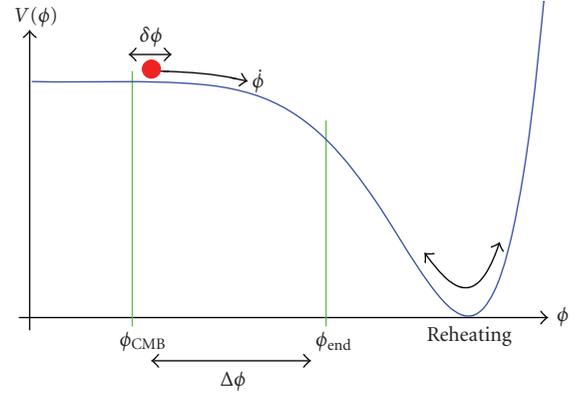


FIGURE 1: A toy scenario for the dynamics of the scalar field during inflation. During the flat part of potential, universe expand exponentially. When field reaches near the minima of the potential, the field oscillates and the radiation is generated.

can be achieved. For a spatially homogeneous scalar field,  $\phi$ , moving in a potential,  $V(\phi)$ , the energy density is given by

$$\rho_\phi = \frac{1}{2} \left( \frac{d\phi}{dt} \right)^2 + V(\phi), \quad (2)$$

and the pressure is given by

$$p_\phi = \frac{1}{2} \left( \frac{d\phi}{dt} \right)^2 - V(\phi). \quad (3)$$

Hence the condition for accelerated expansion of the universe dominated with scalar field  $\phi$  is

$$\left( \frac{d\phi}{dt} \right)^2 < V(\phi). \quad (4)$$

Physically this condition corresponds to situations where kinetic energy of the field is much smaller than its potential energy. This condition is referred to *slowly-rolling* of the scalar field. During such *slow-roll*, the Hubble parameter,  $H(t) = d \ln a / dt$ , is nearly constant in time, and the expansion scale factor  $a(t)$  is given by

$$a(t) = a(t_0) \exp \left( \int_{t_0}^t H(t') dt' \right) \approx a(t_0) \exp [H(t)(t - t_0)]. \quad (5)$$

This exponential expansion drives the observable universe spatially flat, homogeneous, and isotropic.

A toy model is shown in Figure 1. In the *slow-roll* phase,  $\phi$  rolls down on  $V(\phi)$  slowly, satisfying (4) and hence driving the universe to expand exponentially. Near the minima of the potential,  $\phi$  oscillates rapidly and inflation ends. After inflation ends, interactions of  $\phi$  with other particles lead  $\phi$  to decay with a decay rate of  $\Gamma_\phi$ , producing particles and radiation. This is called a *reheating* phase of the universe, as  $\phi$  converts its energy density into heat by the particle production.

Not only does inflation solve the flatness, homogeneity and isotropy problem, but it also gives a mechanism for generating seed perturbations. During inflation the quantum fluctuation in the field  $\phi$  exponentially stretched due to the rapid expansion phase. The proper wavelength of the fluctuations stretched out of the Hubble-horizon scale to that time,  $H^{-1}$ . Once outside the horizon, the characteristic rms amplitude of these fluctuations is  $|\phi|_{\text{rms}} \sim H/(2\pi)$ . These fluctuations do not change in time while outside the horizon. After inflation, and reheating, the standard hot-big scenario starts. As the universe decelerates, at some point the fluctuations reenter the Hubble horizon, seeding matter and radiation fluctuations in the universe. Figure 2 summarizes the evolution of characteristic length scales.

*2.1. Primordial Perturbations.* We use linearly perturbed conformal Friedmann Lematre Robertson Walker (FLRW) metric of the form

$$ds^2 = a^2(\tau) \left\{ - (1 + 2AQ)d\tau^2 - 2BQ_i d\tau dx^i + \left[ (1 + 2H_L Q)\delta_{ij} + 2H_T Q_{ij} \right] dx^i dx^j \right\}, \quad (6)$$

where all the metric perturbations,  $A$ ,  $B$ ,  $H_L$ , and  $H_T$ , are  $\ll 1$ , and functions of conformal time  $\tau$ . The spatial coordinate dependence of the perturbations is described by the scalar harmonic eigenfunctions,  $Q$ ,  $Q_i$ , and  $Q_{ij}$ , that satisfy  $\delta^{ij}Q_{ij} = -k^2Q$ ,  $Q_i = -k^{-1}Q_{,i}$ , and  $Q_{ij} = k^{-2}Q_{,ij} + (1/3)\delta_{ij}Q$ . Note that  $Q_{ij}$  is traceless:  $\delta^{ij}Q_{ij} = 0$ .

Let us consider two new perturbation variables [8, 41];

$$\begin{aligned} u &\equiv \delta\phi - \frac{\dot{\phi}}{aH} \mathcal{R}, \\ \zeta &\equiv -\frac{aH}{\dot{\phi}} u = \mathcal{R} - \frac{aH}{\dot{\phi}} \delta\phi, \end{aligned} \quad (7)$$

which are Gauge invariant. Here  $\mathcal{R} \equiv H_L + (1/3)H_T$  is perturbations in the intrinsic spatial curvature. While  $u$  reduces to  $\delta\phi$  in the spatially flat Gauge ( $\mathcal{R} \equiv 0$ ), or to  $-(\dot{\phi}/aH)\mathcal{R}$  in the comoving gauge ( $\delta\phi \equiv 0$ ), its value is invariant under any gauge transformation. Similarly  $\zeta$ , which reduces to  $\mathcal{R}$  in the comoving gauge, and to  $-(aH/\dot{\phi})\delta\phi$  in the spatially flat gauge, is also gauge invariant. The perturbation variable  $\zeta$  helps the perturbation analysis not only because of being gauge invariant, but also because it is *conserved* on super-horizon scales throughout the cosmic evolution.

The quantum fluctuations generate the gauge-invariant perturbation,  $u$ , that reduces to either  $\delta\phi$  or  $(\dot{\phi}/aH)\mathcal{R}$  depending on which gauge we use, either the spatially flat gauge or the comoving gauge. Hence,  $\delta\phi_{\text{flat}}$  and  $(\dot{\phi}/aH)\mathcal{R}_{\text{com}}$  are equivalent to each other at linear order. The benefit of using  $u$  is that it relates these two variables unambiguously, simplifying the transformation between  $\delta\phi_{\text{flat}}$  and  $\mathcal{R}_{\text{com}}$ .

The solution for  $\zeta$  is valid throughout the cosmic history regardless of whether a scalar field, radiation, or matter dominates the universe; thus, once created and leaving the

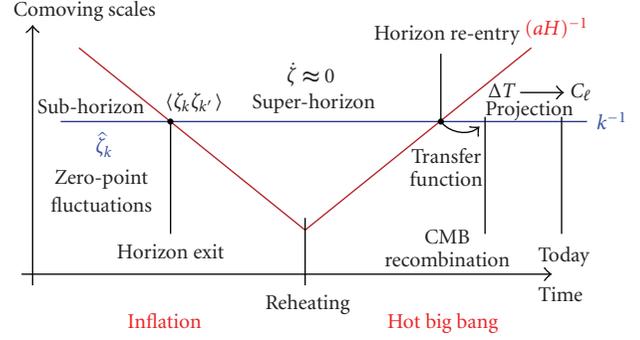


FIGURE 2: Evolution of comoving horizon and generation of perturbations in the inflationary universe. figure is from [40].

Hubble horizon during inflation,  $\zeta$  remains constant in time throughout the subsequent cosmic evolution until reentering the horizon. The amplitude of  $\zeta$  is fixed by the quantum-fluctuation amplitude in  $u$

$$\Delta_\zeta^2(k) = \left( \frac{aH}{\dot{\phi}} \right)^2 \Delta_\phi^2(k) \approx \left( \frac{aH^2}{2\pi\dot{\phi}} \right)^2 = \left[ \frac{H^2}{2\pi \left( \frac{d\phi}{dt} \right)} \right]^2. \quad (8)$$

This is the spectrum of  $\zeta$  on super-horizon scales.

*2.2. From Primordial Perturbations to CMB Anisotropies.* The metric perturbations perturb CMB, producing the CMB anisotropy on the sky. Among the metric perturbation variables, the curvature perturbations play a central role in producing the CMB anisotropy.

As we have shown in the previous subsection, the gauge-invariant perturbation,  $\zeta$ , does not change in time on super-horizon scales throughout the cosmic evolution regardless of whether a scalar field, radiation, or matter dominates the universe. The intrinsic spatial curvature perturbation,  $\mathcal{R}$ , however, does change when equation of state of the universe,  $w \equiv p/\rho$ , changes. Since  $\zeta$  remains constant, it is useful to write the evolution of  $\mathcal{R}$  in terms of  $\zeta$  and  $w$ ; however,  $\mathcal{R}$  is *not* gauge invariant itself, but  $\zeta$  is gauge invariant, so that the relation between  $\mathcal{R}$  and  $\zeta$  may look misleading. In 1980, Bardeen et al. [42] introduced another gauge-invariant variable,  $\Phi$  (or  $\Phi_H$  in the original notation), which reduces to  $\mathcal{R}$  in the zero-shear gauge, or the Newtonian gauge, in which  $B \equiv 0 \equiv H_T$ .  $\Phi$  is given by

$$\Phi = \mathcal{R} - \frac{aH}{k} \left( -B - \frac{\dot{H}_T}{k} \right). \quad (9)$$

Here, the terms in the parenthesis represent the shear, or the anisotropic expansion rate, of the  $\tau = \text{constant}$  hypersurfaces. While  $\Phi$  represents the curvature perturbations in the

zero-shear gauge, it also represents the shear in the spatially flat gauge in which  $\mathcal{R} \equiv 0$ . Using  $\Phi$ , we may write  $\zeta$  as

$$\zeta = \mathcal{R} - \frac{aH}{\dot{\phi}} \delta\phi = \Phi - \frac{aH}{k} \left( v_\phi - \frac{\dot{H}_T}{k} \right), \quad (10)$$

where the terms in the parenthesis represent the gauge-invariant fluid velocity.

We use  $\Phi$  in rest of the paper because it gives the closest analogy to the Newtonian potential, which we have some intuition of.  $\Phi$  reduces to  $\mathcal{R}$  in the zero-shear gauge (or the Newtonian gauge) in which the metric (6) becomes just like the Newtonian limit of the general relativity.

The gauge-invariant velocity term,  $v - k^{-1}\dot{H}_T$ , differentiates  $\zeta$  from  $\Phi$ . Since this velocity term depends on the equation of state of the universe,  $w = p/\rho$ , the velocity and  $\Phi$  change as  $w$  changes, while  $\zeta$  is independent of  $w$ . The evolution of  $\Phi$  on super-horizon scales in cosmological linear perturbation theory gives the following [43]:

$$\Phi = \frac{3 + 3w}{5 + 3w} \zeta, \quad (11)$$

for adiabatic fluctuations, and hence  $\Phi = (2/3)\zeta$  in the radiation era ( $w = 1/3$ ), and  $\Phi = (3/5)\zeta$  in the matter era ( $w = 0$ ).  $\Phi$  then perturbs CMB through the so-called (static) Sachs-Wolfe effect [44]

$$\frac{\Delta T}{T} = -\frac{1+w}{5+3w} \zeta. \quad (12)$$

At the decoupling epoch, the universe has already been in the matter era in which  $w = 0$ , so that we observe adiabatic temperature fluctuations of  $\Delta T/T = -(1/3)\Phi = -(1/5)\zeta$ , and the CMB fluctuation spectrum of the Sachs-Wolfe effect,  $\Delta_{\text{SW}}^2(k)$ , is

$$\Delta_{\text{SW}}^2(k) = \frac{1}{9} \Delta_\Phi^2(k) = \frac{1}{25} \Delta_\zeta^2(k). \quad (13)$$

By projecting the 3-dimensional CMB fluctuation spectrum,  $\Delta_{\text{SW}}^2(k)$ , on the sky, we obtain the angular power spectrum<sup>2</sup>,  $C_l$  [45],

$$\begin{aligned} C_l^{\text{SW}} &= 4\pi \int_0^\infty \frac{dk}{k} \Delta_{\text{SW}}^2(k) j_l^2[k(\tau_0 - \tau_{\text{dec}})] \\ &= C_2^{\text{SW}} \frac{\Gamma[(9 - n_s)/2] \Gamma[l + (n_s - 1)/2]}{\Gamma[(n_s + 3)/2] \Gamma[l + (5 - n_s)/2]}, \end{aligned} \quad (14)$$

where  $\tau_0$  and  $\tau_{\text{dec}}$  denote the conformal time at the present epoch and at the decoupling epoch, respectively, and  $n_s \equiv 1 + [d \ln \Delta^2(k)/d \ln k]$  is a spectral index which is conventionally used in the literature.

On small angular scales ( $\ell > 10$ ), the Sachs-Wolfe approximation breaks down, and the acoustic physics in the photon-baryon fluid system modifies the primordial radiation spectrum [46]. To calculate the anisotropies at all the scales, one has to solve the Boltzmann photon

transfer equation together with the Einstein equations. These equations can be solved numerically with the Boltzmann code such as CMBFAST [47]. The CMB power spectrum then can be written as

$$C_\ell = 4\pi \int_0^\infty \frac{dk}{k} \Delta_\Phi^2(k) g_{T\ell}^2(k). \quad (15)$$

Here  $g_{T\ell}(k)$  is called the radiation transfer function, and it contains all the physics which modifies the primordial power spectrum  $\Delta_\Phi$  to generate CMB power spectrum  $C_\ell$ . For the adiabatic initial conditions, in the Sachs-Wolfe limit,  $g_{T\ell}(k) = -(1/3)j_l[k(\tau_0 - \tau_{\text{dec}})]$ . Often in the literature power spectrum,  $P_\Phi(k)$ , is used instead of  $\Delta_\Phi^2(k)$ . The two are related as  $\Delta_\Phi^2(k) = (2\pi^2)^{-1} k^3 P_\Phi(k)$ .  $\Delta_\Phi^2(k)$  is called the dimensionless power spectrum.

If  $\Phi$  were exactly Gaussian, all the statistical properties of  $\Phi$  would be encoded in the *two-point function* or in  $C_\ell$  in the spherical harmonic space. Since  $\Phi$  is directly related to  $\zeta$  through (11), all the information of  $\zeta$  is also in-coded in  $C_\ell$ . Although  $\zeta$  which is related to a Gaussian variable,  $u$ , through  $\zeta = -(aH/\dot{\phi})u$ , in the linear order  $\zeta$  also obeys Gaussian statistics; however the *nonlinear* relation between  $\zeta$  and  $u$  makes  $\zeta$  (and hence  $\Phi$  and CMB anisotropies) slightly non-Gaussian. The *non-linear* relation between  $\zeta$  and  $\Phi$  is not the only source of non-Gaussianity in the CMB anisotropies. For example, at the second order, the relationship between  $\Phi$  and  $\Delta T/T$  is also *non-linear*.

**2.3. Probes of the Cosmological Initial Conditions.** The main predictions of a canonical inflation model are the following:

- (i) spatial flatness of the observable universe,
- (ii) homogeneity and isotropy on large angular scales of the observable universe,
- (iii) seed scalar and tensor perturbation with primordial density perturbations being
  - (a) nearly scale invariant,
  - (b) nearly adiabatic,
  - (c) very close to Gaussian.

At the time of writing, these predictions are consistent with all current observations. This represents a major success for the inflationary paradigm. On the other hand, the inflationary paradigm can be realized by a large “zoo”<sup>3</sup> of models. In addition, somewhat surprisingly, there exist scenarios where the Universe first contracts and then expands (such as the ekpyrotic/cyclic model), which (up to theoretical uncertainties regarding the precise mechanics of the bounce) also reproduce Universes with the properties described above. What we would like to do is to find observables that allows us to distinguish between members of the inflationary zoo. The exciting fact is that upcoming experiments will have the sensitivity to achieve this goal. *Tilt and Running:* Inflationary models very generically predict a slight deviation from completely flat spectrum. If we write the primordial power spectrum as  $\Delta_\Phi(k) = A(k_0)(k/k_0)^{n_s-1}$ , then  $n_s = 1$  correspond to flat spectrum and the quantity  $|n_s - 1|$  is called

a tilt, which characterizes the deviation from scale invariant spectrum. Although the deviations from the scale invariance are predicted to be small, the exact amount of deviation depends on the details of the inflationary model. For example in most slow roll models  $|n - 1|$  is of order  $1/N_e$ , where  $N_e \sim 60$  is a number of  $e$ -folds to the end of inflation. Ghost inflation, however, predicts negligible tilt. Hence characterizing the tilt of the scalar spectral index is a useful probe of the early universe. Currently the most stringent constraints on tilt come from the WMAP 5-year data,  $n_s = 0.960_{-0.013}^{+0.014}$  [48], which already disfavors inflationary models with “blue spectral index” ( $n_s > 1$ ). The  $1\sigma$  error on  $n_s$  will reduce to  $\Delta n_s = 0.0036$  for upcoming Planck satellite and to  $\Delta n_s = 0.0016$  for futuristic CMBPol-like satellite [49].

Apart from the tilt in the primordial power spectrum, inflationary models also predict  $n_s$  to be slightly scale dependent. This scale dependence is referred to as “running” of the spectral index  $n_s$  and is defined as  $dn_s/d\ln k$ . The constraints on the running from the WMAP 5-year data are  $-0.090 < dn_s/d\ln k < 0.0019$  [48]. The  $1\sigma$  error will reduce to  $\Delta(dn_s/d\ln k) = 0.0052$  for upcoming Planck satellite and to  $\Delta(dn_s/d\ln k) = 0.0036$  for a fourth-generation satellite such as CMBPol [49].

*Primordial Gravitational Waves.* Inflation also generates tensor perturbations (gravitational waves), which although small compared to scalar component, are still detectable, in principle. So far primordial gravitational waves have not been detected. There are upper limits on their amplitude; see [50] for a current observational bounds on the level for primordial gravitational waves. Detection of these tensor perturbations or primordial gravitational waves is considered a “smoking gun” for the inflationary scenario. In contrast to inflation, ekpyrotic (cyclic) models predict an amount of gravitational waves that is much smaller than polarized foreground emission would allow us to see even for an ideal CMB experiment. Primordial scalar perturbations create only E-modes of the CMB<sup>4</sup>, while primordial tensor perturbations generate both parity even E-modes and parity odd B-modes polarization [51–53]. The detection of primordial tensor B-modes in the CMB would confirm the existence of tensor perturbations in the early universe. This primordial *B-mode* signal is directly related to the Hubble parameter  $H$  during inflation, and thus a detection would establish the energy scale at which inflation happened. Various observational efforts are underway to detect such B-mode signal of the CMB [54]. Search for primordial B-modes is challenging. Apart from foreground subtraction challenges, and the challenge of reaching the instrumental sensitivity to detect primordial B-modes, there are several nonprimordial sources such as weak lensing of CMB by the large-scale structure [55, 56], rotation of the CMB polarization [57], and instrumental systematics that generate B-modes which contaminate the inflationary signal [58, 59]. The amplitude of gravitational waves is parametrized as the ratio of the amplitude of tensor and scalar perturbations,  $r$ . The limit from WMAP 5-year data is  $r < 0.22$  ( $2\sigma$ ) [48].

*Isocurvature Modes.* Inflationary models with a single scalar field predict primordial perturbations to be adiabatic. Hence detection of isocurvature density perturbations is a “smoking gun” for multifield models of inflation. A large number of inflationary models with multiple scalar fields predict some amount of isocurvature modes [60–72]. For example, curvaton models predict the primordial perturbations to be a mixture of adiabatic and isocurvature perturbations. Isocurvature initial conditions specify perturbations in the energy densities of two (or more) species that add up to zero. It does not perturb the spatial curvature of comoving slice (i.e.,  $\mathcal{R}$  is zero, hence the name isocurvature). In general, there can be four types of isocurvature modes, namely: baryon isocurvature modes, CDM isocurvature modes, neutrino density isocurvature modes, and neutrino velocity isocurvature modes. These perturbations imprint distinct signatures in the CMB temperature and E-polarization anisotropies [73]. The contribution of isocurvature modes is model dependent, and different models predict different amounts of it. There exists an upper limit on the allowed isocurvature modes using CMB temperature anisotropies [74, 75] a characterization (or detection of any) of isocurvature modes has a potential of discriminating between early Universe models.

*Primordial Non-Gaussianity.* Canonical inflationary models predict primordial perturbations to be very close to Gaussian [5–9], and any non-Gaussianity predicted by the canonical inflation models is very small [14, 15]. However models with nonlinearity [10, 13, 76], interacting scalar fields [12, 77], and deviation from ground state [78, 79] can generate large non-Gaussian perturbations. The amplitude of the non-Gaussian contribution to the perturbation is often referred to as  $f_{\text{NL}}$  even if the nature of the non-Gaussianities can be quite different. Different models of inflation predict different amounts of  $f_{\text{NL}}$ , starting from very close to zero for almost Gaussian perturbations, to  $f_{\text{NL}} \approx 100$  for large non-Gaussian perturbations. For example, the canonical inflation models with slow roll inflation, where only a couple of derivatives of potential, are responsible for inflationary dynamics, predict  $f_{\text{NL}} \sim 0.05$  [15]. In models where higher-order derivatives of the potential are important, the value of  $f_{\text{NL}}$  varies from  $f_{\text{NL}} \sim 0.1$  where higher order derivatives are suppressed by a low UV cutoff [80] to  $f_{\text{NL}} \sim 100$  based on Dirac-Born-Infeld effective action. Ghost inflation, where during inflation, the background has a constant rate of change as opposed to the constant background in conventional inflation, is also capable of giving  $f_{\text{NL}} \sim 100$  [81]. The additional field models generating inhomogeneities in nonthermal species [82] can generate  $f_{\text{NL}} \sim 5$  [83]; while curvaton models, where isocurvature perturbations in second field during the inflation generate adiabatic perturbations after the inflation, can have  $f_{\text{NL}} \sim 10$  [84].

In the following we will see that non-Gaussianity, far from being merely a test of standard inflation, may reveal detailed information about the state and physics of the very early Universe, if it is present at the level suggested by the theoretical arguments above.

### 3. Primordial Non-Gaussianity

Large primordial non-Gaussianity can be generated if any of the following condition is violated [85].

- (i) *Single Field*. Only one scalar field is responsible for driving the inflation and the quantum fluctuations in the same field is responsible for generating the seed classical perturbations.
- (ii) *Canonical Kinetic Energy*. The kinetic energy of the field is such that the perturbations travel at the speed of light.
- (iii) *Slow Roll*. During inflation phase the field evolves much slowly than the Hubble time during inflation.
- (iv) *Initial Vacuum State*. The quantum field was in the Bunch-Davies vacuum state before the quantum fluctuation were generated.

To characterize the non-Gaussianity one has to consider the higher order moments beyond two-point function, which contains all the information for Gaussian perturbations. The 3-point function which is zero for Gaussian perturbations contains the information about non-Gaussianity. The 3-point correlation function of Bardeen's curvature perturbations,  $\Phi(k)$ , can be simplified using the translational symmetry to give

$$\langle \Phi(\mathbf{k}_1)\Phi(\mathbf{k}_2)\Phi(\mathbf{k}_3) \rangle = (2\pi)^3 \delta^3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) f_{\text{NL}} \cdot F(k_1, k_2, k_3), \quad (16)$$

where  $F(k_1, k_2, k_3)$  tells the shape of the bispectrum in momentum space while the amplitude of non-Gaussianity is captured dimensionless non-linearity parameter  $f_{\text{NL}}$ . The shape function  $F(k_1, k_2, k_3)$  correlates fluctuations with three wave-vectors and form a triangle in Fourier space. Depending on the physical mechanism responsible for the bispectrum, the shape of the 3-point function,  $F(k_1, k_2, k_3)$ , can be broadly classified into three classes [86]. The local, ‘‘squeezed,’’ non-Gaussianity, where  $F(k_1, k_2, k_3)$  is large for the configurations in which  $k_1 \ll k_2 \approx k_3$ . Most of the studied inflationary and Ekpyrotic models produce non-Gaussianity of local shape (e.g., [82, 84, 87–104]). Second, the nonlocal, ‘‘equilateral,’’ non-Gaussianity where  $F(k_1, k_2, k_3)$  is large for the configuration when  $k_1 \approx k_2 \approx k_3$ . Finally the folded [105, 106] shape, where  $F(k_1, k_2, k_3)$  is large for the configurations in which  $k_1 \approx 2k_2 \approx 2k_3$ . Figure 3 shows these three shapes.

*Non-Gaussianity of Local Type*. The local form of non-Gaussianity may be parametrized in real space as<sup>5</sup> [13, 107, 108]:

$$\zeta(\mathbf{r}) = \zeta_L(\mathbf{r}) + \frac{3}{5} f_{\text{NL}} (\zeta_L^2(\mathbf{r}) - \langle \zeta_L^2(\mathbf{r}) \rangle), \quad (17)$$

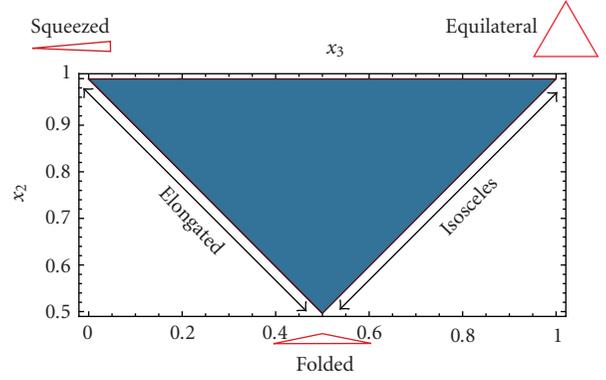


FIGURE 3: Shapes of non-Gaussianity. The shape function  $F(k_1, k_2, k_3)$  forms a triangle in Fourier space. The triangles are parametrized by rescaled Fourier modes,  $x_2 = k_2/k_1$  and  $x_3 = k_3/k_1$ . Figure is from [40].

where  $\zeta_L(r)$  is the linear Gaussian part of the perturbations, and  $f_{\text{NL}}$  characterizes the amplitude of primordial non-Gaussianity. Different inflationary models predict different amounts of  $f_{\text{NL}}$ , starting from  $O(1)$  to  $f_{\text{NL}} \sim 100$ , beyond which values have been excluded by the Cosmic Microwave Background (CMB) bispectrum of WMAP temperature data. The bispectrum in this model can be written as

$$F_{\text{local}}(k_1, k_2, k_3) = 2\Delta_\Phi^2 f_{\text{NL}} \left[ \frac{1}{k_1^{3-(n_s-1)} k_2^{3-(n_s-1)}} + \frac{1}{k_1^{3-(n_s-1)} k_3^{3-(n_s-1)}} + \frac{1}{k_2^{3-(n_s-1)} k_3^{3-(n_s-1)}} \right], \quad (18)$$

where  $\Delta_\Phi$  is the amplitude of the primordial power spectrum.

The local form arises from a non-linear relation between inflaton and curvature perturbations [10, 11, 13], curvaton models [84], or the New Ekpyrotic models [109, 110]. Models with fluctuations in the reheating efficiency [9, 10] and multifield inflationary models [17] also generate non-Gaussianity of local type.

Being local in real space, non-Gaussianity of local type describes correlations among Fourier modes of very different  $k$ . In the limit in which one of the modes becomes of very long wavelength [111],  $k_3 \rightarrow 0$ , (i.e., the other two  $k$ 's become equal and opposite),  $\zeta_{k_3}^-$  freezes out much before  $k_1$  and  $k_2$  and behaves as a background for their evolution. In this limit  $F_{\text{local}}$  is proportional to the power spectrum of the short and long wavelength modes

$$F_{\text{local}} \propto \frac{1}{k_3^3} \frac{1}{k_1^3}. \quad (19)$$

As an example, for canonical single field slow-roll inflationary models, the three-point function is given by [15]

$$F_{\text{slow-roll}}(\vec{k}_1, \vec{k}_2, \vec{k}_3) = \frac{1}{8} \left(\frac{3}{5}\right) \Delta_\Phi^2 \frac{1}{\prod k_i^3} \left[ (3\epsilon - 2\eta) \sum_{\vec{k}_i} k_i^3 + \epsilon \sum_{i \neq j} k_i k_j^2 + 8\epsilon \frac{\sum_{i>j} k_i^2 k_j^2}{k_1 + k_2 + k_3} \right], \quad (20)$$

where  $\epsilon$  and  $\eta$  are the usual slow-roll parameters and are assumed to be much smaller than unity. Taking the limit  $k_3 \rightarrow 0$  gives the local form as in (19). To show this point, Figure 4 compares the non-Gaussianity shape for local type and for slow-roll model. Although in this limit, slow-roll models do predict non-Gaussianity of local type but as evident from (20), the bispectrum of *inflaton perturbations* yields a non-trivial-scale dependence of  $f_{\text{NL}}$  [12, 15]. However in the slow roll limit  $\eta, \epsilon \ll 1$  and hence the amplitude is too small to detect.

*Non-Gaussianity of Equilateral Type.* While vast numbers of inflationary models predict non-Gaussianity of local type, this model, for instance, fails completely when non-Gaussianity is localized in a specific range in  $k$  space, the case that is predicted from inflation models with higher derivative terms [81, 106, 112–115]. In these models the correlation is among modes with comparable wavelengths which go out of the horizon nearly at the same time. The shape function for the equilateral shape can be written as [25]

$$F_{\text{equil.}}(k_1, k_2, k_3) = f_{\text{NL}}^{\text{equil.}} \cdot 6\Delta_\Phi^2 \cdot \left[ -\frac{1}{k_1^{3-(n_s-1)} k_2^{3-(n_s-1)}} + (2 \text{ perm.}) - \frac{2}{(k_1 k_2 k_3)^{2-2(n_s-1)/3}} + \frac{1}{k_1^{1-(n_s-1)/3} k_2^{2-2(n_s-1)/3} k_3^{3-(n_s-1)}} + (5 \text{ perm.}) \right]. \quad (21)$$

The models of this kind have large  $F(k_1, k_2, k_3)$  for the configurations, where  $k_1 \approx k_2 \approx k_3$ . The equilateral form arises from non-canonical kinetic terms such as the Dirac-Born-Infeld (DBI) action [112], the ghost condensation [81], or any other single-field models in which the scalar field acquires a low speed of sound [106, 115].

As an example, models with higher derivative operators in the usual inflation scenario and a model of inflation based

on the Dirac-Born-Infeld (DBI) action produce a bispectrum of the form

$$F_{\text{hd}}(\vec{k}_1, \vec{k}_2, \vec{k}_3) = \frac{3}{40} \Delta_\Phi^2 \frac{\phi^2}{\Lambda^4} \frac{1}{\prod k_i^3} \left[ \frac{1}{k_1 + k_2 + k_3} \times \left( \sum_i k_i^5 + \sum_{i \neq j} (2k_i^4 k_j - 3k_i^3 k_j^2) + \sum_{i \neq j \neq \ell} (k_i^3 k_j k_\ell - 4k_i^2 k_j^2 k_\ell) \right) \right]. \quad (22)$$

The previous model uses  $(1/8\Lambda^4)(\nabla\phi)^2(\nabla\phi)^2$  as a leading order operator. DBI inflation, which can produce large non-Gaussianity,  $f_{\text{NL}} \sim 100$ , also has  $F(k_1, k_2, k_3)$  of a similar form.

Ghost inflation, where an inflationary de Sitter phase is obtained with a ghost condensate, produces a bispectrum of the following form [81]:

$$F_{\text{ghost}}(k_1, k_2, k_3) = -\left(\frac{3}{5}\right)^3 \cdot \frac{2\sqrt{2}\pi^{3/2} H^5 \beta}{\Gamma(1/4)^3 M^2} \left(\frac{M}{\alpha H}\right)^4 \times \frac{1}{\prod_i k_i^3} \int_{-\infty}^0 d\eta \eta^{-1} F^*(\eta) F^*\left(\frac{k_2}{k_1} \eta\right) F'^*\left(\frac{k_3}{k_1} \eta\right) \times k_3 (\vec{k}_1 \cdot \vec{k}_2) + \text{symm.}, \quad (23)$$

where  $\alpha$  and  $\beta$  are free parameters of order unity, and

$$F(x) = \sqrt{\frac{\pi}{8}} (-x)^{3/2} H_{3/4}^{(1)}\left(\frac{x^2}{2}\right). \quad (24)$$

Ghost inflation also produces large non-Gaussianity,  $f_{\text{NL}} \sim 100$ . Figure 3 shows the shape of non-Gaussianity of equilateral type by showing  $F(k_1, k_2, k_3)$  for ghost inflation and for a model with a higher derivative term.

*Folded Shape.* So far the 3-point functions were calculated assuming the regular Bunch-Davis vacuum state, giving rise to either local or equilateral type non-Gaussianity. However if the bispectrum is calculated by dropping the assumption of Bunch-Davis initial state gives rise to bispectrum shape which peaks for the folded shape,  $k_1 \approx 2k_2 \approx 2k_3$ , with shape function given as [105, 106, 116]

$$F_{\text{non-BD}}(k_1, k_2, k_3) = \left(\frac{3}{5}\right)^3 \cdot \frac{1}{M_p^2} \frac{4}{\prod (2k_i^3)} \frac{H^6}{\phi^2} \sum_j \frac{3k_1^2 k_2^2 k_3^2}{k_j^2 \tilde{k}_j} \times \mathcal{R}e(\beta_{k_j}) (\cos(\tilde{k}_j \eta_0) - 1), \quad (25)$$

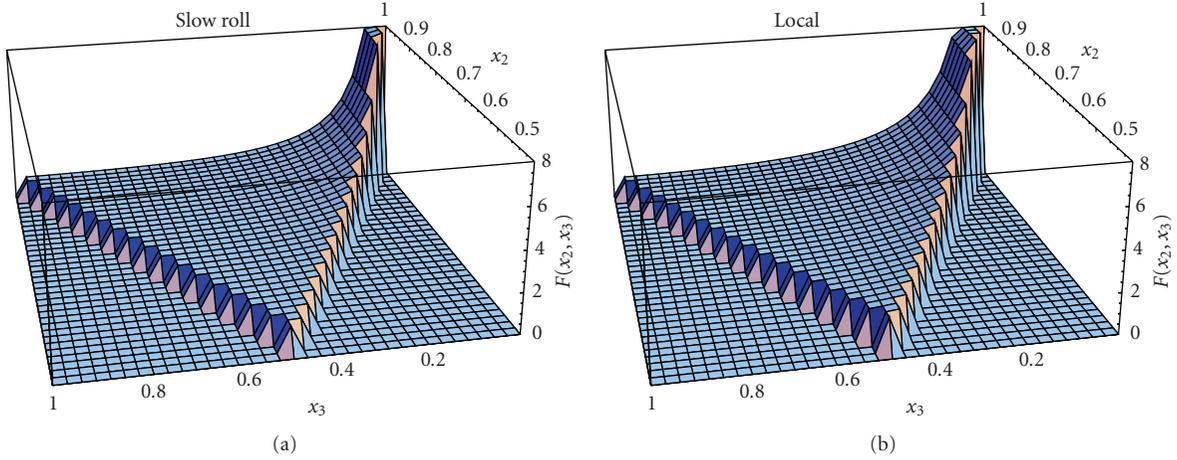


FIGURE 4: Plot of the function  $F(1, x_2, x_3)x_2^2x_3^2$  for the Slow-Roll inflation as given by (20) (a) and the local distribution as given by (18) (b). The figures are normalized to have value 1 for equilateral configurations  $x_2 = x_3 = 1$  and set to zero outside the region  $1 - x_2 \leq x_3 \leq x_2$ . Here  $x_3 \equiv k_3/k_1$ ,  $x_2 \equiv k_2/k_1$ , and  $\epsilon = \eta = 1/30$ . The figures are taken from Babich et al. 2004 [86].

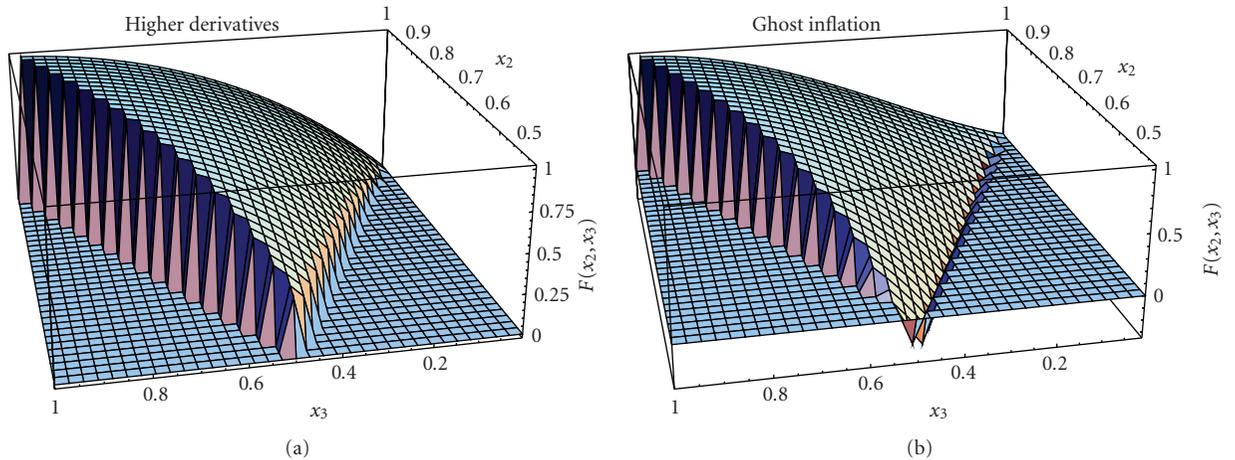


FIGURE 5: Plot of the function  $F(1, x_2, x_3)x_2^2x_3^2$  for the inflation with higher derivatives as given by (22) (a) and the ghost inflation as given by (23) (b). The figures are normalized to have value 1 for equilateral configurations  $x_2 = x_3 = 1$  and set to zero outside the region  $1 - x_2 \leq x_3 \leq x_2$ . Here  $x_3 \equiv k_3/k_1$  and  $x_2 \equiv k_2/k_1$ . The figures are taken from Babich et al. 2003 [86].

where  $\beta_{k_j}$  are the Bogoliubov coefficients which encode information about the initial conditions,  $\eta_0$  is the initial conformal time and  $\tilde{k}_j = \sum_i k_i - 2k_j$ .

#### 4. The Cosmic Microwave Background Bispectrum

Since the discovery of CMB by Penzias and Wilson in 1965 [117] and the first detection of CMB temperature anisotropies on large scales by the COBE DMR [118], the space satellite WMAP and over a dozens of balloon and ground-based experiments have characterized the CMB

temperature anisotropies to a high accuracy and over a wide range of angular scales. The space satellite Planck which launched in 2009 will soon characterize the temperature anisotropies to even higher accuracy up to angular scales of  $\ell_{\max} \approx 2500$ . The CMB power spectrum is obtained by reducing all the information of  $N_{\text{pix}} (\sim 10^6$  for WMAP and  $\sim 10^7$  for Planck). Such reduction is justified to obtain a fiducial model, given that the non-Gaussianities are expected to be small. With high quality data on the way, the field of non-Gaussianity is taking off. CMB bispectrum contains information which is not present in the power-spectrum and as we say in the previous section, is a unique probe of the early universe.

The harmonic coefficients of the CMB anisotropy  $a_{lm} = T^{-1} \int d^2\hat{\mathbf{n}} \Delta T(\hat{\mathbf{n}}) Y_{\ell m}^*$  can be related to the primordial fluctuation  $\Phi$  as

$$a_{\ell m}^p = b_\ell 4\pi (-i)^\ell \int \frac{d^3k}{(2\pi)^3} \Phi(\mathbf{k}) g_\ell^p(k) Y_{\ell m}^*(\hat{\mathbf{k}}) + n_{\ell m}, \quad (26)$$

where  $\Phi(\mathbf{k})$  is the primordial curvature perturbations, for a comoving wavevector  $\mathbf{k}$ ,  $g_\ell^p(r)$  is the radiation transfer function, where the index  $p$  refers to either temperature ( $T$ ) or  $E$ -polarization ( $E$ ) of the CMB. A beam function  $b_\ell$  and the harmonic coefficient of the noise  $n_{\ell m}$  are instrumental effects. Equation (26) is written for a flat background, but can easily be generalized.

Any non-Gaussianity present in the primordial perturbations  $\Phi(\mathbf{k})$  gets transferred to the observed CMB via (26). The most common way to look for non-Gaussianity in the CMB is to study the *bispectrum*, the three-point function of temperature and polarization anisotropies in harmonic space. The CMB angular bispectrum is defined as

$$B_{\ell_1 \ell_2 \ell_3, m_1 m_2 m_3}^{pqr} \equiv \langle a_{\ell_1 m_1}^p a_{\ell_2 m_2}^q a_{\ell_3 m_3}^r \rangle, \quad (27)$$

and the angular-averaged bispectrum is

$$B_{\ell_1 \ell_2 \ell_3}^{pqr} = \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{\ell_1 \ell_2 \ell_3, m_1 m_2 m_3}^{pqr}, \quad (28)$$

where the matrix is the Wigner 3J symbol imposing selection rules which makes bispectrum zero unless

- (i)  $\ell_1 + \ell_2 + \ell_3 = \text{integer}$ ,
- (ii)  $m_1 + m_2 + m_3 = 0$ ,
- (iii)  $|\ell_i - \ell_j| \leq \ell_k \leq \ell_i + \ell_j$  for  $i, j, k = 1, 2, 3$ .

Using (26) the bispectrum can be written as

$$\begin{aligned} B_{\ell_1 \ell_2 \ell_3}^{pqr} &= (4\pi)^3 (-i)^{\ell_1 + \ell_2 + \ell_3} \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \\ &\times \int \frac{d^3k_1}{(2\pi)^3} \frac{d^3k_2}{(2\pi)^3} \frac{d^3k_3}{(2\pi)^3} Y_{\ell_1 m_1}^*(\hat{\mathbf{k}}_1) Y_{\ell_2 m_2}^*(\hat{\mathbf{k}}_2) \\ &\times Y_{\ell_3 m_3}^*(\hat{\mathbf{k}}_3) g_{\ell_1}^p(k_1) g_{\ell_2}^q(k_2) g_{\ell_3}^r(k_3) \\ &\times \langle \Phi(\mathbf{k}_1) \Phi(\mathbf{k}_2) \Phi(\mathbf{k}_3) \rangle, \end{aligned} \quad (29)$$

where  $\langle \Phi(\mathbf{k}_1) \Phi(\mathbf{k}_2) \Phi(\mathbf{k}_3) \rangle$  is the primordial curvature three-point function as defined in (16).

To forecast constraints on non-Gaussianity using CMB data, we will perform a Fisher matrix analysis. The Fisher matrix for the parameters  $p_a$  can be written as [20, 34, 108]

$$\mathcal{F}_{ab} = \sum_{\{ijk, pqr\}} \sum_{\ell_1 \leq \ell_2 \leq \ell_3} \frac{1}{\Delta_{\ell_1 \ell_2 \ell_3}} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{pqr}}{\partial p_a} (\mathbf{Cov}^{-1})_{ijk, pqr} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{ijk}}{\partial p_b}. \quad (30)$$

The indices  $a$  and  $b$  run over all the parameters bispectrum depends on that we will assume all the cosmological parameters except  $f_{\text{NL}}$  to be known. Indices  $ijk$  and  $pqr$  run over all the eight possible ordered combinations of temperature and polarization given by  $TTT, TTE, TET, ETT, TEE, ETE, EET$  and  $EEE$ , the combinatorial factor  $\Delta_{\ell_1 \ell_2 \ell_3}$  equals 1 when all  $\ell$ 's are different, 6 when  $\ell_1 = \ell_2 = \ell_3$ , and 2 otherwise. The covariance matrix  $\mathbf{Cov}$  is obtained in terms of  $C_\ell^{TT}, C_\ell^{EE}$ , and  $C_\ell^{TE}$  (see [21, 34]) by applying Wick's theorem.

For non-Gaussianity of the local type, for which the functional form  $F(k_1, k_2, k_3)$ , is given by (18), we have

$$\begin{aligned} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{ijk}}{\partial f_{\text{NL}}} &= \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \\ &\times 2 \int_0^\infty r^2 dr \left[ -\alpha_{\ell_1}^i \beta_{\ell_2}^j \beta_{\ell_3}^k + 2 \text{ perm.} \right], \end{aligned} \quad (31)$$

where the functions  $\alpha$  and  $\beta$  are given by

$$\begin{aligned} \alpha_\ell^i(r) &\equiv \frac{2}{\pi} \int dk k^2 g_\ell^i(k) j_\ell(kr), \\ \beta_\ell^i(r) &\equiv \frac{2}{\pi} \int dk k^{-1} g_\ell^i(k) j_\ell(kr) \Delta_\Phi k^{n_s - 1}. \end{aligned} \quad (32)$$

In the previous expression we use the dimensionless power spectrum amplitude  $\Delta_\Phi$ , which is defined by  $P_\Phi(k) = \Delta_\Phi k^{-3+(n_s-1)}$ , where  $n_s$  is the tilt of the primordial power spectrum. One can compute the transfer functions  $g_\ell^T(k)$  and  $g_\ell^E(k)$  using publicly available codes such as CMBfast [47] and CAMB [119]

In a similar way, from (21), one can derive the following expressions for the bispectrum derivatives in the equilateral case:

$$\begin{aligned} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{ijk}}{\partial f_{\text{NL}}} &= \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \\ &\times 6 \int r^2 dr \left[ -\alpha_{\ell_1}^i \beta_{\ell_2}^j \beta_{\ell_3}^k + 2 \text{ perm.} + \beta_{\ell_1}^i \gamma_{\ell_2}^j \delta_{\ell_3}^k \right. \\ &\quad \left. + 5 \text{ perm.} - 2\delta_{\ell_1}^i \delta_{\ell_2}^j \delta_{\ell_3}^k \right], \end{aligned} \quad (33)$$

where the functions  $\delta$  and  $\gamma$  are given by

$$\begin{aligned} \gamma_\ell^i(r) &\equiv \frac{2}{\pi} \int dk k g_\ell^i(k) j_\ell(kr) \Delta_\Phi^{1/3} k^{(n_s-1)/3}, \\ \delta_\ell^i(r) &\equiv \frac{2}{\pi} \int dk g_\ell^T(k) j_\ell(kr) \Delta_\Phi^{2/3} k^{2(n_s-1)/3}. \end{aligned} \quad (34)$$

Recently, a new bispectrum template shape, an orthogonal shape, has been introduced [120] which characterizes the size of the signal ( $f_{\text{NL}}^{\text{ortho}}$ ) which peaks both for equilateral and flat-triangle configurations. The shape of non-Gaussianities associated with  $f_{\text{NL}}^{\text{ortho}}$  is orthogonal to the one associated to  $f_{\text{NL}}^{\text{equil}}$ . The bispectrum for orthogonal shape can be written as [120]

$$\begin{aligned} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{ijk}}{\partial f_{\text{NL}}^{\text{ortho}}} &= \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \\ &\times 18 \int r^2 dr \left[ -\alpha_{\ell_1}^i \beta_{\ell_2}^j \beta_{\ell_3}^k + 2 \text{ perm.} + \beta_{\ell_1}^i \gamma_{\ell_2}^j \delta_{\ell_3}^k \right. \\ &\quad \left. + 5 \text{ perm.} - \frac{2}{3} \delta_{\ell_1}^i \delta_{\ell_2}^j \delta_{\ell_3}^k \right]. \end{aligned} \quad (35)$$

*4.1. Estimator.* An unbiased bispectrum-based minimum variance estimator for the nonlinearity parameter in the limit of full sky and homogeneous noise can be written as [17, 20, 25]

$$\hat{f}_{\text{NL}} = \frac{1}{N} \sum_{\ell_i m_i} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \frac{B_{\ell_1 \ell_2 \ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}, \quad (36)$$

where  $B_{\ell_1 \ell_2 \ell_3}$  is angle averaged theoretical CMB bispectrum for the model in consideration. The normalization  $N$  can be calculated to require the estimator to be unbiased,  $\langle \hat{f}_{\text{NL}} \rangle = f_{\text{NL}}$ . If the bispectrum  $B_{\ell_1 \ell_2 \ell_3}$  is calculated for  $f_{\text{NL}} = 1$  then the normalization takes the following form:

$$N = \sum_{\ell_i} \frac{(B_{\ell_1 \ell_2 \ell_3})^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \quad (37)$$

The estimator for non-Gaussianity, (36), can be simplified using (26) to yield

$$\begin{aligned} \hat{f}_{\text{NL}} &= \frac{1}{N} \cdot \sum_{\ell_i m_i} \int d^2 \hat{n} Y_{\ell_1 m_1}(\hat{n}) Y_{\ell_2 m_2}(\hat{n}) Y_{\ell_3 m_3}(\hat{n}) \\ &\times \int_0^\infty r^2 dr j_{\ell_1}(k_1 r) j_{\ell_2}(k_2 r) j_{\ell_3}(k_3 r) C_{\ell_1}^{-1} C_{\ell_2}^{-1} C_{\ell_3}^{-1} \\ &\times \int \frac{2k_1^2 dk_1}{\pi} \frac{2k_2^2 dk_2}{\pi} \frac{2k_3^2 dk_3}{\pi} F(k_1, k_2, k_3) \Delta_{\ell_1}^T(k_1) \\ &\times \Delta_{\ell_2}^T(k_2) \Delta_{\ell_3}^T(k_3) a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}, \end{aligned} \quad (38)$$

where  $F(k_1, k_2, k_3)$  is a shape of 3-point function as defined in (16). Given the shape  $F(k_1, k_2, k_3)$ , one is interested in, it is conceptually straightforward to constrain the non-

linearity parameter from the CMB data. Unfortunately the computation time for the estimate scales as  $N_{\text{pix}}^{5/2}$ , which is computationally challenging as even for the WMAP data the number of pixels, is of order  $N_{\text{pix}} \sim 10^6$ . The scaling can be understood by noting that each spherical harmonic transform scales as  $N_{\text{pix}}^{3/2}$  and the estimator requires  $\ell^2$  ( $\propto N_{\text{pix}}$ ) number of spherical harmonic transforms.

The computational cost decreases if the shape can be factorized as

$$F(k_1, k_2, k_3) = f_1(k_1) f_2(k_2) f_3(k_3), \quad (39)$$

with which the estimator simplifies to

$$\begin{aligned} \hat{f}_{\text{NL}} &= \frac{1}{N} \cdot \int_0^\infty d^2 \hat{n} \int_0^\infty r^2 dr \prod_{i=1}^3 \sum_{\ell_i m_i} \\ &\times \int \frac{2k^2 dk}{\pi} j_{\ell_i}(kr) f_i(k) \Delta_{\ell_i}^T(k) C_{\ell_i}^{-1} a_{\ell_i m_i} Y_{\ell_i m_i}(\hat{n}), \end{aligned} \quad (40)$$

and computational cost now scales as  $N_{\text{pix}}^{3/2}$ . For Planck ( $N_{\text{pix}} \sim 5 \times 10^7$ ) this translates into a speed-up by factors of millions, reducing the required computing time from thousands of years to just hours and thus making  $f_{\text{NL}}$  estimation feasible for future surveys. The speed of the estimator now allows sufficient number of Monte Carlo simulations, to characterize its statistical properties in the presence of real world issues such as instrumental effects, partial sky coverage, and foreground contamination. Using the Monte Carlo simulations it has been shown that estimator is indeed optimal, where optimality is defined by saturation of the Cramer Rao bound, if noise is homogeneous. Note that even for the nonfactorizable shapes, by using the flat sky approximation and interpolating between the modes, one can estimate  $f_{\text{NL}}$  in a computationally efficient way [121].

The extension of the estimator of  $f_{\text{NL}}$  from the temperature data [17] to include both the temperature and polarization data of the CMB is discussed in Babich and Zaldarriaga [34] and Yadav et al. [20, 21, 35]. Summarizing briefly, we construct a cubic statistic as a combination of (appropriately filtered) temperature and polarization maps which is specifically sensitive to the primordial perturbations. This is done by reconstructing a map of primordial perturbations and using that to define a fast estimator. We also show that this fast estimator is equivalent to the optimal estimator by demonstrating that the inverse of the covariance matrix for the optimal estimator [34] is the same as the product of inverses we get in the fast estimator. The estimator still takes only  $N_{\text{pix}}^{3/2}$  operations in comparison to the full bispectrum calculation which takes  $N_{\text{pix}}^{5/2}$  operations.

For a given shape, the estimator for non-linearity parameter can be written as  $\hat{f}_{\text{NL}} = \hat{S}_{\text{shape}}/N_{\text{shape}}$ , where for the

equilateral, local and orthogonal shapes, the  $S_{\text{shape}}$  can be written as

$$\hat{S}_{\text{equilateral}} = \frac{3}{f_{\text{sky}}} \int r^2 dr \int d^2 \hat{n} \times \left[ B(\hat{n}, r) B(\hat{n}, r) A(\hat{n}, r) + \frac{2}{3} D(\hat{n}, r)^3 - 2B(\hat{n}, r) C(\hat{n}, r) D(\hat{n}, r) \right], \quad (41)$$

$$\hat{S}_{\text{local}} = \frac{1}{f_{\text{sky}}} \int r^2 dr \int d^2 \hat{n} B(\hat{n}, r) B(\hat{n}, r) A(\hat{n}, r),$$

$$\hat{S}_{\text{orthogonal}} = \frac{9}{f_{\text{sky}}} \int r^2 dr \int d^2 \hat{n} \times \left[ B(\hat{n}, r) B(\hat{n}, r) A(\hat{n}, r) + \frac{8}{9} D(\hat{n}, r)^3 - 2B(\hat{n}, r) C(\hat{n}, r) D(\hat{n}, r) \right], \quad (42)$$

with

$$\begin{aligned} B(\hat{n}, r) &\equiv \sum_{ip} \sum_{lm} (C^{-1})^{ip} a_{\ell m}^i \beta_{\ell}^p(r) Y_{\ell m}(\hat{n}), \\ C(\hat{n}, r) &\equiv \sum_{ip} \sum_{lm} (C^{-1})^{ip} a_{\ell m}^i \beta_{\ell}^p(r) Y_{\ell m}(\hat{n}), \\ A(\hat{n}, r) &\equiv \sum_{ip} \sum_{lm} (C^{-1})^{ip} a_{\ell m}^i \alpha_{\ell}^p(r) Y_{\ell m}(\hat{n}), \\ D(\hat{n}, r) &\equiv \sum_{ip} \sum_{lm} (C^{-1})^{ip} a_{\ell m}^i \beta_{\ell}^p(r) Y_{\ell m}(\hat{n}), \end{aligned} \quad (43)$$

and  $f_{\text{sky}}$  is a fraction of sky. Index  $i$  and  $p$  can either be  $T$  or  $E$ :

$$\begin{aligned} N &= \sum_{ijkpqr} \sum_{2 \leq \ell_1 \leq \ell_2 \leq \ell_3} \frac{1}{\Delta_{\ell_1 \ell_2 \ell_3}} \\ &\times B_{\ell_1 \ell_2 \ell_3}^{pqr, \text{prim}} (C^{-1})_{\ell_1}^{ip} (C^{-1})_{\ell_2}^{jq} (C^{-1})_{\ell_3}^{kr} B_{\ell_1 \ell_2 \ell_3}^{ijk, \text{prim}}. \end{aligned} \quad (44)$$

Indices  $i, j, k, p, q$  and  $r$  can either be  $T$  or  $E$ . Here,  $\Delta_{\ell_1 \ell_2 \ell_3}$  is 1 when  $\ell_1 \neq \ell_2 \neq \ell_3$ , 6 when  $\ell_1 = \ell_2 = \ell_3$ , and 2 otherwise:  $B_{\ell_1 \ell_2 \ell_3}^{pqr, \text{prim}}$  is the theoretical bispectrum for  $f_{\text{NL}} = 1$  [21].

It has been shown that the previous estimators defined in (42) are minimum variance amongst bispectrum-based estimators for full sky coverage and homogeneous noise [21]. To be able to deal with the realistic data, the estimator has to be able to deal with the inhomogeneous noise and foreground masks. The estimator can be generalized to deal with partial sky coverage as well as inhomogeneous noise by adding a linear term to  $\hat{S}_{\text{prim}}$ :  $\hat{S}_{\text{prim}} \rightarrow \hat{S}_{\text{prim}} + \hat{S}_{\text{prim}}^{\text{linear}}$ . For the temperature only case, this has been done in [25]. Following the same argument, we find that the linear term for

the combined analysis of CMB temperature and polarization data is given by

$$\hat{S}_{\text{prim}}^{\text{linear}} = \frac{-1}{f_{\text{sky}}} \int r^2 dr \int d^2 \hat{n} \left\{ 2B(\hat{n}, r) \langle A_{\text{sim}}(\hat{n}, r) B_{\text{sim}}(\hat{n}, r) \rangle_{\text{MC}} + A(\hat{n}, r) \langle B_{\text{sim}}^2(\hat{n}, r) \rangle_{\text{MC}} \right\}, \quad (45)$$

where  $A_{\text{sim}}(\hat{n}, r)$  and  $B_{\text{sim}}(\hat{n}, r)$  are the  $A$  and  $B$  maps generated from Monte Carlo simulations that contain signal and noise, and  $\langle \dots \rangle$  denotes the average over the Monte Carlo simulations.

The generalized estimator is given by

$$\hat{f}_{\text{NL}} = \frac{\hat{S}_{\text{prim}} + \hat{S}_{\text{prim}}^{\text{linear}}}{N}. \quad (46)$$

Note that  $\langle \hat{S}_{\text{prim}}^{\text{linear}} \rangle_{\text{MC}} = -\langle \hat{S}_{\text{prim}} \rangle_{\text{MC}}$ , and this relation also holds for the equilateral shape. Therefore, it is straightforward to find the generalized estimator for the equilateral shape: first, find the cubic estimator of the equilateral shape,  $\hat{S}_{\text{equil}}$ , and take the Monte Carlo average,  $\langle \hat{S}_{\text{equil}} \rangle_{\text{MC}}$ . Let us suppose that  $\hat{S}_{\text{equil}}$  contains terms in the form of  $ABC$ , where  $A, B$ , and  $C$  are some filtered maps. Use the Wick's theorem to rewrite the average of a cubic product as  $\langle ABC \rangle_{\text{MC}} = \langle A \rangle_{\text{MC}} \langle BC \rangle_{\text{MC}} + \langle B \rangle_{\text{MC}} \langle AC \rangle_{\text{MC}} + \langle C \rangle_{\text{MC}} \langle AB \rangle_{\text{MC}}$ . Finally, remove the MC average from single maps and replace maps in the product with the simulated maps  $\langle A \rangle_{\text{MC}} \langle BC \rangle_{\text{MC}} + \langle B \rangle_{\text{MC}} \langle AC \rangle_{\text{MC}} + \langle C \rangle_{\text{MC}} \langle AB \rangle_{\text{MC}} \rightarrow A \langle B_{\text{sim}} C_{\text{sim}} \rangle_{\text{MC}} + B \langle A_{\text{sim}} C_{\text{sim}} \rangle_{\text{MC}} + C \langle A_{\text{sim}} B_{\text{sim}} \rangle_{\text{MC}}$ . This operation gives the correct expression for the linear term, both for the local form and the equilateral form.

The main contribution to the linear term comes from the inhomogeneous noise and sky cut. For the temperature only case, most of the contribution to the linear term comes from the inhomogeneous noise, and the partial sky coverage does not contribute much to the linear term. This is because the sky-cut induces a monopole contribution outside the mask. In the analysis, one subtracts the monopole from outside the mask before measuring  $\hat{S}_{\text{prim}}$ , which makes the linear contribution from the mask small [25]. For a combined analysis of the temperature and polarization maps, however, the linear term does get a significant contribution from a partial sky coverage. Subtraction of the monopole outside of the mask is of no help for polarization, as the monopole does not exist in the polarization maps by definition. (The lowest relevant multipole for polarization is  $l = 2$ .)

The estimator is still computationally efficient, taking only  $N_{\text{pix}}^{3/2}$  (times the  $r$  sampling, which is of order 100) operations in comparison to the full bispectrum calculation which takes  $N_{\text{pix}}^{5/2}$  operations. Here  $N_{\text{pix}}$  refers to the total number of pixels. For Planck,  $N_{\text{pix}} \sim 5 \times 10^7$ , and so the full bispectrum analysis is not feasible while our analysis is.

## 5. Constraints from the CMB Bispectrum

*5.1. Current Status.* Currently the the Wilkinson Microwave Anisotropy Probe (WMAP) satellite provides the ‘‘best’’

(largest number of signal dominated modes) CMB data for non-Gaussianity analysis. Over the course of WMAP operation the field of non-Gaussianity has made vast progress both in terms of theoretical predictions of non-Gaussianities from inflation and improvement in the bispectrum- based estimators. At the time of WMAP's first data release in 2003 the estimator was suboptimal in the presence of partial sky coverage and/or inhomogeneous noise. With the sub-optimal estimator, one could not use the entirety of WMAP data and only the data up to  $l_{\max} 350$  were used to obtain the constraint  $f_{\text{NL}}^{\text{local}} = 38 \pm 96(2\sigma)$  [23]. These limits were around 30 times better than the previous constraints of  $|f_{\text{NL}}| < 1.5 \times 10^3$  from the Cosmic Background Explorer (COBE) satellite [22].

By the time of second WMAP release in 2007, the estimator was generalized by adding a linear to the KSW estimator which allows to deal with partial sky coverage and inhomogeneous noise. The idea of adding a linear term to reduce excess variance due to noise inhomogeneity was introduced in [25]. Applied to a combination of the Q, V, and W channels of the WMAP 3-year data up to  $l_{\max} \sim 400$ , this estimator had yielded the tightest constraint at the time on  $f_{\text{NL}}$  as  $-36 < f_{\text{NL}} < 100(2\sigma)$  [26]. This estimator was further generalized to utilize both the temperature and  $E$ -polarization information in [20], where it was pointed out that the linear term had been incorrectly implemented in 30 of [25]. Using MonteCarlo simulations it has been shown that this corrected estimator is nearly optimal and enables analysis of the entire WMAP data without suffering from a blow-up in the variance at high multipoles<sup>6</sup>. The first analysis using this estimator shows an evidence of non-Gaussianity of local type at around  $2.8\sigma$  in the WMAP 3-year data. Independent analysis shows the evidence of non-Gaussianity at lower significance, around  $2.5\sigma$  (see Table 1).

By the time of the third WMAP data release (with 5-year observational data) in 2008 the  $f_{\text{NL}}$  estimation technique was improved further by implementing the covariance matrix including inhomogeneous noise to make the estimator completely optimal [122]. Using the optimal estimator and using the entirety of WMAP 3-year data, there is an evidence for non-Gaussianity of local type at around  $2.5\sigma$  level  $f_{\text{NL}} \approx 58 \pm 23(1\sigma)$  [122]. However with WMAP 5-year data the significance goes down from  $\sim 2.5\sigma$  to  $\sim 1.8\sigma$  [122]. Table 2 compares the constraints obtained by different groups using WMAP 3-year and WMAP 5-year data. Figure 6 shows this comparison in more detail, showing the constraints also as a function of maximum multipole  $l_{\max}$  used in the analysis. Few comments are in place: (1) constraints on  $f_{\text{NL}}$  from WMAP 3-year data as a function of  $l_{\max}$  show a trend where the mean value rises at around  $l_{\max} = 450$ , below which data is consistent with Gaussianity and above which there is deviation from Gaussianity at above  $2\sigma$ . The result becomes roughly independent of  $l_{\max} > 550$  with evidence for non-Gaussianity at around  $2.5\sigma$  level (2) independent analysis and using different estimators (optimal and near-optimal with linear term) see this deviation from non-Gaussianity at around  $2.5\sigma$  in WMAP 3-year data, (3) significance of non-Gaussianity goes down to around  $2\sigma$  with WMAP 5-year data. The drop in the mean value between WMAP 3-year and 5-year data can be attributed to statistical shift.

The best constraints on the equilateral and orthogonal shape of non-Gaussianity using the WMAP 5-year data are  $f_{\text{NL}}^{\text{equil}} = 155 \pm 140(1\sigma)$  and  $f_{\text{NL}}^{\text{orthog}} = 149 \pm 110(1\sigma)$  respectively [120].

As we were completing this paper, the WMAP 7-year data was released, with constraints  $f_{\text{NL}}^{\text{local}} = 32 \pm 21(1\sigma)$ ,  $f_{\text{NL}}^{\text{equil}} = 26 \pm 140(1\sigma)$ , and  $f_{\text{NL}}^{\text{orthog}} = -202 \pm 104(1\sigma)$  [123].

*5.2. Future Prospects.* Now we discuss the future prospects of using the bispectrum estimators for constraining the non-linearity parameter  $f_{\text{NL}}$  for local and equilateral shapes. We compute the Fisher bounds for three experimental setups, (1) cosmic variance limited experiment with perfect beam (ideal experiment hereafter), (2) Planck satellite with and noise sensitivity  $\Delta_p = 56 \mu\text{K-arcmin}$  and beam FWHM  $\sigma = 7'$ , and (3) a futuristic CMBPol-like satellite experiment with noise sensitivity  $\Delta_p = 1.4 \mu\text{K-arcmin}$  and beam FWHM  $\sigma = 4'$  (CMBPol hereafter). Beside  $f_{\text{NL}}$  we fix all the other cosmological parameters to a standard fiducial model with a flat  $\Lambda\text{CDM}$  cosmology, with parameters described by the best fit to WMAP 5-year results [48], given by  $\Omega_b = 0.045$ ,  $\Omega_c = 0.23$ ,  $H_0 = 70.5$ ,  $n_s = 0.96$ ,  $n_t = 0.0$ , and  $\tau = 0.08$ . We calculate the theoretical CMB transfer functions and power spectrum from publicly available code CMBFAST [47]. We also neglect any non-Gaussianity which can be generated during recombination or there after. We discuss the importance and effect of these nonprimordial non-Gaussianities in the next section.

The scaling of signal-to-noise as a maximum multipole  $l_{\max}$  for the local [124, 125] and equilateral model [126] are

$$\left(\frac{S}{N}\right)_{\text{local}} \propto \ln \ell_{\max}, \quad \left(\frac{S}{N}\right)_{\text{equil}} \propto \sqrt{\ell_{\max}}. \quad (47)$$

In principle one could go to arbitrary high  $l_{\max}$  but in reality secondary signals will certainly overwhelm primary signal beyond  $l_{\max} > 3000$ : we restrict to the analysis to  $l_{\max} = 3000$ . In Figure 7, we show the  $1\sigma$  Fisher bound as function of maximum multipole  $\ell_{\max}$ , for local and equilateral type of non-Gaussianity. For both local and equilateral cases, we show the Fisher bound for the analysis using only the CMB temperature information ( $TTT$ ), only the CMB polarization information ( $EEE$ ), and the combined temperature and polarization analysis. Note that by having both the temperature and  $E$ -polarization information one can improve the sensitivity by combining the information. Apart from combining the  $T$  and  $E$  signal, one can also do cross-checks and diagnostics by independently analysing the data. Temperature and polarization will have different foregrounds and instrumental systematics.

A CMBPol-like experiment will be able to achieve the sensitivity of  $\Delta f_{\text{NL}}^{\text{local}} \simeq 2(1\sigma)$  for non-Gaussianity of local type and  $\Delta f_{\text{NL}}^{\text{equil}} \simeq 13(1\sigma)$  for non-Gaussianity of equilateral type. For the local type of non-Gaussianity, this amounts to an improvement of about a factor of 2 over the Planck satellite and about a factor of 12 over current best constraints. These estimates assume that foreground cleaning can be done perfectly that is, the effect of residual foregrounds has

TABLE 1: Hint of local non-Gaussianity at  $2\sigma$  level.

Data (mask, estimator)	$f_{\text{NL}}^{\text{local}} \pm 1\sigma$ error	Deviation from Gaussianity	
WMAP 3-year (Kp0, near-optimal)	$87 \pm 31$	$2.8\sigma$	Yadav and Wandelt [28]
WMAP 3-year (KQ75, optimal)	$58 \pm 23$	$2.5\sigma$	Smith et al. [122]
WMAP 3-year (Kp0, near-optimal)	$69 \pm 30$	$2.3\sigma$	Smith et al. [122]
WMAP 5-year (KQ75, near-optimal)	$51 \pm 30$	$1.7\sigma$	Komatsu et al. [48]
WMAP 5-year (KQ75, optimal)	$38 \pm 21$	$1.8\sigma$	Smith et al. [122]

The difference between the results by [28, 122] for WMAP 3-year data using the Kp0 mask can be a result of different choices of weighting in the near-optimal estimator. The optimal estimator has a unique weighting scheme.

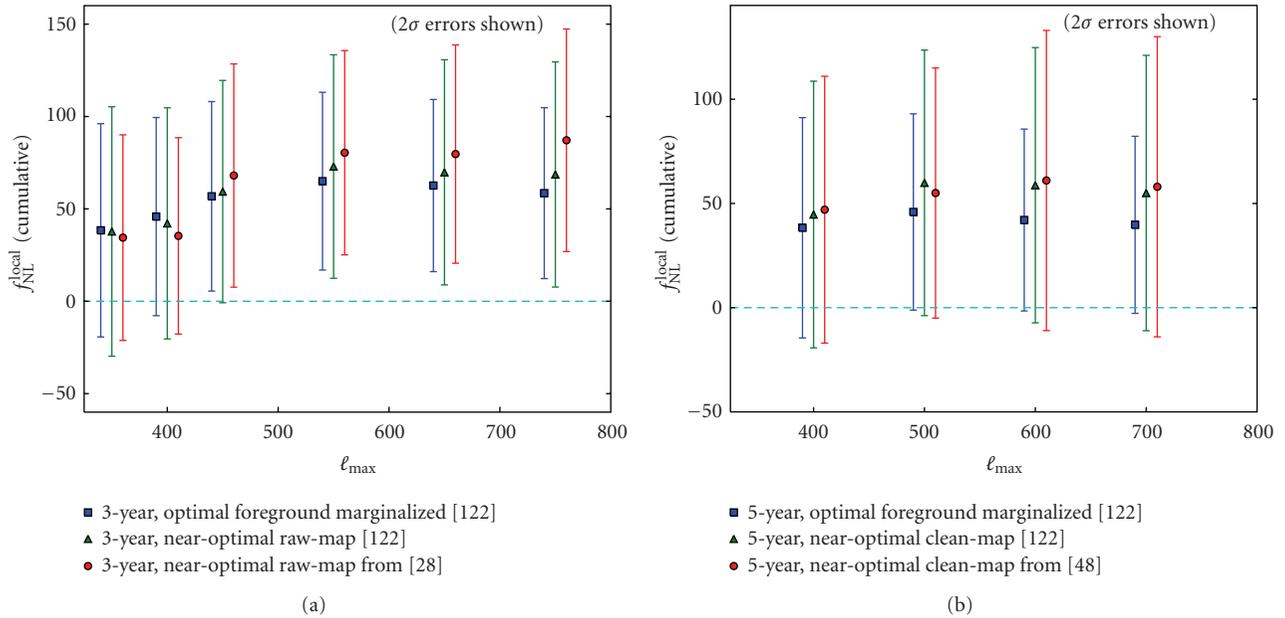


FIGURE 6: (a) Constraints on local  $f_{\text{NL}}$  using WMAP 3-year data as a function of maximum multipole  $\ell_{\text{max}}$  used in the analysis. The red circles are the results obtained using near optimal estimator by Yadav and Wandelt [28]. The green triangles are using the the near-optimal estimator by Smith et al. [122]. The blue square results are obtained using either the optimal estimator by Smith et al. [122]. For all the three analysis Kp0 mask was used. (b) Comparison between the 5-year results (optimal estimator, raw maps) reported in Komatsu et al. [48] and results obtained using the optimal or suboptimal estimator by Smith et al. [122].

been neglected. Also the contribution from unresolved point sources and secondary anisotropies such as ISW-lensing and SZ-lensing has been ignored.

*Running Non-Gaussianity.* The primordial non-Gaussian parameter  $f_{\text{NL}}$  has been shown to be scaledependent in several models of inflation with a variable speed of sound, such as Dirac-Born-Infeld (DBI) models. Starting from a simple ansatz for a scale-dependent amplitude of the primordial curvature bispectrum for primordial non-Gaussianity,

$$f_{\text{NL}} \rightarrow f_{\text{NL}} \left( \frac{K}{k_p} \right)^{n_{\text{NG}}}, \quad (48)$$

where  $K \equiv (k_1 k_2 k_3)^{1/3}$  and  $k_p$  is a pivot point. The primordial bispectrum is therefore determined in terms of two parameters: the amplitude  $f_{\text{NL}}$  and the new parameter  $n_{\text{NG}}$

quantifying its running. One can generalize the Fisher matrix analysis of the bispectra of the temperature and polarization of the CMB radiation and derive the expected constraints on the parameter  $n_{\text{NG}}$  that quantifies the running of  $f_{\text{NL}}(k)$  for current and future CMB missions such as WMAP, Planck, and CMBPol. We will consider some nonzero  $f_{\text{NL}}$  as our fiducial value for the Fisher matrix evaluation. Clearly, in order to be able to constrain a scaledependence of  $f_{\text{NL}}$ , its amplitude must be large enough to produce a detection. If  $f_{\text{NL}}$  is too small to be detected ( $f_{\text{NL}} < 2$  is a lowest theoretical limit even for the ideal experiment), we will obviously not be able to measure any of its features, either. The following we will then always consider a fiducial value of  $f_{\text{NL}}$  large enough to enable a detection. Figure 8 shows the  $1 - \sigma$  joint constraints on  $f_{\text{NL}}$  and  $n_{\text{NG}}$ . In the event of a significant detection of the non-Gaussian component, corresponding to  $f_{\text{NL}} = 50$  for the local model and

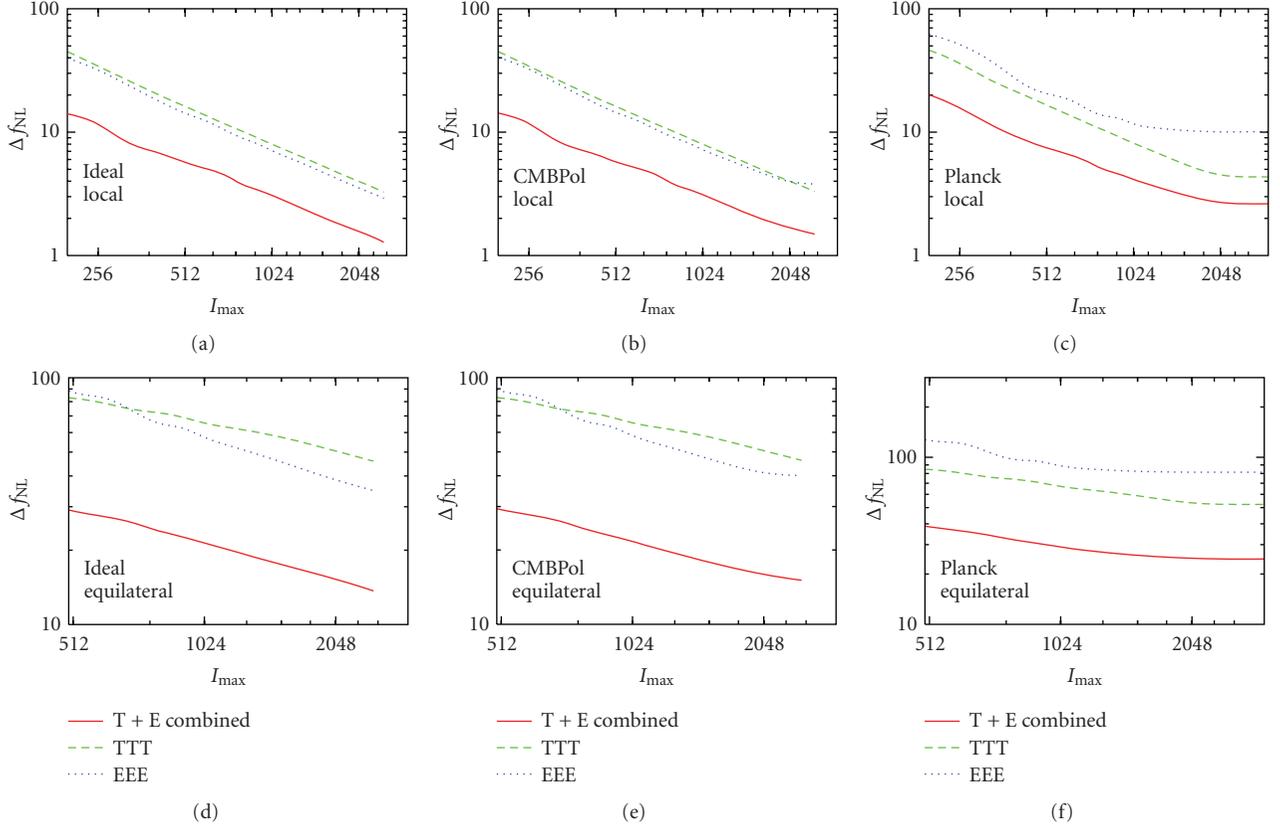


FIGURE 7: Fisher predictions for minimum detectable  $f_{\text{NL}}$  (at  $1\sigma$ ) as a function of maximum multipole  $\ell_{\text{max}}$ . Upper panels are for the local model while ((d),(e), and (f)) are for the equilateral model. ((a) and (d)) show an ideal experiment, ((b) and (e)) are for CMBPol like experiment with noise sensitivity  $\Delta_p = 1.4 \mu\text{K-arcmin}$  and beam FWHM  $\sigma = 4'$  and ((c) and (f)) are for Planck-like satellite with and noise sensitivity  $\Delta_p = 40 \mu\text{K-arcmin}$  and beam FWHM  $\sigma = 5'$ . In all the panels, the solid lines represent temperature and polarization combined analysis; dashed lines represent temperature only analysis; dot-dashed lines represent polarization only analysis.

$f_{\text{NL}} = 100$  for the equilateral model of non-Gaussianity, is able to determine  $n_{\text{NG}}$  with a  $1 - \sigma$  uncertainty of  $\Delta n_{\text{NG}} \simeq 0.1$  and  $\Delta n_{\text{NG}} \simeq 0.3$ , respectively, for the Planck mission and a factor of two better for CMBPol. In addition to CMB, one can include the information of the galaxy power spectrum, galaxy bispectrum, and cluster number counts as a probe of nonGaussianity on small scales to further constrain the two parameters [127].

**5.3. Contaminations.** A detection of non-Gaussianity has profound implications on our understanding of the early Universe. Hence it is important to know and quantify all the possible sources of non-Gaussianities in the CMB. Here we highlight some sources of non-Gaussianities due to second-order anisotropies after last scattering surface and during recombination. The fact that Gaussian initial conditions imply Gaussianity of the CMB is only true at linear order. We will also discuss the effects of instrumental effects and uncertainties in the cosmological parameters on the bispectrum estimate.

**5.3.1. Secondary Non-Gaussianities.** Current analysis of the CMB data ignore the contributions from the secondary

non-Gaussianities. For WMAP resolutions it may not be a bad approximation. Studies of the dominant secondary anisotropies conclude that they are negligible for the analysis of the WMAP data for  $l_{\text{max}} < 800$  [108, 128]. However on smaller angular scales several effects start to kick in, for example, (1) the bispectrum contribution due to unresolved point source like thermal Sunyaev-Zeldovich clusters or standard radio sources, (2) three-way correlations between primary CMB, lensed CMB, and secondary anisotropies. We will refer to the bispectrum generated due to these three-way correlations as  $B^{\text{secondary-}\kappa}$ , where some secondaries are, the integrated Sachs-Wolfe (ISW)  $B^{\text{ISW-}\kappa}$ , Sunyaev-Zeldovich signal and Rees-Sciama [23, 108, 128–131].

For Future experiments such as Planck and CMBPOL the joint estimation of primordial and secondary bispectrum will be required. The observed bispectrum in general would take the following form:

$$\hat{B}_{\ell_1 \ell_2 \ell_3}^{\text{obs}} = f_{\text{NL}} B_{\ell_1 \ell_2 \ell_3}^{\text{prim}} + b_{\text{ps}} B_{\ell_1 \ell_2 \ell_3}^{\text{ps}} + A_{\text{SZ}} B_{\ell_1 \ell_2 \ell_3}^{\text{SZ-}\kappa} + A_{\text{ISW}} B_{\ell_1 \ell_2 \ell_3}^{\text{ISW-}\kappa} + \dots \quad (49)$$

The amplitude of bispectrum due to primary-lensing-secondary cross-correlation is proportional to the product

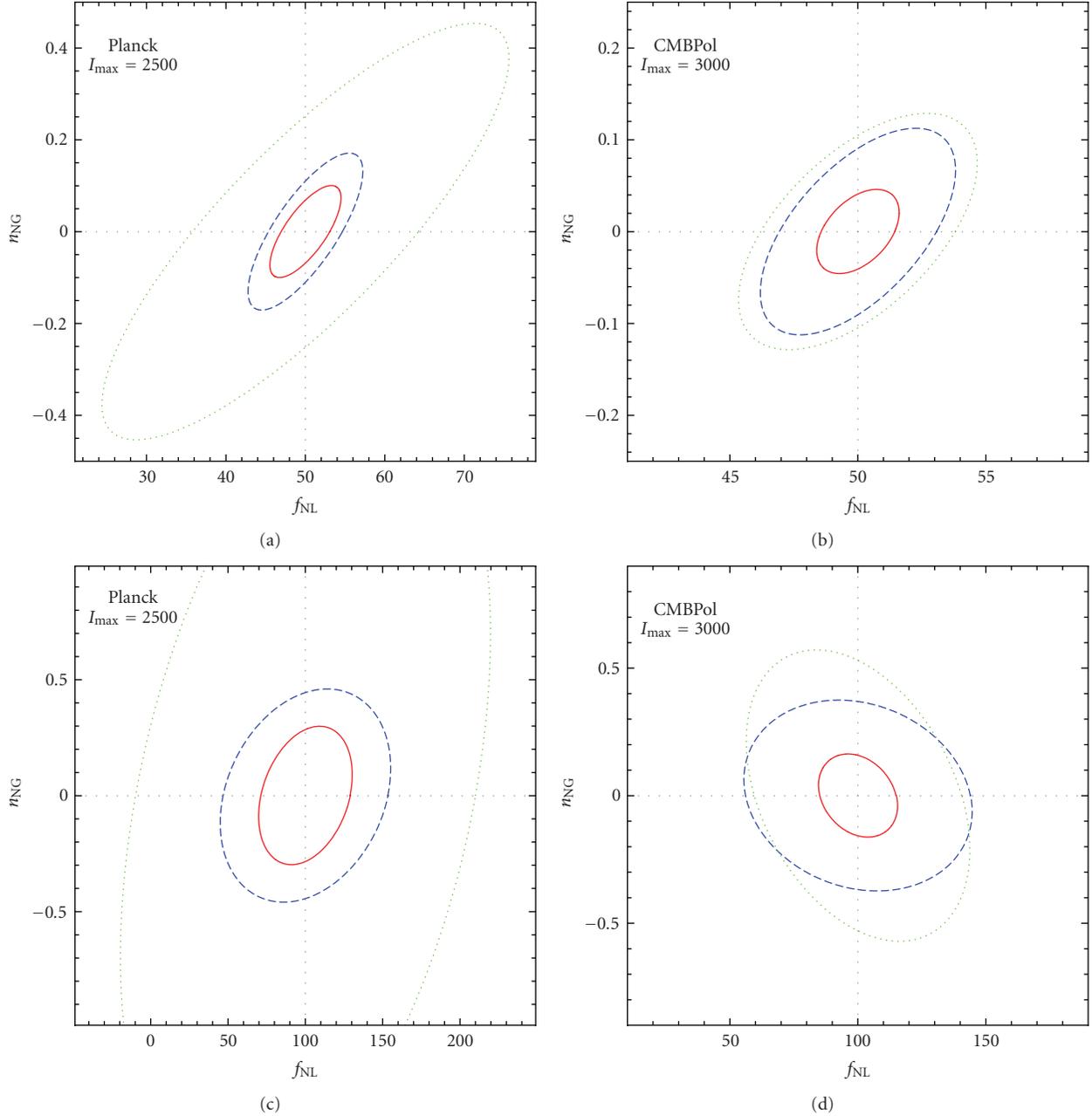


FIGURE 8:  $1\text{-}\sigma$  constraints on  $f_{\text{NL}}$  and  $n_{\text{NG}}$  for local (a) and (b) and equilateral model (c) and (d) assuming  $k_p = 0.04 \text{ Mpc}^{-1}$  and fiducial values  $f_{\text{NL}} = 50$ ,  $n_{\text{NG}} = 0$  for the local case and  $f_{\text{NL}} = 100$ ,  $n_{\text{NG}} = 0$  for the equilateral case. Dashed lines correspond to the limits from the temperature information alone, dotted lines to polarization ( $EEE$ ), while the continuous lines correspond to all bispectrum combinations. We consider Planck (a) and (c) and CMBPOL-like (b) and (d) CMB experiment.

of primary CMB power-spectrum and power spectrum of cross-correlation between secondary and lensing signals.

The reduced bispectrum from the residual point sources (assuming Poisson distributed) is constant, that is,  $b_{\ell_1 \ell_2 \ell_3}^{\text{ps}} = \text{constant}$ . The value of the constant will depend on the flux limit at which the point source can be detected and on assumed flux and frequency distribution of the sources.

Depending on the shape of primordial bispectrum in consideration, some secondary bispectra are more dangerous

than others. For example, ISW-lensing  $B^{\text{ISW-}\kappa}$  peaks at the “local” configurations hence, it is more dangerous for local primordial shape than the equilateral primordial shape. For example for the Planck satellite if the secondary bispectrum is not incorporated in the analysis, the ISW-lensing contribution will bias the estimate for the local  $f_{\text{NL}}$  by around  $\Delta f_{\text{NL}}^{\text{local}} \approx 10$  [19]. The bispectrum contribution from primary-lensing-Rees-Sciama signal also peaks at squeezed limit and contribute to effective local  $f_{\text{NL}}^{\text{local}} \approx 10$  [132].

For Planck sensitivity the point source will contamination the local non-Gaussianity by around  $\Delta f_{\text{NL}}^{\text{local}} \sim 1$  [133]. A recent analysis of the full second-order Boltzmann equation for photons [134] claims that second order effects add a contamination  $\Delta f_{\text{NL}} \sim 5$ .

The generalization of the Fisher matrix given by (30) to include multiple bispectrum contribution is

$$\mathcal{F}_{ab}^{(XY)} = \sum_{\{ijk,\text{pqr}\}} \sum_{\ell_1 \leq \ell_2 \leq \ell_3} \frac{1}{\Delta_{\ell_1 \ell_2 \ell_3}} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{\text{pqr},(X)}}{\partial p_a} \times (\text{Cov}^{-1})_{ijk,\text{pqr}} \frac{\partial B_{\ell_1 \ell_2 \ell_3}^{ijk,(Y)}}{\partial p_b}, \quad (50)$$

where the additional  $X$  and  $Y$  indices denote a component such as primordial, point-sources, and ISW-lensing, and so forth. For fixed cosmological parameters, the signal to noise  $(S/N)_i$  for the component  $i$  is

$$\left(\frac{S}{N}\right)_i = \frac{1}{\sqrt{\mathcal{F}^{(ii)}}}. \quad (51)$$

**5.3.2. Non-Gaussianities from Recombination.** Non-Gaussianities can be generated during recombination. One requires to solve second-order Boltzmann and Einstein equations to evaluate the effect. The second-order effect on CMB is an active field of study [135–153] see [154] for a recent review. The non-Gaussianities produced during recombination comprise of various effects, for example, see [135].

The dominant bispectrum due to perturbative recombination comes from perturbations in the electron density. The amplitude of perturbations of the free electron density  $\delta_e$  is around a factor of 5 larger than the baryon density perturbations [155]. The bispectrum generated due to  $\delta_e$  peaks around the “local” configuration with corresponding effective non-linearity amplitude  $f_{\text{NL}} \sim \text{few}$  [156, 157].

The bispectrum contribution due to second-order terms which are the products of the first-order perturbations is calculated in [158]. The bispectrum contribution from these terms which also peak for the squeezed triangles is small and can be neglected in the analysis. For example, the signal to noise is about 0.4 at  $l_{\text{max}} = 2000$  for a full-sky, cosmic variance limited experiment.

Another contribution to bispectrum which peaks for the equilateral configurations comes from the nonlinear evolution of the second order gravitational potential. Because of this effect, the minimum detectable non-Gaussianity parameter  $f_{\text{NL}}^{\text{equil}}$  changes by  $\Delta f_{\text{NL}}^{\text{equil}} = \mathcal{O}(10)$  for Planck-like experiment [126]. The bispectrum peaks for the equilateral shape because the growth of potential happens on scales smaller than the horizon size.

On large scales, in the absence of primordial non-Gaussianities and assuming matter domination (so that the early and late ISW can be neglected), it has been shown in [151] that for the squeezed limit the effective  $f_{\text{NL}}$  generated by second-order gravitational effects on the CMB  $f_{\text{NL}} = -1/6 - \cos(2\theta)$  (also see [141, 152, 153]). Here  $\theta$  is the angle between the short and the long modes. The angle dependent contribution comes from lensing.

**5.3.3. Effect of Cosmological Parameter Uncertainties.** Impact of uncertainties on the cosmological parameters effect, the error bar on  $f_{\text{NL}}$ . The effect of cosmological parameters have been discussed in [25, 26, 28, 159]. The cosmological parameters are determined using the 2-point statistics of the CMB, and therefore we expect the largest effect of  $f_{\text{NL}}$  would come from those parameters which leave the CMB power spectrum unchanged while change the bispectrum. The expectation value of the estimator

$$\langle \hat{f}_{\text{NL}} \rangle = \frac{1}{N} \sum_{\ell_1, \ell_2, \ell_3} \frac{B_{\ell_1 \ell_2 \ell_3} \hat{B}_{\ell_1 \ell_2 \ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}, \quad (52)$$

changes with the change in cosmological parameters. Here  $\hat{B}_{\ell_1 \ell_2 \ell_3}$  is the true CMB bispectrum. When changing the parameters the normalization  $N$  should be changed to make the estimator unbiased. In general for a set of cosmological parameters  $\{p_i\}$ , the error in  $f_{\text{NL}}$  is given by [159]

$$\delta \hat{f}_{\text{NL}} = \sqrt{\sum_{ij} \left. \frac{\partial f_{\text{NL}}}{\partial p_i} \right|_{p_i = \bar{p}_i} \left. \frac{\partial f_{\text{NL}}}{\partial p_j} \right|_{p_j = \bar{p}_j} \text{Cov}(p_i, p_j)}. \quad (53)$$

Here the average parameter values  $\bar{p}_i$  and their covariance matrix  $\text{Cov}(p_i, p_j)$  can be determined using CMB-likelihood analysis tools.

If the parameters are allowed to vary in the analysis then for WMAP this increases the  $1\sigma$  uncertainty in  $f_{\text{NL}}$  by  $\delta f_{\text{NL}}^{\text{local}}/f_{\text{NL}} \approx 16\%$  for the local shape and  $\delta f_{\text{NL}}^{\text{equil}}/f_{\text{NL}} \approx 14\%$  for the equilateral shape. For Planck experiment, the increases in  $1\sigma$  uncertainty are  $\delta f_{\text{NL}}^{\text{local}}/f_{\text{NL}} \approx 5\%$  for local shape and  $\delta f_{\text{NL}}^{\text{equil}}/f_{\text{NL}} \approx 4\%$  for the equilateral shape. Most of the contribution to the error comes from three cosmological parameters, the amplitude of scalar perturbations  $\Delta_\phi$ , the tilt of the power spectrum of the scalar perturbations  $n_s$ , and re-ionization optical depth  $\tau$ .

For modes inside the horizon during reionization, the reionization optical depth  $\tau$  appears as a multiplicative factor  $e^{-\tau}$  in front of transfer function  $g_\ell^i$ . For local model one of the mode is outside so the effect on bispectrum  $\tilde{b}_{\ell_1 \ell_2 \ell_3}^{\text{local}} = \exp(-2\tau) b_{\ell_1 \ell_2 \ell_3}^{\text{local}}$  and for equilateral model all the modes are inside the horizon so  $\tilde{b}_{\ell_1 \ell_2 \ell_3}^{\text{equil}} = \exp(-3\tau) b_{\ell_1 \ell_2 \ell_3}^{\text{equil}}$ . This reduces to  $\delta f_{\text{NL}}^{\text{local}} \simeq -2 f_{\text{NL}} \tau$  for local model and  $\delta f_{\text{NL}}^{\text{equil}} \simeq -3 f_{\text{NL}} \tau$  for equilateral model.

The effect of amplitude of perturbations can be seen by noting that the level of non-Gaussianity is given by  $f_{\text{NL}} \cdot \Delta_\phi^{1/2}$ . Hence the decreases (increase) in the amplitude of perturbations relax (tighten) the constraints on  $f_{\text{NL}}$ . The effect of red tilt ( $n_s < 1$ ) can be thought of as a reduction in power on at scales shorter than first peak and enhancement of power on larger scales. The effect of blue tilt is just opposite of red tilt. For local shape, the limit on  $f_{\text{NL}}$  becomes tighter proportional to  $\Delta_{\text{long}}^{1/2}$  [26]. Note that [159] shows that the effect of cosmological parameters is negligible if the parameters are allowed to vary in the analysis and then marginalize over.

5.3.4. *Instrumental Effects and Distortions along the Line of Sight.* Here we point out that any cosmological or instrumental effect that can be modelled as a line-of-sight CMB distortions of the primary CMB does not generate new bispectrum contribution. Although they can modify the primordial bispectrum. A general model of line-of-sight distortions of the primary CMB are described in [58, 59, 160], where the changes in the Stokes parameter of the CMB due to distortions along the line-of-sight can be written as

$$\begin{aligned} & \delta[Q \pm iU](\hat{\mathbf{n}}) \\ &= [a \pm i2w](\hat{\mathbf{n}})[\tilde{Q} \pm i\tilde{U}](\hat{\mathbf{n}}) + [f_1 \pm if_2](\hat{\mathbf{n}})[\tilde{Q} \mp i\tilde{U}](\hat{\mathbf{n}}) \\ &+ [\gamma_1 \pm i\gamma_2](\hat{\mathbf{n}})\tilde{T}(\hat{\mathbf{n}}) + \sigma p(\hat{\mathbf{n}}) \cdot \nabla[\tilde{Q} \pm i\tilde{U}](\hat{\mathbf{n}}; \sigma) \\ &+ \sigma[d_1 \pm id_2](\hat{\mathbf{n}})[\partial_1 \pm i\partial_2]\tilde{T}(\hat{\mathbf{n}}; \sigma) \\ &+ \sigma^2 q(\hat{\mathbf{n}})[\partial_1 \pm i\partial_2]^2 \tilde{T}(\hat{\mathbf{n}}; \sigma) + \dots \end{aligned} \quad (54)$$

The first line captures the distortions in a single perfectly known direction  $\hat{\mathbf{n}}$ . The distortions in second line capture mixing of the polarization fields in a local region of length scale  $\sigma$  around  $\hat{\mathbf{n}}$ . We Taylor expand the CMB fields  $Q$ ,  $U$ , and  $T$  around the point  $\hat{\mathbf{n}}$  and consider the leading-order terms. Here  $\tilde{Q}$ ,  $\tilde{U}$ , and  $\tilde{T}$  stand for primordial (undistorted) CMB fields. Since  $(Q \pm iU)(\hat{\mathbf{n}})$  is spin  $\pm 2$  field,  $a(\hat{\mathbf{n}})$  is a scalar field that describes modulation in the amplitude of the fields in a given direction  $\hat{\mathbf{n}}$ ;  $\omega(\hat{\mathbf{n}})$  is also a scalar field that describes the rotation of the plane of polarization,  $(f_1 \pm if_2)$  are spin  $\pm 4$  fields that describe the coupling between two spin states (spin-flip), and  $(\gamma_1 \pm i\gamma_2)(\hat{\mathbf{n}})$  are spin  $\pm 2$  fields that describe leakage from the temperature to polarization (monopole leakage hereon). Distortions in the second line of (54),  $(p_1 \pm p_2)$ ,  $(d_1 \pm d_2)$ , and  $q$  are measured in the units of the length scale  $\sigma$ . The field  $(p_1 \pm ip_2)(\hat{\mathbf{n}})$  is a spin  $\pm 1$  field and describes the change in the photon direction; we will refer to it as a deflection field. Finally  $(d_1 \pm d_2)(\hat{\mathbf{n}})$  and  $q(\hat{\mathbf{n}})$  describe leakage from temperature to polarization,  $(d_1 \pm d_2)(\hat{\mathbf{n}})$  is spin  $\pm 1$  field, and we will refer to it as dipole leakage;  $q(\hat{\mathbf{n}})$  is a scalar field that we will call quadrupole leakage.

These distortions can be produced by various cosmological processes such as weak gravitational lensing of the CMB, screening effects from patchy reionization, rotation of the plane of polarization due to magnetic fields or parity violating physics, and various instrumental systematics such as gain fluctuations, pixel rotation, differential gain, pointing, differential ellipticity are also captured via line-of-sight distortions. All these distortions modify the primordial bispectrum as

$$\begin{aligned} & \tilde{B}_{(\ell_1, \ell_2, \ell_3)} \\ &= B_{(\ell_1, \ell_2, \ell_3)} + \int \frac{d^2 \ell'}{(2\pi)^2} C_{\ell'}^{\mathcal{D}\mathcal{D}} \\ &\times \left[ B_{(\ell_1, \ell_2 - \ell', \ell_3 + \ell')} W^{\mathcal{D}}(\ell_1 - \ell') W^{\mathcal{D}}(\ell_2 - \ell'') + \text{perm.} \right], \end{aligned} \quad (55)$$

where  $W$  is a window which depends on the type of distortion in consideration and tells how the primordial CMB bispectrum modes are coupled to the distortion field power spectrum  $C_{\ell}^{\mathcal{D}\mathcal{D}}$ . The effect of the distortions on the bispectrum is to smooth out the acoustic features. These effects for the case of lensing have been shown to be small and can be neglected [161, 162].

In [163], the impact of the  $1/f$  noise and asymmetric beam on local  $f_{\text{NL}}^{\text{local}}$  has been found insignificant in the context of a Planck-like experiment.

## 6. Other Probes of Non-Gaussianity in the CMB

Although using the full bispectrum is the most sensitive cubic statistic other statistical methods may be sensitive to different aspects of non-Gaussianity and, more importantly, different methods have different systematic effects. Therefore it is important to study various probes. In this section we will discuss some of the methods which have been recently used or developed to test for primordial non-Gaussianities in the CMB.

*Trispectrum.* The four-point function in harmonic space is called trispectrum, which can be written as

$$\begin{aligned} & \langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} a_{l_4 m_4} \rangle \\ &= \sum_{LM} (-1)^M \begin{pmatrix} \ell_1 & \ell_2 & L \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} \ell_3 & \ell_4 & L \\ m_1 & m_2 & m_3 \end{pmatrix} T_{l_3 l_4}^{l_1 l_2}(L), \end{aligned} \quad (56)$$

where  $T_{l_3 l_4}^{l_1 l_2}(L)$  is the angular averaged trispectrum,  $L$  is the length of a diagonal that forms triangles with  $l_1$  and  $l_2$  and with  $l_3$  and  $l_4$ , and the matrix is the Wigner 3- $j$  symbol. The trispectrum contains unconnected part,  $T_G$ ,

$$\begin{aligned} T_{G l_3 l_4}^{l_1 l_2}(L) &= (-1)^{l_1 + l_3} \sqrt{(2l_1 + 1)(2l_3 + 1)} C_{l_1} C_{l_3} \delta_{l_1 l_2} \delta_{l_3 l_4} \delta_{L0} \\ &+ (2L + 1) C_{l_1} C_{l_2} \left[ (-1)^{l_1 + l_2 + L} \delta_{l_1 l_3} \delta_{l_2 l_4} + \delta_{l_1 l_4} \delta_{l_2 l_3} \right]. \end{aligned} \quad (57)$$

which comes from the Gaussian part of the perturbations, and the connected part  $T_c$  which contains non-Gaussian signatures. Using permutation symmetry, one may write the connected part of the trispectrum as

$$\begin{aligned} T_{c l_3 l_4}^{l_1 l_2}(L) &= P_{l_3 l_4}^{l_1 l_2}(L) + (2L + 1) \sum_{L'} \\ &\times \left[ (-1)^{l_2 + l_3} \begin{Bmatrix} \ell_1 & \ell_2 & L \\ \ell_4 & \ell_3 & L' \end{Bmatrix} P_{l_2 l_4}^{l_1 l_3}(L') \right. \\ &\left. + (-1)^{L + L'} \begin{Bmatrix} \ell_1 & \ell_2 & L \\ \ell_3 & \ell_4 & L' \end{Bmatrix} P_{l_3 l_2}^{l_1 l_4}(L') \right]. \end{aligned} \quad (58)$$

where

$$P_{l_3 l_4}^{l_1 l_2}(L) = t_{l_3 l_4}^{l_1 l_2}(L) + (-1)^{2L+l_1+l_2+l_3+l_4} t_{l_4 l_3}^{l_2 l_1}(L) \\ + (-1)^{L+l_3+l_4} t_{l_4 l_3}^{l_1 l_2}(L) + (-1)^{L+l_1+l_2} t_{l_3 l_4}^{l_2 l_1}(L). \quad (59)$$

Here, the matrix is the Wigner 6- $j$  symbol, and  $t_{l_3 l_4}^{l_1 l_2}(L)$  is called the reduced trispectrum, which contains all the physical information about non-Gaussianities. For non-Gaussianity of local-type for which

$$\Phi(x) = \Phi_L(x) + f_{\text{NL}}[\Phi_L^2 - \langle \Phi_L^2 \rangle] + g_{\text{NL}}\Phi_L^3, \quad (60)$$

both  $f_{\text{NL}}$  and  $g_{\text{NL}}$  contribute to the trispectrum, but only  $f_{\text{NL}}$  contributes to the bispectrum. Trispectrum-based estimators for measuring  $f_{\text{NL}}$  and  $g_{\text{NL}}$  have been developed [164–168]. For local template, the bispectrum nearly contains all the information on  $f_{\text{NL}}$  [166] however if the non-Gaussianity is seen in bispectrum, trispectrum can serve as a important cross-check. Generically for single field slow-roll models the trispectrum is small and unobservable [169] however for more general single field models whenever the equilateral bispectrum is large, the trispectrum is large as well [170–172]. For example for equilateral non-Gaussianity [173] study how to tune the model parameters to get large trispectrum and small bispectrum. For multifield inflation one can construct models that predicts small  $f_{\text{NL}}$  but large  $g_{\text{NL}}$ , for example, [174] discussed the local form from a multi-field inflation and briefly mentioned the condition in their class of models to get the large trispectrum and small bispectrum. Joint constraints on both  $f_{\text{NL}}$  and  $g_{\text{NL}}$  have the potential to add to the specificity of the search for primordial non-Gaussianity. For a given model, these two numbers will often be predicted in terms of a single model parameter, such as a coupling constant see, for example, [175] for the case of ekpyrotic models. Using WMAP 5-year data, the constraints on  $g_{\text{NL}}$  using the trispectrum are  $-7.4 < g_{\text{NL}}/10^5 < 8.2$  at  $2\sigma$  [176].

*Minkowski Functionals.* Minkowski Functionals (MFs) describe morphological properties (such as area, circumference, and Euler characteristic) of fluctuating fields [177–180]. For a  $d$ -dimensional fluctuating field,  $f$ , the  $k$ th Minkowski Functionals of weakly non-Gaussian fields in,  $V_k^{(d)}(\nu)$  for a given threshold  $\nu = f/\sigma_0$  can be written as [181, 182]

$$V_k^{(d)}(\nu) \\ = \frac{1}{(2\pi)^{(k+1)/2}} \frac{\omega_d}{\omega_{d-k}\omega_k} \left( \frac{\sigma_1}{\sqrt{d}\sigma_0} \right)^k e^{-\nu^2/2} \\ \times \left\{ H_{k-1}(\nu) + \left[ \frac{1}{6}S^{(0)}H_{k+2}(\nu) + \frac{k}{3}S^{(1)}H_k(\nu) \right. \right. \\ \left. \left. + \frac{k(k-1)}{6}S^{(2)}H_{k-2}(\nu) \right] \sigma_0 + \mathcal{O}(\sigma_0^2) \right\}, \quad (61)$$

where  $\sigma_0 \equiv \langle f^2 \rangle^{1/2}$  is the variance of the fluctuating field,  $H_n(\nu)$  are the Hermite polynomials,  $\omega_k \equiv \pi^{k/2}/\Gamma(k/2 + 1)$ , and finally  $S^{(i)}$  are the “skewness parameters” defined as

$$S^{(0)} \equiv \frac{\langle f^3 \rangle}{\sigma_0^3}, \\ S^{(1)} \equiv -\frac{3}{4} \frac{\langle f^2 (\nabla^2 f) \rangle}{\sigma_0^2 \sigma_1^2}, \quad (62) \\ S^{(2)} \equiv -\frac{3d}{2(d-1)} \frac{\langle (\nabla f) \cdot (\nabla f) (\nabla^2 f) \rangle}{\sigma_1^4},$$

which characterize the skewness of fluctuating fields and their derivatives. Here  $\sigma_i$  characterizes the variance of the fluctuating field and is given by

$$\sigma_i^2 = \frac{1}{4\pi} \sum_l (2\ell + 1) [\ell(\ell + 1)]^i C_\ell^2. \quad (63)$$

For CMB, for which  $d = 2$  and  $f = \Delta T/T$ , the skewness parameters are [183]

$$S^{(0)} = \frac{1}{4\pi\sigma_0^4} \sum_{l_i m_i} B_{l_1 l_2 l_3}^{m_1 m_2 m_3} \mathcal{G}_{l_1 l_2 l_3}^{m_1 m_2 m_3} W_{l_1} W_{l_2} W_{l_3}, \\ S^{(1)} = \frac{3}{16\pi\sigma_0^2 \sigma_1^2} \sum_{l_i m_i} \frac{l_1(l_1 + 1) + l_2(l_2 + 1) + l_3(l_3 + 1)}{3} \\ \times B_{l_1 l_2 l_3}^{m_1 m_2 m_3} \mathcal{G}_{l_1 l_2 l_3}^{m_1 m_2 m_3} W_{l_1} W_{l_2} W_{l_3}, \\ S^{(2)} = \frac{3}{8\pi\sigma_1^4} \sum_{l_i m_i} \left\{ \frac{[l_1(l_1 + 1) + l_2(l_2 + 1) - l_3(l_3 + 1)]}{3} \right. \\ \left. \times [l_3(l_3 + 1) + (\text{cyc.})] \right\} B_{l_1 l_2 l_3}^{m_1 m_2 m_3} \mathcal{G}_{l_1 l_2 l_3}^{m_1 m_2 m_3} \\ \times W_{l_1} W_{l_2} W_{l_3}, \quad (64)$$

where  $B_{l_1 l_2 l_3}$  is the CMB bispectrum,  $W_l$  represents a smoothing kernel which depends on the experiment beam and  $\mathcal{G}_{l_1 l_2 l_3}^{m_1 m_2 m_3}$  is the usual Gaunt function.

Since MFs can be determined as weighted sum of the bispectrum, they contain less information than the bispectrum. MFs can still be useful because they perhaps suffer from different systematics, though they are less specific to primordial non-Gaussianity since they measure a smaller number of independent bispectrum modes. Also, the bispectrum is defined in Fourier (or harmonic) space while the MFs are defined in real space. Limits on non-Gaussianity of local-type from the MFs of the WMAP 5-year temperature data are  $-70 < f_{\text{NL}} < 91(2\sigma)$  [188]. The MFs from the Planck temperature data should be sensitive to  $f_{\text{NL}} \sim 20$  at  $1\sigma$  level [193] in contrast to bispectrum which is sensitive to  $f_{\text{NL}} \sim 5$  at  $1\sigma$  level. Note that polarization data further improves the sensitivity.

*Wavelets.* Several studies have used wavelet representations of the WMAP maps to search for a non-Gaussian signal

TABLE 2: Summary of constraints on local non-Gaussianity.

Year	Data	Method	$f_{\text{NL}}^{\text{local}} \pm 2\sigma$ error	
2002	COBE	Bispectrum sub-optimal	$ f_{\text{NL}}  < 1500$	Komatsu et al. [184]
2003	MAXIMA	Bispectrum sub-optimal	$ f_{\text{NL}}  < 1900$	Santos et al. [185]
2003	WMAP 1-year	Bispectrum sub-optimal	$39.5 \pm 97.5$	Komatsu et al. [23]
2004	VSA	Bispectrum sub-optimal	$f_{\text{NL}} < 5400$	Smith et al. [186]
2005	WMAP 1-year	Bispectrum sub-optimal-v1	$47 \pm 74$	Creminelli et al. [25]
2006	WMAP 3-year	Bispectrum sub-optimal	$30 \pm 84$	Spergel et al. [24]
2006	WMAP 3-year	Bispectrum sub-optimal-v1	$32 \pm 68$	Creminelli et al. [26]
2007	WMAP 3-year	Bispectrum near-optimal	$87 \pm 62$	Yadav and Wandelt [28]
2007	Boomerang	Minkowski Functionals	$110 \pm 910$	de Troia et al. [187]
2008	WMAP 3-year	Minkowski Functionals	$10.5 \pm 80.5$	Hikage et al. [188]
2008	WMAP 5-year	Bispectrum near-optimal	$51 \pm 60$	Komatsu et al. [48]
2008	ARCHEOPS	Minkowski Functionals	$70_{-950}^{1075}$	Curto et al. 2008 [189]
2009	WMAP 3-year	Bispectrum optimal	$58 \pm 46$	Smith et al. [122]
2009	WMAP 5-year	Bispectrum optimal	$38 \pm 42$	Smith et al. [122]
2009	WMAP 5-year	Spherical Mexican hat wavelet	$31 \pm 49$	Curto, et al. [190]
2009	BOOMERanG	Minkowski Functionals	$-315 \pm 705$	P. Natoli et al. [191]
2009	WMAP 5-year	Skewness power spectrum	$11 \pm 47.4$	Smidt, et al. [192]
2010	WMAP 7-year	Bispectrum optimal	$32 \pm 42$	Komatsu et al. [123]

[194–200]. In most of these studies, wavelets were used as a tool for blind searches of non-Gaussian anomalies in a basis with resolution in both scale and location. However, in some more recent studies, wavelets were tuned to look for non-Gaussianity of a particular type. In the context of searches for primordial non-Gaussianity of local type, wavelet-based estimators for  $f_{\text{NL}}$  have been built by extracting a signature of local non-Gaussianity that is cubic in the wavelet coefficients from simulations of non-Gaussian skies and searching for this signature in data. This ability to calibrate on a set of simulations makes the wavelet approach very flexible. While not optimal in a least-squared sense, using a wavelet representation can be thought of as a generalized cubic statistic with a different weighting scheme to the optimal bispectrum estimator. Using such estimators therefore provides a useful exploration of nearly optimal cubic estimators similar to the full bispectrum estimator. Any believable detection of non-Gaussianity should be robust to such changes in the analysis. Similarly, contaminating non-Gaussianity from astrophysical and instrumental systematics will propagate through the analysis in a different way to the bispectrum-based analysis.

There are several constraints on local  $f_{\text{NL}}$  using wavelet-based estimators. For example, using the COBE data the constraints are  $|f_{\text{NL}}| < 2200(1\sigma)$  [201]. Using an estimator based on the skewness of the wavelet coefficients, Mukherjee and Wang constrain the  $f_{\text{NL}}$  value for WMAP 1-yr data obtaining  $f_{\text{NL}} = 50 \pm 160(2\sigma)$  [29]. Using an extension of the previous estimator by combining wavelet coefficients at different contiguous scales, Curto et al. obtain  $-8 < f_{\text{NL}} < 111(2\sigma)$  [202]. Recently, using a generalized third-order estimator based on the wavelet coefficients, Curto et al. obtain  $-18 < f_{\text{NL}} < 80(2\sigma)$  [190].

*Needlet Bispectrum.* Needlets are a family of spherical wavelets which are localized and asymptotically uncorrelated [203, 204]. The needlet-based statistics has been considered for testing Gaussianity and isotropy (see, e.g., [205–211]). Using the bispectrum of needlet coefficient, the constraints on non-Gaussianity of local-type using WMAP 5-year data yield  $f_{\text{NL}} = 73 \pm 62(2\sigma)$  [212, 213]. As is clear, the needlet-based bispectrum is not as sensitive as the CMB bispectrum discussed in Section 4; however again in the event of detection the needlet based methods can be calibrated on simulations and represent a different weighting scheme for handling the sky mask and anisotropic noise. Finally, needlets and wavelets allow for the possibility to analyze spatially localized regions in the sky.

*Probing Non-Gaussianity Using Bayesian Statistics.* A somewhat different approach to searching for non-Gaussianity is provided by the Bayesian approach. Here, the starting point is an explicit physical or statistical model for the data and the goal is to evaluate the posterior density of the parameters of the model and/or the relative probability of the Gaussianity and non-Gaussianity.

On large scales, in the Sachs-Wolfe regime, one can simplify the Bayesian approach by modeling directly the temperature anisotropy. Rocha et al. (2001) [214] discuss a Bayesian exploration of a model, where each spherical harmonic coefficient is drawn from a non-Gaussian distribution. In this regime, the simple form of the non-Gaussian potential for the local model (17) also translates into a simple model for the temperature anisotropy. Reference [215] develop several results for it, including an analytical expression of the evidence ratio of the Gaussian and non-Gaussian models. At the level of current data, this approximation is too restrictive,

since most of the information about  $f_{\text{NL}}$  is contained near the resolution limit of the experiment, where most of the measured perturbation modes are concentrated.

A full implementation of a physical non-Gaussian model must include the effect of Boltzmann transport. In the context of local non-Gaussianity, the model equation (17) suggests that a full Bayesian treatment may be feasible. At the time of writing, no fully Bayesian analysis for local  $f_{\text{NL}}$  has been published. The effort has focused on developing approximations to the full Bayesian problem.

Using a perturbative analysis, [216] relates the frequentist bispectrum estimator to moments of the Bayesian posterior distribution. Reference [217] described approximations to the full Bayesian treatment that simplify the analysis for high signal-to-noise maps and compared these to the full Bayesian treatment for a simple 1D toy model of non-Gaussian sky where this analysis is feasible.

## 7. Summary

The physics of the early universe responsible for generating the seed perturbations in the CMB is not understood. Inflation which is perhaps the most promising paradigm for generating seed perturbations allows for vast number of inflationary models that are compatible with data based on 2-point statistics like CMB power spectrum. Moreover, the alternatives to inflation such as cyclic models are also compatible with the data. Characterizing the non-Gaussianity in the primordial perturbations has emerged as probe for discriminating between different models of the early universe. Models based on slowly rolling single field produce undetectable amount of non-Gaussianity. Single field models without the slow roll can generate large (detectable with future experiments) non-Gaussianities but (1) cannot produce large non-Gaussianity of local type unless inflation started with excited vacuum state, (2) if non-Gaussianity is produced it would naturally be as bispectrum while higher order as trispectrum can be generated, it requires fine tuning.

The bispectrum of the CMB is one of the most promising tool for connecting the non-Gaussianities in the cosmic microwave background and the models of inflation. Bispectrum-based estimator which saturates Cramer-Rao bound has been developed and well characterised using non-Gaussian MonteCarlos. Other statistics although not as sensitive to non-Gaussianity as an optimally weighted bispectrum estimator do provide independent checks and have different systematics. While Bayesian analysis has been applied in the context of nonGaussianity analysis, this still appears to be an open area for fruitful investigation.

Given the importance of detecting primordial non-Gaussianity, it is crucial to characterise any nonprimordial sources of non-Gaussianities. We describe several sources of non-Gaussianities such as from second-order anisotropies after last scattering surface and during recombination.

With Planck launched and taking data, we look forward to the next few years as an exciting time in the exploration of primordial non-Gaussianity in the cosmic microwave background.

## Acknowledgment

A. P.S. Yadav gratefully acknowledges support from the IBM Einstein Fellowship.

## Endnotes

1. Although inflation is the most popular theory for the early universe, other mechanisms, for example, ekpyrotic models [218] and cyclic models [219, 220] have been proposed for generating nearly scale invariant Gaussian perturbations, while retaining homogeneity and flatness. In the cyclic universe, there is no beginning of time, and our expansion of the universe is one out of the infinite number of such cycles. Each cycle consists of the following phases. (1) A hot big bang phase, during which a structure formation takes place. (2) An accelerated expansion phase which dilutes the matter and radiation energy density. Since observations suggest that our universe is going through an accelerated expansion phase, in the cyclic model interpretation, we are presently going through this phase. (3) A decelerating phase, which makes the universe flat, and generates nearly Gaussian and scale invariant density perturbations. (4) A big crunch/bang transition phase during which matter and radiation is created. Although the mechanism is different, the outcome of phase (3) of the cyclic model is in some sense analogous to a slow-roll expansion phase of inflation; and phase (4) will correspond with the reheating phase in the inflationary scenario. As we will discuss in the next section these two scenarios can be distinguished by their different predictions about the gravitational waves, and non-Gaussianity. Cyclic models predict negligible contribution of gravitational waves while inflationary models can produce large gravitational wave contribution, which can be detected by next generation experiments. Second, cyclic models produce much larger non-Gaussianity (of local type) in comparison to the standard slow-roll inflationary scenario.
2. For the scale invariant ( $n = 1$ ) case,  $C_l^{\text{SW}} = [l(l+1)]^{-1} 6C_2^{\text{SW}}$ .
3. Example of some inflationary models are: eternal inflation, hybrid inflation, chaotic, Ghost inflation, Tilted Ghost inflation, DBI inflation, brane inflation, N-inflation, bubble inflation, extended inflation, false vacuum inflation, power law inflation, k-inflation, hyperextended inflation, supersymmetric inflation, Quintessential inflation, Natural inflation, Super inflation, Supernatural inflation, D-term inflation, B -inflation, Thermal inflation, discrete inflation, Assisted inflation, Polar cap inflation, Open inflation, Topological inflation, Double inflation, Multiple inflation, Induced-gravity inflation, Warm inflation, stochastic inflation, Generalized assisted inflation, self-sustained inflation, Graduated inflation, Local inflation, Singular inflation, Slinky inflation, locked inflation, Elastic inflation, Mixed

inflation, Phantom inflation, Boundary inflation, Non-commutative inflation, Tachyonic inflation, Tsunami inflation, Lambda inflation, Steep inflation, Oscillating inflation, Mutated Hybrid inflation, intermediate inflation, and Inhomogeneous inflation.

4. To first order in perturbations, primordial scalar perturbations do not generate B-modes of CMB. However at second (and higher) order in perturbations, scalar perturbations do produce B-modes [135, 221]. The B-modes generated from higher order perturbations are expected to be smaller than the tensor B-mode levels that the upcoming and future experiments (like CMBPol) are sensitive too.
5. Or equivalently  $\Phi(\mathbf{r}) = \Phi_L(\mathbf{r}) + f_{\text{NL}}(\Phi_L^2(\mathbf{r}) - \langle \Phi_L^2(\mathbf{r}) \rangle)$ .
6. We will refer to the estimator in [25] as a near-optimal-v1, while the corrected estimator of [28] as near-optimal.

## References

- [1] A. H. Guth, "Inflationary universe: a possible solution to the horizon and flatness problems," *Physical Review D*, vol. 23, no. 2, pp. 347–356, 1981.
- [2] K. Sato, "Cosmological baryon-number domain structure and the first order phase transition of a vacuum," *Physics Letters B*, vol. 99, no. 1, pp. 66–70, 1981.
- [3] A. D. Linde, "A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems," *Physics Letters B*, vol. 108, no. 6, pp. 389–393, 1982.
- [4] A. Albrecht and P. J. Steinhardt, "Cosmology for grand unified theories with radiatively induced symmetry breaking," *Physical Review Letters*, vol. 48, no. 17, pp. 1220–1223, 1982.
- [5] A. H. Guth and S.-Y. Pi, "Fluctuations in the new inflationary universe," *Physical Review Letters*, vol. 49, no. 15, pp. 1110–1113, 1982.
- [6] A. A. Starobinsky, "Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations," *Physics Letters B*, vol. 117, no. 3-4, pp. 175–178, 1982.
- [7] S. W. Hawking, "The development of irregularities in a single bubble inflationary universe," *Physics Letters B*, vol. 115, no. 4, pp. 295–297, 1982.
- [8] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, "Spontaneous creation of almost scale-free density perturbations in an inflationary universe," *Physical Review D*, vol. 28, no. 4, pp. 679–693, 1983.
- [9] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, "Theory of cosmological perturbations," *Physics Report*, vol. 215, no. 5-6, pp. 203–333, 1992.
- [10] D. S. Salopek and J. R. Bond, "Nonlinear evolution of long-wavelength metric fluctuations in inflationary models," *Physical Review D*, vol. 42, no. 12, pp. 3936–3962, 1990.
- [11] D. S. Salopek and J. R. Bond, "Stochastic inflation and nonlinear gravity," *Physical Review D*, vol. 43, no. 4, pp. 1005–1031, 1991.
- [12] T. Falk, R. Rangarajan, and M. Srednicki, "The angular dependence of the three-point correlation function of the cosmic microwave background radiation as predicted by inflationary cosmologies," *Astrophysical Journal*, vol. 403, no. 1, pp. L1–L3, 1993.
- [13] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, "The three-point correlation function of the cosmic microwave background in inflationary models," *Astrophysical Journal*, vol. 430, no. 2, pp. 447–457, 1994.
- [14] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, "Gauge-invariant second-order perturbations and non-Gaussianity from inflation," *Nuclear Physics B*, vol. 667, no. 1-2, pp. 119–148, 2003.
- [15] J. Maldacena, "Non-gaussian features of primordial fluctuations in single field inflationary models," *Journal of High Energy Physics*, vol. 7, no. 5, pp. 233–264, 2003.
- [16] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, "Non-Gaussianity from inflation: theory and observations," *Physics Reports*, vol. 402, no. 3-4, pp. 103–266, 2004.
- [17] E. Komatsu, D. N. Spergel, and B. D. Wandelt, "Measuring primordial non-Gaussianity in the cosmic microwave background," *Astrophysical Journal*, vol. 634, no. 1, pp. 14–19, 2005.
- [18] P. Cabella, F. K. Hansen, M. Liguori et al., "The integrated bispectrum as a test of cosmic microwave background non-Gaussianity: detection power and limits on fNL with WMAP data," *Monthly Notices of the Royal Astronomical Society*, vol. 369, no. 2, pp. 819–824, 2006.
- [19] K. M. Smith and M. Zaldarriaga, "Algorithms for bispectra: forecasting, optimal analysis, and simulation," <http://arxiv.org/abs/astro-ph/0612571>.
- [20] A. P. S. Yadav, E. Komatsu, B. D. Wandelt, M. Liguori, F. K. Hansen, and S. Matarrese, "Fast estimator of primordial non-Gaussianity from temperature and polarization anisotropies in the cosmic microwave background. II. Partial sky coverage and inhomogeneous noise," *Astrophysical Journal*, vol. 678, no. 2, pp. 578–582, 2008.
- [21] A. P. S. Yadav, E. Komatsu, and B. D. Wandelt, "Fast estimator of primordial non-Gaussianity from temperature and polarization anisotropies in the cosmic microwave background," *Astrophysical Journal*, vol. 664, no. 2, pp. 680–686, 2007.
- [22] E. Komatsu, B. D. Wandelt, D. N. Spergel, A. J. Banday, and K. M. Górski, "Measurement of the cosmic microwave background bispectrum on the COBE DMR sky maps," *Astrophysical Journal*, vol. 566, no. 1, pp. 19–29, 2002.
- [23] E. Komatsu, A. Kogut, M. R. Nolte et al., "First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: tests of Gaussianity," *Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 119–134, 2003.
- [24] D. N. Spergel, R. Bean, O. Dore' et al., "Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: implications for cosmology," *Astrophysical Journal, Supplement Series*, vol. 170, no. 2, pp. 377–408, 2007.
- [25] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, "Limits on non-Gaussianities from WMAP data," *Journal of Cosmology and Astroparticle Physics*, no. 5, article 4, 2006.
- [26] P. Creminelli, L. Senatore, M. Zaldarriaga, and M. Tegmark, "Limits on fNL parameters from WMAP 3yr data," *Journal of Cosmology and Astro-Particle Physics*, vol. 3, article 5, 2007.
- [27] G. Chen and I. Szapudi, "Constraining primordial non-gaussianities from the WMAP2 2-1 cumulant correlator power spectrum," *Astrophysical Journal*, vol. 647, no. 2, pp. L87–L90, 2006.
- [28] A. P. S. Yadav and B. D. Wandelt, "Evidence of primordial non-gaussianity (fNL) in the wilkinson microwave anisotropy probe 3-year data at  $2.8\sigma$ ," *Physical Review Letters*, vol. 100, no. 18, Article ID 181301, 2008.

- [29] P. Mukherjee and Y. Wang, "Wavelets and Wilkinson microwave anisotropy probe non-gaussianity," *Astrophysical Journal*, vol. 613, no. 1, pp. 51–60, 2004.
- [30] D. L. Larson and B. D. Wandelt, "The hot and cold spots in the Wilkinson Microwave Anisotropy Probe data are not hot and cold enough," *Astrophysical Journal*, vol. 613, no. 2, pp. L85–L88, 2004.
- [31] P. Vielva, E. Gonzalez, R. B. Barreiro, J. L. Sanz, and L. Cayón, "Detection of non-gaussianity in the Wilkinson microwave anisotropy probe first-year data using spherical wavelets," *Astrophysical Journal*, vol. 609, no. 1, pp. 22–34, 2004.
- [32] L.-Y. Chiang, P. D. Naselsky, O. V. Verkhodanov, and M. J. Way, "Non-Gaussianity of the derived maps from the first-year Wilkinson Microwave Anisotropy Probe data," *Astrophysical Journal*, vol. 590, no. 2, pp. L65–L68, 2003.
- [33] L.-Y. Chiang, P. D. Naselsky, and P. Coles, "The robustness of phase mapping as a non-Gaussianity test," *Astrophysical Journal*, vol. 602, no. 1, pp. L1–L4, 2004.
- [34] D. Babich and M. Zaldarriaga, "Primordial bispectrum information from CMB polarization," *Physical Review D*, vol. 70, no. 8, Article ID 083005, 2004.
- [35] A. P. S. Yadav and B. D. Wandelt, "CMB tomography: reconstruction of adiabatic primordial scalar potential using temperature and polarization maps," *Physical Review D*, vol. 71, no. 12, Article ID 123004, 2005.
- [36] J. M. Kovac, E. M. Leitch, C. Pryke, J. E. Carlstrom, N. W. Halverson, and W. L. Holzapfel, "Detection of polarization in the cosmic microwave background using DASI," *Nature*, vol. 420, no. 6917, pp. 772–787, 2002.
- [37] A. Kogut, D. N. Spergel, C. Barnes et al., "First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: temperature-polarization correlation," *Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 161–173, 2003.
- [38] L. Page, G. Hinshaw, E. Komatsu et al., "Three-year Wilkinsonmicrowave Anisotropy Probe (WMAP) observations: polarization analysis," *Astrophysical Journal, Supplement Series*, vol. 170, no. 2, pp. 335–376, 2007.
- [39] T. E. Montroy, P. A. R. Ade, J. J. Bock et al., "A Measurement of the CMB (EE) Spectrum from the 2003 flight of BOOMERANG," *Astrophysical Journal*, vol. 647, no. 2, pp. 813–822, 2006.
- [40] D. Baumann, "TASI lectures on inflation," <http://arxiv.org/abs/0907.5424>.
- [41] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, "Theory of cosmological perturbations," *Physics Report*, vol. 215, no. 5-6, pp. 203–333, 1992.
- [42] W. A. Bardeen, R. B. Pearson, and E. Rabinovici, "Hadron masses in quantum chromodynamics on the transverse lattice," *Physical Review D*, vol. 21, no. 4, pp. 1037–1054, 1980.
- [43] H. Kodama and M. Sasaki, "Cosmological perturbation theory," *Progress of Theoretical Physics Supplement*, no. 78, 1984.
- [44] R. K. Sachs and A. M. Wolfe, "Perturbations of a cosmological model and angular variations of the microwave background," *The Astrophysical Journal*, vol. 147, p. 73, 1967.
- [45] J. R. Bond and G. Efstathiou, "The statistics of cosmic background radiation fluctuations," *Monthly Notices of the Royal Astronomical Society*, vol. 226, pp. 655–687, 1987.
- [46] P. J. E. Peebles and J. T. Yu, "Primeval adiabatic perturbation in an expanding universe," *The Astrophysical Journal*, vol. 162, pp. 815–836, 1970.
- [47] U. Seljak and M. Zaldarriaga, "A line-of-sight integration approach to cosmic microwave background anisotropies," *Astrophysical Journal*, vol. 469, no. 2, pp. 437–444, 1996.
- [48] E. Komatsu, J. Dunkley, M. R. Nolte et al., "Five-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation," *The Astrophysical Journal Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [49] D. Baumann, M. G. Jackson, P. Adshead et al., "Probing inflation with CMB polarization," in *Proceedings of the CMB Polarization Workshop. Theory and Foregrounds: CMBPol Mission Concept*, vol. 1141 of *AIP Conference Proceedings*, pp. 10–120, 2009.
- [50] B. P. Abbott, R. Abbott, F. Acernese et al., "An upper limit on the stochastic gravitational-wave background of cosmological origin," *Nature*, vol. 460, no. 7258, pp. 990–994, 2009.
- [51] U. Seljak and M. Zaldarriaga, "Signature of gravity waves in the polarization of the microwave background," *Physical Review Letters*, vol. 78, no. 11, pp. 2054–2057, 1997.
- [52] M. Kamionkowski, A. Kosowsky, and A. Stebbins, "Statistics of cosmic microwave background polarization," *Physical Review D*, vol. 55, no. 12, pp. 7368–7388, 1997.
- [53] M. Kamionkowski, A. Kosowsky, and A. Stebbins, "A probe of primordial gravity waves and vorticity," *Physical Review Letters*, vol. 78, no. 11, pp. 2058–2061, 1997.
- [54] ACTPol, BICEP, CAPMAP, CBI, Clover, CMBPol, EBEX, POLARBEAR, PIPER, QUaD et al., current and upcoming experiments.
- [55] U. Seljak, "Gravitational lensing effect on cosmic microwave background anisotropies: a power spectrum approach," *Astrophysical Journal*, vol. 463, no. 1, pp. 1–7, 1996.
- [56] M. Zaldarriaga and U. Seljak, "Gravitational lensing effect on cosmic microwave background polarization," *Physical Review D*, vol. 58, no. 2, Article ID 023003, 1998.
- [57] A. P. S. Yadav, R. Biswas, M. Su, and M. Zaldarriaga, "Constraining a spatially dependent rotation of the cosmic microwave background polarization," *Physical Review D*, vol. 79, no. 12, Article ID 123009, 2009.
- [58] W. Hu, M. M. Hedman, and M. Zaldarriaga, "Benchmark parameters for CMB polarization experiments," *Physical Review D*, vol. 67, no. 4, Article ID 043004, 2003.
- [59] A. P. S. Yadav, M. Su, and M. Zaldarriaga, "Primordial B-mode diagnostics and self-calibrating the CMB polarization," *Physical Review D*, vol. 81, no. 6, Article ID 063512, 2010.
- [60] A. D. Linde, "Generation of isothermal density perturbations in an inflationary universe," *JETP Letters*, vol. 40, pp. 1333–1336, 1984.
- [61] L. A. Kofman and A. D. Linde, "Generation of density perturbations in inflationary cosmology," *Nuclear Physics B*, vol. 282, pp. 555–588, 1987.
- [62] D. Polarski and A. A. Starobinsky, "Isocurvature perturbations in multiple inflationary models," *Physical Review D*, vol. 50, no. 10, pp. 6123–6129, 1994.
- [63] C. Gordon, D. Wands, B. A. Bassett, and R. Maartens, "Adiabatic and entropy perturbations from inflation," *Physical Review D*, vol. 63, no. 2, Article ID 023506, 2001.
- [64] A. D. Linde, "Generation of isothermal density perturbations in the inflationary universe," *Physics Letters B*, vol. 158, no. 5, pp. 375–380, 1985.
- [65] G. Efstathiou and J. R. Bond, "Isocurvature cold dark matter fluctuations," *Monthly Notices of the Royal Astronomical Society*, vol. 218, pp. 103–121, 1986.

- [66] P. J. E. Peebles, "Origin of the large-scale galaxy peculiar velocity field: a minimal isocurvature model," *Nature*, vol. 327, no. 6119, pp. 210–211, 1987.
- [67] H. Kodama and M. Sasaki, "Evolution of isocurvature perturbations I: Photon–baryon universe," *International Journal of Modern Physics A*, vol. 1, no. 1, pp. 265–301, 1986.
- [68] S. Weinberg, "Can nonadiabatic perturbations arise after single-field inflation?" *Physical Review D*, vol. 70, no. 4, Article ID 043541, 2004.
- [69] J. Garcia-Bellido and D. Wands, "Metric perturbations in two-field inflation," *Physical Review D*, vol. 53, no. 10, pp. 5437–5445, 1996.
- [70] M. Sasaki and E. D. Stewart, "A general analytic formula for the spectral index of the density perturbations produced during inflation," *Progress of Theoretical Physics*, vol. 95, no. 1, pp. 71–78, 1996.
- [71] M. Sasaki and T. Tanaka, "Super-horizon scale dynamics of multi-scalar inflation," *Progress of Theoretical Physics*, vol. 99, no. 5, pp. 763–781, 1998.
- [72] N. Bartolo, S. Matarrese, and A. Riotto, "Adiabatic and isocurvature perturbations from inflation: power spectra and consistency relations," *Physical Review D*, vol. 64, no. 12, Article ID 123504, 2001.
- [73] M. Bucher, K. Moodley, and N. Turok, "Characterizing the primordial cosmic perturbations using MAP and Planck," *Physical Review D*, vol. 66, no. 2, Article ID 023528, 2002.
- [74] M. Bucher, K. Moodley, and N. Turok, "Constraining isocurvature perturbations with cosmic microwave background polarization," *Physical Review Letters*, vol. 87, no. 19, Article ID 191301, 4 pages, 2001.
- [75] R. Bean, J. Dunkley, and E. Pierpaoli, "Constraining isocurvature initial conditions with WMAP 3-year data," *Physical Review D*, vol. 74, no. 6, Article ID 063503, 2006.
- [76] S. Gupta, A. Berera, A. F. Heavens, and S. Matarrese, "Non-Gaussian signatures in the cosmic background radiation from warm inflation," *Physical Review D*, vol. 66, no. 4, Article ID 043510, 2002.
- [77] T. J. Allen, B. Grinstein, and M. B. Wise, "Non-gaussian density perturbations in inflationary cosmologies," *Physics Letters B*, vol. 197, no. 1-2, pp. 66–70, 1987.
- [78] J. Lesgourgues, D. Polarski, and A. A. Starobinsky, "Quantum-to-classical transition of cosmological perturbations for non-vacuum initial states," *Nuclear Physics B*, vol. 497, no. 1-2, pp. 479–508, 1997.
- [79] J. Martin, A. Riazuelo, and M. Sakellariadou, "Nonvacuum initial states for cosmological perturbations of quantum-mechanical origin," *Physical Review D*, vol. 61, no. 8, Article ID 083518, pp. 1–15, 2000.
- [80] P. Creminelli, "On non-Gaussianities in single-field inflation," *Journal of Cosmology and Astroparticle Physics*, no. 10, pp. 53–62, 2003.
- [81] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, "Ghost inflation," *Journal of Cosmology and Astroparticle Physics*, no. 4, pp. 1–18, 2004.
- [82] G. Dvali, A. Gruzinov, and M. Zaldarriaga, "New mechanism for generating density perturbations from inflation," *Physical Review D*, vol. 69, no. 2, Article ID 023505, 2003.
- [83] M. Zaldarriaga, "Non-Gaussianities in models with a varying inflaton decay rate," *Physical Review D*, vol. 69, no. 4, Article ID 043508, 2004.
- [84] D. H. Lyth, C. Ungarelli, and D. Wands, "Primordial density perturbation in the curvaton scenario," *Physical Review D*, vol. 67, no. 2, Article ID 023503, 2003.
- [85] E. Komatsu, N. Afshordi, N. Bartolo et al., "Non-Gaussianity as a probe of the physics of the primordial universe and the astrophysics of the low redshift universe," <http://arxiv.org/abs/0902.4759>.
- [86] D. Babich, P. Creminelli, and M. Zaldarriaga, "The shape of non-Gaussianities," *Journal of Cosmology and Astroparticle Physics*, no. 8, pp. 199–217, 2004.
- [87] N. Bartolo, S. Matarrese, and A. Riotto, "Non-Gaussianity from inflation," *Physical Review D*, vol. 65, no. 10, Article ID 103505, 2002.
- [88] F. Bernardeau and J.-P. Uzan, "Non-Gaussianity in multi-field inflation," *Physical Review D*, vol. 66, Article ID 103506, 2002.
- [89] F. Bernardeau and J.-P. Uzan, "Inflationary models inducing non-Gaussian metric fluctuations," *Physical Review D*, vol. 67, no. 12, Article ID 121301, 2003.
- [90] M. Sasaki, "Multi-brid inflation and non-gaussianity," *Progress of Theoretical Physics*, vol. 120, no. 1, pp. 159–174, 2008.
- [91] A. Naruko and M. Sasaki, "Large non-Gaussianity from multi-brid inflation," *Progress of Theoretical Physics*, vol. 121, no. 1, pp. 193–210, 2009.
- [92] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall, "Conditions for large non-Gaussianity in two-field slow-roll inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 10, article 8, 2008.
- [93] C. T. Byrnes and D. Wands, "Curvature and isocurvature perturbations from two-field inflation in a slow-roll expansion," *Physical Review D*, vol. 74, no. 4, Article ID 043529, 2006.
- [94] D. Langlois, F. Vernizzi, and D. Wands, "Non-linear isocurvature perturbations and non-Gaussianities," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 12, article 4, 2008.
- [95] J. Valiviita, H. Assadullahi, and D. Wands, "Primordial non-gaussianity from multiple curvaton decay," <http://arxiv.org/abs/0806.0623>.
- [96] H. Assadullahi, J. Valiviita, and D. Wands, "Primordial non-Gaussianity from two curvaton decays," *Physical Review D*, vol. 76, no. 10, Article ID 103003, 2007.
- [97] J. Valiviita, M. Sasaki, and D. Wands, "Non-Gaussianity and constraints for the variance of perturbations in the curvaton model," <http://arxiv4.library.cornell.edu/abs/astro-ph/0610001>.
- [98] F. Vernizzi and D. Wands, "Non-Gaussianities in two-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 19, 2006.
- [99] L. E. Allen, S. Gupta, and D. Wands, "Non-Gaussian perturbations from multi-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 1, pp. 93–107, 2006.
- [100] A. Linde and V. Mukhanov, "Non-Gaussian isocurvature perturbations from inflation," *Physical Review D*, vol. 56, no. 2, pp. R535–R539, 1997.
- [101] L. Kofman, "Probing string theory with modulated cosmological fluctuations," <http://arxiv.org/abs/astro-ph/0303614>.
- [102] J.-L. Lehners and P. J. Steinhardt, "Non-Gaussian density fluctuations from entropically generated curvature perturbations in ekpyrotic models," *Physical Review D*, vol. 77, no. 6, Article ID 063533, 2008.
- [103] J.-L. Lehners and P. J. Steinhardt, "Intuitive understanding of non-Gaussianity in ekpyrotic and cyclic models," *Physical Review D*, vol. 78, no. 2, Article ID 023506, 2008.
- [104] K. Koyama, S. Mizuno, and D. Wands, "Curvature perturbations from ekpyrotic collapse with multiple fields," *Classical and Quantum Gravity*, vol. 24, no. 15, pp. 3919–3931, 2007.

- [105] R. Holman and A. J. Tolley, “Enhanced non-Gaussianity from excited initial states,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 5, article 1, 2008.
- [106] X. Chen, M.-X. Huang, S. Kachru, and G. Shiu, “Observational signatures and non-Gaussianities of general single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 1, article 2, 2007.
- [107] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, “Large-scale structure, the cosmic microwave background and primordial non-Gaussianity,” *Monthly Notices of the Royal Astronomical Society*, vol. 313, no. 1, pp. 141–147, 2000.
- [108] E. Komatsu and D. N. Spergel, “Acoustic signatures in the primary microwave background bispectrum,” *Physical Review D*, vol. 63, no. 6, Article ID 063002, 2001.
- [109] K. Koyama, S. Mizuno, F. Vernizzi, and D. Wands, “Non-Gaussianities from ekpyrotic collapse with multiple fields,” *Journal of Cosmology and Astroparticle Physics*, no. 11, pp. 1–21, 2007.
- [110] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, “Non-Gaussianities in new ekpyrotic cosmology,” *Physical Review Letters*, vol. 100, no. 17, Article ID 171302, 2008.
- [111] J. Maldacena, “Non-Gaussian features of primordial fluctuations in single field. Inflationary models,” *Journal of High Energy Physics*, vol. 2003, no. 5, article 13, 2003.
- [112] M. Alishahiha, E. Silverstein, and D. Tong, “DBI in the sky: non-Gaussianity from inflation with a speed limit,” *Physical Review D*, vol. 70, no. 12, Article ID 123505, 2004.
- [113] P. Creminelli, “On non-Gaussianities in single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 10, pp. 53–62, 2003.
- [114] D. Seery and J. E. Lidsey, “Primordial non-Gaussianities in single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 6, pp. 45–70, 2005.
- [115] C. Cheung, P. Creminelli, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, “The effective field theory of inflation,” *Journal of High Energy Physics*, vol. 2008, no. 3, article 14, 2008.
- [116] P. D. Meerburg, J. P. van der Schaar, and P. S. Corasaniti, “Signatures of initial state modifications on bispectrum statistics,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 18, 2009.
- [117] A. Penzias and R. Wilson, “A measurement of excess antenna temperature at 4080 Mc/S,” *The Astrophysical Journal*, vol. 142, pp. 419–421, 1965.
- [118] G. F. Smoot, C. L. Bennett, A. Kogut et al., “Structure in the COBE differential microwave radiometer first-year maps,” *Astrophysical Journal*, vol. 396, no. 1, pp. L1–L5, 1992.
- [119] A. Lewis, A. Challinor, and A. Lasenby, “Efficient computation of cosmic microwave background anisotropies in closed Friedmann-Robertson-Walker models,” *Astrophysical Journal*, vol. 538, no. 2, pp. 473–476, 2000.
- [120] L. Senatore, K. M. Smith, and M. Zaldarriaga, “Non-Gaussianities in single field inflation and their optimal limits from the WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 1, article 28, 2010.
- [121] J. R. Fergusson and E. P. S. Shellard, “Shape of primordial non-Gaussianity and the CMB bispectrum,” *Physical Review D*, vol. 80, no. 4, Article ID 043510, 2009.
- [122] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $\delta N_{\text{local}}$  from WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article 6, 2009.
- [123] E. Komatsu, K. M. Smith, J. Dunkley et al., “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation,” <http://arxiv.org/abs/1001.4538>.
- [124] P. Creminelli, L. Senatore, and M. Zaldarriaga, “Estimators for local non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, no. 3, article 19, 2007.
- [125] M. Liguori, A. Yadav, F. K. Hansen, E. Komatsu, S. Matarrese, and B. Wandelt, “Temperature and polarization CMB maps from primordial non-Gaussianities of the local type,” *Physical Review D*, vol. 76, no. 10, Article ID 105016, 2007.
- [126] N. Bartolo and A. Riotto, “On the non-Gaussianity from recombination,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 3, article 17, 2009.
- [127] E. Sefusatti, M. Liguori, A. P. S. Yadav, M. G. Jackson, and E. Pajer, “Constraining running non-gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article 22, 2009.
- [128] P. Serra and A. Cooray, “Impact of secondary non-Gaussianities on the search for primordial non-Gaussianity with CMB maps,” *Physical Review D*, vol. 77, no. 10, Article ID 107305, 2008.
- [129] D. M. Goldberg and D. N. Spergel, “Microwave background bispectrum. II. A probe of the low redshift universe,” *Physical Review D*, vol. 59, no. 10, Article ID 103002, 6 pages, 1999.
- [130] D. N. Spergel and D. M. Goldberg, “Microwave background bispectrum. I. Basic formalism,” *Physical Review D*, vol. 59, no. 10, Article ID 103001, 8 pages, 1999.
- [131] A. Cooray and W. Hu, “Imprint of reionization on the cosmic microwave background bispectrum,” *Astrophysical Journal*, vol. 534, no. 2, pp. 533–550, 2000.
- [132] A. Mangilli and L. Verde, “Non-Gaussianity and the CMB bispectrum: confusion between primordial and lensing-Rees-Sciama contribution?” *Physical Review D*, vol. 80, no. 12, Article ID 123007, 2009.
- [133] D. Babich and E. Pierpaoli, “Point source contamination in CMB non-Gaussianity analyses,” *Physical Review D*, vol. 77, no. 12, Article ID 123011, 2008.
- [134] C. Pitrou, J. Uzan, and F. Bernardeau, “The cosmic microwave background bispectrum from the non-linear evolution of the cosmological perturbations,” <http://arxiv.org/abs/1003.0481>.
- [135] N. Bartolo, S. Matarrese, and A. Riotto, “CMB anisotropies at second-order II: analytical approach,” *Journal of Cosmology and Astroparticle Physics*, no. 1, article 19, 2007.
- [136] W. Hu, D. Scott, and J. Silk, “Reionization and cosmic microwave background distortions: a complete treatment of second-order Compton scattering,” *Physical Review D*, vol. 49, no. 2, pp. 648–670, 1994.
- [137] S. Dodelson and J. M. Jubas, “Reionization and its imprint on the cosmic microwave background,” *Astrophysical Journal*, vol. 439, no. 2, pp. 503–516, 1995.
- [138] T. Pyne and S. M. Carroll, “Higher-order gravitational perturbations of the cosmic microwave background,” *Physical Review D*, vol. 53, no. 6, pp. 2920–2929, 1996.
- [139] S. Mollerach and S. Matarrese, “Cosmic microwave background anisotropies from second order gravitational perturbations,” *Physical Review D*, vol. 56, no. 8, pp. 4494–4502, 1997.
- [140] S. Matarrese, S. Mollerach, and M. Bruni, “Relativistic second-order perturbations of the Einstein-de Sitter universe,” *Physical Review D*, vol. 58, no. 4, Article ID 043504, 1998.
- [141] P. Creminelli and M. Zaldarriaga, “CMB 3-point functions generated by nonlinearities at recombination,” *Physical Review D*, vol. 70, no. 8, Article ID 083532, 2004.

- [142] N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant temperature anisotropies and primordial non-gaussianity,” *Physical Review Letters*, vol. 93, no. 23, Article ID 231301, 2004.
- [143] K. Tomita, “Relativistic second-order perturbations of nonzero- $\Lambda$  flat cosmological models and CMB anisotropies,” *Physical Review D*, vol. 71, no. 8, Article ID 083504, 11 pages, 2005.
- [144] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity of large-scale cosmic microwave background anisotropies beyond perturbation theory,” *Journal of Cosmology and Astroparticle Physics*, no. 8, pp. 177–196, 2005.
- [145] N. Bartolo, S. Matarrese, and A. Riotto, “The full second-order radiation transfer function for large-scale CMB anisotropies,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 10, 2006.
- [146] K. Tomita, “Second-order power spectra of CMB anisotropies due to primordial random perturbations in flat cosmological models,” *Physical Review D*, vol. 77, no. 10, Article ID 103521, 2008.
- [147] C. Pitrou, J.-P. Uzan, and F. Bernardeau, “Cosmic microwave background bispectrum on small angular scales,” *Physical Review D*, vol. 78, no. 6, Article ID 063526, 2008.
- [148] N. Bartolo, S. Matarrese, and A. Riotto, “Cosmic microwave background anisotropies at second order: I,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 6, article 24, 2006.
- [149] C. Pitrou, “The radiative transfer at second order: a full treatment of the Boltzmann equation with polarization,” *Classical and Quantum Gravity*, vol. 26, no. 6, Article ID 065006, 2009.
- [150] L. Senatore, S. Tassev, and M. Zaldarriaga, “Cosmological perturbations at second order and recombination perturbed,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 31, 2009.
- [151] L. Boubekur, P. Creminelli, G. D’Amico, J. Noreña, and F. Vernizzi, “Sachs-Wolfe at second order: the CMB bispectrum on large angular scales,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 29, 2009.
- [152] N. Bartolo, S. Matarrese, and A. Riotto, “Enhancement of non-gaussianity after inflation,” *Journal of High Energy Physics*, vol. 8, no. 4, pp. 147–159, 2004.
- [153] N. Bartolo, S. Matarrese, and A. Riotto, “Evolution of second-order cosmological perturbations and non-Gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, no. 1, pp. 47–70, 2004.
- [154] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity and the cosmic microwave background anisotropies,” <http://arxiv.org/abs/1001.3957>.
- [155] B. Novosyadlyj, “Perturbations of ionization fractions at the cosmological recombination epoch,” *Monthly Notices of the Royal Astronomical Society*, vol. 370, no. 4, pp. 1771–1782, 2006.
- [156] R. Khatri and B. D. Wandelt, “Crinkles in the last scattering surface: non-Gaussianity from inhomogeneous recombination,” *Physical Review D*, vol. 79, no. 2, Article ID 023501, 2009.
- [157] L. Senatore, S. Tassev, and M. Zaldarriaga, “Non-gaussianities from perturbing recombination,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article 38, 2009.
- [158] D. Nitta, E. Komatsu, N. Bartolo, S. Matarrese, and A. Riotto, “CMB anisotropies at second order III: bispectrum from products of the first-order perturbations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 14, 2009.
- [159] M. Liguori and A. Riotto, “Impact of uncertainties in the cosmological parameters on the measurement of primordial non-Gaussianity,” *Physical Review D*, vol. 78, no. 12, Article ID 123004, 2008.
- [160] M. Su, A. P. S. Yadav, and M. Zaldarriaga, “Impact of instrumental systematic contamination on the lensing mass reconstruction using the CMB polarization,” *Physical Review D*, vol. 79, no. 12, Article ID 123002, 2009.
- [161] D. Hanson, K. M. Smith, A. Challinor, and M. Liguori, “CMB lensing and primordial non-Gaussianity,” *Physical Review D*, vol. 80, no. 8, Article ID 083004, 2009.
- [162] A. Cooray, D. Sarker, and P. Serra, “Weak lensing of the primary CMB bispectrum,” *Physical Review D*, vol. 77, no. 12, Article ID 123006, 2008.
- [163] S. Donzelli, F. K. Hansen, M. Liguori, and D. Maino, “Impact of the  $1/f$  noise and the asymmetric beam on non-gaussianity searches with planck,” *Astrophysical Journal*, vol. 706, no. 2, pp. 1226–1240, 2009.
- [164] T. Okamoto and W. Hu, “Angular trispectra of CMB temperature and polarization,” *Physical Review D*, vol. 66, no. 6, Article ID 063008, 2002.
- [165] N. Kogo and E. Komatsu, “Angular trispectrum of CMB temperature anisotropy from primordial non-Gaussianity with the full radiation transfer function,” *Physical Review D*, vol. 73, no. 8, Article ID 083007, 2006.
- [166] P. Creminelli, L. Senatore, and M. Zaldarriaga, “Estimators for local non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 3, article 19, 2007.
- [167] D. Munshi, A. Heavens, A. Cooray et al., “New optimised estimators for the primordial trispectrum,” <http://arxiv.org/abs/0910.3693>.
- [168] D. M. Regan, E. P. S. Shellard, and J. R. Fergusson, “General CMB and primordial trispectrum estimation,” <http://arxiv.org/abs/1004.2915>.
- [169] D. Seery, J. E. Lidsey, and M. S. Sloth, “The inflationary trispectrum,” *Journal of Cosmology and Astroparticle Physics*, no. 1, article 27, 2007.
- [170] M.-X. Huang and G. Shiu, “Inflationary trispectrum for models with large non-Gaussianities,” *Physical Review D*, vol. 74, no. 12, article 121301, 2006.
- [171] X. Chen, B. Hu, M.-X. Huang, G. Shiu, and Y. Wang, “Large primordial trispectra in general single field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 8, 2009.
- [172] F. Arroja, S. Mizuno, K. Koyama, and T. Tanaka, “Full trispectrum in single field DBI inflation,” *Physical Review D*, vol. 80, no. 4, Article ID 043527, 2009.
- [173] K. T. Engel, K. S. M. Lee, and M. B. Wise, “Trispectrum versus bispectrum in single-field inflation,” *Physical Review D*, vol. 79, no. 10, Article ID 103530, 2009.
- [174] C. T. Byrnes, M. Sasaki, and D. Wands, “Primordial trispectrum from inflation,” *Physical Review D*, vol. 74, no. 12, Article ID 123519, 2006.
- [175] J.-L. Lehners and P. J. Steinhardt, “Non-Gaussianity generated by the entropic mechanism in bouncing cosmologies made simple,” *Physical Review D*, vol. 80, no. 10, Article ID 103520, 2009.
- [176] J. Smidt, A. Amblard, A. Cooray et al., “A measurement of cubic-order primordial non-Gaussianity (gnl and taunl) with WMAP 5-year data,” <http://arxiv.org/abs/1001.5026>.
- [177] K. R. Mecke, T. Buchert, and H. Wagner, “Robust morphological measures for large-scale structure in the universe,” *Astronomy and Astrophysics*, vol. 288, pp. 697–704, 1994.

- [178] J. Schmalzing and T. Buchert, “Beyond genus statistics: a unifying approach to the morphology of cosmic structure,” *Astrophysical Journal*, vol. 482, no. 1, pp. L1–L4, 1997.
- [179] J. Schmalzing and K. M. Górski, “Minkowski functionals used in the morphological analysis of cosmic microwave background anisotropy maps,” *Monthly Notices of the Royal Astronomical Society*, vol. 297, no. 2, pp. 355–365, 1998.
- [180] S. Winitzki and A. Kosowsky, “Minkowski functional description of microwave background Gaussianity,” *New Astronomy*, vol. 3, no. 2, pp. 75–99, 1998.
- [181] T. Matsubara, “Analytic expression of the genus in a weakly non-Gaussian field induced by gravity,” *Astrophysical Journal*, vol. 434, no. 2, pp. L43–L46, 1994.
- [182] T. Matsubara, “Statistics of smoothed cosmic fields in perturbation theory. I. Formulation and useful formulae in second-order perturbation theory,” *Astrophysical Journal*, vol. 584, no. 1, pp. 1–33, 2003.
- [183] C. Hikage, E. Komatsu, and T. Matsubara, “Primordial non-gaussianity and analytical formula for minkowski functionals of the cosmic microwave background and large-scale structure,” *Astrophysical Journal*, vol. 653, no. 1, pp. 11–26, 2006.
- [184] E. Komatsu, B. D. Wandelt, D. N. Spergel, A. J. Banday, and K. M. Górski, “Measurement of the cosmic microwave background bispectrum on the COBE DMR sky maps,” *Astrophysical Journal*, vol. 566, no. 1, pp. 19–29, 2002.
- [185] M. G. Santos, A. Heavens, A. Balbi et al., “Multiple methods for estimating the bispectrum of the cosmic microwave background with application to the MAXIMA data,” *Monthly Notices of the Royal Astronomical Society*, vol. 341, no. 2, pp. 623–643, 2003.
- [186] S. Smith, G. Rocha, A. Challinor et al., “Estimating the bispectrum of the very small array data,” *Monthly Notices of the Royal Astronomical Society*, vol. 352, no. 3, pp. 887–902, 2004.
- [187] G. de Troia, P. A. R. Ade, J. J. Bock et al., “Searching for non-Gaussian signals in the BOOMERanG 2003 CMB map: preliminary results,” *New Astronomy Reviews*, vol. 51, no. 3-4, pp. 250–255, 2007.
- [188] C. Hikage, T. Matsubara, P. Coles, M. Liguori, F. K. Hansen, and S. Matarrese, “Limits on primordial non-Gaussianity from Minkowski Functionals of the WMAP temperature anisotropies,” *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 3, pp. 1439–1446, 2008.
- [189] A. Curto, J.F. Macías-Pérez, E. Martínez-González et al., “Constraints on the non-linear coupling parameter  $f_{\text{nl}}$  with Archeops data,” *Astronomy and Astrophysics*, vol. 486, no. 2, pp. 383–391, 2008.
- [190] A. Curto, E. Martínez-González, and R. B. Barreiro, “Improved constraints on primordial non-gaussianity for the wilkinson microwave anisotropy probe 5-year data,” *Astrophysical Journal*, vol. 706, no. 1, pp. 399–403, 2009.
- [191] P. Natoli, G. De Troia, C. Hikage et al., “BOOMERanG constraints on primordial non-Gaussianity from analytical minkowski functionals,” <http://arxiv.org/abs/0905.4301>.
- [192] J. Smidt, A. Amblard, P. Serra, and A. Cooray, “Measurement of primordial non-Gaussianity using the WMAP 5-year temperature skewness power spectrum,” *Physical Review D*, vol. 80, no. 12, Article ID 123005, 2009.
- [193] C. Hikage, E. Komatsu, and T. Matsubara, “Primordial non-gaussianity and analytical formula for minkowski functionals of the cosmic microwave background and large-scale structure,” *Astrophysical Journal*, vol. 653, no. 1, pp. 11–26, 2006.
- [194] N. Aghanim and O. Forni, “Searching for the non-Gaussian signature of the CMB secondary anisotropies,” *Astronomy and Astrophysics*, vol. 347, no. 2, pp. 409–418, 1999.
- [195] M. P. Hobson, A. W. Jones, and A. N. Lasenby, “Wavelet analysis and the detection of non-Gaussianity in the cosmic microwave background,” *Monthly Notices of the Royal Astronomical Society*, vol. 309, no. 1, pp. 125–140, 1999.
- [196] P. Vielva, E. Martínez-González, R. B. Barreiro, J. L. Sanz, and L. Cayón, “Detection of non-gaussianity in the Wilkinson microwave anisotropy probe first-year data using spherical wavelets,” *Astrophysical Journal*, vol. 609, no. 1, pp. 22–34, 2004.
- [197] J. -L. Starck, N. Aghanim, and O. Forni, “Detection and discrimination of cosmological non-gaussian signatures by multi-scale methods,” *Astronomy and Astrophysics*, vol. 416, no. 1, pp. 9–17, 2004.
- [198] J. Jin, J.-L. Starck, D. L. Donoho, N. Aghanim, and O. Forni, “Cosmological non-Gaussian signature detection: comparing performance of different statistical tests,” *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 15, pp. 2470–2485, 2005.
- [199] P. Vielva, Y. Wiaux, E. Martínez-González, and P. Vandergheynst, “Steerable wavelet analysis of CMB structures alignment,” *New Astronomy Reviews*, vol. 50, no. 11-12, pp. 880–888, 2006.
- [200] J. D. McEwen, P. Vielva, Y. Wiaux et al., “Cosmological applications of a wavelet analysis on the sphere,” *Journal of Fourier Analysis and Applications*, vol. 13, no. 4, pp. 495–510, 2007.
- [201] L. Cayón, E. Martínez-González, F. Argüeso, A. J. Banday, and K. M. Górski, “COBE-DMR constraints on the non-linear coupling parameter: a wavelet based method,” *Monthly Notices of the Royal Astronomical Society*, vol. 339, no. 4, pp. 1189–1194, 2003.
- [202] A. Curto, E. Martínez-González, P. Mukherjee et al., “Wilkinson microwave anisotropy probe 5-yr constraints on  $f_{\text{nl}}$  with wavelets,” *Monthly Notices of the Royal Astronomical Society*, vol. 393, no. 2, pp. 615–622, 2009.
- [203] F. J. Narcowich, P. Petrushev, and J. D. Ward, “Localized tight frames on spheres,” *SIAM Journal on Mathematical Analysis*, vol. 38, no. 2, pp. 574–594, 2006.
- [204] F. Narcowich, P. Petrushev, and J. Ward, “Decomposition of Besov and Triebel-Lizorkin spaces on the sphere,” *Journal of Functional Analysis*, vol. 238, no. 2, pp. 530–564, 2006.
- [205] P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard, “High frequency asymptotics for wavelet-based tests for Gaussianity and isotropy on the torus,” *Journal of Multivariate Analysis*, vol. 99, no. 4, pp. 606–636, 2008.
- [206] P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard, “Asymptotics for spherical needlets,” *Annals of Statistics*, vol. 37, no. 3, pp. 1150–1171, 2009.
- [207] P. Baldi, G. Kerkycharian, D. Marinucci, and D. Picard, “Subsampling needlet coefficients on the sphere,” *Bernoulli*, vol. 15, pp. 438–463, 2009.
- [208] X. Lan and D. Marinucci, “The needlets bispectrum,” *Electronic Journal of Statistics*, vol. 2, pp. 332–367, 2008.
- [209] D. Pietrobon, P. Cabella, A. Balbi, G. De Gasperis, and N. Vittorio, “Constraints on primordial non-Gaussianity from a needlet analysis of the WMAP-5 data,” *Monthly Notices of the Royal Astronomical Society*, vol. 396, no. 3, pp. 1682–1688, 2009.
- [210] D. Pietrobon, P. Cabella, A. Balbi, R. Crittenden, G. de Gasperis, and N. Vittorio, “Needlet bispectrum asymmetries in the WMAP 5-year data,” *Monthly Notices of the Royal Astronomical Society*, vol. 402, p. L34, 2010.

- [211] P. Cabella, D. Pietrobon, M. Veneziani et al., “Foreground influence on primordial non-Gaussianity estimates: needlet analysis of WMAP 5-year data,” *Monthly Notices of the Royal Astronomical Society*, vol. 504, pp. 961–968, 2010.
- [212] Ø. Rudjord, F. K. Hansen, X. Lan, M. Liguori, D. Marinucci, and S. Matarrese, “An estimate of the primordial non-gaussianity parameter  $f_{NL}$  using the needlet bispectrum from wmap,” *Astrophysical Journal*, vol. 701, no. 1, pp. 369–376, 2009.
- [213] Ø. Rudjord, F. K. Hansen, X. Lan, M. Liguori, D. Marinucci, and S. Matarrese, “Directional variations of the non-gaussianity parameter  $f_{NL}$ ,” *Astrophysical Journal*, vol. 708, no. 2, pp. 1321–1325, 2010.
- [214] G. Rocha, J. Magueijo, M. Hobson, and A. Lasenby, “Bayesian joint estimation of non-Gaussianity and the power spectrum,” *Physical Review D*, vol. 64, no. 6, Article ID 063512, 2001.
- [215] P. Vielva and J. L. Sanz, “Analysis of non-Gaussian CMB maps based on the N-pdf. Application to WMAP data,” <http://arxiv.org/abs/0812.1756>.
- [216] T. A. Enßlin, M. Frommert, and F. S. Kitaura, “Information field theory for cosmological perturbation reconstruction and non-linear signal analysis,” *Physical Review D*, vol. 80, Article ID 105005, 2009.
- [217] F. Elsner, B. D. Wandelt, and M. D. Schneider, “Probing local non-Gaussianities within a Bayesian framework,” *Astronomy and Astrophysics*, vol. 513, no. 11, 2010.
- [218] J. Khoury, B. A. Ovrut, P. J. Steinhardt, and N. Turok, “Ekpyrotic universe: colliding branes and the origin of the hot big bang,” *Physical Review D*, vol. 64, no. 12, Article ID 123522, 2001.
- [219] P. J. Steinhardt and N. Turok, “A cyclic model of the universe,” *Science*, vol. 296, no. 5572, pp. 1436–1439, 2002.
- [220] P. J. Steinhardt and N. Turok, “Cosmic evolution in a cyclic universe,” *Physical Review D*, vol. 65, no. 12, Article ID 126003, 2002.
- [221] D. Baumann, P. Steinhardt, K. Takahashi, and K. Ichiki, “Gravitational wave spectrum induced by primordial scalar perturbations,” *Physical Review D*, vol. 76, no. 8, Article ID 084019, 2007.

## Review Article

# Primordial Non-Gaussianity and Bispectrum Measurements in the Cosmic Microwave Background and Large-Scale Structure

Michele Liguori,<sup>1</sup> Emiliano Sefusatti,<sup>2</sup> James R. Fergusson,<sup>1</sup> and E. P. S. Shellard<sup>1</sup>

<sup>1</sup>Department of Applied Mathematics and Theoretical Physics, Centre for Theoretical Cosmology, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK

<sup>2</sup>Institut de Physique Théorique, Commissariat à l'Énergie Atomique, 91191 Gif-sur-Yvette, France

Correspondence should be addressed to Michele Liguori, ml453@damtp.cam.ac.uk

Received 5 February 2010; Accepted 12 May 2010

Academic Editor: Dragan Huterer

Copyright © 2010 Michele Liguori et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The most direct probe of non-Gaussian initial conditions has come from bispectrum measurements of temperature fluctuations in the Cosmic Microwave Background and of the matter and galaxy distribution at large scales. Such bispectrum estimators are expected to continue to provide the best constraints on the non-Gaussian parameters in future observations. We review and compare the theoretical and observational problems, current results, and future prospects for the detection of a nonvanishing primordial component in the bispectrum of the Cosmic Microwave Background and large-scale structure, and the relation to specific predictions from different inflationary models.

## 1. Introduction

The standard inflationary paradigm predicts a flat Universe perturbed by nearly-Gaussian and scale-invariant primordial perturbations. These predictions have been verified to a high degree of accuracy by Cosmic Microwave Background (CMB) and Large-Scale Structure (LSS) measurements, such as those provided by the Wilkinson Microwave Anisotropy Probe (WMAP) [1], the 2dF Galaxy Redshift Survey (2dFGRS) [2], and the Sloan Digital Sky Survey (SDSS) [3]. Despite this success, it has proved to be difficult to discriminate between the vast array of inflationary scenarios that have been proposed by high-energy theoretical investigations or even to rule-out alternatives to inflation. Since most of the present constraints on the Lagrangian of the inflaton field have been obtained from measurements of the two-point function, or power spectrum, of the primordial fluctuations, a natural step to extend the available information is to look at non-Gaussian signatures in higher-order correlators.

The lowest-order additional correlator to take into account is the three-point function or its counterpart in Fourier space, the *bispectrum*. Most models of inflation are characterized by specific predictions for the bispectrum of

the primordial perturbations in the gravitational potential  $\Phi(\mathbf{k})$ . The bispectrum  $B_\Phi(k_1, k_2, k_3)$  of these perturbations is defined as

$$\langle \Phi(\mathbf{k}_1)\Phi(\mathbf{k}_2)\Phi(\mathbf{k}_3) \rangle \equiv (2\pi)^3 \delta_D(\mathbf{k}_{123}) B_\Phi(k_1, k_2, k_3), \quad (1)$$

where we have introduced the notation  $\mathbf{k}_{ij} \equiv \mathbf{k}_i + \mathbf{k}_j$  so that the Dirac delta function here is  $\delta_D(\mathbf{k}_{123}) \equiv \delta_D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3)$ . Together with the assumption of statistical homogeneity and isotropy for the primordial perturbations, this implies that the bispectrum is a function of the triplet defined by the magnitude of the wavenumbers  $k_1$ ,  $k_2$ , and  $k_3$  forming a closed triangular configuration. The current constraints that we are able to derive on the bispectrum  $B_\Phi(k_1, k_2, k_3)$  provide additional information about the early Universe; the possible detection of a non-vanishing primordial bispectrum in future observations would represent a major discovery, especially as it is predicted to be negligible by standard inflation.

The cosmological observable most directly related to the initial curvature bispectrum is given by the bispectrum of the CMB temperature fluctuations, which provide a map of the density perturbations at the time of decoupling, the earliest

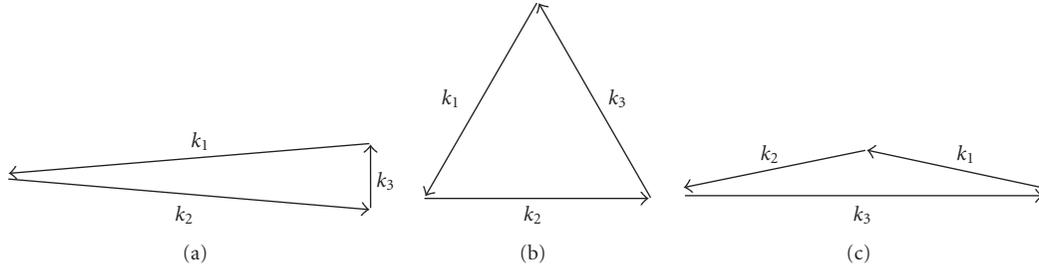


FIGURE 1: Triangle types contributing to the bispectrum corresponding to “squeezed” or local configurations with  $k_3 \ll k_1, k_2$  (a), equilateral configurations with  $k_3 \approx k_1 \approx k_2$  (b), and flattened configurations with  $k_3 \approx k_1 + k_2$  (c).

information we have about the Universe. Current measurements of individual triangular configurations of the CMB bispectrum are, however, consistent with zero. Studies of the primordial bispectrum, therefore, are usually characterized by constraints on a single-*amplitude parameter*, denoted by  $f_{\text{NL}}$ , once a specific model for  $B_\Phi$  is assumed. Since most models predict a curvature bispectrum obeying the hierarchical scaling  $B_\Phi(k, k, k) \sim P_\Phi^2(k)$ , with  $P_\Phi(k)$  being the curvature power spectrum, the non-Gaussian parameter roughly quantifies the ratio  $f_{\text{NL}} \sim B_\Phi(k, k, k)/P_\Phi^2(k)$ , defining the “strength” of the primordial non-Gaussian signal. In addition, we can write

$$B_\Phi(k_1, k_2, k_3) \equiv f_{\text{NL}} F(k_1, k_2, k_3), \quad (2)$$

where  $F(k_1, k_2, k_3)$  encodes the functional dependence of the primordial bispectrum on the specific triangle configurations. For brevity, the characteristic shape-dependence of a given bispectrum is often referred to simply as the *bispectrum shape* (a precise definition of the bispectrum shape function will be given in Section 2.1). Inflationary predictions for both the amplitude  $f_{\text{NL}}$  and the shape of  $B_\Phi$  are strongly model dependent. Notice that the subscript “NL” stands for “nonlinear”, since a common phenomenological model for the non-Gaussianity of the initial conditions can be written as a simple nonlinear transformation of a Gaussian field. Generically, of course, non-Gaussianity is associated with nonlinearities, such as nontrivial dynamics during inflation, resonant behaviour at the end of inflation (“preheating”), or nonlinear postinflationary evolution. At the very least, future CMB and LSS observations are expected to be able to eventually detect the last of the three effects mentioned above.

Perturbations in the CMB provide a particularly convenient test of the primordial density field because CMB temperature and polarization anisotropies are small enough to be studied in the *linear regime* of cosmological perturbations. Once the effects of foregrounds are properly taken into account, a non-vanishing CMB bispectrum at large scales would be a direct consequence of a non-vanishing primordial bispectrum. As we will see, while other CMB probes of primordial non-Gaussianity are available, such as tests of the topological properties of the temperature map based on Minkowski Functionals or measurements of the CMB trispectrum, the estimator for the non-Gaussian parameter  $f_{\text{NL}}$  has been shown to be optimal. We will focus mostly

on this bispectrum estimator in the section of this paper dedicated to the CMB.

In the standard cosmological model, the large-scale structure of the Universe, that is, the distribution of matter and galaxies on large scales, is the result of the nonlinear evolution due to gravitational instability of the same initial density perturbations responsible for the CMB anisotropies. This is, perhaps, the most important prediction of the inflationary framework which provides a common origin for the CMB and large-scale structure perturbations as the result of tiny quantum fluctuations stretched over cosmological scales during a phase of accelerated expansion. The large-scale structure we observe at low redshift, however, is characterized by large voids and small regions with very large-matter density, and it is therefore a much less direct probe of the initial conditions. The *distribution of matter becomes a highly non-Gaussian field* precisely as a result of the nonlinear growth of structures, *even for Gaussian initial conditions*. This non-Gaussianity is expressed, in particular, by a non-vanishing matter bispectrum at *any* measurable scale, including the largest scales probed by current or future redshift surveys. In this context, the effect of primordial non-Gaussianity, that is, of an initial component in the curvature bispectrum, will constitute a *correction* to the galaxy bispectrum. It follows that the possibility of constraining or detecting this initial component is strictly related to our ability to *distinguish* it from other primary sources of non-Gaussianity, that is, the nonlinear gravitational evolution, and, in the case of galaxy surveys, nonlinear bias.

The study of non-Gaussian initial conditions for large-scale structure has a relatively long history, with important contributions going back to the mid eighties. The standard picture that has been developed over the years assumed that, at large scales, the effect of primordial non-Gaussianity on the galaxy distribution is simply given in terms of an additional component to the galaxy bispectrum. This is obtained, in perturbation theory, as the linearly evolved and linearly biased initial matter bispectrum, related to the curvature bispectrum  $B_\Phi(k_1, k_2, k_3)$  by the Poisson equation. Such component becomes subdominant as the gravity-induced non-Gaussian contribution grows in time. In this framework, as one can expect, high-redshift and large-volume galaxy surveys would constitute the best probes of the initial conditions. It has been shown, in fact, that proposed and planned redshift surveys, such as those of Euclid [4],

should be able to provide constraints on the primordial non-Gaussian parameters comparable to, if not better than, those expected from CMB missions such as those of Planck. What is more important, in the event of a detection by Planck, is that confirmation by large-scale structure observations will be required.

Recent results from N-body simulations with non-Gaussian initial conditions, however, have revealed a more complex picture. The effect of primordial non-Gaussianity at large scales is not limited to an additional contribution to the galaxy bispectrum, but it quite dramatically affects the galaxy bias relation itself, that is, the relation between the matter and galaxy distributions. A surprising consequence is that it induces a large correction even for the galaxy power spectrum. Such an effect has attracted considerable recent attention and, remarkably, has placed constraints on the non-Gaussian parameter from current LSS datasets which already appear to marginally improve on CMB limits. However, from a theoretical point of view, a proper understanding of the phenomenon is not fully developed yet. For example, reliable predictions for the galaxy bispectrum are not yet available. Most importantly, as for general cosmological parameter estimation, a complete likelihood analysis aimed at constraining, or detecting, primordial non-Gaussianity in large-volume redshift surveys should involve *joint* measurements of the galaxy power spectrum and bispectrum, as well as possibly higher-order correlation functions. While we are still far from a proper assessment of what such analysis would be able to achieve, current results in this direction are very encouraging.

This review is divided into four parts. In Section 2 we will first discuss initial conditions as defined in terms of the primordial curvature bispectrum and its phenomenology. We will then review the observational consequences of primordial non-Gaussianity on the CMB bispectrum, Section 3, and on the large-scale structure bispectrum as measured in redshift surveys, Section 4. In both cases we will discuss theoretical models for the observed bispectra and technical problems related to the estimation of the non-Gaussian parameters, with the differences that naturally characterize such distinct observables. We also give an example of joint analysis using both CMB and large-scale structure when we consider the possibility of constraining a strongly scale-dependent non-Gaussian parameter  $f_{\text{NL}}(k)$ , emerging in some recently proposed inflationary models.

## 2. Initial Conditions and the Primordial Bispectrum

In this section we will briefly overview the main predictions of inflationary models regarding the non-Gaussianity (NG) of the primordial curvature perturbation field. The link between NG of primordial density fluctuations and NG of CMB and LSS will be shown in following sections. In order to provide a full description of an NG random field, all correlators beyond the 2-point function are in principle necessary. However in this review we will focus on the primordial bispectrum (i.e., three-point function

in Fourier space). This is not only justified by the fact that the bispectrum is the first and simplest higher-order correlator to look at, but also by the fact that most models of inflation predict vanishingly small correlators beyond the bispectrum. In Section 2.1 we will introduce the relevant quantities, their mathematical definitions, and we will provide a general overview and classification of the bispectra predicted in different inflationary scenarios (only from a purely mathematical point of view, without linking them to the Physics originating them at this stage). Finally, a useful eigenmode expansion technique for bispectra will be introduced in Section 2.2 and applied to the calculation of correlations between different bispectra in Section 3.4. In the same section we will also show which kinds of bispectra are predicted by different models of inflation.

*2.1. The Primordial Bispectrum and Shape Function.* The starting point for this discussion is the primordial gravitational potential perturbation  $\Phi(\mathbf{x}, t)$  which was seeded by quantum fluctuations during inflation or by some other mechanism in the very early Universe ( $t \ll t_{\text{dec}}$ ). When characterizing the fluctuations  $\Phi$ , we usually work in Fourier space with the (flat space) transform defined through

$$\Phi(\mathbf{x}, t) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{-i\mathbf{k}\cdot\mathbf{x}} \Phi(\mathbf{k}, t). \quad (3)$$

The primordial power spectrum  $P_\Phi(k)$  of these potential fluctuations is found using an ensemble average:

$$\langle \Phi(\mathbf{k}) \Phi^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_D(\mathbf{k} - \mathbf{k}') P_\Phi(k), \quad (4)$$

where we have assumed that physical processes creating the fluctuations are statistically isotropic so that only the dependence on the wavenumber remains  $k = |\mathbf{k}|$ . Recall that, for nearly scale-invariant perturbations, the fluctuation variance on the horizon scale  $k \approx H$  is almost constant  $\Delta_{k \sim H}^2 \approx k^3 P_\Phi(k) / 2\pi^2 \approx \text{const.}$ , implying that  $P_\Phi(k) \sim k^{-3}$ .

The primordial bispectrum  $B_\Phi(k, k_2, k_3)$  is found from the Fourier transform of the three-point correlator as

$$\langle \Phi(\mathbf{k}_1) \Phi(\mathbf{k}_2) \Phi(\mathbf{k}_3) \rangle = (2\pi)^3 \delta_D(\mathbf{k}_{123}) B_\Phi(k_1, k_2, k_3). \quad (5)$$

Here, the delta function enforces the triangle condition, that is, the constraint that the wavevectors in Fourier space must close to form a triangle:  $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$ . Examples of such triangles are shown in Figure 1, illustrating the basic squeezed, equilateral, and flattened triangles to which we will refer later. Note that a specific triangle can be completely described by the three lengths of its sides and so, in the isotropic case, we are able to describe the bispectrum using only the wavenumbers  $k_1, k_2, k_3$ . The triangle condition restricts the allowed wavenumber configurations  $(k_1, k_2, k_3)$  to the interior of the tetrahedron illustrated in Figure 2.

The most studied primordial bispectrum is the *local model* in which contributions from “squeezed” triangles are dominant, that is, with, for example,  $k_3 \ll k_1, k_2$  (as illustrated in Figure 1(a)). This is well motivated physically as it encompasses “superhorizon” effects during inflation when a large-scale mode  $k_3$  (say) which has exited the

Hubble radius exerts a nonlinear influence on the subsequent evolution of smaller-scale modes  $k_1$ ,  $k_2$ . Although this effect is small in single-field slow-roll inflation, it can be much larger for multifield models. In a weakly coupled regime, the potential can be split into two components: the linear term  $\Phi_L$ , representing a Gaussian field, giving the usual perturbation results, plus a small local non-Gaussian term  $\Phi_{NL}$  [5]:

$$\begin{aligned}\Phi(\mathbf{x}) &= \Phi_L(\mathbf{x}) + \Phi_{NL}(\mathbf{x}) \\ &= \Phi_L(\mathbf{x}) + f_{NL}[\Phi_L^2(\mathbf{x}) - \langle \Phi_L^2(\mathbf{x}) \rangle],\end{aligned}\quad (6)$$

where  $f_{NL}$  is called the nonlinearity parameter. In Fourier space, the nonlinear term is then given by the convolution

$$\Phi_{NL}(\mathbf{k}) = f_{NL} \left[ \int \frac{d^3\mathbf{k}'}{(2\pi)^3} \Phi_L(\mathbf{k}+\mathbf{k}')\Phi_L(\mathbf{k}') - (2\pi)^3 \delta_D(\mathbf{k}) \langle \Phi_L^2 \rangle \right]. \quad (7)$$

From this we can infer, using (4), that the only non-vanishing contributions to the bispectrum (5) take the form

$$\begin{aligned}\langle \Phi(\mathbf{k}_1)\Phi(\mathbf{k}_2)\Phi(\mathbf{k}_3) \rangle \\ = 2(2\pi)^3 \delta_D(\mathbf{k}_{123}) [P_\Phi(k_1)P_\Phi(k_2) + P_\Phi(k_2)P_\Phi(k_3) \\ + P_\Phi(k_3)P_\Phi(k_1)].\end{aligned}\quad (8)$$

In the scale-invariant case the power spectrum of the primordial potential takes the form  $P_\Phi(k) = \Delta_\Phi k^{-3}$ , where  $\Delta_\Phi$  defines the amplitude of primordial fluctuations at the end of inflation. Accounting for permutations, the local bispectrum then becomes

$$\begin{aligned}B_\Phi(k_1, k_2, k_3) \\ = 2f_{NL}[P_\Phi(k_1)P_\Phi(k_2) + P_\Phi(k_2)P_\Phi(k_3) + P_\Phi(k_3)P_\Phi(k_1)] \\ \simeq 2f_{NL} \frac{\Delta_\Phi^2}{(k_1 k_2 k_3)^2} \left( \frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_1 k_3} + \frac{k_3^2}{k_1 k_2} \right).\end{aligned}\quad (9)$$

Although this is a rather pathological function which diverges along the edges of the tetrahedron (i.e., when any  $k_i \rightarrow 0$ ), we can infer from it some basic properties of the bispectrum for any model which is nearly scale invariant. For example, we can observe that the bispectrum at equal  $k_i$  has the characteristic scaling

$$B_\Phi(k, k, k) = \frac{2f_{NL}\Delta_\Phi^2}{k^6}. \quad (10)$$

If we remove this overall  $k^{-6}$  scaling by multiplying (9) by the factor  $(k_1 k_2 k_3)^2$ , then we note that on transverse slices through the tetrahedron defined by  $\tilde{k} \equiv (k_1 + k_2 + k_3)/2 = \text{const.}$  (see Figure 2) the bispectrum only depends on the ratios of the wavenumbers, say,  $k_2/k_1$  and  $k_3/k_1$ . Indeed, it can prove convenient to characterize the bispectrum in terms of the following transverse parameters [6, 7]:

$$\tilde{k} = \frac{1}{2}(k_1 + k_2 + k_3), \quad \tilde{\alpha} = \frac{(k_2 - k_3)}{\tilde{k}}, \quad \tilde{\beta} = \frac{(\tilde{k} - k_1)}{\tilde{k}}, \quad (11)$$

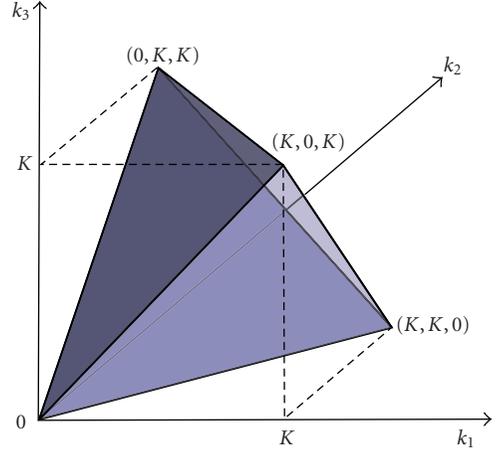


FIGURE 2: Tetrahedral domain for allowed wavenumber configurations  $k_1, k_2, k_3$  contributing to the primordial bispectrum  $B(k_1, k_2, k_3)$ . A regular tetrahedron is shown satisfying  $k_1 + k_2 + k_3 \leq 2k_{\max} \equiv 2K$ .

with the domains  $\tilde{k} \leq k_{\max}$ ,  $0 \leq \tilde{\beta} \leq 1$ , and  $-(1 - \tilde{\beta}) \leq \tilde{\alpha} \leq 1 - \tilde{\beta}$ . The volume element on the regular tetrahedron of allowed wavenumbers then becomes  $dk_1 dk_2 dk_3 = k^2 d\tilde{k} d\tilde{\alpha} d\tilde{\beta}$ .

These considerations lead naturally to the definition of the primordial shape function [8]

$$S(k_1, k_2, k_3) \equiv \frac{1}{N} (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3), \quad (12)$$

where  $N$  is a normalization factor which is often chosen such that  $S$  is unity for the equal  $k_i$  case; that is,  $S(k, k, k) = 1$  (we will discuss alternatives to this rather arbitrary convention later). For example, the canonical ‘‘local’’ model (9) has the shape

$$S^{\text{local}}(k_1, k_2, k_3) = \frac{1}{3} \left( \frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_1 k_3} + \frac{k_3^2}{k_1 k_2} \right). \quad (13)$$

Thus it is usual to describe the primordial bispectrum in terms of an overall amplitude  $f_{NL}$  and a transverse two-dimensional shape  $S(k_1, k_2, k_3) = S(\tilde{\alpha}, \tilde{\beta})$ , which incorporates any distinctive momentum dependence. Of course, if there is a nontrivial scale dependence, then the full three-dimensional dependence of  $S(k_1, k_2, k_3)$  on  $k_i$  must be retained.

There are other physically well-motivated shapes in the literature which have also been extensively studied. The simplest shape is the *constant model*

$$S^{\text{const}}(k_1, k_2, k_3) = 1, \quad (14)$$

which, like the local model, has a large-angle analytic solution for the CMB bispectrum [9]. The local model tends to be the benchmark against which all other models are compared and normalized, but for practical purposes the constant model is much more useful, given its regularity at

both late and early times. The *equilateral shape* is another important case with [8]

$$S^{\text{equil}}(k_1, k_2, k_3) = \frac{(k_1 + k_2 - k_3)(k_2 + k_3 - k_1)(k_3 + k_1 - k_2)}{k_1 k_2 k_3}. \quad (15)$$

While being not derived directly from a physical model, it has been chosen phenomenologically as a separable ansatz for higher-derivative models [10] and DBI inflation [11]. The equilateral shape is contrasted with the local model in Figure 3.

Another important early result was the primordial bispectrum shape for single-field slow-roll inflation derived by Acquaviva et al. [12] and Maldacena [13]:

$$\begin{aligned} S^{\text{Mald}}(k_1, k_2, k_3) & \propto (3\epsilon - 2\eta) \left[ \frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_1 k_3} + \frac{k_3^2}{k_1 k_2} \right] \\ & + \epsilon \left[ (k_1 k_2^2 + 5 \text{ perm.}) + 4 \frac{k_1^2 k_2^2 + k_2^2 k_3^2 + k_3^2 k_1^2}{k_1 k_2 k_3} \right] \\ & \simeq (6\epsilon - 2\eta) S^{\text{local}}(k_1, k_2, k_3) + \frac{5}{3} \epsilon S^{\text{equil}}(k_1, k_2, k_3), \end{aligned} \quad (16)$$

where  $\epsilon$ ,  $\eta$  are the usual slow-roll parameters. In the second line, we have noted that this shape can be accurately represented as the superposition of local and equilateral shapes. The coefficients in (16), which include the scalar spectral index  $n - 1 = -6\epsilon + 2\eta \sim -0.05$ , confirm that  $f_{\text{NL}} \ll 1$ , and so standard single slow-roll inflation cannot produce an observationally significant signal. Nevertheless, it is interesting to determine which shape is dominant in (16) and to what extent other primordial shapes are independent from one another.

Whether two different primordial shapes can be distinguished observationally can be determined from the correlation between the corresponding two CMB bispectra weighted for the anticipated signal-to-noise-ratio, as in the estimator (see next section) and the Fisher matrix analysis (see Section 3.8). However, direct calculations of the CMB bispectrum can be very computationally demanding. A much simpler approach is to determine the independence of the two shape functions  $S$  and  $S'$  from the correlation integral (see [9], and also the study by Babich et al. in [8])

$$F_\epsilon(S, S') = \int_{\mathcal{V}_k} S(k_1, k_2, k_3) S'(k_1, k_2, k_3) \omega_\epsilon(k_1, k_2, k_3) d\mathcal{V}_k, \quad (17)$$

where we choose the weight function to be

$$\omega_\epsilon(k_1, k_2, k_3) = \frac{1}{k_1 + k_2 + k_3}, \quad (18)$$

reflecting the primary scaling of the CMB correlator. The shape correlator is then defined by

$$\bar{\mathcal{C}}(S, S') = \frac{F(S, S')}{\sqrt{F(S, S)F(S', S')}}. \quad (19)$$

Here, the integral is over the tetrahedral region shown in Figure 2 taken out to a maximum wavenumber  $k \lesssim k_{\text{max}}$  corresponding to the experimental range  $l \leq \ell_{\text{max}}$  for which forecasts are sought (with  $\ell_{\text{max}} \approx \tau_0 k_{\text{max}}$ , with  $\tau_0$  being the present-day conformal time). The weight function  $\omega_\epsilon(k_1, k_2, k_3)$  appropriate for mimicking the large-scale structure bispectrum estimator (see Section 4.3.2) would be different with varying scaling laws introduced by the transfer functions for wavenumbers  $k$  above and below  $k_{\text{eq}}$ , the inverse comoving horizon at equal matter-radiation. Nevertheless, the  $1/k$  weight given in (18) provides a compromise between these scalings, and so shape correlation results should offer a useful first approximation.

Below we will survey primordial models in the literature, showing how close the shape correlator comes to a full Fisher matrix analysis. However, here we note that the local shape (13) and the equilateral shape (53) have only a modest 46% correlation. For the natural values of the slow-roll parameters  $\epsilon \approx \eta$  we find the somewhat surprising result that  $S^{\text{Mald}}$  is 99.7% correlated with  $S^{\text{local}}$  (and it cannot be easily tuned otherwise because  $3\epsilon \approx \eta$  is not consistent with deviations from scale invariance favored observationally  $n - 1 < 0$ ). Such strong correspondences are important in defining families of related primordial shapes, thus reducing the number of different cases for which separate observational constraints must be sought.

*2.2. General Primordial Bispectra and Separable Mode Expansions.* The three shape functions (13), (14), and (53) quoted above share the important property of *separability*; that is, they can be written in the form

$$S(k_1, k_2, k_3) = X(k_1)Y(k_2)Z(k_3) + 5 \text{ perms}, \quad (20)$$

or as the sum of just a few such terms. As we will see, if a shape  $S$  is separable, then the computational cost of evaluating the corresponding CMB bispectrum  $B_{\ell_1 \ell_2 \ell_3}$  is dramatically reduced. In fact, without this property, the task of estimating whether a nonseparable bispectrum is consistent with observation appears to be intractable (for large  $\ell_{\text{max}}$ ). Of course, the number of models which can be expressed directly in the form of (20) is very limited, despite the usefulness of approximate ansätze such as the equilateral shape (53). Indeed, approximating nonseparable shapes by educated guesses for the separable functions  $X$ ,  $Y$ ,  $Z$  is neither systematic nor computationally efficient (because arbitrary nonscaling functions create numerical difficulties, as we will explain later).

Instead, we will present a separable mode expansion approach for efficient calculations with any nonseparable bispectrum, as described in detail by Fergusson et al. in [14] (and originally proposed in [6]). Our aim will be to express any shape function as an expansion in mode functions

$$\begin{aligned} S(k_1, k_2, k_3) & = \sum_p \sum_r \sum_s \alpha_{prs} q_{\{p}(k_1) q_r(k_2) q_{\{s}(k_3) \\ & \equiv \sum_n \alpha_n^{\mathcal{Q}} \mathcal{Q}_n(k_1, k_2, k_3), \end{aligned} \quad (21)$$

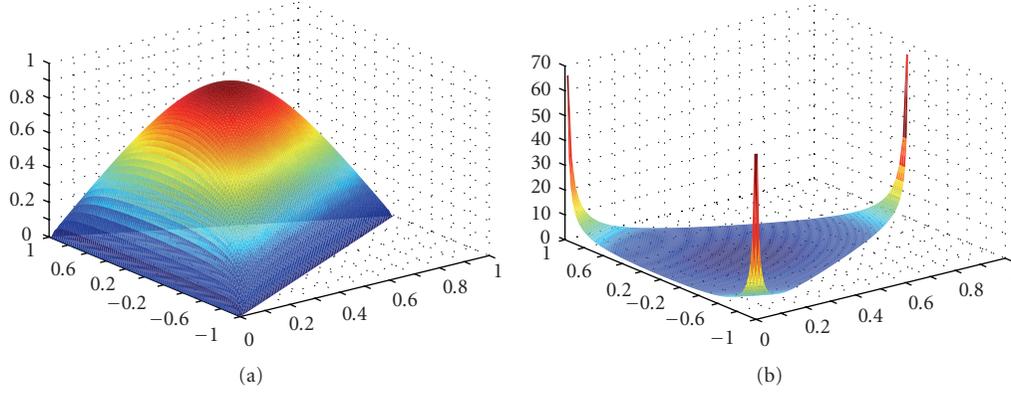


FIGURE 3: Shape functions for the scale-invariant equilateral (a) and local (b) models:  $S(k_1, k_2, k_3) = S(\tilde{\alpha}, \tilde{\beta})$  on transverse slices with  $2\tilde{k} = k_1 + k_2 + k_3 = \text{const}$ . See main text for the definition of the coordinate reparametrization in terms of  $\tilde{\alpha}$  ( $y$ -axis) and  $\tilde{\beta}$  ( $x$ -axis).

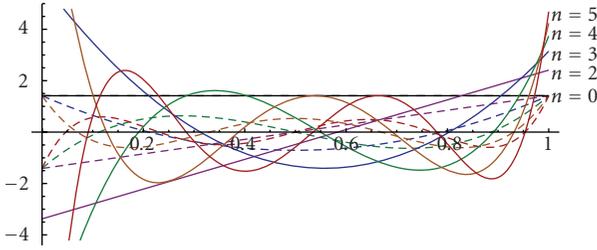


FIGURE 4: The one-dimensional tetrahedral polynomials  $q_n(k)$  on the domain (23), rescaled to the unit interval for  $n = 0 - 5$ . Also plotted are the shifted Legendre polynomials  $P_n(2x - 1)$  (dashed lines) which share qualitative features such as  $n$  nodal points.

where, here, for convenience, we have represented the symmetrized products of the separable basis functions  $q_p(k)$  as

$$\begin{aligned} \mathcal{Q}_n(k_1, k_2, k_3) &= \frac{1}{6} [q_p(x)q_r(y)q_s(z) + 5 \text{ perms}] \\ &\equiv q_{\{pqr\}}, \end{aligned} \quad (22)$$

with a one-to-one mapping ordering the products as  $n \leftrightarrow \{prs\}$ . The important point is that  $q_p(k)$  must be an independent set of well-behaved basis functions which can be used to construct complete and orthogonal three-dimensional eigenfunctions on the tetrahedral region  $\mathcal{V}_{\mathcal{T}}$  defined by (see Figure 2)

$$k_1, k_2, k_3 \leq k_{\max}, \quad k_1 \leq k_2 + k_3 \quad \text{for } k_1 \geq k_2, k_3, +2 \text{ perms.} \quad (23)$$

The introduction of the cutoff at  $k_{\max}$  is motivated by both separability and the correspondence with the observational domain  $l \leq l_{\max}$ . In the shape correlator (19), we have already seen what is essentially an inner product between two shapes on this tetrahedral region, which we can define for two functions  $f, g$  as

$$\langle f, g \rangle = \int_{\mathcal{V}_{\mathcal{T}}} f(k_1, k_2, k_3)g(k_1, k_2, k_3)\omega(k_1, k_2, k_3)d\mathcal{V}_{\mathcal{T}}, \quad (24)$$

with weight function  $w$ .

Satisfactory convergence for known bispectra can be found by using simple polynomials  $q_p(k)$  in the expansion (21), that is, using analogues of the Legendre polynomials on the domain (23). With unit weight, the polynomials satisfying  $\langle q_p(k_1), q_r(k_1) \rangle = \delta_{pr}$  can be found by generating functions with the first three given by [14]

$$q_0(x) = \sqrt{2}, \quad q_1(x) = 5.79 \left( -\frac{7}{12} + x \right), \quad (25)$$

$$q_2(x) = 23.3 \left( \frac{54}{215} - \frac{48}{43}x + x^2 \right), \dots$$

The first few polynomials  $q_p(k)$  are plotted in Figure 4, where they are contrasted with the Legendre polynomials.

The three-dimensional separable basis functions  $\mathcal{Q}_n$  in (22) reflect the six symmetries of the bispectrum through the permuted sum of the product terms. They could have been constructed directly from simpler polynomials, such as  $1, k_1 + k_2 + k_3, k_1^2 + k_2^2 + k_3^2, \dots$ ; however, the  $q_p$  polynomials have two distinct advantages. First, the  $q_p$ 's confer partial orthogonality on the  $\mathcal{Q}_n$  and, secondly, these remain well behaved when convolved with transfer functions.

In order to rapidly decompose an arbitrary shape function  $S$  into the coefficients  $\alpha_n^{\mathcal{Q}} \leftrightarrow \alpha_{prs}^{\mathcal{Q}}$ , it is more convenient to work in a nonseparable orthonormal basis  $\mathcal{R}_n$  ( $\langle \mathcal{R}_n, \mathcal{R}_m \rangle = \delta_{nm}$ ). These can be derived directly from  $\mathcal{Q}_n$  through Gram-Schmidt orthogonalization, so that  $\mathcal{R}_n = \sum_{p=0}^n \lambda_{mp} \mathcal{Q}_p$  with  $\lambda_{mp}$  being a lower triangular matrix (see [14]). Thus we can find the unique shape function decomposition

$$\begin{aligned} S(k_1, k_2, k_3) &= \sum_n^N \alpha_n^{\mathcal{R}} \mathcal{R}_n(k_1, k_2, k_3) \\ &= \sum_n^N \alpha_n^{\mathcal{Q}} \mathcal{Q}_n(k_1, k_2, k_3), \end{aligned} \quad (26)$$

$$\text{with } \alpha_n^{\mathcal{R}} = \langle S, \mathcal{R}_n \rangle, \alpha_n^{\mathcal{Q}} = \sum_p^N (\lambda^\top)_{np} \alpha_p^{\mathcal{R}}.$$

In the orthonormal  $\mathcal{R}_n$  frame, Parseval's theorem ensures that the autocorrelator is simply  $\langle S, S \rangle = \sum_n \alpha_n^{\mathcal{R}^2}$ . Hence,

TABLE 1: Shape correlations (19) and CMB correlations (61) between the equilateral family of primordial models.

	DBI		Ghost		Single	
	$\mathcal{C}(S, S')$	$\mathcal{C}(B, B')$	$\mathcal{C}(S, S')$	$\mathcal{C}(B, B')$	$\mathcal{C}(S, S')$	$\mathcal{C}(B, B')$
Equilateral	0.99	0.99	0.98	0.98	0.95	0.96
DBI			0.94	0.95	0.98	0.99
Ghost					0.86	0.89

with a simple and efficient prescription we can construct separable and complete basis functions on the tetrahedral domain (23) providing rapidly convergent expansions for any well-behaved shape function  $S$ . These eigenmode expansions will prove to be of great utility in subsequent sections. Examples of this bispectral decomposition and its rapid convergence for the equilateral and DBI models are shown in Figure 5.

*2.3. Families of Primordial Models and Their Correlations.* We will now briefly survey the main categories of primordial models in the literature and their relative independence, closely following the discussion by Fergusson and Shellard in [9].

*2.3.1. The Constant Model.* The constant model (14) is the simplest possible primordial shape with triangles of every configuration contributing equally to the bispectrum  $B(k_1, k_2, k_3)$ ; it is the equipartition model. The constant model was motivated initially by its simplicity [9] leading to an analytic solution for the large-angle CMB bispectrum, as well as due to its close correlation with equilateral models. However, the shape does have a more explicit physical motivation in at least one context [15], during multifield inflation for a slowly turning trajectory (denoted as quasi-single-field inflation). For multifield inflation, it is well known that the conversion of isocurvature fluctuations into curvature fluctuations during “corner-turning” can source significant non-Gaussianity (see, e.g., [7, 16]). In the quasi-single-field case with mass  $m \sim H$  isocurvature modes, a detailed investigation of the ongoing conversion into the curvature mode demonstrated that novel shapes could be generated [15], amongst them are the shapes which were very nearly constant. Generically, these model-dependent shapes belonged to a one-parameter family which interpolated nontrivially between equilateral (53) and local (13) shapes (see also [17, 18]). This is an important caveat for the present discussion, because non-Gaussian searches could uncover shapes intermediate between the categories we will discuss below.

*2.3.2. Equilateral Triangles—Centre-Weighted Models.* Bispectra dominated by contributions from nearly equilateral triangle configurations,  $k_1 \approx k_2 \approx k_3$ , can be fairly easily characterized analytically and are the most amenable to CMB searches. However, equilateral non-Gaussianity requires that the amplification of nonlinear effects around the time

modes exit the horizon, which is not possible in a slow-roll single-field inflation. Instead, the kinetic terms in the effective action must be modified as in the Dirac-Born-Infeld (DBI) model [11] or by explicitly adding higher-derivative terms, such as in K-inflation (see, e.g., [19]). The resulting corrections modify the sound speed  $c_s$ , and inflation is able to take place in steep potentials. For DBI inflation, this leads to non-Gaussianity being produced with a shape function of the form [10, 11]

$$S(k_1, k_2, k_3) = \frac{1}{k_1 k_2 k_3 (k_1 + k_2 + k_3)^2} \times \left[ \sum_i k_i^5 + \sum_{i \neq j} (2k_i^4 k_j - 3k_i^3 k_j^2) + \sum_{i \neq j \neq l} (k_i^3 k_j k_l - 4k_i^2 k_j^2 k_l) \right]. \quad (27)$$

Another example of a model with nonstandard kinetic terms is ghost inflation [20] with a derivatively coupled field driving inflation and a trilinear term in the Lagrangian creating a nonzero equilateral-type shape  $S^{\text{ghost}}$  tending towards constant.

General non-Gaussian shapes arising from modifications to single-field inflation have been extensively reviewed in [19]. Using a Lagrangian that was an arbitrary function of the field and its first derivative, they were able to identify six distinct shapes describing the possible non-Gaussian contributions. Half of these had negligible amplitude being of the order of slow-roll parameters (with two already given in (16)). Of the remaining three shapes (see [19], and also [21]), one was believed to be subdominant and the second recovered the DBI shape (27), leaving a third distinct single-field shape which is the inverse of the local shape (13):  $S^{\text{single}} \propto S^{\text{local}^{-1}}$ . Finally, we recall the original equilateral shape (53), noting that it was introduced not because of a fundamental physical motivation, but as a separable approximation to the DBI shape (27) [8].

Despite the apparent visual differences between these shapes (see [9]), particularly near the edges of the tetrahedral domain, the shape correlator (19) reveals at least a 95% or greater correlation of the DBI, ghost, and single shapes to the equilateral shape (53) (consistent with results in [8, 22]). Comparative results between the shape correlator are given in Table 1 (together with the corresponding CMB correlation results brought forward and showing the efficacy of these estimates). These particular centre-weighted shapes must be regarded as a single class which would be extremely differentiate observationally, without a bispectrum detection of very high significance.

Finally, we comment on the “orthogonal” shape  $S^{\text{orthog}}$  proposed by Smith et al. in [18], together with  $S^{\text{equil}}$ , for characterizing single-field inflation models with an approximate shift symmetry (see also [19]). This shape is approximately  $S^{\text{orthog}} \propto S^{\text{equil}} - 2/3$ , which means that it is very similar to an earlier study of flattened shapes [23] which proposed an “enfolded” shape with  $S^{\text{enfold}} \propto S^{\text{equil}} - 1$ . From

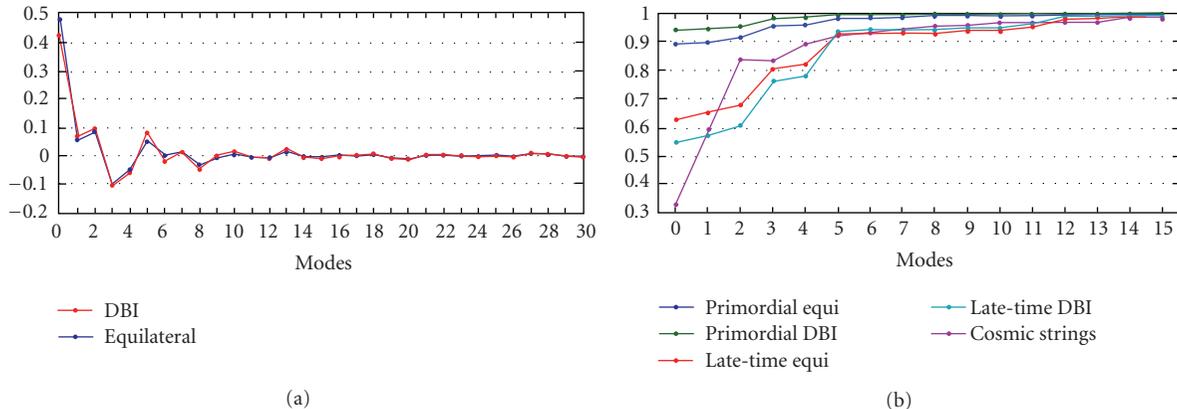


FIGURE 5: Orthonormal eigenmode decomposition coefficients (26) for the equilateral and DBI models (a) and shape correlations (19) of the original bispectrum against the partial sum up to a given mode  $n$  (b). The correlation plot includes both primordial and late-time CMB bispectra for the equilateral and DBI models, as well as the late-time CMB bispectrum from cosmic strings (refer to Section 3). In all cases, we find that we need at most 15 three-dimensional modes to obtain a correlation greater than 98% (primordial convergence without the acoustic peaks requires only 6 modes).

TABLE 2: Shape correlations (19) and CMB correlations (61) for 5 distinct families of primordial non-Gaussian models.

	Local		Warm		Flat		Feature	
	$\mathcal{C}(S, S')$	$\mathcal{C}(B, B')$						
Equilateral	0.46	0.51	0.44	0.42	0.30	0.39	-0.36	-0.43
Local			0.30	0.52	0.62	0.79	-0.41	-0.39
Warm					0.01	0.21	-0.05	-0.27
Flat							-0.44	-0.32

the eigenmode decomposition (26) of the equilateral model shown in Figure 5, it is clear how the degree of correlation can be altered by subtracting out the important constant term. With the specific choice of constant term used in the analysis by Smith et al. [18], one gets a correlation of about 30% between both local and equilateral shapes and the orthogonal ansatz.

**2.3.3. Squeezed Triangles—Corner-Weighted Models.** The local shape covers a wide range of models where the non-Gaussianity is produced by local interactions. These models have their peak signal in “squeezed” states where one  $k_i$  is much smaller than the other two due to non-Gaussianity typically being produced on superhorizon scales. We have already observed that single-field slow-roll inflation (16) is dominated by the local shape [24], though  $f_{\text{NL}}^{\text{loc}}$  is tiny [12, 13, 24]. The production of non-Gaussianity during multiple-field inflation [7, 16, 21, 25–30] shows much greater promise through conversion of isocurvature into adiabatic perturbations (see, e.g., recent works in [15, 17, 31, 32] and references therein). The magnitude of the non-Gaussianity generated is normally around  $f_{\text{NL}}^{\text{loc}} \approx O(1)$ , which is at the limit for Planck detection, but models can be tuned to create larger signals. Significant  $f_{\text{NL}}^{\text{loc}}$  can be produced in curvaton models with  $f_{\text{NL}}^{\text{loc}} \approx O(100)$  [33–35]. Large  $f_{\text{NL}}^{\text{loc}}$  can also be generated at the end of inflation from massless preheating or other reheating mechanisms [36–38].

We note that local non-Gaussianity can also be created in more exotic scenarios. Models based on nonlocal field theory, such as  $p$ -adic inflation, can have inflation in very steep potentials. Like single-field slow-roll inflation, the predicted “nonlocal” shape function is a combination of a dominant local shape (13) and an equilateral shape (53) (see, e.g., [39–42]). The ekpyrotic model can also generate significant  $f_{\text{NL}}^{\text{loc}}$  [43–47]. Here the density perturbations are generated by a scalar field rolling in a negative exponential potential, so nonlinear interactions are important with  $f_{\text{NL}}^{\text{loc}} \approx O(100)$ .

In using the shape correlator for the local model, we must introduce a small-wavenumber cutoff, taken to be  $k_{\text{min}} = 2/\tau_0$ ; otherwise the shape correlator  $\bar{\mathcal{C}}(S^{\text{local}}, S^{\text{local}})$  becomes infinite. This logarithmic divergence does not afflict the CMB bispectrum  $b_{l_1 l_2 l_3}$  because we do not consider contributions below the quadrupole  $l = 2$  (a threshold which is approximated by the primordial cutoff). The local shape is modestly correlated at the 40%–55% level with the equilateral shapes, mainly through the constant term in the expansion (26). As can be seen in Table 2, this somewhat underestimates the CMB correlator. Nevertheless, a NG signal of only modest significance should be able to distinguish between these independent models.

Finally, warm inflation scenarios, that is, models in which dissipative effects play a dynamical role, are also predicted to produce significant non-Gaussianity [48, 49]. Contributions are again dominated by squeezed configurations but with a

different more complex shape possessing a sign flip as the corner is approached (see Figure 6). This makes the warm  $S^{\text{warm}}$  and local  $S^{\text{local}}$  shapes essentially orthogonal with only a 33% correlation. Again, in using the shape correlator, we need to introduce the same phenomenological cutoff  $k_{\text{min}}$  as for the local model, but we also note the more serious concern which is the apparent breakdown of the approximations used to calculate the warm inflation shape near the corners and edges.

**2.3.4. Flattened Triangles—Edge-Weighted Models.** It is possible to consider inflationary vacuum states which are more general than the Bunch-Davies vacuum, such as an excited Gaussian (and Hadamard) state (see [50], and also discussions in [19, 23]). Observations of non-Gaussianity in this case might provide insight into trans-Planckian physics. The proposed shape for the bispectrum is

$$S^{\text{flat}}(k_1, k_2, k_3) \propto 6 \left( \frac{k_1^2 + k_2^2 - k_3^2}{k_2 k_3} + 2 \text{ perms} \right) + \frac{2(k_1^2 + k_2^2 + k_3^2)s}{(k_1 + k_2 - k_3)^2(k_2 + k_3 - k_1)^2(k_3 + k_1 - k_2)^2}. \quad (28)$$

The bispectrum contribution from early times is dominated by flattened triangles, with, for example,  $k_3 \approx k_1 + k_2$ , and for a small sound speed  $c_s \ll 1$  can be large. Unfortunately, as the divergent analytic approximation breaks down at the boundary of the allowed tetrahedron, some form of cutoff must be imposed, as shown for the smoothed shape in Figure 6 where an edge truncation has been imposed together with a Gaussian filter. The lack of compelling physical motivation and ill-defined asymptotics make predictions for this model uncertain.

**2.3.5. Features—Scale-Dependent Models.** There are also models in which the inflation potential has a feature, providing a break from scale invariance. This can take the form of either a step [51] or a small oscillation superimposed onto the potential [52]. Analytic forms for both by these three-point functions have been presented by Chen et al. in [53] with one approximation taking the form

$$S^{\text{feat}}(k_1, k_2, k_3) \propto \sin\left(\frac{k_1 + k_2 + k_3}{k^*} + P\right), \quad (29)$$

where  $k^*$  is the associated scale of the feature in question and  $P$  is a phase factor. Results for the shape correlator for a particular feature model (with  $k^* \approx \ell^*/\tau_0$  and  $\ell^* = 50$ ) are given in Table 2, showing that it is essentially independent of all of the other shapes. Clearly, scale-dependent feature models form a distinct fifth category beyond equilateral, local, warm, and flat shapes.

### 3. Cosmic Microwave Background

Non-Gaussian initial conditions at the end of inflation (or produced by alternative models for the generation of

primordial perturbations) can produce observable signatures in both the CMB and LSS. It is clear that an eventual detection of such signatures would be of great scientific interest. Cosmological measurements of primordial NG would indeed allow to constrain and discriminate between the different candidate scenarios of primordial inflation that have been briefly reviewed in the previous section in terms of their bispectrum prediction. This section is devoted to study the CMB bispectrum produced by a non-Gaussian primordial curvature perturbation field. We will start in Section 3.1 by determining how the primordial curvature bispectrum propagates to the observed bispectra of CMB temperature and polarization anisotropies. At the end of this section, we will obtain a formula expressing the CMB bispectrum as a convolution of the primordial one with suitable *radiation transfer functions*. The latter encode all of the radiative and gravitational effects producing the observed pattern of CMB anisotropies starting from a given primordial potential. This result will not come as a surprise to the reader familiar with CMB theory, since it is completely analogous to the formula relating the power spectrum of primordial curvature perturbations to the CMB angular power spectrum. We will be simply recasting the same formalism in terms of higher-order correlator. Armed with this useful relation, we will revisit the various models described in the previous section and calculate their corresponding CMB bispectra. This will be done in Sections 3.2 and 3.3 where we will employ the useful distinction between separable and nonseparable bispectra already introduced before and discuss its various implications. We will then compare the bispectra predicted in different scenarios in order to find out whether different models produce observationally distinguishable shapes (and thus whether primordial NG is a viable tool to discriminate between different theories of the Early Universe). This will be done in Section 3.4 by means of a suitably defined *shape correlator* that will tell us “how similar” two CMB bispectrum shapes are. After this preliminary work of definition and classification of various bispectra, in Section 3.5 we will finally deal with the problem of extracting the CMB bispectrum from the data in order to produce statistical estimates of the level of primordial non-Gaussianity and compare the NG measurements to theoretical predictions.

**3.1. The CMB Bispectrum.** In this section we will study the connection between the primordial bispectrum at the end of inflation and the observed bispectrum of CMB anisotropies  $B_{\ell_1 \ell_2 \ell_3}$ . Our work will be primarily concerned with the analysis of the three-point function induced by a NG primordial gravitational potential  $\Phi(\mathbf{k})$  in the CMB temperature fluctuation field. Temperature anisotropies are represented using the  $a_{\ell m}$  coefficients of a spherical harmonic decomposition of the cosmic microwave sky:

$$\frac{\Delta T}{T}(\hat{\mathbf{n}}) = \sum_{\ell m} a_{\ell m}^T Y_{\ell m}(\hat{\mathbf{n}}). \quad (30)$$

Analogous expansions are performed for the  $E$ -mode polarization field in order to produce polarization multipoles  $a_{\ell m}^E$ .

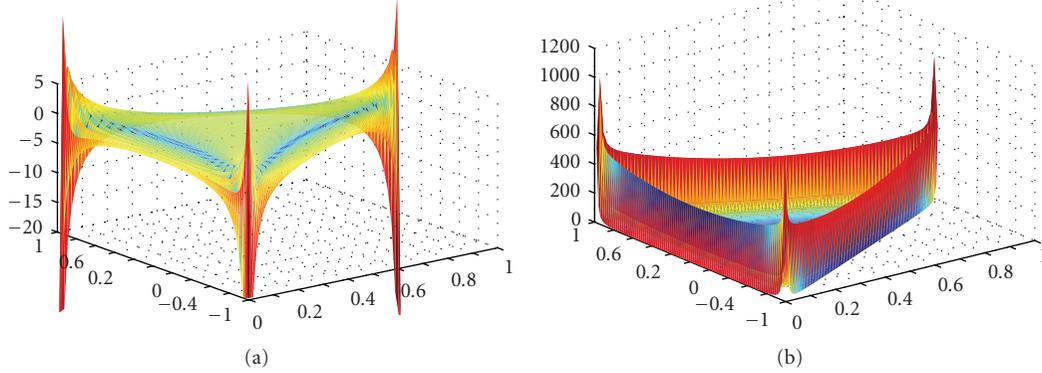


FIGURE 6: Shape functions for the nearly scale-invariant “warm” and “flat” NG models:  $S(k_1, k_2, k_3) = S(\tilde{\alpha}, \tilde{\beta})$  on transverse slices with  $2\tilde{k} = k_1 + k_2 + k_3 = \text{const}$ . These distinct and independent shapes prove to be largely uncorrelated with each other and the local and equilateral models illustrated in figure; from the study by Fergusson and Shellard in [9]. See main text for the definition of the coordinate reparametrization in terms of  $\tilde{\alpha}$  ( $y$ -axis) and  $\tilde{\beta}$  ( $x$ -axis).

For simplicity and clarity, throughout most of this review we will focus on the temperature multipoles  $a_{\ell m}^T$  and omit the superscript  $T$  for convenience of notation. However we stress here that all of the considerations we make in the following can be readily applied to polarization multipoles and related bispectra. More discussion about this subject can be found in Section 3.8.2.

The primordial potential  $\Phi$  is imprinted on the CMB multipoles  $a_{\ell m}$  by a convolution with transfer functions  $\Delta_\ell(k)$  representing the linear perturbation evolution, through the integral

$$a_{\ell m} = 4\pi(-i)^l \int \frac{d^3k}{(2\pi)^3} \Delta_\ell(k) \Phi(\mathbf{k}) Y_{\ell m}(\hat{\mathbf{k}}). \quad (31)$$

The radiation transfer functions  $\Delta_\ell(k)$  encode all of the typical effects observed in the CMB power spectrum at linear order, that is, the Sachs-Wolfe effect, Integrated Sachs-Wolfe effect, acoustic peaks, and silk damping (see, e.g., [54, 55]). An equation identical to (31) produces the E-mode polarization CMB multipoles starting from the primordial temperature fluctuation field, provided that polarization transfer functions replace temperature transfer functions in the convolution above. It is sometimes useful to rewrite (31) in position, rather than Fourier, space. In this case it is straightforward to show that (31) becomes

$$a_{\ell m} = \int dr r^2 \alpha_\ell(r) \Phi_{\ell m}(r), \quad (32)$$

where, starting from the primordial potential  $\Phi(\mathbf{x})$ , we transform from Cartesian into polar coordinates  $\mathbf{x} = (r, \hat{\mathbf{x}})$  and defined

$$\begin{aligned} \Phi_{\ell m}(r) &= \int d\Omega_{\hat{\mathbf{x}}} \Phi(\mathbf{x}) Y_{\ell m}(\hat{\mathbf{x}}), \\ \alpha_\ell(r) &= \frac{2}{\pi} \int dk k^2 \Delta_\ell(k) j_\ell(kr). \end{aligned} \quad (33)$$

In this expression  $j_\ell$  is the spherical Bessel function of order  $\ell$ . The CMB bispectrum is the three-point correlator of  $a_{\ell m}$ , so by substituting, we obtain

$$B_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} = \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \quad (34)$$

$$= (4\pi)^3 (-i)^{l_1+l_2+l_3} \int \frac{d^3k_1}{(2\pi)^3} \frac{d^3k_2}{(2\pi)^3} \frac{d^3k_3}{(2\pi)^3} \quad (35)$$

$$\times \Delta_{\ell_1}(k_1) \Delta_{\ell_2}(k_2) \Delta_{\ell_3}(k_3)$$

$$\langle \Phi(\mathbf{k}_1) \Phi(\mathbf{k}_2) \Phi(\mathbf{k}_3) \rangle Y_{\ell_1 m_1}(\hat{\mathbf{k}}_1) Y_{\ell_2 m_2}(\hat{\mathbf{k}}_2) Y_{\ell_3 m_3}(\hat{\mathbf{k}}_3) \quad (36)$$

$$= \left(\frac{2}{\pi}\right)^3 \int x^2 dx \int dk_1 dk_2 dk_3 (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3)$$

$$\times \Delta_{\ell_1}(k_1) \Delta_{\ell_2}(k_2) \Delta_{\ell_3}(k_3) \quad (37)$$

$$\times j_{\ell_1}(k_1 x) j_{\ell_2}(k_2 x) j_{\ell_3}(k_3 x) \int d\Omega_{\hat{\mathbf{x}}} Y_{\ell_1 m_1}(\hat{\mathbf{x}}) \quad (38)$$

$$\times Y_{\ell_2 m_2}(\hat{\mathbf{x}}) Y_{\ell_3 m_3}(\hat{\mathbf{x}}),$$

where in the last line we have integrated over the angular parts of the three  $\mathbf{k}_i$ , having inserted the exponential integral form for the delta function in the bispectrum definition (5). The last integral over the angular part of  $\mathbf{x}$  is known as the Gaunt integral, which can be expressed in terms of Wigner-3j symbols as (for more details on these functions and their properties, see, e.g., [56] and references therein)

$$\begin{aligned} \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} &\equiv \int d\Omega_{\hat{\mathbf{x}}} Y_{\ell_1 m_1}(\hat{\mathbf{x}}) Y_{\ell_2 m_2}(\hat{\mathbf{x}}) Y_{\ell_3 m_3}(\hat{\mathbf{x}}) \\ &= \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \\ &\times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \end{aligned} \quad (39)$$

Given that most theories we will consider are assumed to be isotropic, the  $m$ -dependence can be factorized out of the physically relevant part of the bispectrum [57]. It is then usual to work with the angle-averaged bispectrum,

$$B_{\ell_1 \ell_2 \ell_3} = \sum_{m_i} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle. \quad (40)$$

Or the even more convenient reduced bispectrum which removes the geometric factors associated with the Gaunt integral

$$B_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} = \mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} b_{\ell_1 \ell_2 \ell_3}. \quad (41)$$

From the previous two formulae we also derive the following useful relations between the full, averaged, and reduced bispectra:

$$\begin{aligned} B_{\ell_1 \ell_2 \ell_3} &= \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} b_{\ell_1 \ell_2 \ell_3} D_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} \\ &= \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{\ell_1 \ell_2 \ell_3}. \end{aligned} \quad (42)$$

The reduced bispectrum from (34) then takes the much simpler form

$$\begin{aligned} b_{\ell_1 \ell_2 \ell_3} &= \left(\frac{2}{\pi}\right)^3 \int x^2 dx \int dk_1 dk_2 dk_3 (k_1 k_2 k_3)^2 B_\Phi(k_1, k_2, k_3) \\ &\quad \times \Delta_{\ell_1}(k_1) \Delta_{\ell_2}(k_2) \Delta_{\ell_3}(k_3) j_{\ell_1}(k_1 x) j_{\ell_2}(k_2 x) j_{\ell_3}(k_3 x). \end{aligned} \quad (43)$$

This is the key equation in this section, since it explicitly relates the primordial bispectrum, predicted by inflationary theories, to the reduced bispectrum observed in the cosmic microwave sky. This formula is entirely analogous to the well-known relation linking the primordial curvature power spectrum  $P_\Phi(k)$  and the CMB angular power spectrum  $C_\ell$ ; that is,

$$C_\ell = \frac{2}{\pi} \int dk k^2 P_\Phi(k) \Delta_\ell^2(k). \quad (44)$$

Finally, it is important to note that the Gaunt integral in (41) encodes several constraints on the angle-averaged bispectrum  $B_{\ell_1 \ell_2 \ell_3}$  which are no longer transparent in the reduced bispectrum  $b_{\ell_1 \ell_2 \ell_3}$ . These are as follows.

- (1) The sum of the three multipoles  $\ell_i$  must be even (to ensure parity invariance).
- (2) The  $\ell_i$ 's satisfy the triangle condition  $|\ell_i - \ell_j| < \ell_k < \ell_i + \ell_j$  (to enforce rotational invariance).

Analogous to the wavenumber constraint (23), the second condition tells us that the only multipole configurations giving nonzero contributions to the bispectrum are those that form a closed triangle in harmonic ( $\ell$ -)space. For

wavenumbers, the triangle condition is enforced through the  $x$ -integral over the three spherical Bessel functions  $j_\ell(k_i x)$  which evaluates to zero if the  $k_i$ 's cannot form a triangle, whereas in multipole space it is enforced by the angular integration  $d\Omega_x$  over the spherical harmonics  $Y_{\ell_i m_i}$  in (39).

**3.2. Separable Primordial Shapes and CMB Bispectrum Solutions.** In terms of the shape function (12), the reduced bispectrum (43) can be rewritten as

$$\begin{aligned} b_{\ell_1 \ell_2 \ell_3} &= \frac{1}{N} \left(\frac{2}{\pi}\right)^3 \int x^2 dx \int dk_1 dk_2 dk_3 S(k_1, k_2, k_3) \Delta_{\ell_1}(k_1) \\ &\quad \times \Delta_{\ell_2}(k_2) \Delta_{\ell_3}(k_3) j_{\ell_1}(k_1 x) j_{\ell_2}(k_2 x) j_{\ell_3}(k_3 x). \end{aligned} \quad (45)$$

The expression above can be simplified, and simple analytic solutions can sometimes be obtained for the very important class of separable shapes obeying the ansatz  $S = XYZ$ , as in (20). Substituting (20) into (45), we find that

$$b_{\ell_1 \ell_2 \ell_3} = \int dr r^2 X_{\ell_1}(r) Y_{\ell_2}(r) Z_{\ell_3}(r) + 5 \text{ perms}, \quad (46)$$

where we have defined the quantities

$$\begin{aligned} X_\ell(r) &\equiv \int dk k^2 X(k) j_\ell(kr) \Delta_\ell, \\ Y_\ell(r) &\equiv \int dk k^2 Y(k) j_\ell(kr) \Delta_\ell, \\ Z_\ell(r) &\equiv \int dk k^2 Z(k) j_\ell(kr) \Delta_\ell. \end{aligned} \quad (47)$$

Instead of the three-dimensional integral of (45), we now have to deal with a much more tractable product of three one-dimensional integrals. Moreover, if we work at large angular scales in the Sachs-Wolfe approximation, the transfer functions become  $\Delta_l(k) = (1/3) j_l[(\tau_0 - \tau_{\text{dec}})k]$ , where  $\tau_0$  and  $\tau_{\text{dec}}$  represents respectively, the present-day conformal time and the conformal time at decoupling. The presence of a product of spherical Bessel functions in the integrals above can lead in some cases to simple analytic solutions.

Let us demonstrate this for the separable primordial shapes considered in Section 2. The simplest possible shape, the constant model (14) with  $S(k_1, k_2, k_3) = 1$ , has a large-angle analytic solution for the reduced bispectrum [9]:

$$\begin{aligned} b_{\ell_1 \ell_2 \ell_3}^{\text{const}} &= \frac{\Delta_\Phi^2}{27N} \frac{1}{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)} \\ &\quad \times \left[ \frac{1}{\ell_1 + \ell_2 + \ell_3 + 3} + \frac{1}{\ell_1 + \ell_2 + \ell_3} \right] \quad (l \ll 200). \end{aligned} \quad (48)$$

The large-angle solution (48) is an important benchmark with which to compare the shape of late-time CMB bispectra from other models  $b_{\ell_1 \ell_2 \ell_3}$  (note the  $l^{-4}$  scaling). The more general constant solution does not have an analytic solution because the transfer functions cannot be expressed in

a simple form, but it can be evaluated numerically from the expression

$$b_{\ell_1 \ell_2 \ell_3}^{\text{const}} = \frac{\Delta_{\Phi}^2}{N} \int x^2 dx \mathcal{I}_{\ell_1}(x) \mathcal{I}_{\ell_2}(x) \mathcal{I}_{\ell_3}(x), \quad (49)$$

where  $\mathcal{I}_{\ell}(x) = \frac{2}{\pi} \int dk \Delta_{\ell}(k) j_{\ell}(kx)$ .

The numerical solution is shown in Figure 7, exhibiting a regular pattern of acoustic peaks introduced by the oscillating transfer functions.

For the local shape (13), the Sachs-Wolfe approximation also yields a large-angle analytic solution

$$b_{\ell_1 \ell_2 \ell_3}^{\text{local}} = \frac{2\Delta_{\Phi}^2}{27\pi^2} \left( \frac{1}{\ell_1(\ell_1+1)\ell_2(\ell_2+1)} + \frac{1}{\ell_2(\ell_2+1)\ell_3(\ell_3+1)} + \frac{1}{\ell_3(\ell_3+1)\ell_1(\ell_1+1)} \right), \quad (50)$$

where the divergences for the squeezed triangles ( $k_1 \ll k_2, k_3, \dots$ ) in the primordial shape (13) are also reflected in  $b_{\ell_1 \ell_2 \ell_3}^{\text{local}}$ . It is straightforward, in principle, to calculate the full bispectrum from the separable expressions arising from (13):

$$b_{\ell_1 \ell_2 \ell_3}^{\text{local}} = \int x^2 dx \left[ \alpha_{\ell_1}(x) \beta_{\ell_2}(x) \beta_{\ell_3}(x) + 2 \text{ perm.} \right], \quad (51)$$

where the separated integrals analogous to (49) become

$$\alpha_{\ell}(x) = \frac{2}{\pi} \int dk k^2 \Delta_{\ell}(k) j_{\ell}(kx), \quad (52)$$

$$\beta_{\ell}(x) = \frac{2}{\pi} \int dk k^2 P_{\Phi}(k) \Delta_{\ell}(k) j_{\ell}(kx).$$

However, we note that these highly oscillatory integrals must be evaluated numerically with considerable care.

For the equilateral shape (53) we first make its separability explicit by expanding the expression in the form:

$$S(k_1, k_2, k_3) = -2 - \left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right) + \left( \frac{k_1}{k_2} + 5 \text{ perm.} \right). \quad (53)$$

While there is no simple large-angle analytic solution known for the equilateral model, it can be evaluated from the simplified expression

$$b_{\ell_1 \ell_2 \ell_3}^{\text{equil}} = \int x^2 dx \left[ 2\delta_{\ell_1} \delta_{\ell_2} \delta_{\ell_3} + (\alpha_{\ell_1} \beta_{\ell_2} \beta_{\ell_3} + 2 \text{ perm.}) + (\beta_{\ell_1} \gamma_{\ell_2} \delta_{\ell_3} + 5 \text{ perm.}) \right], \quad (54)$$

where  $\alpha_i$ ,  $\beta_i$  are given in (52) and  $\gamma_i$ ,  $\delta_i$ , in the scale-invariant case, are defined by (compare with the local case)

$$\gamma_i(x) = \frac{2}{\pi} \int dk k^2 P_{\Phi}(k)^{1/3} \Delta_i(k) j_i(kx), \quad (55)$$

$$\delta_i(x) = \frac{2}{\pi} \int dk k^2 P_{\Phi}(k)^{2/3} \Delta_i(k) j_i(kx).$$

Before concluding this section we would like to note how all of the solutions presented here present a characteristic  $1/\ell^4$  scaling. This is just a direct consequence of the  $1/k^6$  scaling of the primordial bispectrum in the scale-invariant case, and it is a model-independent result. It is analogous to the typical  $1/\ell^2$  scaling displayed by the angular power spectrum in correspondence to a scale-invariant primordial power spectrum  $P(k)$ .

**3.3. Nonseparable Bispectra Revisited.** Recall the mode expansion (21) of a general nonseparable primordial shape. If we substitute this into the expression for the reduced bispectrum (45), then the separability of the expansion leads to the same efficient calculation route discussed in the previous section through [14]:

$$b_{\ell_1 \ell_2 \ell_3} = \left( \frac{2}{\pi} \right)^3 \Delta_{\Phi}^2 f_{\text{NL}} \int x^2 dx dk_1 dk_2 dk_3$$

$$\times 6 \sum_n \alpha_n^{\mathcal{Q}} \mathcal{Q}_n(k_1, k_2, k_3) \Delta_{\ell_1}(k_1) \Delta_{\ell_2}(k_2) \Delta_{\ell_3}(k_3)$$

$$\times j_{\ell_1}(k_1 x) j_{\ell_2}(k_2 x) j_{\ell_3}(k_3 x)$$

$$= \Delta_{\Phi}^2 f_{\text{NL}} \sum_{n \rightarrow prs} \alpha_{prs} \int x^2 dx$$

$$\times \left\{ \left[ \frac{2}{\pi} \int dk_1 q_p(k_1) \Delta_{\ell_1}(k_1) j_{\ell_1}(k_1 x) \right] \right.$$

$$\times \left[ \frac{2}{\pi} \int dk_2 q_r(k_2) \Delta_{\ell_2}(k_2) j_{\ell_2}(k_2 x) \right]$$

$$\times \left[ \frac{2}{\pi} \int dk_3 q_s(k_3) \Delta_{\ell_3}(k_3) j_{\ell_3}(k_3 x) \right] + 5 \text{ perm.} \left. \right\}$$

$$= \Delta_{\Phi}^2 f_{\text{NL}} \sum_{prs} \alpha_{prs} \int x^2 dx q_p^{\ell_1} q_r^{\ell_2} q_s^{\ell_3}, \quad (56)$$

where  $q_p^{\ell}$  simply result from convolving the basis functions  $q_p(k)$  with the transfer functions

$$q_p^{\ell}(x) = \frac{2}{\pi} \int dk q_p(k) \Delta_{\ell}(k) j_{\ell}(kx). \quad (57)$$

The computationally costly 3D integrals have again reduced to a sum over products of 1D integrals; we note that this economy arises because the triangle condition is enforced in (56) through the product of Bessel functions, resulting in a manifestly separable form in which we can interchange orders of integration. With this mode expansion, all non-separable theoretical CMB bispectra  $b_{\ell_1 \ell_2 \ell_3}$  become efficiently calculable provided that there is a convergent expansion for the shape function.

In the same way that we decomposed an arbitrary primordial shape  $S(k_1, k_2, k_3)$  in Section 2.2, it is possible to construct analogous late-time separable basis functions  $\mathcal{Q}_n$  and orthonormal modes  $\mathcal{R}_n$  with which to describe the CMB bispectrum  $\beta_{\ell_1 \ell_2 \ell_3}$  [6, 9]. The tetrahedral domain

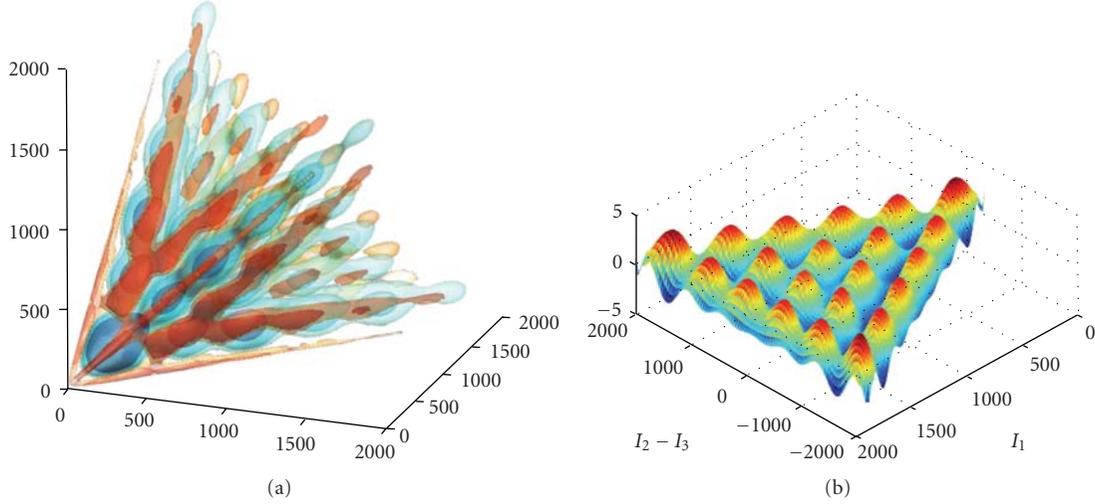


FIGURE 7: The reduced CMB bispectrum for the constant model  $b_{\ell_1 \ell_2 \ell_3}^{\text{const}}$  normalized relative to the large-angle constant solution (48). On (a), the bispectrum is plotted over the allowed tetrahedral region (see Figure 2) using several density contours (light blue positive and magenta negative) out to  $\ell_i \leq 2000$  and, on (b), transverse triangular slices are shown at  $\ell_1 + \ell_2 + \ell_3 = 2000$ . Note the coherent pattern of acoustic peaks induced by the transfer functions, from the study by Fergusson et al. in [14].

$\mathcal{V}_{\mathcal{T}}$  defined by the triangle condition for multipole configurations  $\{\ell_1, \ell_2, \ell_3\}$  is essentially identical to that for wavenumbers (23), except that only even cases contribute  $\sum \ell_1 + \ell_2 + \ell_3 = 2n$ ,  $n \in \mathbb{N}$ . However, the appropriate weight function now incorporates Wigner-3j symbols arising from bispectrum products:

$$w_{\ell_1 \ell_2 \ell_3} = \frac{1}{4\pi} (2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1) \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}^2, \quad (58)$$

$$w_{\ell_1 \ell_2 \ell_3}^s = \frac{w_{\ell_1 \ell_2 \ell_3}}{v_{\ell_1}^2 v_{\ell_2}^2 v_{\ell_3}^2},$$

where in the second expression we have exploited the freedom to divide by a separable function  $v_{\ell} = (2\ell + 1)^{1/6}$  and use a weight which makes the bispectrum functions more scale invariant (eliminating an  $\ell^{-1/2}$  factor—see below). The inner product between two functions  $f_{\ell_1 \ell_2 \ell_3}$  and  $g_{\ell_1 \ell_2 \ell_3}$  is altered from the primordial wavenumber integral (59) into a sum over multipoles on the tetrahedral domain; that is,

$$\langle f, g \rangle \equiv \sum_{\ell_1, \ell_2, \ell_3 \in \mathcal{V}_{\mathcal{T}}} w_{\ell_1 \ell_2 \ell_3}^s f_{\ell_1 \ell_2 \ell_3} g_{\ell_1 \ell_2 \ell_3}. \quad (59)$$

But for the change in the weight (which only affects configurations near the edges of the tetrahedron), the 1D polynomials  $\bar{q}_p(\ell)$  and the 3D separable product basis functions  $\bar{\mathcal{Q}}_n(\ell_1, \ell_2, \ell_3) = \bar{q}_{\{pqr\}}(\ell_1, \ell_2, \ell_3)$  ( $n \leftrightarrow \{prs\}$ ), as well as the resulting orthonormal modes  $\bar{\mathcal{R}}_n$ , are nearly identical to their primordial counterparts  $q_p(k)$ ,  $\mathcal{Q}_n(k_1, k_2, k_3)$ , and  $\mathcal{R}_n(k_1, k_2, k_3)$  defined in Section 2.2.

We can now expand an arbitrary CMB bispectrum  $b_{\ell_1 \ell_2 \ell_3}$  in both the separable and orthonormal mode expansions, which is achieved in the following form:

$$\frac{v_{\ell_1} v_{\ell_2} v_{\ell_3}}{\sqrt{C_{\ell_1} C_{\ell_2} C_{\ell_3}}} b_{\ell_1 \ell_2 \ell_3} = \sum_n \bar{\alpha}_n^{\mathcal{Q}} \bar{\mathcal{Q}}_n(\ell_1, \ell_2, \ell_3) = \sum_n \bar{\alpha}_n^{\mathcal{R}} \bar{\mathcal{R}}_n(\ell_1, \ell_2, \ell_3), \quad (60)$$

where the variance term  $\sqrt{C_{\ell} C_{\ell} C_{\ell}}$  reflects the signal-to-noise weighting expected in the CMB estimator (see Section 3.5). Again, the coefficients in the expansions are determined, first, from the orthonormal inner products  $\bar{\alpha}_n^{\mathcal{R}} = \langle \bar{\mathcal{R}}_n, \cdot \rangle$ , and, secondly, the separable  $\bar{\alpha}_n^{\mathcal{Q}}$  are found with the transformation matrix analogous to (26). Examples of the convergence of these mode expansions for equilateral, DBI, and cosmic string CMB bispectra are given in Figure 5.

**3.4. CMB Bispectrum Calculations and Correlations.** Prior to the systematic mode expansion approach (56) being implemented, robust hierarchical schemes were developed to calculate any nonseparable CMB bispectrum (45) directly [6, 9]. These use the transverse coordinate system  $(\tilde{k}, \tilde{\alpha}, \tilde{\beta})$  given in (11) and employ adaptive methods on a triangular grid to accurately determine the oscillatory 2D  $\alpha\beta$ -integrations, with important efficiencies also coming from the flat sky approximation, binning, and interpolation schemes. Precision to greater than 1% across the full Planck domain  $\ell \leq 2000$  was established by direct comparison with analytic solutions such as (48) and (50). Examples of nonseparable (and separable) CMB bispectra found using these hierarchical coarse-graining methods are shown in Figures 7 and 8. While the CMB bispectra  $b_{\ell_1 \ell_2 \ell_3}$  retain the qualitative features of the primordial shape functions  $S(k_1, k_2, k_3)$ , they are overlaid with the oscillatory transfer functions which give rise to a coherent pattern of acoustic

peaks. These direct bispectrum calculations revealed that typical primordial models could be described by eigenmode or other expansions using only a limited number of terms.

Motivated by the form of the CMB estimator, we can define the following correlator to determine whether or not two competing theoretical bispectra can be distinguished by an ideal experiment:

$$\mathcal{C}(B, B') = \frac{1}{N} \sum_{l_i} \frac{B_{l_1 l_2 l_3} B'_{l_1 l_2 l_3}}{C_{l_1} C_{l_2} C_{l_3}} = \frac{1}{N} \sum_{l_i} w_{l_1 l_2 l_3} \frac{b_{l_1 l_2 l_3} b'_{l_1 l_2 l_3}}{C_{l_1} C_{l_2} C_{l_3}}, \quad (61)$$

where the normalization  $N$  is defined as

$$N = \sqrt{\sum_{l_i} \frac{B_{l_1 l_2 l_3}^2}{C_{l_1} C_{l_2} C_{l_3}}} \sqrt{\sum_{l_i} \frac{B'_{l_1 l_2 l_3}^2}{C_{l_1} C_{l_2} C_{l_3}}}. \quad (62)$$

The emergence of the inner product (59) in the expression (61) means that substitution of the mode expansions (60) for the theoretical bispectra reduces the correlator to

$$\mathcal{C}(B, B') = \sum_n \bar{\alpha}_n^{\mathcal{R}} \bar{\alpha}_n^{\mathcal{R}'}. \quad (63)$$

While the late-time correlator (61) is the best measure of whether two CMB bispectra are truly independent, it can be demonstrated that for the majority of models the shape correlator (19) introduced earlier is sufficient to determine independence.

On the basis of the direct calculation of the bispectrum results and the CMB correlator, we can now quantitatively check the forecasting accuracy of the primordial shape correlator proposed previously (again closely following the discussion in [9]).

**3.4.1. Nearly Scale-Invariant Models.** For nearly scale-invariant models, the centre values for the bispectrum  $b_{lll}$  all have roughly the same profile but with different normalisations. As we see from Figure 8, the oscillatory properties of the transfer functions for the CMB power spectrum create a series of acoustic peaks for any combinations involving the following multipole values:  $l = 200, 500, 800, \dots$ . Of course, to observe the key differences between the scale-invariant models we must study the bispectrum in the plane orthogonal to the  $(l, l, l)$ -direction, that is, the directions reflecting changes in the primordial shape functions. To plot the bispectrum (see Figures 7 and 8), we consistently divide  $b_{lll}$  by the large-angle CMB bispectrum solution for the constant model (14). This is analogous to multiplying the power spectrum  $C_l$ 's by  $l(l+1)$ , because it serves to remove the overall  $\ell^{-4}$  scaling of the bispectrum, flattening while preserving the transverse momentum-dependence primordial shape, and the effects of the oscillating transfer functions.

The starting point is the constant model (14) which, despite its apparent simplicity, has a CMB bispectrum  $b_{l_1 l_2 l_3}^{\text{const}}$  revealing a nontrivial and coherent pattern of acoustic peaks that we have already noted (see Figure 7). Given that the constant model has no momentum dependence, we stress that

the resulting bispectrum is the three-dimensional analogue of the angular power spectrum  $\ell(\ell+1)C_\ell$  for a scale-invariant model. The largest (primary) peak, for example, is located where all three  $l_i = 220$  (corresponding to the large blue region near the origin). We can interpret Figure 7, therefore, as the pure window function or beam effect of convolving any model with the radiation transfer functions  $\Delta_\ell(k)$  while transforming from Fourier to harmonic space.

The CMB bispectrum for the equilateral model is plotted in Figure 8, showing how the centre weighting from the primordial shape is well preserved despite the convolution with the oscillating transfer functions. For the full CMB correlator (61), the DBI, ghost, and single shapes are generally even more closely correlated with the equilateral model, presumably because distinctive features are “washed out” by the transformation from Fourier to harmonic space. Comparative results between the shape correlator and the Fisher matrix analysis are given in Table 1, establishing that these models are highly correlated and difficult to set apart observationally.

The CMB bispectrum for the local model is also shown in Figure 8, demonstrating a marked contrast with the equilateral model which reflects their different primordial shapes shown in Figure 3. The dominance of the signal in the squeezed limit creates strong parallel ridges of acoustic peaks which connect up and emanate along the corner edges of the tetrahedron (see [58] for further details). The 51% CMB correlation between the local and equilateral models is underestimated by the shape correlator at 41%, presumably because of effective smoothing due to the harmonic analysis. Reflecting their distinctive primordial properties, the CMB bispectra for the flat and warm models are poorly correlated with most of the other models, though the flat shape could be susceptible to confusion with the local CMB bispectrum with which it has a larger correlation (see Table 2). It is clear that the local, equilateral, warm, and flat shapes form four distinguishable categories among the scale-invariant models.

**3.4.2. Scale-Dependent Models, Cosmic Strings, and Other Late-Time Phenomena.** Models which have a nontrivial scaling, such as the feature models, can have starkly contrasting bispectra as illustrated in Figure 8. For example, instead of having the same pattern of acoustic peaks which characterise the scale-invariant models, the feature model can become entirely anticorrelated so that the primary peak has the opposite sign. Later, for this particular choice of  $k^*$  in (29), for increasing  $l$  the phase of the oscillations becomes positively correlated by the second and third peaks. This can lead to small correlation with the other primordial shapes, all below 45% as shown in Table 2 for this  $k^*$  and  $\ell_{\text{max}}$ . Clearly, these nonseparable feature models form a distinct fifth category beyond the four scale-invariant shapes noted above and, of course, there are many possible model dependencies which can lead to further subdivision.

By way of further illustration of the breadth of other possible nonseparable CMB bispectra, we present the late-time CMB bispectrum predicted analytically for cosmic

strings [59] as

$$b_{\ell_1 \ell_2 \ell_3}^{\text{string}} = \frac{A}{(\zeta \ell_1 \ell_2 \ell_3)^2} \left[ (\ell_3^2 - \ell_1^2 - \ell_2^2) \left( \frac{L}{2\ell_3} + \frac{\ell_3}{50L} \right) \times \sqrt{\frac{\ell_*}{500}} \operatorname{erf}(0.3\zeta \ell_3) + 2 \operatorname{perm.} \right] \quad (\ell \leq 2000), \quad (64)$$

where  $\ell_{\min} = \min(\ell_1, \ell_2, \ell_3)$ ,  $\ell_* = \min(500, \ell_{\min})$ ,  $\zeta = \min(1/500, 1/\ell_{\min})$ , and

$$L = \zeta \sqrt{\frac{1}{2}(\ell_1^2 \ell_2^2 + \ell_2^2 \ell_3^2 + \ell_3^2 \ell_1^2) - \frac{1}{4}(\ell_1^4 + \ell_2^4 + \ell_3^4)}. \quad (65)$$

Here,  $A \sim (8\pi G\mu)^3$  is a model-dependent amplitude with  $G\mu = \mu/m_{\text{pl}}^2$  measuring the string tension  $\mu$  relative to the Planck scale. The cutoffs around  $\ell \approx 500$  in (50) are associated with the string correlation length at decoupling (perturbations with  $\ell \gtrsim 500$  can only be causally seeded after last scattering). Here, the nonseparable nature and very different scaling of the string CMB bispectrum are clear from a comparison with (50). Moreover, given the late-time origin of this signal from string metric perturbations, the modulating effect of acoustic peaks from the transfer functions is absent, as is clear from Figure 8. This is just one example of late-time phenomena such as gravitational lensing, secondary anisotropies, and contaminants which are accessible to analysis using the more general CMB mode expansions (60).

**3.5. The Estimation of  $f_{\text{NL}}$  from CMB Bispectra.** In light of the previous discussion, it is evident how measurements of the bispectrum from CMB experimental datasets are able to provide information about the primordial three-point function of the cosmological curvature perturbation field at the end of inflation. This in turn allows us to put significant constraints on inflationary models or on alternative models for the generation of cosmological perturbations. We will now start dealing with the problem of bispectrum estimation in the CMB as a test of primordial non-Gaussianity.

Let us assume that we have measured the three-point function of a given CMB dataset. There are now two general ways to exploit this information.

- (1) *Tests of the Gaussian Hypothesis.* By comparing the measured three-point function to its expected distribution obtained from Gaussian simulation we can detect whether some configurations present a significant deviation from Gaussian expectations. The issue with this approach is that it is sensitive not only to primordial non-Gaussianity, but also to any other possible source of NG, including those of noncosmological origin. Original bispectrum tests of this kind on COBE maps [60] revealed significant deviations from Gaussianity in the data. This NG

signature in the three-point function seemed to be localized in harmonic space around multipoles  $\ell = 16$  and was object of much scrutiny (see, e.g., [61–65]). It was then finally ascertained that the detected signal was not cosmological in origin, but due to a systematic artifact [66]. Moreover, the overall statistical significance of the result disappeared in a later analysis involving the measurement of all of the bispectrum modes available in the map [67] (only a subset of all the configurations had been studied before). General tests of Gaussianity are very useful to identify unexpected effects in the data, and to monitor systematics. However, as long as we are interested in a *primordial* NG signal, it is better to follow the approach of making an ansatz for the bispectrum we expect from the theory under study and obtain a *quantitative* constraint on a given model. This approach is outlined in the point that follows.

- (2)  *$f_{\text{NL}}$  Estimation.* In this case we choose the primordial model that we want to test, characterizing it through its bispectrum shape. We then estimate the corresponding amplitude  $f_{\text{NL}}^{\text{model}}$  from the data. If the final estimate is consistent with  $f_{\text{NL}}^{\text{model}} = 0$ , then we conclude that no significant detection of the given shape is produced by the data, but we still determine important constraints on the allowed range of  $f_{\text{NL}}^{\text{model}}$ . Note that ideally we would like to do more than just constrain the overall amplitude, and reconstruct the entire shape from the data by measuring single configurations of the bispectrum. However, the expected primordial signal is too small to allow the signal from a single-bispectrum triangle to emerge over the noise. For this reason we study the cumulative signal from all of the configurations that are sensitive to  $f_{\text{NL}}^{\text{model}}$ .

Since in this review we are concerned with the study of the primordial bispectrum, we will take the latter approach and deal with the problem of  $f_{\text{NL}}$  estimation from measurements of the bispectrum in CMB maps. We will first present a cubic estimator that optimally extracts the  $f_{\text{NL}}$  information from the data contained in the bispectrum (Section 3.5.1). We will then address the issue of understanding whether this optimal cubic statistic extracts all of the possible information available on  $f_{\text{NL}}$  *in the data* or whether there is enough additional information beyond the three-point function to allow more precise  $f_{\text{NL}}$  measurements using non-bispectrum-based estimators of  $f_{\text{NL}}$  (Section 3.5.2). We will then discuss concrete numerical implementations of bispectrum estimators (Section 3.6) and review the experimental constraints on  $f_{\text{NL}}$  obtained from bispectrum analysis of WMAP data (Section 3.7). Using a standard Fisher matrix analysis, forecasts on the  $f_{\text{NL}}$  error bars are achievable for future CMB surveys (Section 3.8). Following, we will study the NG signals in the map that could contaminate the primordial NG measurement and how they are dealt with when analyzing the data. Finally we will describe algorithms for the simulation of primordial

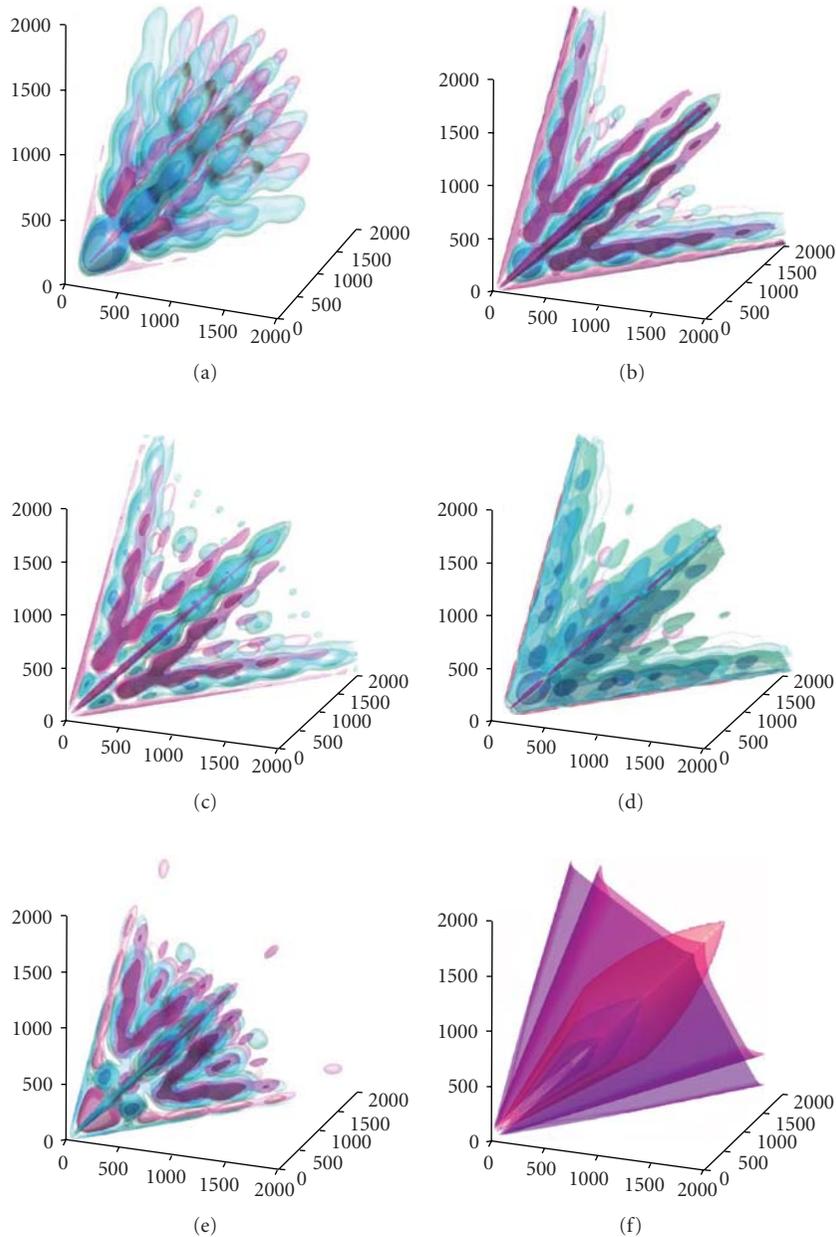


FIGURE 8: The reduced CMB bispectra for several non-Gaussian models, including ((a), (b) and (c)) equilateral, local, and flattened models and ((d), (e), and (f)) warm, feature, cosmic string models (see main text). All five primordial models are normalised relative to the constant solution (48) and are taken from the study by Fergusson and Shellard in [9]. The analytic cosmic string bispectrum (64) is multiplied by  $(\ell_1 \ell_2 \ell_3)^{4/3}$  and is taken from the study by Regan and Shellard in [59].

NG CMB maps that are useful for testing and validation of estimators before applying them to real data.

In the following, we assume that the reader is familiar with essential concepts in statistical estimation theory, such as the definition of a statistical estimator, the role played by maximum-likelihood estimators in statistics, the definitions of unbiasedness and optimality, and the definition and main applications of the Fisher information matrix. The reader unfamiliar with these concepts can consult the appendix of this review and references therein.

**3.5.1. Bispectrum Estimator of  $f_{NL}$ .** In this section we are concerned with the statistical inference of  $f_{NL}$  from measurements of the bispectrum of the CMB anisotropies. We recall that we defined  $f_{NL}$  earlier as the amplitude of the bispectrum of the primordial potential. In principle, we can include both temperature and polarization multipoles  $a_{\ell m}^{T,E}$  in the analysis, in order to maximize the available data. However, for clarity we will consider only temperature multipoles in the following and omit the superscript  $T$  in  $a_{\ell m}$ , for simplicity of notation. The extension to polarization

is conceptually straightforward and will be discussed in a following paragraph. We will start by considering a simple cubic (The estimator is dubbed cubic due to the fact that it contains the third power of the random variable  $a_{\ell m}$ .) statistic written in the form

$$\hat{f}_{\text{NL}} = \frac{1}{\mathcal{N}} \sum_{\{\ell_i, m_i\}} W_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}. \quad (66)$$

In the previous equation  $\hat{f}_{\text{NL}}$  represents the statistical estimate of  $f_{\text{NL}}$  from the data,  $a_{\ell m}$  are the multipoles of the observed CMB temperature fluctuations,  $W_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3}$  are some weight functions, and  $\mathcal{N}$  is a normalization factor that has to be chosen to make the estimator *unbiased*, that is, to ensure that

$$\langle \hat{f}_{\text{NL}} \rangle = f_{\text{NL}}. \quad (67)$$

We now want to find the weights  $W_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3}$  that provide the best estimator (i.e., the minimum error bar estimator) *within the class of cubic statistics* written in the form of (66). It is a well-known result (see Appendix) that the best unbiased estimator of a parameter from a given dataset is the maximum-likelihood estimator. In order to answer our question we then have to write the bispectrum likelihood as a function of the parameter  $f_{\text{NL}}$  and maximize with respect to  $f_{\text{NL}}$ .

In the assumption that the bispectrum configurations are characterized by a Gaussian distribution (This is not strictly true, but it is a good approximation. The same approach applies to most cosmological observables.), maximizing the likelihood is equivalent to minimizing the following  $\chi^2$ :

$$\chi^2 = \sum_{\ell_1 \ell_2 \ell_3} \frac{\left( f_{\text{NL}} B_{\ell_1 \ell_2 \ell_3}^{\text{fNL}=1} - B_{\ell_1 \ell_2 \ell_3}^{\text{obs}} \right)^2}{\sigma^2}, \quad (68)$$

where  $B_{\ell_1 \ell_2 \ell_3}^{\text{obs}}$  is the observed angular averaged bispectrum, that is, by definition

$$B_{\ell_1 \ell_2 \ell_3}^{\text{obs}} = \sum_{m_1 m_2 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} a_{\ell_1 m_1}^{\text{obs}} a_{\ell_2 m_2}^{\text{obs}} a_{\ell_3 m_3}^{\text{obs}}, \quad (69)$$

and  $\sigma^2$  is the bispectrum variance, that is, the  $a_{\ell m}$  six-point function

$$\sigma^2 = \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} a_{\ell_4 m_4} a_{\ell_5 m_5} a_{\ell_6 m_6} \rangle. \quad (70)$$

We will now make the assumption that we are working in the *weak non-Gaussian limit*; that is,  $f_{\text{NL}}$  is small and the distribution of  $a_{\ell m}$  can be approximated as Gaussian in the calculation of the variance. The implications of this approximation will be discussed in greater detail in the following sections; for the moment it will suffice to point out that the weak non-Gaussian approximation is generally a good one since most inflationary models predict  $f_{\text{NL}}$  to be small, and because the level of primordial non-Gaussianity is already constrained to be small by WMAP measurements [1, 68]. After restricting indices so that  $\ell_1 \leq \ell_2 \leq \ell_3$  and

$\ell_4 \leq \ell_5 \leq \ell_6$ , the six-point function above can be calculated using Wick's theorem, yielding [57]

$$\begin{aligned} & \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} a_{\ell_4 m_4} a_{\ell_5 m_5} a_{\ell_6 m_6} \rangle \\ & = \Delta C_{\ell_1} C_{\ell_2} C_{\ell_3} \delta_{\ell_1}^{\ell_4} \delta_{\ell_2}^{\ell_5} \delta_{\ell_3}^{\ell_6} \delta_{m_1}^{m_4} \delta_{m_2}^{m_5} \delta_{m_3}^{m_6}. \end{aligned} \quad (71)$$

In the last formula  $\Delta$  is a permutation factor that takes the value of 1 when all  $\ell$ 's are different, 2 when two  $\ell$ 's are equal, and 6 when all  $\ell$ 's are equal. We can now substitute (71) into (68) and differentiate with respect to  $f_{\text{NL}}$  to get an explicit expression for the optimal cubic statistic we were looking for:

$$\hat{f}_{\text{NL}} = \frac{1}{\mathcal{N}} \sum_{\{\ell_i, m_i\}} \frac{\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{\text{fNL}=1}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}, \quad (72)$$

$$\mathcal{N} = \sum_{\{\ell_i, m_i\}} \frac{\left( \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{\text{fNL}=1} \right)^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}, \quad (73)$$

where  $b_{\ell_1 \ell_2 \ell_3}$  is the reduced bispectrum and  $\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3}$  is the Gaunt integral defined by equation (39);  $\mathcal{N}$  is the normalization factor mentioned at the beginning of the paragraph that guarantees the unbiasedness of the estimator.

Note that the noise and window function of the experiment are included in the  $C_\ell$  and  $b_{\ell_1 \ell_2 \ell_3}$  that appear in the formula above, with the following replacements:

$$C_\ell \rightarrow C_\ell \mathcal{W}_\ell^2 + N_\ell, \quad b_{\ell_1 \ell_2 \ell_3} \rightarrow b_{\ell_1 \ell_2 \ell_3} \mathcal{W}_{\ell_1} \mathcal{W}_{\ell_2} \mathcal{W}_{\ell_3}, \quad (74)$$

where  $\mathcal{W}$  is the window function (not to be confused with the weights  $W$ ) and  $N_\ell$  is the noise power spectrum (constant for uncorrelated white noise). The noise is assumed to be Gaussian, thus characterized by a vanishing three-point function. Comparing our result (72) to the initial ansatz (66), we then see that the optimal weights are

$$W_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} = \frac{\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \quad (75)$$

In other words we are weighting the observed bispectrum by its expected signal-to-noise ratio.

We have now constructed a statistic that optimally extracts the information about  $f_{\text{NL}}$  from the bispectrum of the map. The question now is the following: is there additional information about  $f_{\text{NL}}$  in the map that is not contained in the bispectrum? This issue will be investigated in the following sections. For the impatient reader we anticipate that the answer is no: the bispectrum statistic built here is actually the minimum error bar estimator of  $f_{\text{NL}}$  from CMB data.

**3.5.2. Optimality of the Cubic Estimator.** In this section we address the issue of whether the cubic statistic (72) optimally extracts *all* of the  $f_{\text{NL}}$  information contained in  $a_{\ell m}$  or whether other statistical estimators (e.g., four-point function, or pixel space statistics such as the Minkowski functionals, or again wavelet estimators, just to mention a few among many possible examples) are able to produce

smaller error bars and are thus more efficient than the bispectrum.

In a non-Gaussian primordial CMB map, the  $a_{\ell m}$  likelihood depends on the NG parameter  $f_{\text{NL}}$ . We will indicate it with  $p(\mathbf{a} | f_{\text{NL}})$ , where  $\mathbf{a}$  indicates a vector including all of the  $a_{\ell m}$ 's (we will assume that all other cosmological parameters are fixed and concentrate on  $f_{\text{NL}}$ ). It is a well-known result in parameter estimation theory that there is a *lower limit* on the error bars that can be assigned to a given parameter (in our case  $f_{\text{NL}}$ ). Such lower limit, also known as the *Rao-Cramer bound*, is defined in terms of the *Fisher matrix*  $F$  as (We again refer the reader unfamiliar with these concepts to the brief summary provided in appendix.)

$$\Delta f_{\text{NL}} \geq \frac{1}{\sqrt{F_{f_{\text{NL}} f_{\text{NL}}}}}. \quad (76)$$

We remind the reader that the Fisher matrix is defined as

$$F_{f_{\text{NL}} f_{\text{NL}}}(\mathbf{a}) = \left\langle \frac{\partial^2 \ln p(\mathbf{a} | f_{\text{NL}})}{\partial^2 f_{\text{NL}}} \right\rangle. \quad (77)$$

If we can show that the bispectrum estimator of the previous section saturates the Rao-Cramer bound for the  $a_{\ell m}$  Fisher matrix above, then we conclude that it provides the best (i.e., minimum variance) estimate of  $f_{\text{NL}}$  from the data, rather than just the best  $f_{\text{NL}}$  estimate from the bispectrum of the data. In other words, no more information about  $f_{\text{NL}}$  could be extracted from  $a_{\ell m}$  than the information contained in the bispectrum. The aim of this section is to show that this is actually the case.

The issue of the optimality of bispectrum estimators of  $f_{\text{NL}}$  was addressed in great detail by Babich in [69]. In this section we will basically review the main results of that study, referring the reader to the original paper for their complete derivation.

As we mention in appendix, there is a sufficient and necessary condition for an estimator  $\mathcal{E}$  to saturate the Rao-Cramer bound, expressed by formula (A.5). This condition, applied to our case, reads

$$\frac{\partial \ln p(\mathbf{a} | f_{\text{NL}})}{\partial f_{\text{NL}}} = F_{f_{\text{NL}} f_{\text{NL}}}(\mathcal{E}(\mathbf{a}) - f_{\text{NL}}). \quad (78)$$

Our aim is to show that the bispectrum statistic (72) satisfies this condition. We then need to start from a computation of the full likelihood  $p(\mathbf{a} | f_{\text{NL}})$  for a general primordial non-Gaussian model. Following Babich [69], we will start by limiting ourselves to the particular case of the local model.

We recall from a previous section that local NG is the only case for which an explicit expression for the primordial potential is provided. In real space,

$$\Phi(\mathbf{x}) = \Phi_L(\mathbf{x}) + f_{\text{NL}}^{\text{loc}} [\Phi_L^2(\mathbf{x}) - \langle \Phi_L^2(\mathbf{x}) \rangle]. \quad (79)$$

Starting from this formula, it is possible to obtain a likelihood function for  $\Phi$ , dependent on the parameter  $f_{\text{NL}}^{\text{loc}}$ . This is done by means of an *expansion* in terms of the order parameter  $f_{\text{NL}}^{\text{loc}} \langle \Phi_L^2(\mathbf{x}) \rangle$ . The full expression for the Probability Density Function (PDF)  $P(\Phi | f_{\text{NL}}^{\text{loc}})$  (see [69])

can be expanded around its Gaussian expectation for  $f_{\text{NL}}^{\text{loc}} = 0$  and schematically written as

$$\begin{aligned} \ln P(\Phi | f_{\text{NL}}^{\text{loc}}) &= \ln P_G(\Phi | C) + f_{\text{NL}}^{\text{loc}} \ln P_{\text{NG}}(\Phi | C) \\ &+ \mathcal{O}(f_{\text{NL}}^2 \langle \Phi_L^2(\mathbf{x}) \rangle^2), \end{aligned} \quad (80)$$

where  $C$  is the covariance matrix of the Gaussian part of the potential  $\Phi$  that is,

$$C \equiv \langle \Phi_L(\mathbf{x}_1) \Phi_L(\mathbf{x}_2) \rangle. \quad (81)$$

Formula (80) is then telling us that the logarithm of the full likelihood can be decomposed into the sum of a Gaussian likelihood  $P_G$ , plus a NG term that depends linearly on  $f_{\text{NL}}^{\text{loc}}$ , and that this decomposition is accurate up to terms of order  $\mathcal{O}(f_{\text{NL}}^2 \langle \Phi_L^2(\mathbf{x}) \rangle^2)$ ; that is, we are assuming that NG is weak, as we did in the previous section.

After computing  $P(\Phi | f_{\text{NL}}^{\text{loc}})$ , one has to account for 2D projection and radiative transfer in order to obtain the required likelihood  $P(\mathbf{a} | f_{\text{NL}}^{\text{loc}})$ . As shown by Babich in [69], this can be achieved by expanding the PDF (80) in spherical harmonics and performing the functional integration:

$$\begin{aligned} P(\mathbf{a} | f_{\text{NL}}^{\text{loc}}) &= \int d^N \Phi \delta_D^{(M)} \left[ a_{\ell m} - \int dr r^2 \alpha_\ell(r) \Phi_{\ell m}(r) \right] P(\Phi | f_{\text{NL}}^{\text{loc}}), \end{aligned} \quad (82)$$

where  $\delta_D^{(M)}$  is the Dirac delta function of dimension  $M$  and  $M < N$  due to the 2D projection (As noted by Babich in [69], the additional degrees of freedom do not affect the CMB anisotropies and can therefore be integrated out.) The previous formula can be derived by recalling (32) together with the well-known formula in probability theory as follows:

$$P(\mathbf{y}) = \int d\mathbf{x} P(\mathbf{x}) \delta_D(\mathbf{y} - \mathbf{F}(\mathbf{x})), \quad (83)$$

where  $\delta_D$  is again the Dirac delta function,  $\mathbf{x}$  and  $\mathbf{y}$  are random variables linked by the functional relation  $\mathbf{y} = \mathbf{F}(\mathbf{x})$ , and  $P(\mathbf{x})$ ,  $P(\mathbf{y})$  are the PDFs of  $\mathbf{x}$  and  $\mathbf{y}$ , respectively. Solving the functional integral (82) yields [69]

$$\begin{aligned} \ln P(\mathbf{a} | f_{\text{NL}}^{\text{loc}}) &= -\frac{1}{2} \sum_{\ell m} \frac{a_{\ell m}^* a_{\ell m}}{C_\ell} + f_{\text{NL}}^{\text{loc}} \sum_{\{\ell_1, m_1\}} \frac{\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}^{\text{loc}}=1}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \\ &\times a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} + I_2(\mathbf{a}, f_{\text{NL}}) + \mathcal{O}(f_{\text{NL}}^3 \langle \Phi_L^3(\mathbf{x}) \rangle). \end{aligned} \quad (84)$$

In the previous formula we can recognize the standard  $a_{\ell m}$  PDF, valid in the standard Gaussian case, in the first term on the r.h.s. Added to this, we find a first-order  $f_{\text{NL}}$ -correction proportional to the CMB angular bispectrum. Higher-order correlators are not present at order  $\mathcal{O}(f_{\text{NL}} \langle \Phi_L(\mathbf{x}) \rangle)$ . For reasons that will become clear shortly, although we have not

computed it, we have explicitly denoted the  $\mathcal{O}(f_{\text{NL}}^2 \langle \Phi_L^2(\mathbf{x}) \rangle)$  term in the expansion with  $I_2(\mathbf{a}, f_{\text{NL}})$ . Note that, besides assuming weak NG in this formula, we are also assuming *rotational invariance* (this is evident from the fact that the  $a_{\ell m}$  covariance matrix appearing in the Gaussian piece of (84) is diagonal and equal to  $C_\ell$ ). Rotational invariance is a general property of the CMB sky, but it is broken when we deal with real CMB measurement characterized by inhomogeneous noise patterns and sky cuts. We will investigate these effects in the following section. For the moment we consider the purely ideal case described by (84). Armed with the PDF expression for local NG and recalling the necessary and sufficient condition (78), we can finally determine whether the estimator (72) is optimal or not. First of all we see that

$$\begin{aligned} \frac{\partial \ln p(\mathbf{a} | f_{\text{NL}}^{\text{loc}})}{\partial f_{\text{NL}}^{\text{loc}}} &= \sum_{\{\ell_i, m_i\}} \frac{g_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}^{\text{loc}}=1}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \\ &+ \frac{\partial I_2(\mathbf{a}, f_{\text{NL}})}{\partial f_{\text{NL}}} + \mathcal{O}(f_{\text{NL}}^2 \langle \Phi_L^2(\mathbf{x}) \rangle). \end{aligned} \quad (85)$$

We then see from combining (85) and (72) that

$$\frac{\partial \ln p(\mathbf{a} | f_{\text{NL}}^{\text{loc}})}{\partial f_{\text{NL}}^{\text{loc}}} \propto \left[ \hat{f}_{\text{NL}}(a) + \frac{\partial I_2(\mathbf{a}, f_{\text{NL}})}{\partial f_{\text{NL}}} + \mathcal{O}(f_{\text{NL}}^2 \langle \Phi_L^2(\mathbf{x}) \rangle) \right]. \quad (86)$$

We now see that, in order for the necessary and sufficient condition for optimality (78) to be verified, we need the  $(\partial I_2 / \partial f_{\text{NL}})$  term to be exactly equal to  $-f_{\text{NL}}$ . The second-order quantity  $I_2$  should then be calculated explicitly in the expansion (84) in order to complete the calculation and verify whether, or under which conditions, this is true. However this turns out not to be necessary if we consider the following ‘‘regularity condition for a PDF’’. (Condition (87) can be easily derived remembering that, for a given random variable  $\mathbf{x}$  with probability density  $p(\mathbf{x})$ , we have by definition  $\langle F(\mathbf{x}) \rangle \equiv \int d\mathbf{x} F(\mathbf{x}) p(\mathbf{x})$ , and substituting  $F(\mathbf{x}) \rightarrow \partial \ln p(\mathbf{x} | \lambda) / \partial \lambda$  in the previous expression, one then finds that the regularity condition (87) holds, provided the order of integration and differentiation can be exchanged (hence the ‘‘regularity condition’’ qualification).) For a general PDF of a random variable  $\mathbf{x}$  depending on a parameter  $\lambda$ , we have

$$\left\langle \frac{\partial \ln p(\mathbf{x} | \lambda)}{\partial \lambda} \right\rangle = 0. \quad (87)$$

Since this regularity condition must be valid for each value of the parameter  $\lambda$  ( $\lambda \rightarrow f_{\text{NL}}^{\text{loc}}$  in our case), it is clear that it must hold *term-by-term*, that is, at each order, in the expansion (84). By taking the average value of equation (86), keeping in mind that the estimator is unbiased and imposing (87), we then find that the average value of  $\partial I_2(\mathbf{a}, f_{\text{NL}}) / \partial f_{\text{NL}}$  must be exactly equal to  $-f_{\text{NL}}$ . If we could then replace  $\partial I_2(\mathbf{a}, f_{\text{NL}}) / \partial f_{\text{NL}}$  in (86) with its average value, then we would exactly obtain the condition for optimality and conclude that the cubic estimator (72) saturates the Rao-Cramer bound. For present CMB experiments, the terms in the expansion (84) are evaluated summing over a large number of  $\ell$ -modes

( $\ell_{\text{max}} \sim 500$  for WMAP,  $\ell_{\text{max}} \simeq 2000$  for Planck in the signal-dominated regime), or, equivalently in pixel space, averaging over a large number of pixels ( $\sim 10^6$  and  $10^7$  for WMAP and Planck, resp.). For this reason we expect that the error made by replacing the  $f_{\text{NL}}$ -order term in (86) with its average value will be very small. In [69], an estimate of this error has been done in the approximation of neglecting radiative transfer and projection effects (i.e., working in 3D with the primordial potential, rather than with a CMB map). The conclusion was that for a number of observations  $N > 30$  the approximation above works very well. Moreover the variance of the  $f_{\text{NL}}$ -order term scales like  $1/N$ . In the full radiative transfer case we expect the scaling to be unchanged, although the coefficients in front of it that led to the  $N > 30$  estimate might change. However, as noted above, the number of pixels in present-day experiments is many orders of magnitude larger than 30. That leads us to conclude that the approximation of replacing the average of the first-order-term in equation (86) is a very good one. We then reach the following important conclusion.

*For a rotational invariant CMB sky, in the limit of weak NG, the cubic estimator defined by formula (72) is the best unbiased CMB estimator of  $f_{\text{NL}}^{\text{loc}}$ .*

Let us now move to problem of generalizing the last conclusion to shapes different from local. In this case a full expression of the primordial potential  $\Phi(\mathbf{x})$  is not available. The steps that lead to the conclusion that the local  $f_{\text{NL}}$  estimator is optimal can thus not be reproduced. However it was pointed out by Babich in [69] that, in the limit of weak NG, the full CMB NG likelihood can still be expressed in terms of its power spectrum and bispectrum by means of an Edgeworth expansion, regardless of its full expression. The Edgeworth expansion is basically a way to express a NG PDF as a series expansion around its Gaussian part [70–72]. For CMB anisotropies one finds, at the end of the calculation, that

$$\begin{aligned} P(\mathbf{a} | f_{\text{NL}}) &= \prod_{\ell m} \frac{e^{-a_{\ell m} a_{\ell m}^* / 2C_\ell}}{\sqrt{2\pi C_\ell}} \left[ 1 + \sum_{\{\ell_i, m_i\}} b_{\ell_1 \ell_2 \ell_3} g_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} \frac{a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \right]. \end{aligned} \quad (88)$$

It is easy to see that  $\ln P(\mathbf{a} | f_{\text{NL}})$  takes the same form as in (84). For this reason all the previous derivation applies also to the present case and the following conclusion holds.

*In the weak non-Gaussian limit and assuming rotational invariance of the CMB sky, the cubic estimator (72) is the best unbiased CMB estimator of  $f_{\text{NL}}$  for any non-Gaussian shape.*

Before concluding this section, we would like to stress that, despite the technical complications arising in the detailed probe of the bispectrum estimator’s optimality, the physical reason behind this result is quite clear. We can always expand the  $a_{\ell m}$  PDF in series of its momenta. The order parameter of this expansion is  $(f_{\text{NL}} \langle \Phi \rangle)$ . This parameter is the natural measure of the amplitude of primordial NG, and it is actually predicted by inflation to be very small. For this reason higher-order momenta in the primordial non-Gaussian  $a_{\ell m}$  PDF are suppressed with respect to the

bispectrum. Basically, the information on  $f_{\text{NL}}$  in a CMB map is entirely contained in the three-point function.

**3.5.3. Breaking Rotational Invariance.** So far, the assumption of rotational invariance of the CMB sky has been made when probing the optimality of the cubic estimator (72). In an ideal situation, the CMB sky is clearly rotationally invariant. However, two elements break rotational invariance in a CMB map derived from a real experiment: an anisotropic distribution of noise in pixel space and a galactic mask. Anisotropic noise comes from the fact that the CMB sky is generally scanned in a nonuniform way: regions that are less contaminated by astrophysical foreground emission are generally observed more times and are thus characterized by a lower noise level (see Figure 9, e.g.). A sky cut has also to be introduced in order to remove the regions on the galactic plane that are most contaminated by foregrounds. When rotational invariance is broken, the considerations of the previous two sections do not strictly apply anymore and the estimator (72) becomes suboptimal. However, the same Edgeworth expansion approach that was adopted in the previous section can still be applied, but this time keeping rotation-invariance breaking terms in the calculation, in order to find the new more general form of the optimal estimator. The general estimator turns out to be the sum of two terms: the first term is cubic in  $a_{\ell m}$  and is analogous to the one appearing in the rotationally invariant case, while the second term is linear in  $a_{\ell m}$  and accounts for breaking of rotational invariance. The explicit expression of this general optimal  $f_{\text{NL}}$  estimator is [73]

$$\begin{aligned} \mathcal{E}(a) &= \frac{1}{\mathcal{N}} \sum_{\{\ell_i, m_i\}} \left( \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3} \left( C_{\ell_1 m_1, \ell_4 m_4}^{-1} a_{\ell_1 m_1} \right) \right. \\ &\quad \times \left( C_{\ell_2 m_2, \ell_5 m_5}^{-1} a_{\ell_2 m_2} \right) \left( C_{\ell_3 m_3, \ell_6 m_6}^{-1} a_{\ell_3 m_3} \right) \quad (89) \\ &\quad - 3 \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \\ &\quad \times \left. C_{\ell_1 m_1, \ell_2 m_2}^{-1} C_{\ell_3 m_3, \ell_4 m_4}^{-1} a_{\ell_4 m_4} \right), \\ \mathcal{N} &= \sum_{\{\ell_i, m_i\}} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle C_{\ell_1 m_1, \ell_4 m_4}^{-1} C_{\ell_2 m_2, \ell_5 m_5}^{-1} C_{\ell_3 m_3, \ell_6 m_6}^{-1} \\ &\quad \times \langle a_{\ell_4 m_4} a_{\ell_5 m_5} a_{\ell_6 m_6} \rangle. \quad (90) \end{aligned}$$

In the rotationally invariant case the  $a_{\ell m}$  covariance matrix  $C_{\ell_1 m_1, \ell_2 m_2}$  is diagonal and equal to  $C_\ell$ , while the linear term is proportional to a monopole. We then recover the form of the cubic estimator (72) as expected. Note that in the signal-dominated regime of the experiment under study (e.g.,  $\ell \lesssim 300$  for WMAP and  $\ell \lesssim 1000$  for Planck), and if the mask is not too large, then the simple cubic estimator (72) is still basically optimal, since we are in a nearly rotationally invariant case. For small masks it has been shown by Komatsu and Spergel in [74] that the bispectrum and power spectrum of the map are, to a good approximation,

just rescaled by a factor  $f_{\text{sky}}$ , representing the fraction of the sky left free by the mask; that is,

$$b_{\ell_1 \ell_2 \ell_3}^{\text{mask}} = f_{\text{sky}} b_{\ell_1 \ell_2 \ell_3}^{\text{fullsky}}, \quad C_\ell^{\text{mask}} = f_{\text{sky}} C_\ell^{\text{fullsky}}. \quad (91)$$

In this case one can then assume the covariance matrix to be diagonal and account for the effects of the mask by correctly rescaling the normalization term in order to keep the estimator unbiased. This nearly rotationally invariant estimator then takes the form

$$\begin{aligned} \hat{f}_{\text{NL}} &= \frac{1}{\mathcal{N}} \sum_{\{\ell_i, m_i\}} \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}, \\ \mathcal{N} &= f_{\text{sky}} \sum_{\{\ell_i, m_i\}} \frac{\left( \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1} \right)^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \quad (92) \end{aligned}$$

**3.5.4. Large  $f_{\text{NL}}$  Regime.** The approximation of weak non-Gaussianity is the basis for all of the results derived so far. One can then ask at which point (i.e., for which values of  $f_{\text{NL}}$ ) this approximation breaks down. As we observed earlier, the Edgeworth expansion (88) shows that the likelihood of a generic primordial NG distribution can be expanded in series of its momenta, with order parameter  $\mathcal{O}(f_{\text{NL}} \langle \Phi_L^2(\mathbf{x}) \rangle)$ . We know that  $\Phi_L \sim 10^{-5}$ , while WMAP observations already constrain  $f_{\text{NL}} \lesssim 100$ . That means that the order parameter of the PDF expansion is  $\sim 10^{-3}$  and thus the weak NG approximation seems to be a very good one in the entire range of allowed and predicted  $f_{\text{NL}}$ . However a subtle effect has been pointed out by Creminelli et al. [75], which changes the previous conclusions in certain cases. Let us quickly summarize their main results. We already saw that, for the angular averaged bispectrum of a Gaussian temperature field,

$$\langle B_{\ell_1 \ell_2 \ell_3}^2 \rangle \propto C_{\ell_1} C_{\ell_2} C_{\ell_3}. \quad (93)$$

We then included this expression for the variance in the weights of the optimal estimator (72) and in the normalization factor  $\mathcal{N}$ . It is easy to see that in this approximation the variance of the estimator can be predicted as

$$\langle (\Delta \mathcal{E})^2 \rangle = \sum_{\{\ell_i, m_i\}} \frac{\left( \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3} \right)^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \quad (94)$$

However the approximation of taking  $f_{\text{NL}} = 0$  in the calculation of the estimator variance is not always a good one if  $\Delta f_{\text{NL}}$  is dominated by squeezed configurations (We recall that by squeezed configurations we mean triangles in which one of the sides is much smaller than the other two; that is,  $\ell_1 \ll \ell_2, \ell_3$ .) or more in general by configurations in which one of the  $\ell$ 's is small. It turns out that, in these cases, the  $f_{\text{NL}}$ -dependent corrections to the Gaussian expectation of the bispectrum variance become important when  $f_{\text{NL}}$  gets large enough. This effect increases the variance of the estimator with respect to the expectation for  $f_{\text{NL}} = 0$ . There is a clear physical interpretation for this. One can see, for example, by calculating (94) in the simple Sachs-Wolfe case (see also Section 3.8) that the variance of the estimator scales roughly

like  $1/\ell_{\max}$ , or equivalently like  $1/N_{\text{pix}}$  in pixel space, with  $N_{\text{pix}}$  being the number of pixel in the observed map. This increase of the signal-to-noise ratio with the number of pixels is due to the fact that more and more bispectrum configurations are included into the sum over modes to estimate  $f_{\text{NL}}$ . However, if the signal is completely dominated by low- $\ell$  modes, as in the local case, then there is an intrinsic large cosmic variance, due to the small number of low- $\ell$  configurations. Clearly, cosmic variance cannot be beaten by increasing the resolution of the map. Creminelli et al. [75] then found that, for  $N_{\text{pix}}$  getting large, the  $S/N$  of the estimator for local NG grows asymptotically as  $(\ln N_{\text{pix}})$ , that is, much slower than the expected  $N_{\text{pix}}$ , which one would obtain by neglecting  $f_{\text{NL}}$ -dependent corrections in the calculation of the estimator variance. They carried out a calculation of the estimator variance in the flat-sky approximation, and neglected the transfer functions, to find the following expression:

$$\langle (\Delta\mathcal{E})^2 \rangle = \frac{1}{4AN_{\text{pix}} \ln N_{\text{pix}}} \left( 1 + \frac{8f_{\text{NL}}^2 f_{\text{NL}} A N_{\text{pix}}}{\pi \ln N_{\text{pix}}} \right), \quad (95)$$

where  $A$  is the bispectrum amplitude. We clearly see from this formula what we were stating above; that is, when  $f_{\text{NL}}$  gets large, the variance starts scaling like  $(1/\ln N_{\text{pix}})^2$ . The same formula also shows the technical point behind this behaviour: in the correction term, the order parameter is actually *not*  $f_{\text{NL}} A^{1/2}$  anymore but rather  $f_{\text{NL}} A^{1/2} N_{\text{pix}}$ . This enhancement by a factor  $N_{\text{pix}}$  can make the first-order corrections nonnegligible anymore. The natural question is now how large an  $f_{\text{NL}}$  we need to make the correction term important in (95). Following Creminelli et al. [75], let us call  $\sigma_0$  the standard deviation of the estimator computed for  $f_{\text{NL}}^{\text{loc}} = 0$ . Let us say that for an experiment at a given angular resolution (defined by  $\ell_{\max}$  in harmonic space or by  $N_{\text{pix}}$  in pixel space) a value of  $f_{\text{NL}}^{\text{loc}}$  is measured, equal to  $n\sigma_0$ . Substituting this value into formula (95) behavior and calling  $\sigma_{f_{\text{NL}}}^2$  the real estimator variance, one finds the following relative correction to  $\sigma_0$ :

$$\frac{\sigma_{f_{\text{NL}}}^2}{\sigma_0^2} - 1 = \frac{2n^2}{\pi \ln^2 N_{\text{pix}}}. \quad (96)$$

For an experiment like WMAP, the r.h.s. term becomes  $\sim 1$  when  $f_{\text{NL}}$  is about  $6\sigma_0$  away from the origin. For an experiment like that of Planck,  $f_{\text{NL}}$  has to be about  $3.5\sigma_0$ . The definition of a high- $f_{\text{NL}}$  regime is thus dependent on the experiment under study, as a consequence of the fact that the enhancement of first-order terms in the variance expression  $\langle (\Delta\mathcal{E}) \rangle$  depends on  $N_{\text{pix}}$ . In other words we can conclude the following.

*If  $f_{\text{NL}}^{\text{loc}}$  will be detected at several  $\sigma$  (in terms of the Gaussian expectation for the standard deviation), then  $f_{\text{NL}}$ -dependent correction terms in the estimator variance will have to be taken into account, and the simple expansion (84) of the CMB likelihood does not constitute a valid approximation anymore.*

One caveat in all of this discussion is that formula (95) was obtained in flat sky, neglecting the transfer functions, and it should be checked how dependent the final results are on these approximations. Since the scaling argument is

based on a very general physical reason, that is, the weight of squeezed configurations in the local  $f_{\text{NL}}$  estimator discussed earlier in this section, one expects that the general scaling with  $N_{\text{pix}}$  obtained in (95) does not depend on the details of radiative transfer and 2D projection. Liguori et al. [76] actually checked the results of this section numerically, by applying an implementation of the optimal cubic estimator (72) to full-sky simulations of CMB local NG maps with different  $N_{\text{pix}}$  and  $f_{\text{NL}}^{\text{loc}}$ , including the full radiative transfer (For details about the numerical implementation of the optimal cubic estimator, and about the generation of NG CMB maps, see Sections 3.6 and 3.7. ). Although, as expected, the coefficients in formula (95) change with respect to the simple flat sky no radiative transfer approximation, the scaling of the error bars with  $N_{\text{pix}}$  follows very well the expectations, going from  $\sim 1/N_{\text{pix}} \ln N_{\text{pix}}$  at low  $f_{\text{NL}}$  to  $1/\ln N_{\text{pix}}$  when  $f_{\text{NL}}$  is detected at high significance. Since in the large  $f_{\text{NL}}$  regime the variance starts to scale very slowly, like  $1/\ln^2 N_{\text{pix}}$ , one is led to wonder whether the estimator discussed in the previous sections becomes suboptimal at this point and whether a better one can be found. The answer to this question is not immediate. In order to check for the optimality of an estimator, as we have seen, one has to see whether it saturates the Rao-Cramer bound. However, also the local  $f_{\text{NL}}$  likelihood and Rao-Cramer bound estimated in the previous sections have to be recomputed, since they were obtained neglecting higher-order terms in  $f_{\text{NL}} A^{1/2}$ . In order to account for the  $f_{\text{NL}} A^{1/2} N_{\text{pix}}$ -enhanced terms, it is necessary to produce an exact expression of the full likelihood. This can be extremely challenging in the full radiative transfer case, but it is feasible in the flat sky no radiative transfer approach that we are considering (and that we showed earlier to be a good approximation as long as scaling arguments are involved). Creminelli et al. [75] proceed to calculate the full likelihood in this approximation and conclude that the optimal cubic estimator of weak local NG indeed *does not* saturate the Rao-Cramer bound in the high- $f_{\text{NL}}$  regime. The estimator (72) is thus no longer optimal in this case. They then proceed (always in the flat sky no transfer function case) to derive a cubic estimator that saturates the Rao-Cramer bound also for large  $f_{\text{NL}}^{\text{loc}}$ . We will not enter into the details of this derivation here, referring the reader to the study by Creminelli et al. in [75] for a complete discussion. The main aim of this section was to show under which conditions the optimality of the cubic estimator that we discussed in previous sections is valid. Since current bispectrum analysis of WMAP data [1, 68] finds that  $f_{\text{NL}}^{\text{loc}} \sim 2\sigma_0$ , the weak NG approximation applies, and the cubic estimator we derived earlier is indeed an optimal estimator in this case. However, if future Planck measurement will produce a detection of  $f_{\text{NL}}^{\text{loc}}$  at high significance, then the estimator will have to be modified in order to account for the enhanced variance in the high- $f_{\text{NL}}$  regime. This is not necessarily a remote possibility if one considers that the present central value of  $f_{\text{NL}}$  from WMAP estimates would produce a  $\sim 8998\sigma_0$  detection with Planck. Before concluding this section we would like to remark once again that the variance enhancement discussed here applies *only* to non-Gaussianity of the *local* type, whose bispectrum is

dominated by squeezed configurations, affected by a large cosmic variance. For the other shapes the cubic estimator (72) is optimal in both the small- and high- $f_{\text{NL}}$  regimes.

**3.6. Numerical Implementation of the Bispectrum Estimator.** In this section we turn to the problem of finding an efficient numerical implementation of the optimal bispectrum estimator (89). Let us repeat its expression here for convenience:

$$\begin{aligned} \mathcal{E}(a) &= \frac{1}{\mathcal{N}} \sum_{\{\ell_i, m_i\}} \left[ \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3} \left( C_{\ell_1 m_1, \ell_4 m_4}^{-1} a_{\ell_1 m_1} \right) \right. \\ &\quad \times \left( C_{\ell_2 m_2, \ell_5 m_5}^{-1} a_{\ell_2 m_2} \right) \left( C_{\ell_3 m_3, \ell_6 m_6}^{-1} a_{\ell_3 m_3} \right) \\ &\quad - 3 \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \\ &\quad \left. \times C_{\ell_1 m_1, \ell_2 m_2}^{-1} C_{\ell_3 m_3, \ell_4 m_4}^{-1} a_{\ell_4 m_4} \right], \\ \mathcal{N} &= \sum_{\{\ell_i, m_i\}} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle C_{\ell_1 m_1, \ell_4 m_4}^{-1} C_{\ell_2 m_2, \ell_5 m_5}^{-1} C_{\ell_3 m_3, \ell_6 m_6}^{-1} \\ &\quad \times \langle a_{\ell_4 m_4} a_{\ell_5 m_5} a_{\ell_6 m_6} \rangle. \end{aligned} \quad (97)$$

We remind the reader that this is the full expression, valid for the general nonrotationally invariant case. For a rotationally invariant CMB sky the linear term in the formula above vanishes, and the covariance matrix is diagonal and reduces to  $C_\ell$ , giving the simplified expression (72) that we reproduce again here for convenience:

$$\mathcal{E}(a) = \frac{1}{\mathcal{N}} \sum_{\{\ell_i, m_i\}} \frac{\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}, \quad (99)$$

$$\mathcal{N} = \sum_{\{\ell_i, m_i\}} \frac{\left( \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1} \right)^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \quad (100)$$

In a schematic way, the full estimator can be written as

$$\mathcal{E}(a) = \frac{\mathcal{E}^{\text{cubic}}(a) + \mathcal{E}^{\text{linear}}(a)}{\mathcal{N}}, \quad (101)$$

where the ‘‘cubic’’ term is the one containing the product  $a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}$ , while the linear term is the one dependent on a single  $a_{\ell m}$  and vanishing in the rotationally invariant case, where it is proportional to a monopole. It was shown before in a formal way that a pure cubic estimator becomes suboptimal when rotational invariance is broken, and adding the linear term is necessary to restore optimality. It is useful to try to understand the reason of this effect qualitatively and in a more intuitive way. Let us assume that we have a map characterized by nonstationary noise, and we are observing a region of the sky that was sampled many times so that the noise level in this area is low. That implies that the level of *small-scale* power in this *large* region is lower than average. Now, for a specific realization of the CMB sky, this modulation of small-scale power on a large region can look like a non-Gaussian signal sourcing a *squeezed* configuration of the bispectrum. On average, this effect must

cancel if the underlying noise model is Gaussian. However, this ‘‘confusion’’ between signal and noise increases the variance of any estimator of a primordial NG signal that is peaked on squeezed configurations. We know that this happens for the local model. This heuristic argument thus shows that, even though in principle a linear term must always be included when rotational invariance is broken, for a realistic noise model only local non-Gaussian estimates will be affected.

**3.6.1. Primary Cubic Term for  $f_{\text{NL}}$ .** Let us focus for the moment on the rotationally invariant case, where the linear term vanishes, and the covariance matrix is simply  $C = C_\ell \delta_{\ell_1 \ell_2} \delta_{m_1 m_2}$ . We immediately see that a brute force implementation of equation (99), consisting in computing and summing over all of the bispectrum configurations, would take  $\mathcal{O}(\ell_{\text{max}}^5)$  operations, where  $\ell_{\text{max}}$ , the maximum multipole in the calculation, depends on the resolution of the experiment. As mentioned earlier,  $\ell_{\text{max}} \sim 500$  for WMAP and  $\ell_{\text{max}} \sim 2000$  for Planck in the signal-dominated regime. At these resolutions, a brute force approach would be absolutely unfeasible for a general shape. If however we assume that the primordial shape under study is *separable*, then the dimensionality of the problem can be reduced and the overall number of operations scaled down significantly, making the computational cost affordable. Let us illustrate this point more in detail. Substituting (46) into the estimator expression (72) and remembering the identity (39),

$$\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} = \int d\Omega_{\hat{n}} Y_{\ell_1 m_1}(\hat{n}) Y_{\ell_2 m_2}(\hat{n}) Y_{\ell_3 m_3}(\hat{n}), \quad (102)$$

it is possible to rewrite (99) as follows (or more in general as a linear combination of terms of the following kind):

$$\begin{aligned} \mathcal{E}(a) &= \frac{1}{\mathcal{N}} \int dr r^2 \int d\Omega_{\hat{n}} \sum_{\ell_1 m_1} \frac{a_{\ell_1 m_1} X_{\ell_1}(r) Y_{\ell_1 m_1}(\hat{n})}{C_{\ell_1}} \\ &\quad \times \sum_{\ell_2 m_2} \frac{a_{\ell_2 m_2} X_{\ell_2}(r) Y_{\ell_2 m_2}(\hat{n})}{C_{\ell_2}} \\ &\quad \times \sum_{\ell_3 m_3} \frac{a_{\ell_3 m_3} X_{\ell_3}(r) Y_{\ell_3 m_3}(\hat{n})}{C_{\ell_3}} + \text{perm.} \end{aligned} \quad (103)$$

From an inspection of previous formula we see how, as a direct consequence of separability, the initial sum over the indices  $\ell_1 \ell_2 \ell_3, m_1 m_2 m_3$  has been factorized in the product of three sums, each running over two indices  $\ell, m$ . This greatly reduces the computational cost from  $\mathcal{O}(\ell_{\text{max}}^5)$  to  $\mathcal{O}(\ell_{\text{max}}^3)$  operations. If we define the new quantities

$$\begin{aligned} M_X(r, \hat{n}) &\equiv \sum_{\ell m} \frac{a_{\ell m} X_\ell(r)}{C_\ell} Y_{\ell m}(\hat{n}), \\ M_Y(r, \hat{n}) &\equiv \sum_{\ell m} \frac{a_{\ell m} Y_\ell(r)}{C_\ell} Y_{\ell m}(\hat{n}), \\ M_Z(r, \hat{n}) &\equiv \sum_{\ell m} \frac{a_{\ell m} Z_\ell(r)}{C_\ell} Y_{\ell m}(\hat{n}), \end{aligned} \quad (104)$$

then we can recast the estimator expression above in the following form:

$$\begin{aligned} \mathcal{E}(a) &= \frac{1}{\mathcal{N}} \int dr r^2 \int d\Omega_{\hat{n}} M_X(r, \hat{n}) M_Y(r, \hat{n}) M_Z(r, \hat{n}) + \text{perms}, \\ & \quad (105) \end{aligned}$$

where it is evident that we are now calculating our statistic in position space rather than in pixel space. Note how the filtered maps  $M_X$ ,  $M_Y$ ,  $M_Z$  can be efficiently calculated using a fast harmonic transform algorithms such as those included in the HEALPix package. This fast position space algorithm was initially introduced by Komatsu et al. in [77] in the context of local  $f_{\text{NL}}$  estimation and applied to the estimation of WMAP 1-year data by the WMAP team by Komatsu et al. in [78]. It was then applied to equilateral  $f_{\text{NL}}$  estimation for the first time in study of Creminelli et al. in [73]. An alternative numerical implementation with respect to the one used by the aforementioned authors was introduced by Smith and Zaldarriaga in [22]. Although different under many technical aspects, this second algorithm is still based on the calculation of the position space statistic (105); we refer the reader to the original work for additional details. This second implementation has been used to produce alternative estimates of  $f_{\text{NL}}^{\text{loc}}$ , and  $f_{\text{NL}}^{\text{eq}}$  from WMAP data, and to estimate the amplitude of the orthogonal shape, recently introduced by Smith et al. in [18].

Let us now discuss the possible limitations of this numerical approach. As noted in Section 2, the separability condition is in principle quite restrictive: the only separable shape arising directly from primordial models of inflation is the local one. On the other hand, it is still possible to study nonseparable models by finding separable shapes that are highly correlated to the primordial one. As observed in studies by Creminelli et al. in [73], Fergusson and Shellard in [9], and Smith and Zaldarriaga in [22], the  $f_{\text{NL}}$  limits obtained from a highly correlated separable shape in this way will be very close to those that would have been obtained using the original nonseparable model (see again Sections 2, 3.2, and 2.2 for a detailed discussion of this issue). We know from earlier sections that the other two shapes mentioned so far in this section besides local, namely, the equilateral and orthogonal shapes, have actually been derived as separable approximations of theoretical inflationary shapes. These approximations were obtained in an heuristic way; that is, an educated guess of a good separable approximation of the shape under study was made, and the correlation was checked *a posteriori*. There is obviously no a priori guarantee that this approach would be easily repeatable for all of the shapes of interest. The eigenmode expansion method introduced in [14] and summarized by equation (26), however, provides a general and rigorous method to find separable approximations of any shape, thus enabling the estimation of *any* possible primordial model. In this case, recall that we expand our (nonseparable) primordial shape function in terms of the separable basis functions  $\mathcal{Q}_n$

(see (26)), constructed from symmetric polynomial products  $q_p(k)$ , as

$$\begin{aligned} S(k_1, k_2, k_3) &= \sum_{prs} \alpha_{prs} q_p(k_1) q_r(k_2) q_s(k_3), \rightarrow b_{l_1 l_2 l_3} \\ &= \Delta_{\Phi}^2 f_{\text{NL}} \sum_{prs} \alpha_{prs} \int x^2 dx q_{\{p}^{\ell_1} q_r^{\ell_2} q_{s\}}^{\ell_3}, \end{aligned} \quad (106)$$

where the second expression for the reduced bispectrum  $b_{\ell_1 \ell_2 \ell_3}$  (56) expands in convolved basis functions (57) in harmonic space with

$$q_p^l(x) = \frac{2}{\pi} \int dk q_p(k) \Delta_l(k) j_l(kx). \quad (107)$$

In the mode expansion approach, then, the  $f_{\text{NL}}$  estimator for a specific model generalises to the following:

$$\mathcal{E}(a) = \frac{1}{\mathcal{N}} \sum_{prs} \alpha_{prs} \int dr r^2 \int d\Omega_{\hat{n}} M_{\{p}(r, \hat{n}) M_r(r, \hat{n}) M_{s\}}(r, \hat{n}), \quad (108)$$

where the filtered maps or shells  $M_p(r, \hat{n})$  are defined by

$$M_p(r, \hat{n}) = \sum_{lm} q_p^l \frac{a_{lm} Y_{lm}}{Cl}. \quad (109)$$

Defining the integral  $\beta_{prs} \equiv \int dr r^2 \int d\Omega_{\hat{n}} M_{\{p} M_r M_{s\}}$ , the estimator collapses into the compact form

$$\mathcal{E}(a) = \frac{1}{\mathcal{N}} \sum_{prs} \alpha_{prs} \beta_{prs}, \quad (110)$$

where it is possible to show a precise relationship between the theoretical bispectrum expansion coefficients  $\alpha_{prs}$  and expectations for the observed coefficients  $\beta_{prs}$ .

It was also pointed out by Fergusson et al. in [14] (see also [6]) and summarized in formula (60) that the separation can be performed directly in harmonic space on the reduced bispectrum  $b_{\ell_1 \ell_2 \ell_3}$ , rather than on the primordial shape  $S(k_1, k_2, k_3)$ . This provides an alternative, but equivalent, late-time  $f_{\text{NL}}$ -estimation pipeline with respect to the primordial shape separation approach given above as (110). In fact, since orthonormality is more direct on the harmonic domain without the intervention of transfer functions, the approach is considerably more straightforward conceptually. In this case, expectations for the observational expansion coefficients in the orthonormal frame  $\mathcal{R}_n$  (with  $n \rightarrow \{prs\}$ , see (60)) become simply  $\langle \bar{\beta}_n^{\mathcal{R}} \rangle = \bar{\alpha}_n^{\mathcal{R}}$ , that is, for an ensemble of maps possessing the theoretical bispectrum described by the coefficients. This means that for a NG bispectrum signal of sufficient significance we can consider directly and efficiently reconstruction of the bispectrum from the observed coefficients  $\bar{\beta}_n^{\mathcal{R}}$  using (60). We also note that the harmonic space separation scheme also allows for the estimation of noninflationary late-time bispectra, such as the bispectrum of cosmic strings, as well as other secondary anisotropies.

We can then conclude, in light of these developments, that the fast cubic statistic (105) can be applied in complete generality to any model of primordial NG, as well as to any other potential source of CMB NG. We also point out that alternative approaches have been considered for harmonic space  $f_{\text{NL}}$  analysis using wavelets and binning. For example, Bucher et al. [58] recently proposed using a suitable binning scheme in which the full expression for  $b_{\ell_1, \ell_2, \ell_3}$  is calculated in a subset of all of the triples  $\ell_1, \ell_2, \ell_3$ , small enough to make the calculation feasible while maintaining calculation accuracy. Approaches based on a harmonic space separation scheme, of course, require the full calculation of the reduced bispectrum  $b_{\ell_1, \ell_2, \ell_3}$  in order to determine the correlation between the theoretical prediction and the final expanded or binned bispectrum. The calculation of  $b_{\ell_1, \ell_2, \ell_3}$  implies the necessity of numerically solving the radiative transfer integral (31) for all of the configurations  $\ell_1, \ell_2, \ell_3$  which appears to be intractable in the nonseparable case, since the dimensionality of the problem cannot be reduced. However, this can be achieved efficiently in the general case using either the separable mode expansion integral (56) or else the hierarchical adaptive approach of Fergusson and Shellard [6] discussed in Section 3.4.

**3.6.2. Linear Correction Term for  $f_{\text{NL}}$ .** Let us now consider the realistic situation in which inhomogeneous noise and a sky-cut break rotational invariance (see Figure 9). In this case two complications arise as

- (1) A linear term in  $a_{\ell m}$  has to be added as follows.
- (2) The  $a_{\ell m}$  covariance matrix is now no longer diagonal. The inverse covariance weighting  $C^{-1}a$  that appears in expression (97) is hard to compute numerically for high angular resolution experiment, since its size makes a brute force numerical inversion impossible.

A first approach, introduced by Creminelli et al. in [73], is to simplify the problem by assuming that the covariance matrix is diagonal in the cubic term of the estimator, and then finding the linear term that minimizes the variance under this assumption. In other words, we keep the cubic term in the form of (99) and compute the variance of this term, relaxing the assumption of isotropy at this point (This means that, when we apply Wick's theorem to the  $a_{\ell m}$  six-point function in the calculation of the cubic term variance, we take  $\langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle = C_{\ell_1 m_1, \ell_2 m_2}$  instead of  $\langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle = C_{\ell_1 m_1} \delta_{\ell_1 \ell_2} \delta_{m_1 m_2}$ ).

It turns out that the variance is minimized (while leaving the estimator unbiased) for the following choice of the linear term:

$$\mathcal{E}_{\text{lin}} = -\frac{3}{\mathcal{N}} \sum_{\{\ell_i, m_i\}} \frac{\mathcal{Q}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} C_{\ell_1 m_1, \ell_2 m_2} a_{\ell_3 m_3}, \quad (111)$$

where  $\mathcal{N} = f_{\text{sky}} \sum_{\ell_1, \ell_2, \ell_3} (B_{\ell_1 \ell_2 \ell_3}^2 / C_{\ell_1} C_{\ell_2} C_{\ell_3})$  is the normalization term. Despite being suboptimal with respect to a full implementation of (97), this choice of linear term has been shown to significantly improve the error bars with respect to the simple cubic statistic (99) in presence of

anisotropic noise. At the same time, the simplicity of this implementation in comparison to the full optimal statistic (97) is manifest, since no  $C^{-1}$  terms appear in (111). Let us consider again a separable primordial bispectrum shape that can be written as  $S(k_1, k_2, k_3) = X(k_1)Y(k_2)Z(k_3) + \text{perm.}$  Applying the same procedure as we did for the cubic term, the linear term can be recast in the form

$$\begin{aligned} \mathcal{E}_{\text{lin}} = & -\frac{3}{\mathcal{N}} \int dr r^2 \int d\Omega_{\hat{n}} \\ & \times \left( \sum_{\ell_1 m_1} \frac{X_{\ell_1}(r)}{C_{\ell_1}} Y_{\ell_1 m_1}(\hat{\mathbf{n}}) \sum_{\ell_2 m_2} \frac{Y_{\ell_2}(r)}{C_{\ell_2}} Y_{\ell_2 m_2}(\hat{\mathbf{n}}) \right. \\ & \left. \times \sum_{\ell_3 m_3} \frac{Z_{\ell_3}(r)}{C_{\ell_3}} Y_{\ell_3 m_3}(\hat{\mathbf{n}}) \langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle a_{\ell_3 m_3} + \text{perm.} \right), \end{aligned} \quad (112)$$

where we explicitly wrote  $C_{\ell_1 m_1, \ell_2 m_2}$  as  $\langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle$ . This last formula can be rewritten as

$$\begin{aligned} \mathcal{E}_{\text{lin}} = & -\frac{6}{\mathcal{N}} \int dr r^2 \int d\Omega_{\hat{n}} \sum_{\ell m} \left[ \frac{X_{\ell}(r)}{C_{\ell}} \langle M_Y(r, \mathbf{n}) M_Z(r, \mathbf{n}) \rangle \right. \\ & + \frac{Y_{\ell}(r)}{C_{\ell}} \langle M_X(r, \mathbf{n}) M_Z(r, \mathbf{n}) \rangle \\ & \left. + \frac{Z_{\ell}(r)}{C_{\ell}} \langle M_X(r, \mathbf{n}) M_Y(r, \mathbf{n}) \rangle \right]. \end{aligned} \quad (113)$$

Like for the cubic part of the estimator, we have rewritten the linear term as a fast position space integral. The ensemble averages appearing in the last formula can be computed as Monte Carlo averages over a large number of Gaussian realizations of the CMB sky, characterized by the same beam, mask, and noise properties as in the experiment under study. This pseudooptimal, but relatively straightforward, implementation of the linear term has been adopted by a number of groups in order to estimate  $f_{\text{NL}}^{\text{loc}}$  from WMAP data [1, 73, 75, 79]. The full optimal estimator (89) was implemented only quite recently by Smith et al. in [68], where the authors developed an efficient conjugate gradient inversion (see e.g., [80]) algorithm based on earlier results from the study of Smith et al. in [81], in order to compute the  $C^{-1}a$  prefiltering in reasonable CPU time. Note that after the inverse covariance matrix prefiltering is calculated, the numerical implementation of the estimator is very similar to the one outlined above for the pseudooptimal case. The new position space statistic is obtained from formulae (105), (113), by making the following replacements, wherever the corresponding quantities appear:

$$\begin{aligned} a_{\ell m} & \longrightarrow a_{\ell m}^{\text{filtered}} \equiv (C^{-1}a)_{\ell m}, \\ M_X(r, \hat{\mathbf{n}}) & \longrightarrow \widetilde{M}_X(r, \hat{\mathbf{n}}) \equiv \sum_{\ell m} a_{\ell m} X_{\ell}(r) Y_{\ell m}(\hat{\mathbf{n}}), \\ \frac{X_{\ell}(r)}{C_{\ell}} & \longrightarrow X_{\ell}(r), \end{aligned} \quad (114)$$

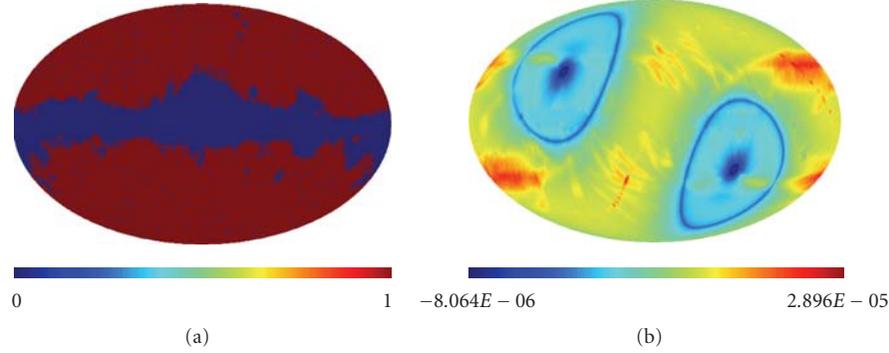


FIGURE 9: (a) The KQ75 galactic and point-source mask used for non-Gaussian analysis of WMAP data. (b) Anisotropic distribution of the noise for a coadded map of the WMAP V and W frequency channels. These two features break rotational invariance of the observed CMB sky and spoil the optimality of the standard cubic statistic (72), unless an additional “linear term” is included, as explained in the main text. Data are obtained from the LAMBDA website, <http://lambda.gsfc.nasa.gov/index.cfm>.

with analogous substitutions to be made for the  $Y, \mathcal{Y}, Z, \mathcal{Z}$  terms appearing in the same equations. The improvement in error bars from the pure cubic suboptimal estimator to the pseudooptimal and optimal statistics is shown in Figure 10.

**3.7. Experimental Constraints on  $f_{\text{NL}}$ .** In order to obtain an estimate of  $f_{\text{NL}}$  from a given dataset, one has first to generate sets of Gaussian CMB maps and obtain the MC averages that appear in the linear term expression (113), after an inverse covariance prefiltering of the full-optimal estimator is implemented. The normalization term  $\mathcal{N}$  can be precomputed using formula (46) to evaluate numerically the theoretical bispectrum shape for the model we want to estimate. The statistic (97) can then be computed for the experimental data  $a_{\text{obs}}$  to get our result:

$$\hat{f}_{\text{NL}}(a_{\text{obs}}) \equiv \frac{\mathcal{E}^{\text{cubic}}(a_{\text{obs}}) + \mathcal{E}^{\text{linear}}(a_{\text{obs}})}{\mathcal{N}}. \quad (115)$$

The error bars are then obtained by running the estimator on simulated Gaussian maps (The error bars can be obtained from Gaussian simulations as long as the weak NG approximation applies. As we saw earlier, this works at any  $f_{\text{NL}}$  for any shape, except for the local shape when a large  $f_{\text{NL}}^{\text{loc}}$  makes the error bars  $f_{\text{NL}}$  dependent. In this case the error bars would need to be calculated from NG simulations of  $f_{\text{NL}}^{\text{loc}}$ . So far, no high-significance detection of  $f_{\text{NL}}^{\text{loc}}$  has been reported, so working with G maps is at this stage sufficient to get accurate error bars.):

$$\sigma_{\hat{f}_{\text{NL}}} = \sqrt{\left\langle \left( \hat{f}_{\text{NL}}(a_{\text{sim}}) \right)^2 \right\rangle_{\text{MC}}}, \quad (116)$$

where  $\langle \cdot \rangle_{\text{MC}}$  indicates the MC average and  $a_{\text{sim}}$  a vector of simulated multipoles (obviously including mask, beam, and noise features of the experiment). For an accurate step-by-step description of an  $f_{\text{NL}}$  analysis of WMAP data, including details about channel coadding, noise model, beams, and pixel weighting schemes, we refer the reader to the explanations contained in the study of Komatsu et al. in [1]. The most stringent limits so far have been obtained by

applying the bispectrum estimator to the WMAP datasets. Constraints have been put on the local, equilateral, and orthogonal shapes. The best constraints come from the full implementation of the optimal estimator done in the study of Smith et al. in [68] and in that of Senatore et al. [18], and applied to the WMAP 7-year data release as in the study of Komatsu et al. in [83]. They are, at 95% C.L.,

$$-10 < f_{\text{NL}}^{\text{loc}} < 74, \quad (117)$$

$$-214 < f_{\text{NL}}^{\text{equil.}} < 266, \quad (118)$$

$$-410 < f_{\text{NL}}^{\text{ortho.}} < 6. \quad (119)$$

Since the first release of WMAP data, different groups have used the cubic statistic described in the previous paragraph, either in its pure cubic form (105) or in the improved version including the pseudooptimal linear term implementation (113). The results of different analyses of the WMAP 1-year, 3-year, 5-year, and 7-year datasets are summarized and commented in Table 3, where just the local and equilateral shapes have been included since the only two constraints on the orthogonal shape have been produced to date. The most recent orthogonal constraint has been already mentioned in (117). The other was obtained by Smith et al. [18] on WMAP 5-year data and it is  $-369 < f_{\text{NL}}^{\text{ortho.}} < 71$ .

**3.8. Fisher Matrix Forecasts.** The fisher matrix, defined as the curvature of the likelihood function calculated in its peak reassessment (see equation (A.3) in the appendix), plays a very important and well-known role in parameter estimation theory, not only because it defines the optimality of estimators through the Rao-Cramer bound, but also because it allows us to estimate *a priori* what the smallest error bars attainable will be for a given parameter (see again appendix). In other words, using the Fisher matrix we can forecast how well a parameter will be measured by a given experiment. This is very useful in order to optimize the experimental design to the detection of the parameters of interest. In our specific case, a Fisher matrix analysis will help us to understand what the smallest  $f_{\text{NL}}$  detectable in principle

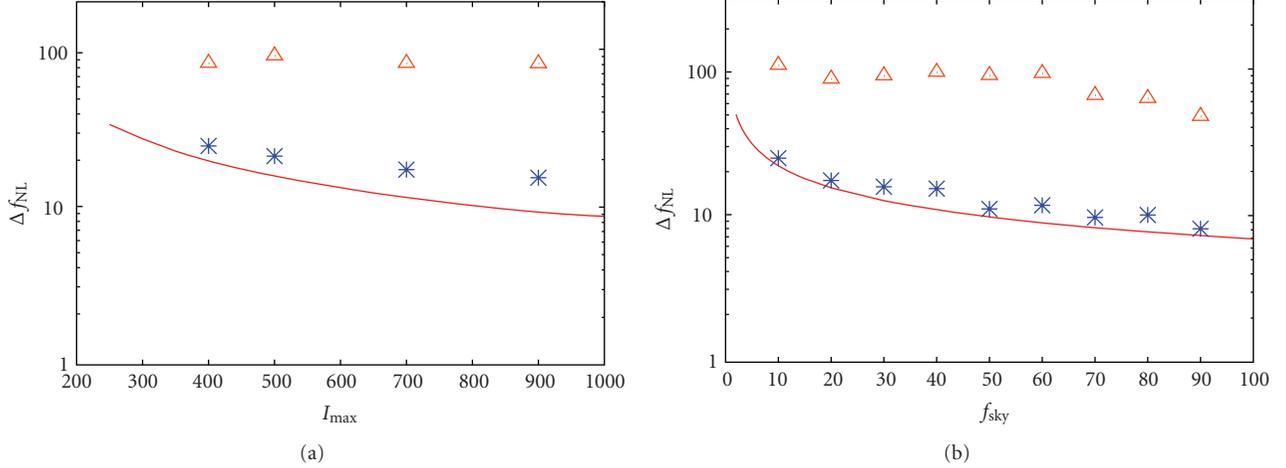


FIGURE 10: Error bars (obtained from Gaussian simulations) for the pure cubic (*triangles*) and pseudo-optimal (*stars*) implementations of the bispectrum estimator, to be compared to the solid red line, representing the Fisher matrix (Rao-Cramer) bound, saturated by the full-optimal statistic described in the text. (a) Error bars as a function of the maximum multipole included in the analysis. (b) Error bars as a function of the fraction of the sky considered in the analysis. This analysis included both temperature and polarization data, from the study by Yadav et al. in [82].

using different CMB datasets is, and which experimental features can be improved in order to increase the sensitivity to  $f_{\text{NL}}$ .

**3.8.1. A General Derivation.** Formula (A.12) from appendix, when applied to our case, yields

$$F_{f_{\text{NL}}, f_{\text{NL}}} = \frac{1}{6} \sum_{\ell_1, \ell_2, \ell_3=2}^{\ell_{\text{max}}} \frac{(B_{\ell_1, \ell_2, \ell_3}^{f_{\text{NL}}=1})^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}, \quad (120)$$

where  $B_{\ell_1, \ell_2, \ell_3}$  is the angular averaged bispectrum (i.e., the measured quantity). This can be rewritten in terms of the reduced bispectrum as

$$F_{f_{\text{NL}}, f_{\text{NL}}} = \frac{1}{6} \sum_{\ell_1, \ell_2, \ell_3=2}^{\ell_{\text{max}}} I_{\ell_1, \ell_2, \ell_3}^2 \frac{(b_{\ell_1, \ell_2, \ell_3}^{f_{\text{NL}}=1})^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}, \quad (121)$$

where we have defined (see also  $w_{l_1, l_2, l_3}$  in (58))

$$I_{\ell_1, \ell_2, \ell_3} = \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}. \quad (122)$$

Note how the features of the experiment enter the Fisher matrix through the parameter  $\ell_{\text{max}}$ , defining the angular resolution, and in the angular power spectrum expression in the denominator, which contains the angular beam and experimental noise:

$$\tilde{C}_\ell = C_\ell W_\ell^2 + N_\ell, \quad (123)$$

where  $C_\ell$  is the theoretical power spectrum for a given set of cosmological parameters,  $W_\ell$  is the beam of the experiment, and  $N_\ell$  is the experimental noise.  $N_\ell$  is a constant for uncorrelated noise. Likewise, the theoretical bispectrum will be convolved by the experimental beam

$$B_{\ell_1, \ell_2, \ell_3} = \mathcal{B}_{\ell_1, \ell_2, \ell_3} W_{\ell_1} W_{\ell_2} W_{\ell_3}. \quad (124)$$

Note that, since the noise is generally Gaussian, its three-point function vanishes. The experimental noise thus only enters in the denominator of the Fisher matrix expression. The effects of partial sky coverage can be easily accounted for. From (91) it follows that if only a fraction  $f_{\text{sky}}$  of the full sky is covered then the Fisher matrix takes an  $f_{\text{sky}}$  factor in front, which produces a degradation of the error bars of  $\sqrt{f_{\text{sky}}}$ .

We saw previously that for separable shapes the reduced bispectrum can be calculated either analytically, under some simplifying assumptions on the transfer functions (e.g., the Sachs-Wolfe approximation), or numerically through formula (43). It is then possible to evaluate numerically the Fisher matrix and the corresponding error  $\Delta f_{\text{NL}} \equiv \sqrt{1/F}$ . In the context of  $f_{\text{NL}}$  estimation, the first calculation of this kind was done for  $f_{\text{NL}}^{\text{loc}}$  by Komatsu and Spergel in [74], where it was found that WMAP could reach a sensitivity  $\Delta f_{\text{NL}} = 20$  (note how this bound is actually saturated by the optimal estimator results presented in Table 3), while Planck [85] could go down to  $\Delta f_{\text{NL}} = 5$ . (Note that all of the errors quoted in this section are at  $1 - \sigma$ .) What allows Planck to improve on WMAP is that it has a much better angular resolution and that it is cosmic variance dominated in a very large range of scales; that is, the power spectrum signal  $C_\ell W_\ell^2$  is larger than the noise  $N_\ell$  up to  $\ell_{\text{max}} = 2000$ . Angular resolution and sensitivity are the two factors that increase the ability of a CMB experiment to constrain  $f_{\text{NL}}$ . This information is provided by the Fisher matrix expression (121). Looking at such expression, we notice how the signal-to-noise ratio is obtained by adding over all of the bispectrum configurations up to  $\ell_{\text{max}}$ , weighted by their variance. Thus, the higher  $\ell_{\text{max}}$  is, the more configurations are included in the sum and the larger is the final sensitivity to  $f_{\text{NL}}$ . On the other hand, we see that, if the power spectrum of the instrumental noise appearing in the variance term in the denominator dominates from a certain  $\ell_{S=N}$ , then the

TABLE 3: Constraints on  $f_{\text{NL}}^{\text{local}}, f_{\text{NL}}^{\text{equil.}}$ , obtained by different groups on the one-year (W1), three-year (W3), five-year (W5), and seven-year (W7) WMAP data releases. Different rows correspond to the different implementations of the  $f_{\text{NL}}$  estimator described in the text: the “pure cubic” implementation (99) in which no linear term is included, the “pseudo-optimal” implementation (113) in which a linear term is added but the covariance matrix is assumed to be diagonal in the cubic term, and the fully “optimal” implementation (97). As we noted in the text, the linear term is important mostly for estimates of local NG, since anisotropic noise “mimics” squeezed configuration. For this reason “pure cubic” estimates of equilateral NG in the table are nearly optimal, while local ones are significantly suboptimal, especially because they have to be confined to the pure signal-dominated region  $l \lesssim 300$ , where the assumption of rotational invariance is correct. There is a certain degree of friction between some of the results shown. In particular the  $27 < f_{\text{NL}}^{\text{loc}} < 147$  WMAP 3-year estimate obtained by Yadav and Wandelt in [79], corresponding to a “nearly  $3\text{-}\sigma$ ” detection of local NG, seems not to agree well with the  $9 < f_{\text{NL}}^{\text{loc}} < 129, \sim 2.3\sigma$  result obtained on the *same* dataset by Smith et al. in [68]. The origin of the discrepancy is unclear, although it is argued by Smith et al. in [68] that it might be due to differences in the coadding scheme of different data channels, or analogous differences in the choice of some weights. As pointed out in Smith et al. [68], one additional advantage of the fully optimal implementation of the estimator is actually that all of the ambiguity related to the use of different coadding schemes disappears, since the optimal coadding strategy is automatically selected in the inverse covariance filtering process. Another discrepancy is that between the two equilateral constraints on WMAP 5-year data. It seems that the pseudo-optimal estimator produces better constraints than the optimal one. This is clearly not possible. Smith et al. [68] claim that their numerical pipeline calculates the theoretical ansatz for the bispectrum shape more accurately than it was done before. That is due to a subtlety that went unnoticed in previous works, consisting in the necessity to extend above the horizon the upper integration limit in the calculation of the equilateral shape-related quantities  $\beta_\ell(r), \gamma_\ell(r)$ , and  $\delta_\ell(r)$  (see (54)). This is required in order to obtain stable numerical solutions, and it calls for a reassessment of the expected and measured error bars, which actually increase with respect to previous calculations.

	Local		Equilateral	
Pure cubic	$-58 < f_{\text{NL}} < 134$	[78], W1	$-366 < f_{\text{NL}} < 238$	[73], W1
	$-54 < f_{\text{NL}} < 114$	[84], W3	$-256 < f_{\text{NL}} < 332$	[73], W3
Pseudo-optimal	$-27 < f_{\text{NL}} < 121$	[73], W1	$-151 < f_{\text{NL}} < 253$	[1], W5
	$-36 < f_{\text{NL}} < 100$	[73], W3		
	$27 < f_{\text{NL}} < 147$	[79], W3		
	$9 < f_{\text{NL}} < 129$	[68], W3		
	$-9 < f_{\text{NL}} < 111$	[1], W5		
Optimal	$12 < f_{\text{NL}} < 104$	[68], W3	$-125 < f_{\text{NL}} < 435$	[68], W5
	$-4 < f_{\text{NL}} < 80$	[68], W5	$-214 < f_{\text{NL}} < 266$	[83], W7
	$10 < f_{\text{NL}} < 74$	[83] W7		

signal contribution is suppressed above that threshold by the noise power spectra appearing in the denominator of (121). So what determines the sensitivity of a CMB experiment to  $f_{\text{NL}}$  is the range of  $\ell$  over which the instrumental noise is low, so the experiment is cosmic variance dominated. This range is  $\ell \lesssim 2000$  for Planck and  $\ell \lesssim 500$  for WMAP, hence Planck can obtain tighter constraints than WMAP. This is shown in Figure 11, where the Fisher matrix forecasts of  $f_{\text{NL}}$  are plotted for different CMB experiments: the predicted error bars decrease with  $\ell$  up to the angular scale at which the measurements start to be noise dominated, after which the  $f_{\text{NL}}$  signal-to-noise ratio saturates. A simple calculation done by Babich and Zaldarriaga in [86] taking the Sachs-Wolfe approximation, and working in flat sky, showed that, before noise dominates, the signal-to-noise ratio for the local shape grows as

$$\frac{S}{N} \propto \ell_{\text{max}} \ln\left(\frac{\ell_{\text{max}}}{\ell_{\text{min}}}\right), \quad (125)$$

where the  $(\ln)$  is dictated by the coupling between large and small scales introduced by squeezed configurations, from which most of the local signal comes.

Note also how, in absence of experimental noise, the beams in the numerator and in the denominator of (121) cancel each other out. An ideal noiseless CMB experiment

would then have a signal-to-noise ratio indefinitely growing. However, this would not imply infinite sensitivity to  $f_{\text{NL}}$ , because, above a certain  $\ell_{\text{max}}$ , secondary anisotropies would start to dominate. The Fisher matrix analysis of the equilateral shape (see [22, 87], e.g.) showed that the minimum achievable error bars in this case are  $\Delta f_{\text{NL}} \sim 100$  and  $\Delta f_{\text{NL}} \sim 60$ , for WMAP and Planck, respectively. (Note how the larger error bars in this case with respect to the local constraints do not reflect a higher sensitivity of CMB measurement to  $f_{\text{NL}}^{\text{loc}}$ , but only the conventional choice of the normalization of the bispectrum amplitude in the definition of  $f_{\text{NL}}$ . The normalizations are in fact chosen in such a way that the bispectra have the same value for equilateral configurations  $\ell_1 = \ell_2 = \ell_3$ , where the local bispectrum is suppressed and the equilateral bispectrum is peaked.) Additional shapes are studied by Smith and Zaldarriaga in [22].

**3.8.2. Polarization.** Babich and Zaldarriaga [86] showed with a Fisher matrix analysis that the CMB E-mode polarization measurements can be used to improve the sensitivity to  $f_{\text{NL}}$ . Although we have dealt so far only with temperature bispectra and related estimators, including polarization is fairly straightforward. As usual, the calculation starts from formula (31) linking the multipoles of CMB anisotropies to the primordial potential  $\Phi$ , but this time including the polarization radiative transfer  $\Delta_\ell^E(k)$  in the convolution

integral:

$$a_{\ell m}^E = 4\pi(-i)^l \int \frac{d^3k}{(2\pi)^3} \Delta_\ell^E(k) \Phi(\mathbf{k}) Y_{\ell m}(\hat{\mathbf{k}}). \quad (126)$$

The bispectrum is then defined in the usual way, but this time more configurations can be built by correlating temperature and polarization multipoles:

$$\begin{aligned} B_{\ell_1 \ell_2 \ell_3}^{TTT} &\equiv \langle a_{\ell_1 m_1}^T a_{\ell_2 m_2}^T a_{\ell_3 m_3}^T \rangle, \\ B_{\ell_1 \ell_2 \ell_3}^{TTE} &\equiv \langle a_{\ell_1 m_1}^T a_{\ell_2 m_2}^T a_{\ell_3 m_3}^E \rangle, \\ B_{\ell_1 \ell_2 \ell_3}^{TET} &\equiv \langle a_{\ell_1 m_1}^T a_{\ell_2 m_2}^E a_{\ell_3 m_3}^T \rangle, \\ &\vdots \\ B_{\ell_1 \ell_2 \ell_3}^{EET} &\equiv \langle a_{\ell_1 m_1}^E a_{\ell_2 m_2}^E a_{\ell_3 m_3}^T \rangle, \\ &\vdots \\ B_{\ell_1 \ell_2 \ell_3}^{EEE} &\equiv \langle a_{\ell_1 m_1}^E a_{\ell_2 m_2}^E a_{\ell_3 m_3}^E \rangle. \end{aligned} \quad (127)$$

The point to emphasize is that the polarization signal is generated on scales where the temperature signal is suppressed by Silk damping. The reason behind this can be briefly illustrated as follows: both the temperature and the polarization patterns that we observe in the CMB are produced by Thomson scattering of photons by electrons in the primordial plasma. Due to the Physics of Thomson scattering, in order for polarization to be generated it is necessary for the incident radiation field to be anisotropic (see, e.g., [88]). More precisely, the angular distribution of the intensity of the incoming radiation must present a non-vanishing quadrupole. However the Thomson scattering itself tends to isotropize the incoming radiation field. For this reason, on scales where the Thomson scattering is efficient (i.e., in the tight coupling regime of the photon-baryon fluid), polarization is not produced. In order to seed a significant quadrupole in the incoming photon distribution it is necessary to go below the free streaming scale of the photons in the primordial plasma, where the weak coupling between photons and electrons makes the isotropization process inefficient. However at these scales the free diffusion of photons with different temperatures *damps* the temperature fluctuations (a phenomenon known as Silk damping). For this reason we conclude that, on scales where temperature anisotropies are generated, polarization anisotropies are basically absent (because in the tight coupling regime the radiation intensity distribution does not present a quadrupole), and vice versa on scales where polarization is produced, temperature anisotropies are damped by diffusion processes. (This is strictly true only if we consider anisotropies generated at recombination. If we include a period of Early Reionization, then this picture is slightly changed, since after reionization polarization is generated on very large scales, where temperature anisotropies are present as well. For more details see, for example, [54] and references therein.)

The polarization bispectra thus open a window over a new  $\mathbf{k}$ -range in the  $3D \rightarrow 2D$  projection  $\mathbf{k} \rightarrow \ell$  and increase the overall information available. In other words, since the new configurations  $TTE$ ,  $TEE$ , and so forth, including polarization, are partially independent of the pure temperature ( $TTT$ ) bispectrum, adding those additional configurations to the Fisher matrix (and to the actual  $f_{\text{NL}}$  estimation from data) increases the total signal available. The Fisher matrix expression now becomes

$$F = \sum_{pqr} \sum_{ijk} \sum_{\ell_1 \ell_2 \ell_3} B_{\ell_1 \ell_2 \ell_3}^{pqr} [\text{cov}^{-1}]_{pqr|ijk}^{\ell_1 \ell_2 \ell_3} B_{\ell_1 \ell_2 \ell_3}^{ijk}, \quad (128)$$

where  $i, j, k, p, q$ , and  $r$  run over the  $T$  and  $E$  superscripts. We still work in the assumption that all of the quantities involved are Gaussian, but now the different bispectra of temperature and polarization are correlated for a given configuration  $\ell_1, \ell_2, \ell_3$ , thus defining a multivariate Gaussian distribution. The full covariance matrix between bispectra (indicated by  $\text{cov}$  in the formula above) has then to be evaluated. A numerical evaluation of (128) shows [86] that, for an ideal (i.e., noiseless) experiment, adding the polarization signal produces an improvement of a factor  $\sim 2$  on  $f_{\text{NL}}$  constraints. For WMAP, adding polarization bispectra produces very little improvement, since polarization data are mostly noise dominated. For Planck, however, including polarization does generate a significant improvement, bringing the forecasted error bars from  $\Delta f_{\text{NL}} \simeq 5$  to  $\Delta f_{\text{NL}} \simeq 3.5$ . Some error bar forecasts from temperature and polarization bispectra as a function of  $\ell_{\text{max}}$  for different experimental designs including WMAP and Planck are shown in Figure 11. Motivated by this analysis, Yadav et al. [82, 89] have implemented a bispectrum estimator of  $f_{\text{NL}}$  including both temperature and polarization bispectra. All of the general considerations about optimality and the numerical implementation techniques described in previous sections apply in an analogous way to the temperature + polarization case, although the presence of additional bispectra with a nontrivial covariance matrix introduces a few additional technical complications. We refer the reader to the study by Yadav et al. in [82, 89] for further discussion.

**3.9. Non-Gaussian Contaminants.** So far, we have considered only primordial non-Gaussianity as a source of the three-point function of the CMB. However many other astrophysical and cosmological effects can produce an observable angular bispectrum. Among these, diffuse astrophysical foreground emission (see, e.g., [91, 92] and references therein) unresolved point sources (see, e.g., [78]) and secondary anisotropies are probably the most important NG sources. Since the main focus in this review is on the primordial bispectrum, we will not describe these NG sources in great detail. We will however outline in this section their main effects in order to understand whether, and how, they could contaminate an estimate of primordial NG. Let us consider a number  $N_s$  of sources of a CMB bispectrum signal and call  $B_{\ell_1 \ell_2 \ell_3}^i$  the bispectrum produced by the  $i$ th source. Let us also indicate with  $B_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1}$  the primordial component of the bispectrum calculated for  $f_{\text{NL}} = 1$ . For our purposes,

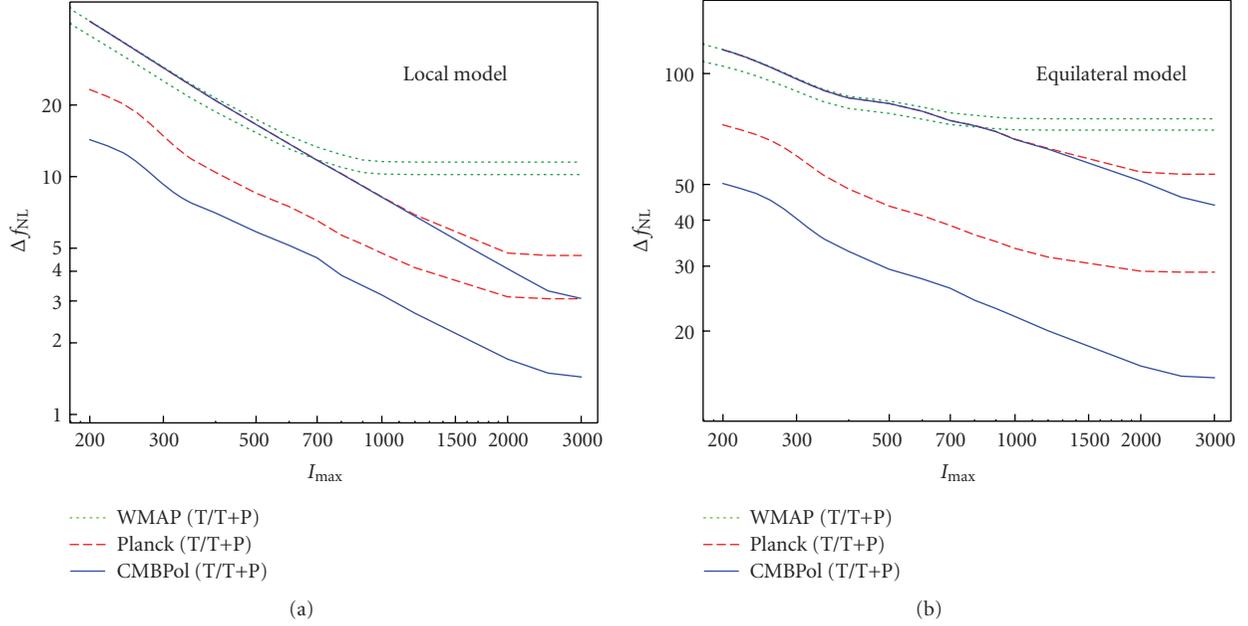


FIGURE 11: The Fisher matrix forecasts on  $\Delta f_{\text{NL}}$ , featured for different experiments: WMAP (green, dotted lines), Planck (red, dashed lines), and the proposed CMBPol [90] survey (blue, solid lines). (a) shows results for the local shape, while (b) refers to the equilateral shape. Thin lines are obtained from temperature data only, and thick lines show the improvement in the error bars coming from adding polarization datasets to the analysis for the various experiments.

$B_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1}$  is the signal that we want to measure, while the other bispectra are contaminants that we would like to eliminate. The total bispectrum of the map in presence of these contaminants is then

$$B_{\ell_1 \ell_2 \ell_3} = f_{\text{NL}} B_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1} + \sum_{i=1}^{N_i} A_i B_{\ell_1 \ell_2 \ell_3}^i, \quad (129)$$

where  $A_i$  is the amplitude of the  $i$ th bispectrum. If we have a precise prediction of the bispectra generated by the contaminants, we can then think of extending our  $f_{\text{NL}}$  estimator to a joint estimator of all of the amplitude parameters. The optimal cubic  $f_{\text{NL}}$  estimator defined in (72) would then be generalized to the multiparameter case by minimizing the following  $\chi^2$ :

$$\chi^2(f_{\text{NL}}, A_i) = \sum_{\ell_1 \ell_2 \ell_3} \frac{\left( f_{\text{NL}} B_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1} + \sum_{i=1}^{N_i} A_i B_{\ell_1 \ell_2 \ell_3}^i - B_{\ell_1 \ell_2 \ell_3}^{\text{obs}} \right)^2}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \quad (130)$$

The new errors on  $f_{\text{NL}}$  in this case can be forecasted as usual by means of a Fisher matrix analysis. The Fisher matrix described in the previous paragraph can be generalized straightforwardly to the multiparameter case. In this case,  $F$  becomes an array whose entries are defined as

$$F_{ij} = \sum_{\ell_1 \ell_2 \ell_3} \frac{B_{\ell_1 \ell_2 \ell_3}^i B_{\ell_1 \ell_2 \ell_3}^j}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \quad (131)$$

The optimal errors on a given amplitude  $A_i$  (including  $f_{\text{NL}}$ ) then become, according to the multidimensional generalization of the Rao-Cramer bound,

$$\Delta A_i = \sqrt{(F^{-1})_{ii}}, \quad (132)$$

where the crucial point to notice is that we now *first* invert the Fisher matrix and then we take the square root of the diagonal elements to find the errors. This is the error that is obtained when the full joint-parameter likelihood is calculated and then the 1-dimensional likelihood for a given parameter is obtained by integrating out all of the other degrees of freedom: a process defined in statistics as *marginalization*. One can see that the inverse of the Fisher matrix defines the *covariance matrix* of the parameters under study. If the various parameters are completely uncorrelated, then the Fisher matrix is diagonal and we would have  $F_{ii}^{-1} = (F^{-1})_{ii}$ , showing that the parameters can obviously be estimated independently and the marginalization process does not change the error bars on a given parameter of interest (in our case  $f_{\text{NL}}$ ). If the different parameters are correlated, however, then off-diagonal terms appear in the Fisher matrix, and the error bars after marginalization (i.e., the “real” error bars to quote in the results) are larger than those that would have been obtained by naively neglecting contaminants. An obvious but useful observation is that two bispectral amplitudes will be strongly correlated when the respective shapes are similar. To make a practical example, the bispectrum generated by correlating weak lensing of CMB anisotropies with the Integrated Sachs-Wolfe (ISW) effect can be shown to be peaked on squeezed

configurations. For this reason the presence of this effect can be a significant contaminant for estimates of local non-Gaussianity.

So far, in this section we have described the degradation effects on the error bars if a hypothetical joint estimator of all of the CMB bispectrum amplitudes was built, and the amplitudes of contaminants were marginalized over to estimate  $f_{\text{NL}}$ . However a joint estimation might be difficult, due to factors like the presence of theoretical uncertainties on the shapes of contaminant bispectra or possible practical difficulties in finding an efficient implementation of this full bispectrum-likelihood estimator (e.g., if the additional secondary bispectra are nonseparable). As a result, the practical approach so far has been to estimate only  $f_{\text{NL}}$  using the techniques described in previous sections and neglect possible nonprimordial contaminants. In this case the possible effect of contaminants would not show up as a degradation of the error bars but in an even worse way, by introducing a *bias* in the  $f_{\text{NL}}$  measurements. Let us see this by assuming that the CMB three-point function takes contributions both from a primordial NG component and from a contaminant bispectrum with amplitude  $A_i$ . Let us also assume that we can produce a set of NG Monte Carlo simulations of CMB maps including both bispectra. We assign a given  $f_{\text{NL}}$  in input to the primordial component of our simulated maps. Finally we estimate the average  $f_{\text{NL}}$  obtained from the simulations by applying the usual optimal cubic statistic described so far. The result of our MC average will be

$$\begin{aligned} \langle \hat{f}_{\text{NL}} \rangle &= \frac{1}{\mathcal{N}} \sum_{\ell_1 \ell_2 \ell_3} \frac{B_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1} B_{\ell_1 \ell_2 \ell_3}^{\text{observed}}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}} \\ &= f_{\text{NL}} + \frac{1}{\mathcal{N}} \sum_{\ell_1 \ell_2 \ell_3} \frac{B_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1} B_{\ell_1 \ell_2 \ell_3}^{\text{cont.}}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}, \end{aligned} \quad (133)$$

where  $B^{\text{observed}}$  is the averaged bispectrum extracted from the map. The  $f_{\text{NL}}$  term on the r.h.s. of the second line comes from the fact that the normalization  $\mathcal{N}$  is chosen in such a way as to obtain an unbiased estimator of the primordial component. However a second term is present, which accounts for the fact that a contaminant bispectrum,  $B_{\ell_1 \ell_2 \ell_3}^{\text{cont.}}$ , is in the map; this term clearly biases the estimator. (Let us remember with that by definition an estimator of a parameter  $\lambda$  (in our case  $f_{\text{NL}}$ ) is unbiased if  $\langle \hat{\lambda} \rangle = \lambda$ ,  $\hat{\lambda}$  being the estimate from data and  $\lambda$  being the true parameter of the underlying model.) The magnitude of the bias will depend again on how similar the shape of the contaminant bispectrum is to the primordial one. If, for example, the contaminant bispectrum is strongly peaked on equilateral configurations and suppressed on squeezed ones, a local estimator of NG will then not be significantly biased by it, since the second term in equation (133) will cancel out. However, an estimate of  $f_{\text{NL}}^{\text{equil.}}$  will in this case be significantly biased.

In general we can define the correlation coefficients between two bispectra, labeled  $i$  and  $j$ , as

$$\begin{aligned} r_{ij} &= \frac{\sum_{\ell_1 \ell_2 \ell_3} \left( (B_{\ell_1 \ell_2 \ell_3}^{(i)} B_{\ell_1 \ell_2 \ell_3}^{(j)}) / (C_{\ell_1} C_{\ell_2} C_{\ell_3}) \right)}{\sqrt{\left( \sum_{\ell_1 \ell_2 \ell_3} B_{\ell_1 \ell_2 \ell_3}^{(i)2} \right) / C_{\ell_1} C_{\ell_2} C_{\ell_3}} \sqrt{\left( \sum_{\ell_1 \ell_2 \ell_3} B_{\ell_1 \ell_2 \ell_3}^{(j)2} \right) / C_{\ell_1} C_{\ell_2} C_{\ell_3}}}. \end{aligned} \quad (134)$$

The definition of ‘‘correlation coefficient’’ becomes completely transparent if we rewrite the previous formula in terms of the Fisher matrix and keep in mind that  $F^{-1}$  defines the covariance matrix of the bispectrum amplitudes:

$$r_{ij} = \frac{(F^{-1})_{ij}}{\sqrt{(F^{-1})_{ii} (F^{-1})_{jj}}}. \quad (135)$$

The correlation coefficient varies by definition from 0, for totally uncorrelated shapes, to 1, for identical shapes, or  $-1$  for totally anticorrelated shapes. The more a given contaminant bispectrum is correlated to the primordial bispectrum that we want to measure, the larger will be the induced bias. At this point we distinguish between three possibilities. The first is that the contaminant bispectrum shape and amplitude are perfectly known. In that case we can compute the expected bias from formula (173) and subtract from our estimate. The second possibility is that the shape of the contaminant bispectrum is known, but its amplitude is defined with a given uncertainty. In this case we can propagate this uncertainty by quoting it in addition to the statistical error bars on  $f_{\text{NL}}$  obtained in the usual way. The third and worst possibility is that we are unaware of the presence of some contaminant effect, or we know nothing about its bispectrum. In this case we might obtain a biased estimate of  $f_{\text{NL}}$  without knowing it and thus eventually misinterpret a spurious NG effect as primordial NG. Contaminants are then very dangerous, because, if not properly taken into account, they can lead to spurious claim of detection of primordial NG. For this reason, if a positive detection of  $f_{\text{NL}}$  were to be made at some point for a certain model, all possible tests for the presence of contaminant effects should be performed. Moreover, since we cannot be absolutely sure that we are considering all possible sources of NG contamination, *cross-validation* of the result using other non-bispectrum-based estimators will be very important. These other estimators (Minkowski Functionals, wavelets, needlets, higher-order correlators are just some examples among those considered in the literature) are by construction suboptimal estimators of the primordial component. However, in principle they are expected to produce a totally different response to NG contaminants than the primordial bispectrum. A cross-detection of  $f_{\text{NL}}$  with many different statistics would then be much less likely due to some unknown spurious effect. Another way to test the primordial origin of an observed NG signal, recently proposed by Munshi and Heavens in [93], is to modify the optimal bispectrum estimator in order to evaluate a function of  $\ell$  rather than a single amplitude  $f_{\text{NL}}$ . The point is that, if a

clear detection of  $f_{\text{NL}}$  is achieved at several  $\sigma$ , then the signal is large enough to allow a less radical data compression. Munshi and Heavens [93] have then recently proposed to estimate the “bispectrum-related power spectrum”  $C_\ell^{\text{skew}}$  defined as

$$C_{\ell_3}^{\text{skew}} = \frac{1}{2(\ell_3 + 1)} \sum_{\ell_1 \ell_2} \frac{B_{\ell_1 \ell_2 \ell_3}^{f_{\text{NL}}=1} B_{\ell_1 \ell_2 \ell_3}^{\text{obs}}}{C_{\ell_1} C_{\ell_2} C_{\ell_3}}. \quad (136)$$

Like in the usual  $f_{\text{NL}}$  estimator, the optimal  $S/N$  weighting is included, and the observed bispectrum from the map is correlated to the theoretical shape. However in this case we do not measure the overall amplitude, but rather the amplitude for each  $\ell$ -bin. Note that

$$\hat{f}_{\text{NL}} \equiv \frac{1}{\mathcal{N}} \sum_{\ell} C_\ell^{\text{skew}}. \quad (137)$$

By construction of  $C_\ell^{\text{skew}}$ , the usual  $f_{\text{NL}}$  estimator is then retrieved by summing the bispectrum-related power spectrum over all  $\ell$ . The general idea is now that the functional dependence of this skew power spectrum on  $\ell$  will show significant variation between different sources of NG, allowing a clearer test of the hypothesis that the origin of the observed signal is primordial. A number of investigations of WMAP data have already been performed using this statistic in order to look for primordial and secondary signals [94, 95], and related pseudo- $C_l$  statistics have been developed by Munshi et al. in [96].

In any case, as long as bispectrum estimators are considered, independently of the specific statistic or implementation, the best way to deal with NG contaminants is to make sure to list all of them and study their bispectra, or at least find ways to assess their potential impact on the final results. In the following paragraphs we will then turn our attention to a classification of the most important potential sources of spurious NG and see how they are treated in the primordial bispectrum analysis. Finally, we will consider some effects that interact with the  $f_{\text{NL}}$  measurement not necessarily by directly producing a secondary bispectrum, but rather by changing the normalization of the estimator or by increasing the error bars without producing any bias.

**3.9.1. Diffuse Foreground Emission.** There are three main astrophysical effects producing a galactic microwave emission from our galaxy in the typical frequency range of a CMB experiment [54, 91]: free-free emission from electron-ion scattering, synchrotron emission from acceleration of cosmic ray electrons in magnetic fields, and thermal dust emission.

Since these sources produce signals with a peculiar spectral and spatial distribution, multifrequency observations allow the separation of them from the primordial component of the CMB signal by suitable *component separation* algorithms. In the resulting “cleaned” map the foreground contribution to  $a_{\ell m}$  is minimized, although obviously it can never be completely eliminated. The remaining foreground contamination after cleaning is called the *foreground residual*. Note that the emission from the galactic plane of the CMB map is so strong that a clean separation of the primordial

CMB component from the foregrounds is impossible. The galactic regions that are too contaminated to produce a clean component separation have to be masked out in the analysis. The size of the galactic mask will depend on the choice of the foreground flux level above which the pixel is considered too contaminated to be included in the analysis. The choice of the cutoff will depend on the specific analysis that one wants to perform on the data. Since the primordial NG signal is much smaller than the Gaussian component, more conservative masks (i.e., larger) need generally to be used for  $f_{\text{NL}}$  estimates than those applied to  $C_\ell$  estimation. Direct information about the spatial distribution of foreground emission in the sky (i.e., free-free, synchrotron, or dust) is provided in the form of templates, obtained either from the most foreground contaminated channels of the CMB experiment itself, or from external astrophysical surveys (e.g., observations of radioemission, maps of H $\alpha$  emission). Templates are affected by several sources of uncertainties and errors (see, e.g., [91]), and using them in assessing the possible impact on  $f_{\text{NL}}$  of foreground emission or residuals has both advantages and disadvantages. The safest approach is probably to combine internal consistency tests on the data with analysis involving the use of templates.

The first extensive tests of possible foreground contamination in  $f_{\text{NL}}$  measurements were performed by Yadav and Wandelt in [79], where a detection of a primordial local signal at above 99.5% level on WMAP 3-year data was claimed. As explained earlier (see caption of Table 3), further analysis on more recent datasets and/or using more optimal estimators have led to an updated  $f_{\text{NL}}^{\text{loc}}$  estimate that is about  $2\text{-}\sigma$  away from the origin, that is, just a “hint” of a possible local signal, rather than a detection. However, as long as a detection was claimed by Yadav and Wandelt in [79], tests to exclude a possible contamination from diffuse foregrounds had to be carried out. In this case the authors relied mostly on the “internal consistency test” approach. Their analysis included the following.

- (1) Expanding the original galactic mask in order to see whether the estimated value of  $f_{\text{NL}}$  is stable for different choice of the mask. A significantly lower value of  $f_{\text{NL}}$  for a larger mask might mean that some unmasked noise contribution is affecting the measurement with the original mask.
- (2) Comparing  $f_{\text{NL}}$  estimates from foreground-reduced maps to estimates from “raw” maps that include a galactic mask, but have not gone through a component separation process. If foregrounds have a significant impact on  $f_{\text{NL}}$ , then one expects the measurements from raw and reduced maps to differ significantly.
- (3) Comparing different frequency channels. If foregrounds significantly contaminate measurements at given frequencies, then different channels should produce different results.

Analyses involving some kind of prior information about foreground emission were carried on by both Yadav and

Wandelt [79] and Smith et al. [68]. The two approaches adopted in this case were the following.

- (1) Producing simulations including both a Gaussian primordial CMB signal and the foreground emission. The latter has in this case to be generated according to a model that allows for a good reconstruction of the observed templates. The  $f_{\text{NL}}$  estimator can then be applied to these simulations in order to check whether the measured  $f_{\text{NL}}$  is consistent with 0 (as it should be, in absence of significant foreground contamination, since the primordial input is Gaussian).
- (2) For an optimal estimator including full  $C^{-1}$  pre-filtering [68], adding the foreground templates to the noise covariance, by assigning *infinite* variance to each template  $T_{\text{templ}}^i(\hat{\mathbf{n}})$ . In this way the estimate is “blind” to the template amplitudes. This produces a loss of information that in turn determines an increase of the variance. The larger the contamination from foreground is, the more the variance increases. For negligible contamination, the variance stays the same. In any case, the effect of foregrounds is entirely included in the error bars, provided that the assumed templates are accurate enough. This method of analysis, called *template marginalization*, is adopted by Smith et al. in [68]. A complete mathematical derivation of this method is provided by Rybicki and Press in [97].

In addition to the methods outlined above, there is also the possibility of using the foreground templates for a joint estimation of  $f_{\text{NL}}$  and of the templates amplitudes (see equation (130)). This approach has been recently used by Cabella et al. in [98] for a needlet estimator. It could be obviously reapplied in the same form to a bispectrum estimator.

In conclusion, all of the tests above have been applied to WMAP 3-year and 5-year data releases. No evidence for the presence of a significant contamination of the local  $f_{\text{NL}}$  measurement from diffuse foreground was produced. Other shapes of  $f_{\text{NL}}$  were not considered since the only type of non-Gaussianity that has produced a marginal detection is so far the local one. Although diffuse foregrounds and foreground residuals do not seem to contaminate primordial NG measurements in WMAP, this is not guaranteed to hold true for Planck, due to its much higher sensitivity.

**3.9.2. Unresolved Point Sources.** Extragalactic point sources are the most important foreground at small angular scales (see [99]). Sources are identified by searching the maps for bright spots that fit the beam profile and then masked out. However not all of the sources can be resolved and eliminated in this way. Unresolved point sources contaminate the map and are a source of a NG signal that can potentially interfere with primordial NG measurements. Unclustered extragalactic point sources have a Poisson distribution and their bispectrum is then simply a constant:

$$b_{\ell_1 \ell_2 \ell_3}^{ps} = b^{ps}, \quad (138)$$

with an amplitude that has to be estimated from the data and depends on the level of contamination from unresolved sources. We can now use (135) to estimate the correlation between primordial shapes and the point source bispectrum. For a given choice of the amplitude we can also estimate the expected bias on the  $f_{\text{NL}}$  estimator. Simulations of NG maps including the bispectrum from point sources can also be produced and the primordial  $f_{\text{NL}}$  estimator for different shapes applied to them in order to estimate the bias. Finally, since  $b_{\ell_1 \ell_2 \ell_3}^{ps}$  is manifestly separable, an estimator of  $b^{ps}$  can be built. All of these analyses were performed by Komatsu et al. in [1, 78] on local and equilateral shapes to conclude that point sources do not contaminate significantly the estimate of  $f_{\text{NL}}^{\text{loc}}$ . On the other hand, they have a larger impact on  $f_{\text{NL}}^{\text{equil}}$ : their induced bias from MC simulations is  $\Delta f_{\text{NL}}^{\text{equil}} = 22 \pm 4$ , to be compared to the statistical error bar  $\Delta f_{\text{NL}}^{\text{equil}} \sim 100$ . Additional tests were performed by Smith et al. in [68] to account for the possible presence of clustered unresolved point sources. No significant contamination on  $f_{\text{NL}}^{\text{loc}}$  was found in this case. As for the diffuse foreground case, the enhanced  $f_{\text{NL}}$  sensitivity that Planck can achieve with respect to WMAP might increase the impact of these effects.

**3.9.3. Secondary Anisotropies.** One big advantage of using CMB anisotropies to test primordial NG is that they are small and can then be treated in the linear regime. The CMB temperature fluctuation field is thus linked to the primordial potential through a *linear* convolution with radiation transfer functions, as we saw earlier. At this level, the Gaussianity of the primordial potential is conserved in the CMB temperature fluctuation field. If, however, we work at second-order in perturbation theory, the initial conditions are propagated *nonlinearly* into the observed CMB anisotropies, and the resulting CMB fluctuations are mildly non-Gaussian even starting from a Gaussian primordial curvature field. Second order effects are clearly very small. However they may well be of the same order of magnitude as primordial NG, since the NG component of the primordial potential is  $O(f_{\text{NL}} \langle \Phi_{\text{I}}^2(\mathbf{x}) \rangle)$ . In conclusion, *secondary anisotropies are a potential source of CMB NG, at a level that could in principle contaminate estimates of primordial non-Gaussianity.* To fully account for these effects, it is necessary to obtain a relation analogous to equation (31), but to second-order in perturbation theory. Radiation transfer functions are obtained at first order by solving the linearized system of Boltzmann-Einstein equations (see, e.g., [54, 55]). The same equations will then have to be expanded and numerically integrated at second order in this case. Having obtained second order transfer functions, the full angular bispectrum of secondary anisotropies can be calculated and correlated to the primordial one in order to check for the presence of contaminant effects. The full system of second order Einstein-Boltzmann equations has been derived in [100–103] and partially integrated numerically in [104] including only the source terms that can be written as product of first-order perturbations. These terms have been shown to produce a totally negligible NG contamination. Very recently, the study in [105] has

integrated numerically the *full* Einstein-Boltzmann system of equations at second order, including “genuine” second-order terms, but neglecting late-time effects, that is, contribution to the temperature anisotropies coming from the evolution of the gravitational potentials due to the late-time acceleration of the Universe or to reionization effects. According to this calculation, second-order early-time effects (i.e., Sachs-Wolfe and acoustic oscillations at second order) are able to bias the estimator at level of  $f_{\text{NL}}^{\text{bias}} \sim 5$  at the angular resolution achieved by the Planck satellite ( $\ell_{\text{max}} \sim 2000$ ), while a negligible contamination is expected for WMAP. Although a formal solution of the full system of equations has not been obtained yet for the “late-time” source, many late-time secondary effects are known and have been modeled for some time. Among these there are, for example, weak lensing, Sunyaev-Zeldovich (SZ) effect, Rees-Sciama (RS) effect, and so on. Therefore, a natural approach that was adopted in the literature consisted in studying the bispectra arising from these well-known effects and from their correlations taken *one-by-one* (e.g., ISW-lensing correlation, SZ-lensing correlation, and so on). It goes beyond the purpose of this review to discuss in detail these results and their implications. Let us just mention them briefly. A Fisher matrix analysis in the study of Serra and Cooray in [106] showed that the combination of bispectra arising from ISW-lensing, SZ-lensing, and unresolved point sources produced a negligible contamination at the angular resolution and sensitivity of WMAP, but a significant one for an experiment with the characteristics of Planck. It was in particular shown that estimates of local NG would be biased, especially by ISW-lensing correlation, with  $f_{\text{NL}}^{\text{bias}} \sim 10$  for local NG. A similar result on ISW-lensing was obtained in another Fisher matrix analysis by Smith and Zaldarriaga [22], and a similar level of contamination was found by Mangilli and Verde in [107] by adding to the ISW-lensing signal also the analogous RS-lensing bispectrum. A bispectrum estimator of local and equilateral NG was applied to simulated lensed primordial NG CMB maps by Hanson et al. [108], and three main effects were studied: a possible bias induced by neglecting the lensing of primordial bispectrum in the normalization and weights of the estimator, an increase of the variance due to lensing-produced higher-order correlators, and ISW-lensing bias. The only significant effect turned out to be the ISW-lensing bias on  $f_{\text{NL}}^{\text{loc}}$ , at a level confirming the Fisher matrix predictions. Note that this bias, being well known and expected, can be simply calculated and subtracted from future Planck estimates, as well as the early-time bias discussed above. The reason why the coupling between lensing and ISW tends to bias the local estimate can be understood physically: *large-scale* potential fluctuations source the ISW effect and produce a lensing signal on *small* scales, generating a NG signal on squeezed triangles. Although both the primordial local bispectrum and the ISW-lensing bispectrum are peaked on squeezed triangles, the presence of acoustic oscillations in the primordial configurations reduces the overall correlation between the two shapes, thus making the final bias significant, but not too large. Another recently studied effect is that of inhomogeneous recombination caused by perturbation in the electron number density as in

the studies of Khatri and Wandelt in [109], and in Senatore et al. [110]. Also in this case the contaminant shape is close to local. Although smaller than the ISW-lensing induced bias, also this effect seems to be able to affect the primordial estimate at a level marginally detectable by Planck. In order to conclude our brief survey of studies of secondary bispectra, let us finally mention the work done by Babich and Pierpaoli in [111], where point source density modulation bispectra induced by lensing magnification and selection effects, as well as SZ modulation from lensing magnification, were studied. The conclusion was once again that these effects are negligible for WMAP but close to the sensitivity level of Planck for local NG. Despite the great attention received so far in the literature, more has yet to be done in the area of assessing NG contamination from secondary sources. Note in particular that all of the predictions in this section concern temperature anisotropies, and estimates on secondary polarization bispectra are yet unavailable. It is clear that a complete and accurate description of secondary bispectra will be crucial for analysis of the future Planck dataset. A summary of the contribution of the various effects described in this paragraph, as well as of the other sources of contamination considered in this section, is presented in Table 4 and relative caption.

**3.9.4. Non-Gaussian Noise.** Systematics are another potential cause of contamination beyond astrophysical and cosmological sources. The noise in the experiment is generally well described as Gaussian. However possible non-Gaussian properties have to be tested in our context. This was done by Yadav and Wandelt in [79] by taking differences of yearly WMAP data in order to create jackknife realizations of WMAP noise maps for different detectors, including instrument systematics. The estimator can then be applied to these realizations in order to check that a negligible  $f_{\text{NL}}$  is measured. This was the result obtained on the WMAP 3-year dataset.

**3.9.5. Other Effects.** In this section we quickly summarize other effects that could interfere with estimates of primordial non-Gaussianity, but did not fit the classification above in the sense that they do not correspond to NG effects contaminating the CMB sky or the instrument noise.

One of these effects is  $1/f$  noise, expected to affect especially the low-frequency channels of Planck. The  $1/f$  noise component is generally removed from the map using “destriping” algorithms (see, e.g., [114, 115]). The unsubtracted “destriping residuals” form a Gaussian-correlated random field in pixel space. Their nontrivial covariance matrix should in principle be included in the inverse covariance prefiltering of the optimal estimator. If not included in the prefiltering, this effect could in principle enhance the estimator error bars (although it cannot generate any bias, since it is Gaussian). Unfortunately, a full numerical evaluation of this covariance matrix is quite challenging. Donzelli et al. [113] applied the estimator in its pseudo-optimal implementation to maps of Gaussian CMB signal + noise, accounting only for anisotropic noise in the linear

TABLE 4: Expected contamination on  $f_{NL}$  measurements from secondary bispectra (see Section 3.9.3) and from the effects described in Section 3.9.5. The expected bias of the primordial  $f_{NL}$  estimate is given for local and equilateral shapes (if not reported, that means that the corresponding shape has not been studied). All of the estimates above are for an experiment with the angular resolution of the Planck satellite (i.e.,  $\ell_{max} \sim 2000$ ). All of the effects above have been proved to be negligible, when compared to primordial  $f_{NL}$ -error bars, in the WMAP case. The expected biases have to be compared to the primordial error bars estimated from Fisher matrix forecasts for a Planck-like experiment. These are, as reported in the table,  $\Delta_{f_{NL}}^{local} \simeq 5$  and  $\Delta_{f_{NL}}^{equil} \simeq 60$ . Some effects, namely, ISW-lensing bispectra and the three-point function from inhomogenous recombination, have been studied by different authors. In these cases all of the results obtained in different works are reported. While a good agreement is found for ISW-lensing estimates, some discrepancy between different studies is present for inhomogenous recombination calculations. Note how asymmetric beams and residuals of destripping change the correlation properties of the CMB temperature field, but leave it Gaussian. For this reason they can in principle affect the final error bars, but they cannot produce any bias. In addition to the effects summarized in this table, another potential source of contamination comes from foreground residuals (see Section 3.9.1). This has been shown to be negligible for WMAP, while its impact for Planck has not been discussed yet in the literature and will be assessed in the forthcoming Planck data release. The effect of point sources on WMAP  $f_{NL}$  estimates has been studied in [1], where it was found that bias from unresolved point sources generates an additional contribution to the error bars of order 5 and 22 for local and equilateral shapes, respectively (to be compared to WMAP primordial  $f_{NL}$ -error bars  $\Delta_{f_{NL}}^{local} \simeq 20$  and  $\Delta_{f_{NL}}^{equilateral} \simeq 100$ ). Note finally how the contribution from the propagation of cosmological parameter errors on the  $f_{NL}$  estimate is dependent on the measured value of  $f_{NL}$  (like for point sources, the effect of cosmological parameters error propagation is to bias the estimator; however the sign and exact magnitude of this bias cannot be computed; that produces an additional uncertainty and correspondingly an additional contribution to the error bars.).

		Local ( $\Delta_{f_{NL}} \simeq 5$ )	Equil. ( $\Delta_{f_{NL}} \simeq 60$ )
ISW-lensing	Serra and Cooray [106]	$f_{bias} \simeq 10$	
	Hanson et al. [108]	$f_{bias} \simeq 10$	$f_{bias} \simeq -3$
ISW+RS-Lensing	Mangilli and Verde [107]	$f_{bias} \simeq 10$	
Unres. Point Sources (PS)	Serra and Cooray [106]	$f_{bias} \simeq 1$	
SZ number density modulation	Babich and Pierpaoli [111]	$f_{bias} \simeq -1.0$	$f_{bias} \simeq 0$
PS density modulation	Babich and Pierpaoli [111]	$f_{bias} \simeq -0.4$	$f_{bias} \simeq 0$
PS lensing magnification	Babich and Pierpaoli [111]	$f_{bias} \simeq 0.3$	$f_{bias} \simeq 0$
SZ lensing magnification	Babich and Pierpaoli [111]	$f_{bias} \simeq 0.02$	$f_{bias} \simeq 0$
Inhomogenous recombination	Khatri and Wandelt [109]	$f_{bias} \simeq -0.1$	$f_{bias} \simeq 0$
	Senatore et al. [110]	$f_{bias} \simeq -3.5$	$f_{bias} \simeq 0$
Boltmann-Einstein (BE) “1st $\times$ 1st”	Nitta et al. [104]	$f_{bias} \simeq 0.5$	
BE “Early times”	Pitrou et al. [105]	$f_{bias} \simeq 5$	$f_{bias} \simeq 5$
Cosm. parameters uncert.	Liguori and Riotto [112]	$ f_{bias}  \simeq 0.05 f_{NL}^{local}$	$ f_{bias}  \simeq 0.05 f_{NL}^{equil}$
Asymmetric beams	Donzelli et al. [113]	$\Delta_{f_{NL}} \simeq 0$	$\Delta_{f_{NL}} \simeq 0$
Residuals of Destripping	Donzelli et al. [113]	$\Delta_{f_{NL}} \simeq 0$	$\Delta_{f_{NL}} \simeq 0$

term, but including destripping residuals in the noise model adopted for the simulations. The final result shows that the error bars do not increase when  $1/f$  noise effects are included in the simulations, even though they are neglected in the covariance matrix appearing in the estimator.

Another effect to take into account for Planck is that of an asymmetric beam. The beam in the estimator normalization term is approximated as a circular beam. However Planck optical simulations (see, e.g., [116]) show that in reality we have to deal with elliptic beams, characterized by a nontrivial azimuthal dependence. If the circular beam approximation in the normalization of the estimate is not accurate enough, a bias could be introduced. Moreover the anisotropy of the beam could cause an increase of the variance if neglected in the inverse covariance prefiltering. Again, these effects were found to be negligible in tests on realistic simulations performed by Donzelli et al. [113].

Finally, the estimate of  $f_{NL}$  is done assuming a given cosmological model, that is, by fixing all of the other cosmological parameters to their best-fit value obtained from a likelihood analysis of the angular power spectrum. Since they

are themselves the product of a statistical estimation process, these values obviously present uncertainties that should be propagated into the final  $f_{NL}$ -error bars. (In particular, since we are not doing a joint-likelihood estimation of all of the parameters and marginalizing to get  $f_{NL}$  (that would be the optimal but time consuming approach), the effect of uncertainties in the parameters propagate onto the  $f_{NL}$  measure as a *bias*. This bias has to be evaluated and quoted in addition to the usual statistical  $f_{NL}$ -error bar.). This calculation was done by Liguori and Riotto in [112], where it was found that the propagated error is  $f_{NL}$  dependent and it can become important only if a large  $f_{NL}$  will be detected in the data at some point.

**3.10. Generation of Simulated Non-Gaussian CMB Maps.** In this section we will describe algorithms for the generation of non-Gaussian CMB maps with a given bispectrum. There are three main reasons why primordial NG simulations of the CMB are useful in the context of bispectrum estimation of  $f_{NL}$  as follows.

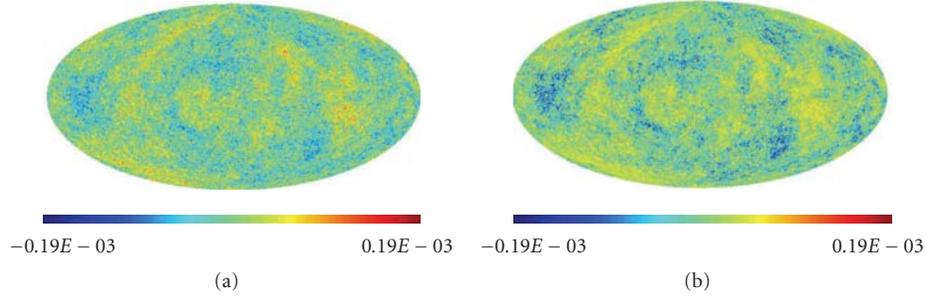


FIGURE 12: (a), (b) A Gaussian realization of the CMB sky (a) and a non-Gaussian local CMB map (b), obtained by adding to the Gaussian one a NG component with  $f_{\text{NL}}^{\text{loc}} = 3000$ , from the study by Liguori et al. in [117]. (b) and (d) A Gaussian CMB map (c) and a non-Gaussian DBI map (d) with  $f_{\text{NL}}^{\text{DBI}} = 4000$ , from [9]. The maps in the upper panel have been obtained using the local algorithm described in Section 3.10.1. The maps in (b) have been produced with the bispectrum algorithm of Section 3.10.2, after having separated the primordial DBI shape using the eigenmode expansion defined in (26).

- (1) To test the unbiasedness of the  $f_{\text{NL}}$  bispectrum estimator (by checking that the Monte Carlo average of the recovered  $f_{\text{NL}}$  reproduces the  $f_{\text{NL}}$  set in input).
- (2) To study how the expected primordial NG signal imprinted in the CMB is modified by the presence of other effects, like those considered in Section 3.1. For example, weak lensing of primordial NG might in principle change the observed bispectrum and affect the estimates. This can be studied again by testing the estimator on NG-lensed simulations, as it was done by Hanson et al. in [108].
- (3) For local NG, to obtain the error bars of the  $f_{\text{NL}}$  estimator if a large  $f_{\text{NL}}$  is detected at several  $\sigma$  (see Section 3.5.4). We have previously seen that for a several-sigma detection of local NG the bispectrum variance is  $f_{\text{NL}}$  dependent. The Monte Carlo average (116) thus has to be evaluated on NG simulations with the measured  $f_{\text{NL}}$  in input.

Unless we are in the situation described at point (3) of the list above, all we need to produce is then maps with given power spectrum and bispectrum, since higher-order correlators can be neglected. In the large local  $f_{\text{NL}}$  case higher-order correlators are instead important and have to be included. Fortunately the local case is the only one for which we have a full expression of the primordial potential  $\Phi(\mathbf{x})$  that allows us to produce exact simulations.

We will divide this section into two parts. In the first we will describe exact simulation algorithms of local NG, while in the second we will describe methods to generate maps with given power spectrum and bispectrum, starting from an arbitrary primordial shape.

**3.10.1. Algorithms for Local Non-Gaussianity.** First of all, let us recall that the CMB multipoles  $a_{\ell m}$  are related to the primordial gravitational potential  $\Phi$  through the well-known formula

$$a_{\ell m} = \int \frac{d^3 k}{(2\pi)^3} \Phi(\mathbf{k}) Y_{\ell m}(\hat{k}) \Delta_{\ell}(k), \quad (139)$$

where  $\Delta_{\ell}(k)$  are the radiation transfer functions and the potential is written in Fourier space. We already met this formula when we calculated the relation between the primordial and CMB bispectrums in Section 3.1. We also recall that the local non-Gaussian primordial potential takes a very simple expression in real space, where

$$\Phi(\mathbf{x}) = \Phi_L(\mathbf{x}) + f_{\text{NL}} [\Phi_L^2(\mathbf{x}) - \langle \Phi_L^2(\mathbf{x}) \rangle]. \quad (140)$$

In the previous expression  $\Phi_L$  is a Gaussian random field, characterized by a primordial power spectrum  $P_{\Phi}(k) = Ak^{n-4}$ ; in the following we will refer to  $\Phi_L(\mathbf{x})$  as the Gaussian part of the primordial potential. The remaining non-Gaussian part of the potential is simply the square of the Gaussian part point-by-point (modulo a constant term, necessary to enforce the condition  $\langle \Phi(\mathbf{x}) \rangle = 0$ ; however it is clear that this term only affects the CMB monopole). It is then convenient to work directly in real space and recast formula (139) in the following form:

$$a_{\ell m} = \int d^3 r \Phi(\mathbf{r}) Y_{\ell m}(\hat{r}) \alpha_{\ell}(r), \quad (141)$$

where  $\alpha_{\ell}(r) \equiv \int dk k^2 j_{\ell}(kr) \Delta_{\ell}(k)$ , also used in (32), is the real-space counterpart of the radiation transfer functions  $\Delta_{\ell}(k)$ ,  $j_{\ell}(kr)$  is a spherical Bessel function, and  $r$  is a look-back conformal distance. This formula suggests to structure an algorithm for the generation of local CMB NG maps in the following steps.

- (1) Generate the Gaussian part  $\Phi_L$  of the potential in a box whose side is the present cosmic horizon.
- (2) Square the Gaussian part point-by-point to get the non-Gaussian part.
- (3) Expand in spherical harmonics the Gaussian and non-Gaussian parts of the potential for different values of the radial coordinate  $r$  in the simulation box.
- (4) Convolve the spherical harmonic expansions of  $\Phi_L$  and  $\Phi_{\text{NL}}$  with the radiation transfer function  $\Delta_{\ell}(r)$  in order to obtain the Gaussian and non-Gaussian parts of the multipoles of the final NG CMB simulation.

For a given choice of the non-Gaussian parameter  $f_{\text{NL}}$ , a CMB map is then obtained simply through the linear combination  $a_{\ell m} = a_{\ell m}^L + f_{\text{NL}} a_{\ell m}^{\text{NL}}$  (the superscripts  $L$  and  $\text{NL}$  are always indicating Gaussian and non-Gaussian, resp.).

The most difficult and time-consuming part in this process is actually the generation of the Gaussian part of the potential  $\Phi$ . One possibility is to generate the Gaussian part of the potential in a cubic box in Fourier space, where different modes are uncorrelated and have variance given by the primordial power spectrum  $P_{\Phi}(k)$ , then apply a Fast Fourier Transform (FFT) algorithm to go to real space. Cartesian coordinates are then transformed into spherical coordinates by means of an interpolation algorithm in order to transform  $\Phi_L(\mathbf{x})$  into  $\Phi_L(r, \hat{n})$ . Finally, the Gaussian potential in spherical coordinates is squared point-by-point to get the NG part, and the spherical harmonic expansion and radiation transfer function convolution at point 4 of the list above are performed in order to obtain the multipoles of the final CMB map. The aforementioned algorithm was implemented by Komatsu et al. in [78] to generate NG local CMB maps at the resolution of the Planck satellite.

The difficulty with this approach arises from the fact that we are working in a box of the size of the present cosmic horizon (about 15 Gpc in conformal time), but at the same time a cell in this box must have a side no bigger than 20 Mpc in order to resolve the last scattering surface, where most of the CMB signal is generated. A more convenient and accurate way to produce the local NG  $a_{\ell m}$  was found in [117, 118]: the idea is to work directly in spherical coordinates, use a nonuniform discretization of the simulation box (since no sample points are needed in a large region of the box where photons are just free streaming, while many sample points are needed at last scattering, as we just pointed out above), and generate the multipoles of the expansion of  $\Phi_L(\mathbf{x})$  through the following two-step approach.

- (1) Generate uncorrelated radial multipoles  $n_{\ell m}(r)$ , Gaussianly distributed and characterized by the following spectrum:

$$\langle n_{\ell_1 m_1}(r_1) n_{\ell_2 m_2}^*(r_2) \rangle = \frac{\delta_D(r_1 - r_2)}{r^2} \delta_{\ell_1}^{\ell_2} \delta_{m_1}^{m_2}, \quad (142)$$

where  $\delta_D$  is the Dirac delta function.

- (2) Filter the multipoles  $n_{\ell m}$  with suitable functions in order to produce a Gaussian random field with the properties of the multipole expansion of the primordial Gaussian potential  $\Phi_L$ . It can be shown that the expression of the filter functions is

$$W_{\ell}(r, r_1) = \frac{2}{\pi} \int dk k^2 \sqrt{P_{\Phi}(k)} j_{\ell}(kr) j_{\ell}(kr_1), \quad (143)$$

where  $P_{\Phi}$  is the primordial curvature power spectrum, and the filtering operation takes the form

$$\Phi_{\ell m}^L(r) = \int dr_1 r_1^2 n_{\ell m}(r_1) W_{\ell}(r, r_1). \quad (144)$$

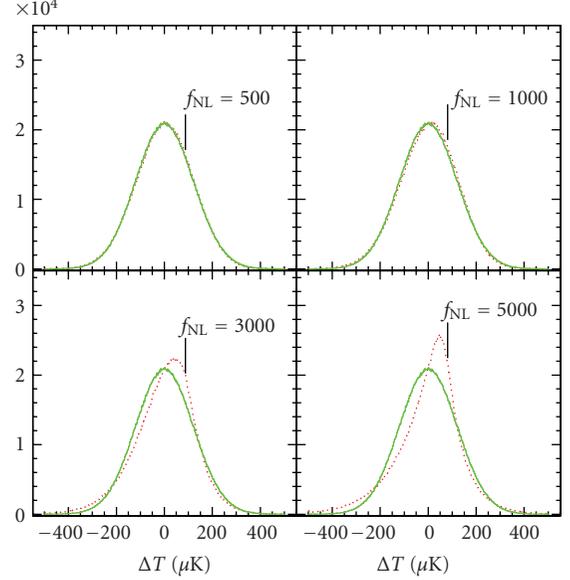


FIGURE 13: Probability density function of temperature pixel from local primordial non-Gaussian CMB maps, obtained with the “exact” simulation algorithm described in Section 3.10.1. Different panels show the result for different values of  $f_{\text{NL}}^{\text{loc}}$ , in order to give an idea of the order of magnitude of the signal that one wants to detect. For  $f_{\text{NL}} < 1000$  the non-Gaussianity is too small to be seen in these plots. Note that WMAP constrains  $f_{\text{NL}}^{\text{loc}}$  to be  $\lesssim 100$ .

In the last expression  $\Phi_{\ell m}^L(r)$  are the desired quantities, that is, the multipoles of the expansion of the Gaussian part of the primordial potential for a given  $r$ .

This algorithm, recently improved by Elsner and Wandelt in [119], was used to produce NG local maps at the resolution of WMAP and Planck in temperature and polarization. An example of its results is shown in Figure 12(a). Figure 13 shows 1-point PDFs of temperature anisotropies for different values of  $f_{\text{NL}}^{\text{loc}}$ , extracted from these simulations.

**3.10.2. Algorithms for Arbitrary Bispectra.** In the limit of weak non-Gaussianity, an algorithm to produce non-Gaussian CMB simulations with a given power spectrum and bispectrum for separable primordial shapes was described by Smith and Zaldarriaga in [22]. In this algorithm the non-Gaussian components of the CMB multipoles are obtained using the following formula:

$$a_{\ell m}^{\text{NG}} = \frac{1}{6} \sum_{\ell_1 \ell_2 \ell_3} B_{\ell_1 \ell_2 \ell_3} \begin{pmatrix} \ell & \ell_2 & \ell_3 \\ m & m_2 & m_3 \end{pmatrix} \frac{a_{\ell_2 m_2}^{G*}}{C_{\ell_2}} \frac{a_{\ell_3 m_3}^{G*}}{C_{\ell_3}}, \quad (145)$$

where  $a_{\ell m}^G$  is the Gaussian part of the CMB multipoles, generated using the angular power spectrum  $C_{\ell}$ , while  $B_{\ell_1 \ell_2 \ell_3}$  is the given bispectrum of the theoretical model for which simulations are required. Note that alternative algorithms to generate CMB maps with given bispectrum have been proposed in the literature [120, 121], but they are less general than the one introduced by (145). Although (145)

is completely general, as before its numerical evaluation is only computationally affordable for bispectra that can be written in separable form. We have emphasized already that separability results in a reduction of the computational cost of the estimator (99) from  $\mathcal{O}(\ell_{\max}^5)$  to  $\mathcal{O}(\ell_{\max}^3)$  operations; the same argument applies here and allows to rewrite (145) into an equivalent form in pixel space. Starting from formula (46), and substituting it in (145), we find that

$$\begin{aligned} a_{\ell m}^{\text{NG}} = & \int dr r^2 \int d\Omega_{\hat{n}} (2X_{\ell}(r)M_Y(r, \hat{n})M_Z(r, \hat{n}) \\ & + 2Y_{\ell}(r)M_X(r, \hat{n})M_Z(r, \hat{n}) \\ & + 2Z_{\ell}(r)M_X(r, \hat{n})M_Y(r, \hat{n})). \end{aligned} \quad (146)$$

As already discussed in the  $f_{\text{NL}}$ -estimator section, the limitation dictated by separability is clearly overcome by using the eigenfunction representations for the bispectrums (26) and (60) introduced by Fergusson et al. in [14]. As usual, the basic idea is to start by expanding an arbitrary bispectrum shape  $S$  (either primordial or in the CMB) using a separable polynomial decomposition until a good level of convergence is achieved and then to substitute the mode decomposition into (145) to get a linear combination of numerically tractable terms written in the form of (146). Using the separable mode coefficients  $\alpha_{prs}$  for the reduced bispectrum (56) and the filtered map expressions  $M_p(r, \hat{n})$  (109) as the starting point, we find that the expression (146) generalises to

$$\begin{aligned} a_{\ell m}^{\text{NG}} \\ = \frac{1}{18} \sum_{prs} \alpha_{prs} \int dx x^2 q_p^l(x) \int d\Omega_{\hat{n}} Y_l^{m*}(\hat{n}) M_r^G(r, \hat{n}) M_s^G(r, \hat{n}), \end{aligned} \quad (147)$$

where the  $M_p^G(r, \hat{n})$  are found by summing using a set of Gaussian  $a_{\ell m}^G$ 's convolved with the  $q_p^l$ 's (refer to (107)):

$$M_p^G(\hat{n}, x) = \sum_{lm} q_p^l \frac{a_{lm}^G Y_{lm}}{C_l}. \quad (148)$$

Here, the accuracy of convergence with  $\alpha_{prs}$  is parametrized in terms of the correlation  $\mathcal{C}(S, S_N)$  between the original nonseparable shape and the eigenmode expansion, as defined previously (19). Note that this convergence can also be checked more accurately using the full Fisher matrix correlation on the CMB bispectra  $\mathcal{C}(b_{\ell_1 \ell_2 \ell_3}, b_{\ell_1 \ell_2 \ell_3}^N)$ , described in Sections 3.8 and 3.9.

In addition to the bispectrum separability requirement, there is an important further caveat which can prevent the straightforward implementation of the algorithm (145). By construction, terms  $\mathcal{O}(f_{\text{NL}}^2)$  and higher are not explicitly

controlled. Following the discussion in [108], we can write the connected N-point functions as

$$\langle a_{\ell_1 m_1}^* a_{\ell_2 m_2} \rangle = [C_{\ell_1} + f_{\text{NL}}^2 C_{\ell_1}^{\text{NG}}], \quad (149)$$

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = [f_{\text{NL}} B_{\ell_1 \ell_2 \ell_3} + \mathcal{O}(f_{\text{NL}}^3)], \quad (150)$$

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \dots a_{\ell_N m_N} \rangle = \mathcal{O}(f_{\text{NL}}^3). \quad (151)$$

Thus the condition that the map has the power spectrum  $C_l$  specified in input will only be satisfied if the power spectrum of the non-Gaussian component in (149) remains small. Since this method does not control  $\mathcal{O}(f_{\text{NL}}^2)$  terms, one has to ascertain that spuriously large  $C_l^{\text{NG}}$  contributions do not affect the overall power spectrum significantly. It turns out that this effect plagues current map simulations if the standard separable expressions for the local and equilateral bispectra are directly substituted into (145). However a slight modification of (145), described by Hanson et al. in [108] and Fergusson et al. in [14], allows us to overcome this problem at no computational cost. Moreover, it was shown by Fergusson et al. [14] that maps obtained from the eigenmode expansions (26) and (60) are stable independently of the shape under study, thus making this map-making generating algorithm robust and fully general. Examples of DBI NG maps produced by combining the eigenmode expansion method with the map-making algorithm described in this section are shown in Figure 12(b).

#### 4. Large-Scale Structure

In the standard scenario, early perturbations produced during inflation are responsible for the common origin of the CMB temperature fluctuations and the large-scale matter and galaxy distributions in the Universe, that is, the large-scale structure. The Cosmic Microwave Background provides a remarkable example of a Gaussian random field in nature. Information on cosmological parameters is in fact derived from measurements of its power spectrum, the  $C_l$ 's, while bispectrum measurements from WMAP data remain consistent with zero. The distribution of matter, as we can infer today from shear or galaxy observations, unlike the CMB, can be described as a highly non-Gaussian random field, *even* for Gaussian initial conditions.

The matter overdensity  $\delta(\mathbf{x})$  is defined in terms of the matter density  $\rho(\mathbf{x})$  and its mean value  $\bar{\rho}$  by

$$\delta(\mathbf{x}) \equiv \frac{\rho(\mathbf{x}) - \bar{\rho}}{\bar{\rho}}, \quad (152)$$

with zero mean by construction. Such quantity assumes, at late times, the limiting value  $\delta = -1$  in voids, accounting for a large fraction of the volume of the Universe, while it achieves values  $\delta \gg 1$  in collapsed objects such as dark-matter halos. Its probability distribution function is therefore expected, at low redshift, to depart strongly from a Gaussian distribution centred at  $\delta = 0$ , even though it could be well approximated by it at decoupling, when perturbations around  $\delta = 0$  were of the order of  $\delta \sim 10^{-5}$ . Such

non-Gaussianity is the result of the nonlinear evolution of structures subject to *gravitational instability*.

In addition, nonlinearities in the bias relation between the galaxy and matter distributions constitute a *second source of non-Gaussianity* in the large-scale structure mapped out by redshift surveys. Non-Gaussian initial conditions would therefore provide a third component in the non-Gaussianity of the galaxy distribution. The question regarding the detection of effects due to *primordial* non-Gaussianity, is therefore strictly related to our ability to *distinguish* between these different contributions and, ultimately, it will depend on the robustness of our theoretical predictions in the linear and mildly nonlinear regimes. From this respect, cosmological Perturbation Theory (PT), and its more recent developments, is very important for providing the tools to study the evolution of non-Gaussianities and how to differentiate their origin.

Considering only the *matter* distribution, the leading-order prediction in standard PT for the matter bispectrum at large scales is given by the *sum* of a primordial component and a component due to gravitational instability, which is present also for Gaussian initial conditions. Until fairly recently it was assumed that this picture could be easily extended to the galaxy distribution, with the galaxy bispectrum receiving an additional contribution due to nonlinear bias. Following the historical development of the subject, in Section 4.1 we will discuss early work on higher-order *moments* of the matter and galaxy distribution, starting with the *skewness*. We will then consider in Section 4.2 the matter bispectrum and its description in the *Eulerian* perturbation theory, with specific attention given to effects at large scales due to a primordial component, as well as at small-scale, nonlinear corrections in presence of non-Gaussian initial conditions. Here, most of the theoretical results on higher-order correlation functions are developed. In Section 4.3, we will deal with the galaxy bispectrum. We will first introduce the simple model based on local bias and discuss problems related to bispectrum measurements in redshift surveys with specific attention given to the detection of primordial non-Gaussianity. We will see how early results indicated that the galaxy bispectrum could be used as a tool to constrain non-Gaussian initial conditions which is, in principle, *competitive* with the CMB, illustrating this with actual results from current datasets. We will then consider the outcome of recent N-body simulations with non-Gaussian initial conditions showing that the simple prediction for the galaxy bispectrum assumed in most of the previous literature on the subject fails to describe not only the measured halo bispectrum, but even the halo *power spectrum*, even at large scales! We now know that correlators of biased populations such as galaxies and dark-matter halos receive large corrections, at large scales, from local primordial non-Gaussianity. These results opened up new and promising opportunities for detection in future large-scale structure observations. Although, in our view, a proper understanding of these effects remains to be adequately developed at the time of writing, particularly with respect to higher-order galaxy correlation functions, we will describe the different descriptions proposed so far in the literature and the prospects for detection of

primordial non-Gaussianity in measurements of the galaxy bispectrum.

From a historical perspective, non-Gaussian initial conditions have been studied for quite a long time. For instance, early works on the clustering of density peaks and rare objects can be found in the study of Grinstein and Wise in [122], Lucchin and Matarrese in [123], Matarrese et al. in [124], while early N-body simulations with non-Gaussian initial conditions go back to the early eighties [125–129]. In the early days, a large variety of non-Gaussian models, often defined in terms of a nonlinear transformation of a Gaussian field, were considered. In some cases, large non-Gaussian components were studied because, on one hand, they could be used to falsify some models and, on the other, as a way to reconcile contradictory observational results with theoretical frameworks. In this review, however, we will consider only models predicting small departures from Gaussian initial conditions which are consistent with CMB observations.

While we focus in this review on direct bispectrum measurements, it should be stressed that the effects of primordial non-Gaussianity on large-scale structure are not limited to corrections to its higher-order correlation functions. Aside from the recent results on the galaxy power spectrum mentioned above, significant departures from Gaussian initial conditions are expected to have important effects on the halo mass function and therefore on the observed cluster number density. See Section 2.1 in the study of Sefusatti et al. in [130] for a brief overview of previous work and the studies of Afshordi and Tolley in [131], Dalal et al. in [132], Desjacques et al. in [133], Fedeli et al. in [134], Grossi et al. in [135], Lam and Sheth in [136], Lo Verde et al. in [137], Maggiore and Riotto in [138], Oguri in [139], Pillepich et al. in [140], and Valageas in [141], for recent theoretical and N-body results. In addition, the corresponding effect on the abundance of voids has been studied by Kamionkowski et al. [142], while the possibility of constraining primordial non-Gaussianity from measurements of Minkowski Functionals in large-scale structure has been explored by Hikage et al. [143, 144]. Further effects on the intergalactic medium and reionization [145, 146] or on future 21 cm observations [147, 148] have also been investigated. We refer the reader to other papers in this issue for a more complete discussion of these alternative approaches.

Finally, we should warn the reader that this section will *not* discuss analytical tools for the estimation of the non-Gaussian parameters corresponding or comparable to those described in the previous section for the CMB bispectrum, with the simple reason been that such tools have not been developed yet! In the first place, the physics of the CMB is simpler in the sense that the bispectrum of the temperature fluctuations at large scales is expected to provide the direct measurement of the initial bispectrum of the curvature perturbations, while the large-scale galaxy distribution is characterized, as mentioned above, by additional sources of non-Gaussianity for which we do not even have, at the moment, a proper model. In the second place, the optimal estimator for the  $f_{\text{NL}}$  parameter presented in Section 3.5 has been developed over several years to tackle the data provided

by the WMAP satellite, which represented, so far, the best test of the Gaussianity of the initial conditions. The analysis of the galaxy bispectrum, on the other hand, did not have such timely and compelling motivations as it was understood, up until a couple of years ago, that large-scale structure observations will be able to provide results comparable to the CMB only in future, large-volume redshift surveys. We hope, nevertheless, that this review might provide a starting point for the development of a proper  $f_{\text{NL}}$  estimator from large-scale structure bispectrum measurements, possibly taking advantage of the techniques already introduced in the context of CMB observations.

**4.1. The Skewness.** This section serves as a brief historical overview. Since the first large-scale observations did not allow an accurate determination of individual bispectrum or trispectrum configurations, most of the attention in the early literature focused on the moments of the galaxy distribution, and, in the first place, on the third- and fourth-order moments, that is, the *skewness* and *kurtosis*, respectively. The “normalized” moment of order  $p$  can be defined in terms of the smoothed density field  $\delta_R(\mathbf{x})$  as

$$s_{p,R} \equiv \frac{\langle \delta_R^p(\mathbf{x}) \rangle_c}{\langle \delta_R^2(\mathbf{x}) \rangle^{p/2}}, \quad (153)$$

where the subscript “c” indicates the *connected* correlations. For Gaussian initial conditions, a perturbative treatment of the equations of gravitational instability predicts at leading order [149]

$$s_{3,R} = \frac{34}{7} \sigma_R, \quad (154)$$

with  $\sigma_R^2 = \langle \delta_R^2 \rangle$ , computed in linear theory. Notice that we are neglecting here, for simplicity, additional and relevant contributions due to the smoothing of the density field (see [150, 151]). When non-Gaussian initial conditions are present, one expects an extra contribution to the skewness, typically with a different relation with  $\sigma_R$ , whose value depends on the non-Gaussian model. Comparisons between the second- and third-order moments,  $S_{3,R}$  and  $\sigma_R$ , (as well as higher-order moments such as the kurtosis) measured in redshift surveys have been early recognized as a tool to test the Gaussianity of primordial perturbations [125, 152–158]. These works recognized as well the importance of reliable predictions in the nonlinear regime and of a proper modeling of the effects of galaxy bias. In this respect, Fry and Scherrer [154] proposed a more quantitative prediction for the contribution to the galaxy skewness due to galaxy bias-based perturbation theory and on the local bias expansion of Fry and Gaztañaga [159]. They derived, for the skewness of the galaxy distribution, an expression of the form

$$s_{3,R} = s_{3,R}^{(0)} + \frac{34}{7} \sigma_R + \frac{6b_2}{b_1} \sigma_R, \quad (155)$$

where we assumed non-Gaussian initial conditions described by a non-vanishing initial skewness  $s_{3,R}^{(0)}$  (but vanishing higher-order moments) and where  $b_1$  and  $b_2$  represent constant bias parameters typical of the galaxy population (which we will discuss explicitly in Section 4.3). This relatively simple expression describes the skewness measured in galaxy surveys, as the sum of *three components corresponding to three sources of non-Gaussianity for the galaxy distribution*: one primordial, one due to gravitational instability, and the last due to nonlinear bias. Further studies in perturbation theory can be found in [160, 161] while an alternative derivation of the smoothed moments of the density field based on the spherical collapse model has been studied in [162]. The skewness predicted by texture models has been studied in simulations as a function of the smoothing scale  $R$  by Gaztanaga and Mähönen [163] and compared to measurements of the same quantities in the APM Galaxy Survey as in the study by Gaztañaga [164]; see Figure 14. The differences between the  $s_{3,R}$  in the non-Gaussian texture model with respect to the Gaussian case provide a qualitative example of the typical effects that we expect for non-Gaussian initial conditions as a function of the smoothing scale  $R$  and redshift. On the other hand, it should be kept in mind that early works focused on models of primordial non-Gaussianity characterized by a scaling of higher-order correlation functions quite different from the one induced by the  $f_{\text{NL}}$  parametrization.

The measured skewness, as higher-order moments, corresponds to a single number. Despite the possibility to study its peculiar dependence on the smoothing scale  $R$ , it is nevertheless difficult to separate the different components, particularly with respect to bias effects. However, this possibility is offered in principle by direct measurements of the galaxy bispectrum, relying on its dependence on the shape of triangular configurations. In the next sections we will discuss in details first the bispectrum of the *matter* distribution then the bispectrum of the *galaxy* distribution, a direct observable in redshift surveys.

**4.2. The Matter Bispectrum.** In this review we will focus on the predictions for correlation functions in the Fourier space from the Eulerian Perturbation Theory (PT). This approach solves perturbatively the equations for the matter density and velocity field evolution governed by gravitational instability. These are the continuity equation, the Euler equation, and the Poisson equation relating the matter density and the gravitational potential. In the PT framework, the relation between the results and the initial conditions, given in terms of the initial correlators of the density field, is particularly transparent. Moreover, recent works have significantly extended, as we will discuss later, the predicting power of this specific tool. Different approaches are also available: see, for instance, the study by Scoccimarro in [165] for a comparison between bispectrum measurements in N-body simulations and predictions in the Lagrangian Perturbation Theory. We refer the reader to the study by Bernardeau et al. in [70] for a comprehensive review of cosmological perturbation theory of the large-scale structure.

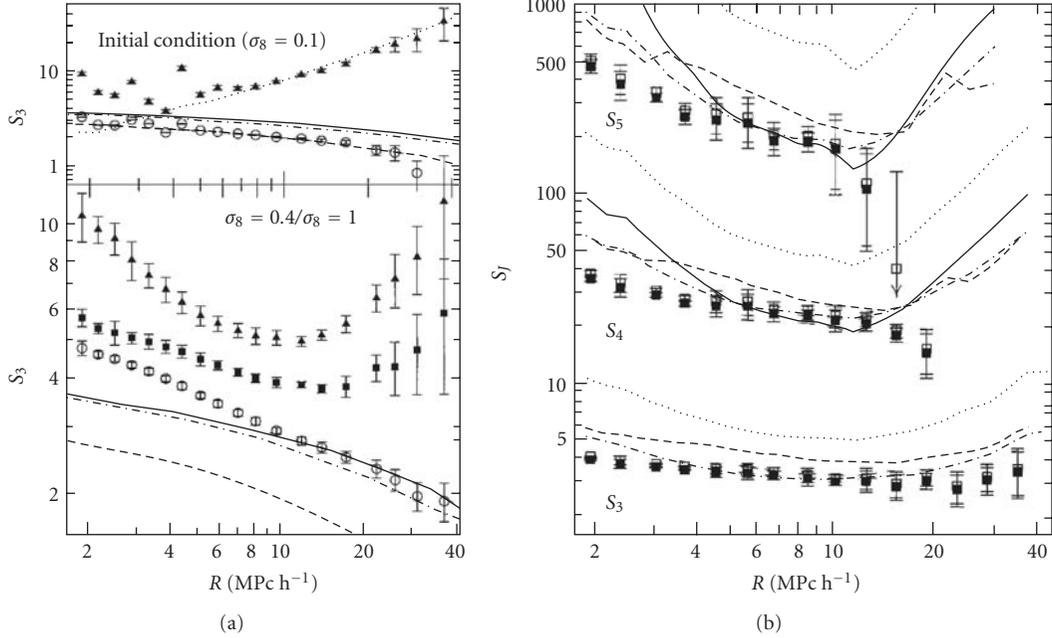


FIGURE 14: (a) Measurements of the skewness of the matter distribution in N-body simulations as a function of the smoothing scale  $R$  for the non-Gaussian texture model (*filled symbols*) and for Gaussian initial conditions (*open circles*). The time evolution is parametrized by the value of  $\sigma_8$ , with the non-Gaussian results shown at  $\sigma_8 = 0.4$  (*triangles*) and  $\sigma_8 = 1$  (*squares*). Lines show different theoretical predictions. (b) Measurements of the third-, fourth- and fifth-order moments of the galaxy distribution in the APM Galaxy Survey, compared with the simulation results with non-Gaussian initial conditions with different bias assumptions, from the study by Gaztanaga and Mähönen in [163] (see the reference for further details).

**4.2.1. Leading-Order Results in Perturbation Theory.** As mentioned before, we consider specifically models where non-Gaussian initial conditions are completely given in terms of the correlators of the curvature perturbations at early times, and the mechanism responsible for the *extra* non-Gaussian properties of the density field is not active during the subsequent evolution of matter perturbations, governed only by gravitational instability. In PT, the solution for the evolved matter density contrast is expressed as a series of corrections to the linear solution  $\delta^{(1)}$  [166]:

$$\delta_{\mathbf{k}} = \delta_{\mathbf{k}}^{(1)} + \delta_{\mathbf{k}}^{(2)} + \delta_{\mathbf{k}}^{(3)} + \dots, \quad (156)$$

where each term can be written formally as (From now on, we will adopt a different convention for the Fourier transform with respect to the one used for the formulae in previous section. The present convention is more common in the large-scale structure literature and conforms with the one adopted in the classical paper by Bernardeau et al. in [70].)

$$\delta_{\mathbf{k}}^{(n)} \equiv \int d^3\mathbf{q}_1 \dots d^3\mathbf{q}_n F_n(\mathbf{q}_1, \dots, \mathbf{q}_n) \delta_{\mathbf{q}_1}^{(1)} \dots \delta_{\mathbf{q}_n}^{(1)}, \quad (157)$$

with  $F_n(\mathbf{q}_1, \dots, \mathbf{q}_n)$  representing the symmetrized  $n$ -order kernel in PT. The initial conditions in the Gaussian case are completely specified by the linear power spectrum  $P_0(k)$ , with  $\langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \rangle = \delta_D(\mathbf{k}_{12}) P_0(k_1)$ , where we adopt the notation  $\mathbf{k}_{ij} \equiv \mathbf{k}_i + \mathbf{k}_j$ . Non-Gaussian initial conditions are described, in the first place, by a nonzero expression for the three-point function of the linear solution, that is,  $\langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(1)} \rangle$ . In turn,

the *initial* matter correlators, that is, the correlators of the *linear* solution  $\delta^{(1)}$ , are given in terms of the correlators of the curvature perturbations as

$$\langle \delta_{\mathbf{k}_1} \dots \delta_{\mathbf{k}_n} \rangle = M(\mathbf{k}_1, z) \dots M(\mathbf{k}_n, z) \langle \Phi_{\mathbf{k}_1} \dots \Phi_{\mathbf{k}_n} \rangle, \quad (158)$$

where we introduce the function

$$M(k, z) = \frac{2}{3} \frac{k^2 T(k) D(z)}{\Omega_m H_0^2}, \quad (159)$$

with  $T(k)$  being the matter transfer function and  $D(z)$  the growth factor, expressing Poisson's equation in the Fourier space as

$$\delta_{\mathbf{k}}(z) = M(k, z) \Phi_{\mathbf{k}}. \quad (160)$$

Notice that we denote with  $\Phi$  the *primordial* curvature perturbations, that is, evaluated during the matter-dominated era, not their value linearly extrapolated at present time. (This choice, not unique in the literature, is particularly convenient since curvature perturbations are constant during matter domination. Also, it conforms to the definition of  $f_{\text{NL}}$  in terms of  $\Phi$  assumed in the CMB literature on observational constraints and specifically in the study by Komatsu and Spergel in [74].) The linear, that is, *initial*, power spectrum is given by

$$P_0(k) = M^2(k, z) P_{\Phi}(k), \quad (161)$$

while the initial bispectrum and trispectrum are

$$\begin{aligned} B_0(k_1, k_2, k_3) &= M(k_1)M(k_2)M(k_3)B_\Phi(k_1, k_2, k_3), \\ T_0(k_1, k_2, k_3, k_4) &= M(k_1)M(k_2)M(k_3)M(k_4)T_\Phi(k_1, k_2, k_3, k_4). \end{aligned} \quad (162)$$

Notice that, given these simple relations between curvature and primordial matter correlators, issues such as the property of separability discussed in Section 2.2 for the CMB bispectrum are not present in the case of three-dimensional, large-scale structure observables.

The nonlinear power spectrum is obtained perturbatively from the expansion

$$\begin{aligned} \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \rangle &= \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \rangle + \left( \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(2)} \rangle + \text{perm.} \right) + \langle \delta_{\mathbf{k}_1}^{(2)} \delta_{\mathbf{k}_2}^{(2)} \rangle \\ &+ \left( \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(3)} \rangle + \text{perm.} \right) + \dots, \end{aligned} \quad (163)$$

where the term  $\langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \rangle$  corresponds to the linear solution,  $P_0(k)$ , while the other terms represent, in analogy with perturbation theory in quantum field theory, *one-* and *higher-loop corrections* as they involve integrations over internal momenta. In particular, the term  $\langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(2)} \rangle$  vanishes for Gaussian initial conditions as it depends on the initial bispectrum  $B_0$  (see [167] for an analysis of nonlinear corrections to the matter power spectrum *due to* primordial non-Gaussianity).

In a similar fashion, nonlinear corrections in (156) provide a perturbative expansion for the matter bispectrum:

$$\begin{aligned} \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \delta_{\mathbf{k}_3} \rangle &= \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(1)} \rangle + \left( \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(2)} \rangle + \text{perm.} \right) \\ &+ \left( \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(2)} \delta_{\mathbf{k}_3}^{(2)} \rangle + \text{perm.} \right) + \left( \langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(3)} \rangle + \text{perm.} \right) + \dots \end{aligned} \quad (164)$$

In this case, the leading-order contributions are given by the *tree-level* terms  $\langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(1)} \rangle$  and  $\langle \delta_{\mathbf{k}_1}^{(1)} \delta_{\mathbf{k}_2}^{(1)} \delta_{\mathbf{k}_3}^{(2)} \rangle$ , with the first being the initial component and the second corresponding to a contribution to the matter bispectrum due to *gravity alone*, of the form

$$B_G^{\text{tree}}(k_1, k_2, k_3) = 2F_2(\mathbf{k}_1, \mathbf{k}_2)P_0(k_1)P_0(k_2) + 2 \text{ perm.} \quad (165)$$

Notice that this contribution is present even for Gaussian initial conditions as it depends only on the initial power spectrum  $P_0$  and describes the emergence of non-Gaussianity due to gravitational instability. The leading-order, *tree-level* expression of the matter bispectrum with non-Gaussian initial conditions is therefore given in terms of the *sum*

$$B^{\text{tree}}(k_1, k_2, k_3) = B_0(k_1, k_2, k_3; z) + B_G^{\text{tree}}(k_1, k_2, k_3; z). \quad (166)$$

This expression corresponds to the first two terms on the r.h.s. of (155) for the skewness, which can be obtained from (166) by integration. For instance, the contribution to the

skewness induced by gravity, in the unsmoothed case, is obtained from  $B_G^{\text{tree}}$  as

$$\begin{aligned} \langle \delta^3 \rangle &= \int d^3 k_1 d^3 k_2 B_G^{\text{tree}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \\ &= 6 \int d^3 k_1 d^3 k_2 P(k_1)P(k_2)F_2(\mathbf{k}_1, \mathbf{k}_2), \end{aligned} \quad (167)$$

and, integrating over the angles,

$$\langle \delta^3 \rangle = \frac{34}{7} \left[ 4\pi \int dk k^2 P(k) \right]^2 = \frac{34}{7} \langle \delta^2 \rangle^2, \quad (168)$$

corresponding to the result of (154).

The possibility of distinguishing the primordial component  $B_0$  from the gravity-induced one  $B_G$  relies on their specific and distinct dependence on *scale*, on the *triangular configuration*, *shape* and on *redshift*. For a primordial non-Gaussianity described by a curvature bispectrum obeying the hierarchical scaling  $B_\Phi \sim P_\Phi^2$ , typical of weakly non-Gaussian models such as the local and equilateral ones, the different redshift and scale dependence of the two contributions are evident in their ratio for *equilateral triangles* ( $k_1 = k_2 = k_3 = k$ ), given by (The first equality is in fact *identical* for local, equilateral, and orthogonal non-Gaussianity, simply by definition of the equilateral bispectrum, (53), introduced in [8] and of the orthogonal bispectrum introduced in [18], where  $f_{\text{NL}}^{\text{eq.}}$  and  $f_{\text{NL}}^{\text{orthog.}}$  are precisely the amplitudes that provide the same value for the curvature bispectrum as the local model for equilateral configurations.)

$$\frac{B_0(k, k, k; z)}{B_G^{\text{tree}}(k, k, k; z)} = \frac{7}{4} \frac{f_{\text{NL}}}{M(k; z)} \underset{k \rightarrow 0}{\sim} \frac{f_{\text{NL}}}{k^2 D(z)}. \quad (169)$$

We therefore expect, for a wide range of non-Gaussian models, the initial contribution  $B_0$  to be larger at large scales and at high redshift. Figure 15(a) shows the two contributions and their sum for equilateral configurations  $B(k, k, k)$  as a function of  $k$ . Moreover Figure 15(b), 15(c), and 15(d) show the effect of the primordial component for different non-Gaussian models, for values of the respective parameters  $f_{\text{NL}}$  corresponding to the current 95% C.L. limits [18, 68] and with the shaded area indicating the allowed region.

In addition,  $B_G^{\text{tree}}$  presents a specific dependence on triangle shapes, determined by gravitational instability and described by (165) at tree level. The shape dependence of  $B_0$ , determined by the specific non-Gaussian model under consideration, is generically different. Such differences can be explicitly shown in plots of the *reduced* bispectrum, defined as

$$Q(k_1, k_2, k_3) = \frac{B(k_1, k_2, k_3)}{P(k_1)P(k_2) + 2 \text{ perm.}}, \quad (170)$$

which removes the redshift and scale dependencies of the gravity contribution. Figure 16 shows the reduced bispectrum  $Q(k_1, k_2, k_3)$  at tree-level in perturbation theory, at  $z = 1$  for  $k_1 = 0.01 \text{ hMpc}^{-1}$ ,  $k_2 = 1.5k_1$  as a function of the angle  $\theta$  between  $k_1$  and  $k_2$ . In all panels, the continuous

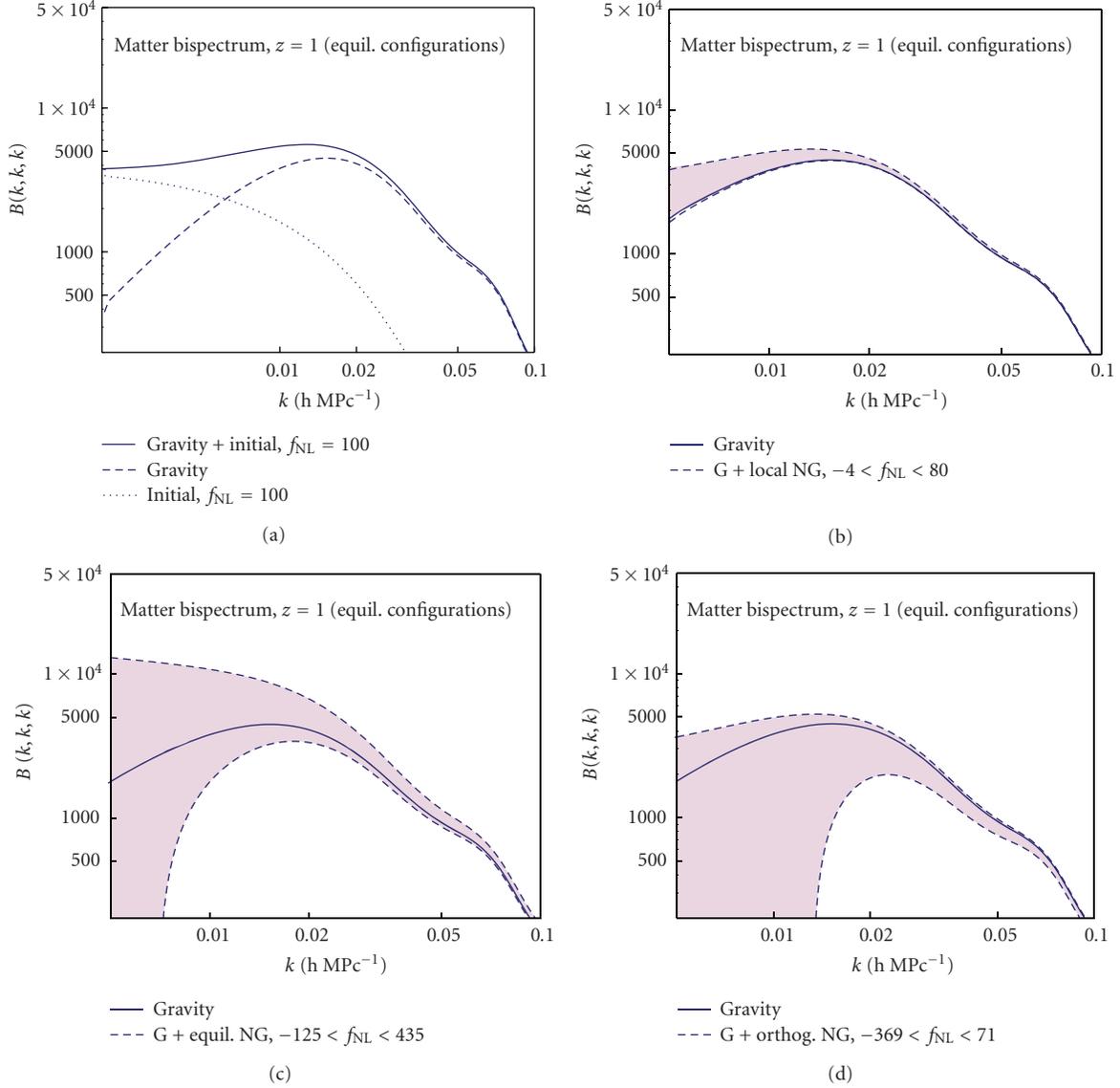


FIGURE 15: Effect of the primordial component for different non-Gaussian models on the equilateral configurations of matter bispectrum,  $B(k, k, k)$ , at redshift  $z = 1$ , as a function of scale, at tree level in PT. In (a) the continuous line shows the initial component  $B_0$  (dotted line), the gravity-induced component,  $B_G^{\text{ecc}}$  (dashed line) and their sum (continuous line). For equilateral configurations the initial component coincides for the local, equilateral, and orthogonal models while it vanishes in the folded model. In (b), (c), and (d), continuous lines show the gravity component alone while dashed lines show the tree-level bispectrum including the primordial component for the local (b), equilateral (c), and orthogonal (d) models assuming the values of  $f_{\text{NL}}$  corresponding to the 95% C.L. limits as determined by Smith et al. [68] and Senatore et al. [18] from WMAP observations. The shaded area indicates the currently allowed region.

line represents the gravity-induced term which assumes larger values for nearly collapsed triangles, that is, for  $\theta \simeq 0$  or  $\pi$ . This indicates that the *probability* of finding larger values for the matter density in triplets of points forming a squeezed or folded triangle is larger than that for nearly equilateral triangles. This prediction is confirmed by the typical *filamentary* nature of the large-scale structure, evident from snapshots of N-body simulations or images of redshift surveys, since along these filaments it is easier to form collapsed triangles than equilateral ones. It should be stressed that the bispectrum is, in fact, the lowest-order

statistic sensitive to the three dimensionality of structures and that these features are not captured by the information contained in the power spectrum alone. The effects of the primordial component on the matter bispectrum are shown by the dashed lines which correspond, as in Figure 15, to the  $2\text{-}\sigma$  limits from CMB observations, in the case of the local (a), equilateral (b) and orthogonal (c) models while they correspond to the values  $f_{\text{NL}} = \pm 300$  in the folded case, for which no experimental bounds are available. Although the large scales  $k_1 = 0.01 \text{ hMpc}^{-1}$  and  $k_2 = 0.015 \text{ hMpc}^{-1}$  and the relatively high redshift  $z = 1$  have been chosen to enhance

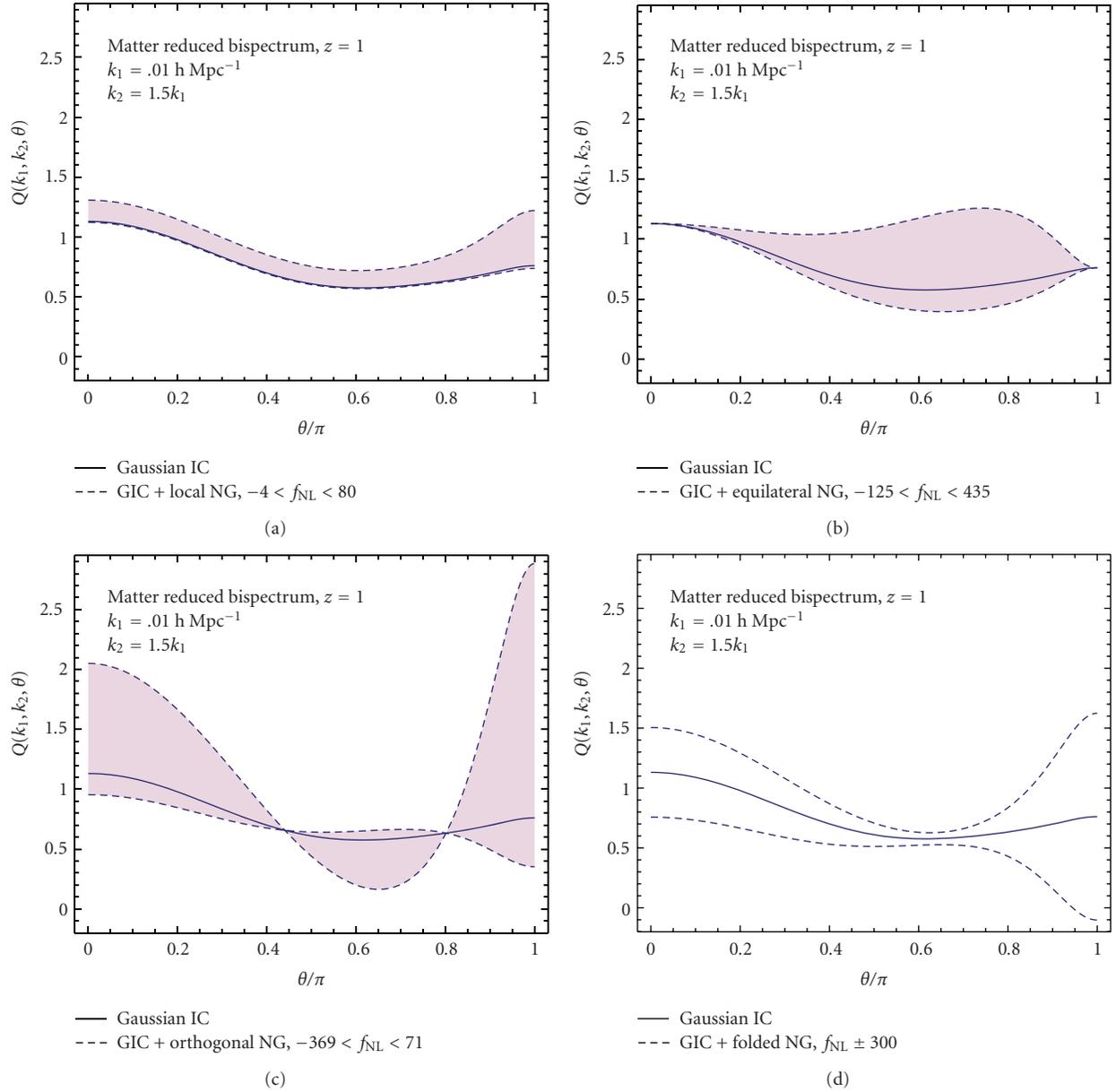


FIGURE 16: Effect of the primordial component for different non-Gaussian models on the matter reduced bispectrum, as a function of the triangle shape. The continuous line shows the reduced bispectrum  $Q(k_1, k_2, k_3)$  at tree level in PT for Gaussian initial conditions at redshift  $z = 1$  assuming  $k_1 = 0.01 \text{ h Mpc}^{-1}$ ,  $k_2 = 1.5 k_1$  as a function of the angle  $\theta$  between  $k_1$  and  $k_2$ . Dashed lines show the reduced bispectrum including the primordial component for the local (a), equilateral (b), orthogonal (c), and folded (d) models. For the local, equilateral, and orthogonal models we assume the values of  $f_{\text{NL}}$  corresponding to the 95% C.L. limits as determined by Smith et al. [68] and Senatore et al. [18] from WMAP observations. The shaded area indicates the currently allowed region. For the folded model, for which no observational constraints are available, the values  $f_{\text{NL}} = \pm 300$  are considered.

the effect of the non-Gaussian component, these triangles are not completely out of reach for future, large-volume surveys. Primordial non-Gaussianity modifies, in very specific ways, the shape dependence of the matter bispectrum produced by gravitational instability.

While the dependence of the matter bispectrum on scale and redshift is responsible for the specific behavior of the skewness of the matter density field on the smoothing scale  $R$  and redshift, the sensitivity to the triangle shape is completely

lost in analysis of the density higher-order moments. Instead, accurate measurements of the bispectrum, when achievable, offer in principle the possibility to disentangle the different contributions when triangles of different size *and* shape are included in the analysis.

The matter bispectrum is not, unfortunately, a direct observable. While we will discuss later how the statistical properties of the matter distribution can be inferred from galaxy redshift surveys, we should mention that the shear

field in weak lensing surveys is another observable directly related to the matter distribution. The observational consequences on the *weak lensing* bispectrum, of a primordial non-Gaussian component (of the *local* type) such as the one in (166), have been explored by Takada and Jain in [168]. The authors find that the primordial component *alone* (i.e., *without* contamination from the gravitational one) could be detected if  $f_{\text{NL}}^{\text{loc}} > 150 f_{\text{sky}}^{1/2}$ , assuming  $l_{\text{max}} \simeq 500$  and a tomography over four redshift bins for a galaxy number density of  $\bar{n}_g = 100 \text{ arcmin}^{-2}$ . The large cosmic variance for low  $\ell$ 's makes difficult the detection of the primordial component, prominent instead at larger scales. As we will see in the next section, primordial non-Gaussianity has some effect on small scales as well, due to the nonlinear evolution of structures.

**4.2.2. Second-Order Corrections.** The simple prediction of (166) for the matter bispectrum is expected to be valid at the largest observable scales and at high redshift, where nonlinear evolution is subdominant. Despite the fact that such conditions correspond as well to the regime where a detection of the initial component  $B_0$  is favored, the effects of non-Gaussian initial conditions can be significant even at smaller scales and at low redshift. Since these effects are the result of nonlinear gravitational evolution *and* non-Gaussian initial conditions, it is no longer possible to identify distinct contributions resulting from distinct sources of non-Gaussianity, as it is the case for the tree-level expression of (166). Nevertheless, it is possible to distinguish individual corrections in PT to the matter bispectrum depending exclusively on the initial power spectrum  $P_0$ , and therefore present as well for Gaussian initial conditions, and corrections depending instead on higher-order initial correlators, such as the initial bispectrum  $B_0$  and trispectrum  $T_0$ , which can be interpreted as small-scale effects due to non-Gaussian initial conditions. One-loop corrections in PT for Gaussian initial conditions have been studied by Scoccimarro in [169] and Scoccimarro et al. in [170], while the extension of these results to non-Gaussian initial conditions is studied by Sefusatti in [171].

A comparison of these results with measurements of the matter bispectrum in N-body simulations [133] with non-Gaussian initial conditions of the local kind can be found in the study by Sefusatti et al. in [172]. Figure 17 shows the equilateral configurations of the matter bispectrum measured in N-body simulations together with predictions from perturbation theory at tree level (*dashed line*) and one loop (*continuous line*). In particular, Figure 17(a) considers  $B(k, k, k)$  for Gaussian initial conditions while Figure 17(b) shows the same quantity divided by the tree-level prediction in PT to highlight the small-scales nonlinear behavior. Figures 17(c) and 17(d) show, respectively, the ratio and the difference between the matter bispectrum with an initial local component corresponding to  $f_{\text{NL}} = 100$  and the Gaussian case. The agreement between one-loop predictions and the simulations results is quite remarkable, while we notice that the tree-level prediction fails to accurately describe the effect of primordial non-Gaussianity already at relatively large scales.

The significance of these relatively small corrections to individual configurations is to be considered in relation to the much larger number of configurations that can be measured as we include smaller and smaller scales, and they could lead to a measurable effect when considered in terms of the cumulative signal-to-noise ratio. On the other hand, these effects loose in part the shape dependence of the original initial bispectrum and require an accurate model (perhaps beyond standard perturbation theory) and strong priors on the underlying cosmological parameters to be distinguished from the nonlinear, “Gaussian” component. A step in the direction of improved predictions is offered by the promising results of the Renormalized Perturbation Theory [173–175] and of the Renormalization Group approach [176, 177]. The extension of the latter to the case of non-Gaussian initial conditions has been recently considered by Bartolo et al. in [178], which studies specific predictions for the matter power spectrum and bispectrum.

**4.3. The Galaxy Bispectrum.** From the discussion above, we could expect that future, *large-volume* and *high-redshift* galaxy surveys will be able to *directly* detect a possible, large primordial component to the matter bispectrum by measurements of the galaxy bispectrum, or at least provide constraints on the non-Gaussian parameters *comparable* to the constraints from measurements of the CMB bispectrum. Such an expectation is motivated by the simple observation that the number of *Fourier modes* available in a *three-dimensional*, ideal, all-sky galaxy survey is in principle much larger than the number of modes available in *two-dimensional* CMB maps.

The galaxy distribution is, however, a less direct probe of the early Universe than the CMB temperature fluctuations. On top of the nonlinear evolution of structures and its contribution to higher-order correlation functions, one has to take into account the *nonlinear* nature of *galaxy bias*, being itself responsible for additional non-Gaussianity. An analysis of the galaxy bispectrum should therefore be able to detect a small primordial component by separating it from these primary contributions.

In this respect, an even more complex picture, due to additional and somehow unexpected effects of primordial non-Gaussianity *on* galaxy bias, has been emerging in the last couple of years, following the results of Dalal et al. [132]. N-body simulations have shown, in fact, that nonlinear bias and an initial bispectrum are not two distinct sources of non-Gaussianity for the galaxy bispectrum, *not even at large scales!* Instead a *local* initial component can significantly affect the bias relation precisely at large scales, adding extra corrections. In the spirit of a review and since *we do not have*, at the time of writing, *a satisfactory model of the galaxy bispectrum in presence of non-Gaussian initial conditions*, in this section we will summarize earlier results, while in Section 4.3.5 we will present the recent developments that radically changed our understanding of the effects of local non-Gaussianity on the large-scale structure and finally comment, in Section 4.3.6, on some consequences for galaxy bispectrum measurements as far as current research provides.

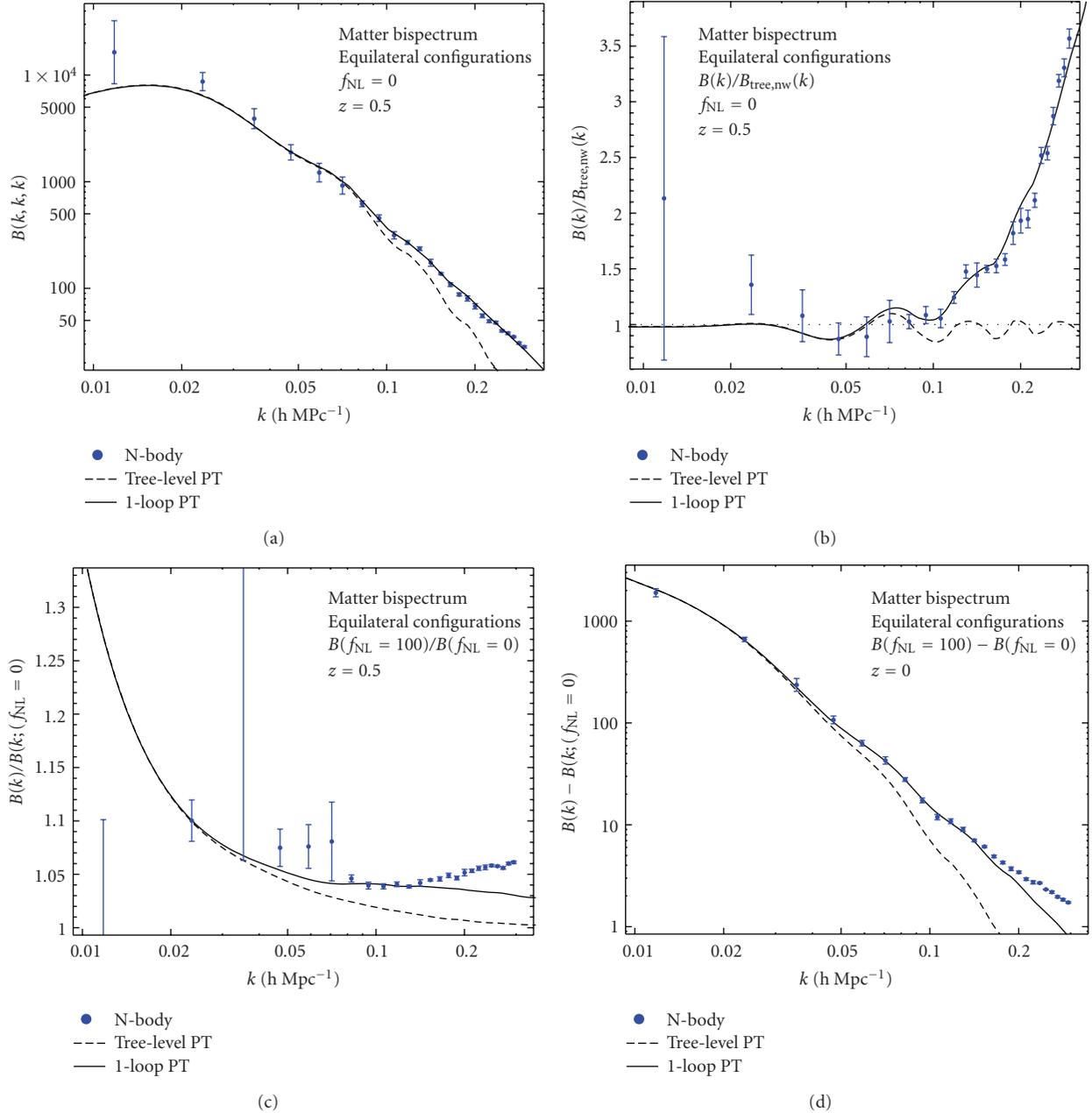


FIGURE 17: (a), (b) Equilateral configurations of the matter bispectrum measured in N-body simulations with Gaussian initial conditions (*data points*) and tree-level (*dashed lines*) and one-loop (*continuous lines*) predictions in perturbation theory. (b) shows the ratio to the tree-level prediction with acoustic oscillations removed. (c), (d) Ratio (c) and difference (d) between the matter bispectrum measured in realizations with local non-Gaussian initial conditions ( $f_{\text{NL}} = 100$ ) and the Gaussian case, compared with PT predictions, from [172].

4.3.1. *The Galaxy Bispectrum and Local Bias.* Until recently, it was commonly assumed, even for non-Gaussian initial conditions, that the galaxy overdensity  $\delta_g(\mathbf{x})$ , defined in terms of the galaxy density  $n_g(\mathbf{x})$  and its mean  $\bar{n}_g$  as

$$\delta_g(\mathbf{x}) \equiv \frac{n_g(\mathbf{x}) - \bar{n}_g}{\bar{n}_g}, \quad (171)$$

can be expressed, at large-scales, as a *local* function of the matter density contrast,  $\delta(\mathbf{x})$ , (Properly speaking we should

consider here the *smoothed* matter density contrast, that is,  $\delta_R(\mathbf{x}) = \int d^3\mathbf{x}' W_R(\mathbf{x} - \mathbf{x}')\delta(\mathbf{x}')$  with  $W_R$  being a top-hat filter function. For simplicity, we implicitly assume a *smooth* density field, so that, for large enough filtering scale, for example,  $R \sim 10 \text{ h}^{-1} \text{Mpc}$ , matter perturbations are small,  $\delta \lesssim 1$ .) that is,

$$\delta_g(\mathbf{x}) = f[\delta(\mathbf{x})]. \quad (172)$$

Such a reasonable expectation is based on the fact that the physics of galaxy formation operates on much smaller scales,

below the typical halo size than those we are interested in. At large scales, where fluctuations are small,  $\delta_R \lesssim 1$ , we can consider the Taylor expansion [159]

$$\delta_g(\mathbf{x}) = b_1 \delta(\mathbf{x}) + \frac{1}{2} b_2 \delta^2(\mathbf{x}) + \frac{1}{3!} b_3 \delta^3(\mathbf{x}) + \dots, \quad (173)$$

describing the bias relation between galaxy and matter in terms of a series of *constant bias parameters*,  $b_i$ . This expansion allows for a consistent extension of the perturbative expressions for the matter correlators to the galaxy ones. In fact, from (173) we can derive the galaxy three-point function in position space

$$\begin{aligned} & \langle \delta_g(\mathbf{x}_1) \delta_g(\mathbf{x}_2) \delta_g(\mathbf{x}_3) \rangle \\ &= b_1^3 \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \delta(\mathbf{x}_3) \rangle \\ &+ b_1^2 b_2 \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \delta^2(\mathbf{x}_3) \rangle + \text{perm.} + \dots, \end{aligned} \quad (174)$$

and the *tree-level* expression for the *galaxy bispectrum* given by

$$\begin{aligned} B_g(k_1, k_2, k_3) \\ = b_1^3 B^{\text{tree}}(k_1, k_2, k_3) + b_1^2 b_2 [P_0(k_1)P_0(k_2) + 2 \text{ perm.}], \end{aligned} \quad (175)$$

where the second term on the r.h.s., proportional to the *quadratic* bias parameter  $b_2$ , is of the same order of the gravity-induced contribution to the matter bispectrum  $B_G^{\text{tree}}$ , (165). Relying on this simple result, measurements of the galaxy bispectrum have been considered in the first place, in the context of Gaussian initial conditions, as a way to determine the bias parameters and break the degeneracy between linear bias ( $b_1$ ) and the amplitude of matter fluctuations (e.g.,  $\sigma_8$ ), otherwise affecting power spectrum measurements [165, 179–185]. In this respect, the corresponding reduced galaxy bispectrum is

$$\begin{aligned} Q_g(k_1, k_2, k_3) &\equiv \frac{B_g(k_1, k_2, k_3)}{P_g(k_1)P_g(k_2) + 2 \text{ perm.}} \\ &= \frac{1}{b_1} Q(k_1, k_2, k_3) + \frac{b_2}{b_1^2}, \end{aligned} \quad (176)$$

where  $Q$  is the reduced *matter* bispectrum (including a possible initial contribution) and the effect of nonlinear bias is simply given by an additive constant term. As already mentioned, measurements of triangular configurations different in shape and size allow to disentangle the different sources of non-Gaussianity and determine independently  $b_1$  and  $b_2$ , provided that accurate predictions for the matter bispectrum, from PT or N-body simulations, are available [186, 187] and the effects of redshift distortions and the survey geometry are properly taken into account [165, 188].

In particular, if we allow the possibility of non-Gaussian initial conditions, then the matter bispectrum includes an initial contribution, so that we can rewrite (176) at tree level

explicitly as

$$Q_g^{\text{tree}} = \frac{1}{b_1} [Q_I(f_{\text{NL}}) + Q_G^{\text{tree}}] + \frac{b_2}{b_1^2}, \quad (177)$$

and we can extend the analysis to obtain simultaneous constraints on the bias parameters *and* on the parameter determining the amplitude of the primordial bispectrum, that is,  $f_{\text{NL}}$ . A first conservative estimate of the possibilities offered by this method in measurements of the galaxy bispectrum in the 2dF Galaxy Redshift Survey [189] and in the Sloan Digital Sky Survey (SDSS), [190] is given by Verde et al. in [191] as a simple extension of previous results for the bias alone [181] suggesting that a primordial component could be detected for values of a *local*  $f_{\text{NL}}$  of the order of  $10^3$ - $10^4$ . As we will see in the next sections, a complete analysis of the galaxy bispectrum, including all measurable configurations, can improve this estimate by more than an order of magnitude: Scoccimarro et al. [192] forecast in fact for the SDSS limits of the order of  $f_{\text{NL}} \simeq 100$ .

Among the various observational issues in analyses of galaxy correlators, for example, finite volume effects or completeness of the galaxy samples, we stress that particularly relevance has the problem of redshift distortions. Redshift distortions have in fact a significant impact on the shape dependence of the galaxy bispectrum, particularly at small scales [165, 183, 193]. A recent treatment of redshift distortions in bispectrum predictions (with Gaussian initial conditions) can be found in the study by Smith et al. in [188].

**4.3.2. A Bispectrum Estimator.** In this section, we define a simple estimator for the measurement of the galaxy bispectrum in N-body simulations as well as actual data. This allows us to derive an expression for the bispectrum variance and define a Fisher matrix for an analysis of the galaxy bispectrum in terms of the non-Gaussian (and bias) parameters. In the next section we will consider a proper likelihood analysis and the effects of the bispectrum covariance. Since what follows can be applied in general to bispectrum measurements, we will consider, to simplify the notation, the case of the *matter* density field in Fourier space, described by the density contrast  $\delta_{\mathbf{k}}$ . We will point out relevant differences in the application to the galaxy distribution.

For a cubic box of volume  $V$ , a bispectrum estimator can be defined as [170]

$$\begin{aligned} \hat{B}(k_1, k_2, k_3) \\ \equiv \frac{V_f}{V_B(k_1, k_2, k_3)} \int_{k_1} d^3 q_1 \int_{k_2} d^3 q_2 \int_{k_3} d^3 q_3 \delta_D(\mathbf{q}_{123}) \delta_{\mathbf{q}_1} \delta_{\mathbf{q}_2} \delta_{\mathbf{q}_3}, \end{aligned} \quad (178)$$

where  $V_f \equiv k_f^3 = (2\pi)^3/V$  is the volume of the fundamental cell and where each integration is defined over the bin  $q_i \in [k_i - \Delta k/2, k_i + \Delta k/2]$  centered at  $k_i$  and of size  $\Delta k$  equal to a multiple of the fundamental frequency  $k_f$ . The Dirac delta function  $\delta_D(\mathbf{q}_{123})$  ensures that the wavenumbers  $\mathbf{q}_1$ ,  $\mathbf{q}_2$ , and  $\mathbf{q}_3$  indeed form a closed triangle, as imposed

by translational invariance, while the normalization factor  $V_B(k_1, k_2, k_3)$ , given by

$$\begin{aligned} V_B(k_1, k_2, k_3) & \\ & \equiv \int_{k_1} d^3 \mathbf{q}_1 \int_{k_2} d^3 \mathbf{q}_2 \int_{k_3} d^3 \mathbf{q}_3 \delta_D(\mathbf{q}_{123}) \simeq 8\pi^2 k_1 k_2 k_3 \Delta k^3, \end{aligned} \quad (179)$$

represents the number of *fundamental* triangular configurations (given by the triplet  $\mathbf{q}_1$ ,  $\mathbf{q}_2$ , and  $\mathbf{q}_3$ ) that belong to the triangular configuration *bin* defined by the triangle sizes  $k_1$ ,  $k_2$ , and  $k_3$  with uncertainty  $\Delta k$ .

The leading contribution to the bispectrum variance following from this estimator, in analogy with the power spectrum case [194], is given by Scoccimarro et al. in [170] as (This expression (see [192]) corrects a typo in equation (A.16) of Scoccimarro et al. in [170].)

$$\Delta B^2(k_1, k_2, k_3) = V_f \frac{s_{123}}{V_B(k_1, k_2, k_3)} P_{\text{tot}}(k_1) P_{\text{tot}}(k_2) P_{\text{tot}}(k_3), \quad (180)$$

with the factor  $s_{123} = 6, 2, 1$ , respectively, for equilateral, isosceles, and general triangles and where

$$P_{\text{tot}}(k) \equiv P(k) + \frac{1}{(2\pi)^3} \frac{1}{\bar{n}}, \quad (181)$$

with the particle (or galaxy) number density  $\bar{n}$  accounting for the shot noise contribution. In the case of a galaxy distribution, the matter power spectrum  $P(k)$  on the r.h.s. should be replaced with the galaxy power spectrum, expressed, at large scales, by  $P_g(k) = b_1^2 P(k)$ , under the local bias assumption of (173). Equation (180) constitutes the *Gaussian limit* to the bispectrum variance, as it neglects higher-order corrections dependent on the three-, four-, and six-point, connected, correlation functions.

We will not discuss here the theory of the Fourier-space correlation functions estimation in redshift surveys as this would take us quite far away from our topic. For the power spectrum estimation we refer the reader to [195, 196] for a pedagogical and historical introduction to a rather extensive literature. It should be noted that only a limited fraction of these results relative to the power spectrum has been extended to the bispectrum [165, 181]. No optimal estimator for  $f_{\text{NL}}$ , given a specific model, has been studied so far.

**4.3.3. The Fisher Matrix Forecasts.** In this section we consider simple forecasts for the constraints on the non-Gaussian parameters from measurements of the galaxy bispectrum in future redshift surveys. Specifically, we will consider a Fisher matrix for reduced galaxy bispectrum  $Q_g$  in terms of the non-Gaussian parameter  $f_{\text{NL}}$  and the linear and quadratic bias parameters  $b_1$  and  $b_2$ . These three parameters characterize the relative weight of the different non-Gaussian contributions to the galaxy bispectrum. Since the possibility to detect a primordial component relies on our ability to separate the three contributions, a robust result should, *at least*, involve a marginalization over bias. On the other

hand, we will assume all cosmological parameters as known. This is in part justified by the weak dependence of the matter-reduced bispectrum on cosmology discussed in the previous section. In this respect, it can be shown that the reduced bispectrum has the same signal-to-noise ratio as the bispectrum. For given triangular configurations, in fact,

$$\left(\frac{S}{N}\right)_{(k_1, k_2, k_3)} \equiv \frac{Q_g(k_1, k_2, k_3)}{\Delta Q_g(k_1, k_2, k_3)} \simeq \frac{B_g(k_1, k_2, k_3)}{\Delta B_g(k_1, k_2, k_3)}, \quad (182)$$

since the variance of  $Q$  is dominated by the variance of  $B$  (see, for instance, [192]).

The Fisher matrix can be written as

$$F_{\alpha\beta} \equiv \sum_{\text{triangles}} \frac{\partial Q_g}{\partial p_\alpha} \frac{\partial Q_g}{\partial p_\beta} \frac{1}{\Delta Q_g^2}, \quad (183)$$

where the indices  $\alpha$  and  $\beta$  run over the parameters of interest  $f_{\text{NL}}$ ,  $b_1$ , and  $b_2$ , while the reduced bispectrum variance, as mentioned above, can be expressed in first approximation as

$$\Delta Q_g^2(k_1, k_2, k_3) \simeq \frac{\Delta B_g^2(k_1, k_2, k_3)}{\left[P_g(k_1)P_g(k_2) + 2 \text{ perm.}\right]^2}, \quad (184)$$

with  $\Delta B_g^2$  given by (180). Notice that  $\Delta Q_g^2$  depends on the linear bias parameter  $b_1$ . The sum over the triangles configurations can be explicitly defined in terms of three sums over the wavenumbers  $k_1$ ,  $k_2$ , and  $k_3$  in steps of  $\Delta k$ :

$$\sum_{\text{triangles}} \equiv \sum_{k_1=k_{\min}}^{k_{\max}} \sum_{k_2=k_{\min}}^{k_1} \sum_{k_3=k_{\min}}^{k_2}, \quad (185)$$

with  $k_{\min}^* = \max(k_{\min}, |k_1 - k_2|)$  to ensure that a close triangle can be formed and with  $k_{\max}$  representing the minimal physical scale included in the analysis. Clearly, larger values of  $k_{\max}$  correspond to a much larger number of available configurations. For this reason, in fact, the cumulative signal-to-noise ratio for the bispectrum, that is, the sum of the signal-to-noise ratio over all measurable configurations, grows more rapidly with  $k_{\max}$  than it does for the power spectrum. On the other hand, we expect the primordial component to decrease significantly at small scales (high- $k$ ). In practice, however,  $k_{\max}$  can be defined as the smallest scale at which we can trust our model for the galaxy bispectrum, in our case, the tree-level expression in (177).

In Figure 18, the forecasted errors on bias parameters and non-Gaussian parameters as a function of  $k_{\max}$  for an ideal geometry galaxy survey of volume  $V = 10 h^{-3} \text{ Gpc}^3$  and a galaxy number density of  $n_g = 5 \times 10^{-3} h^3 \text{ Mpc}^{-3}$  at redshift  $z = 1$  (dashed, red lines) and  $z = 3$  (continuous, blue lines) are shown. The negligible difference between the results for the non-Gaussian parameters at different redshift is a consequence of the fact that the signal-to-noise ratio of the primordial component to the matter and galaxy bispectrum for a single triangular configuration,  $B_0/\Delta B$ , is, in our approximation, *constant*, both as a function of redshift and scale. This is not the case for the contributions due to gravitational instability and bias. It is clear that

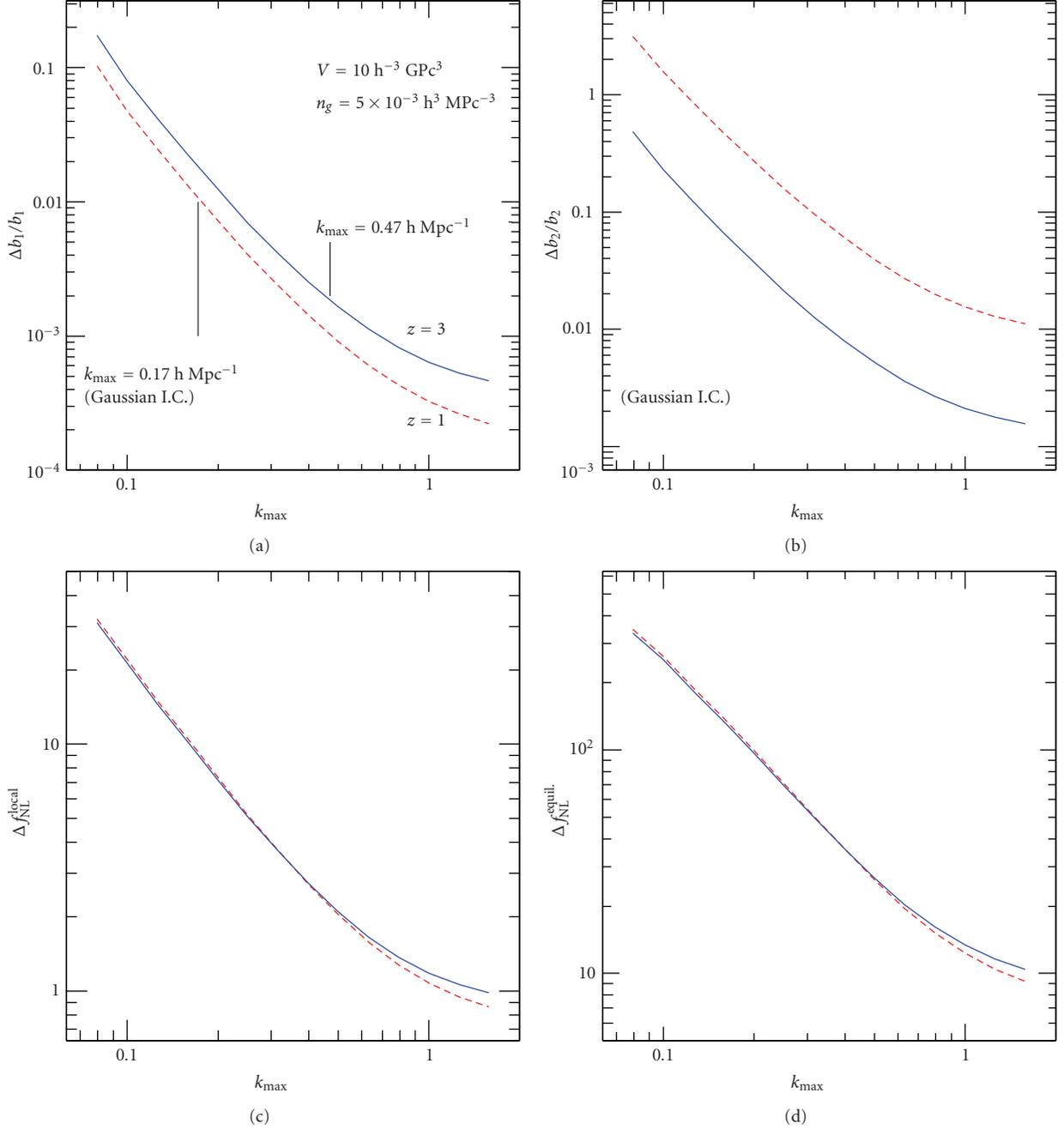


FIGURE 18: (a), (b) Predicted errors on galaxy bias parameters  $b_1$  (a) and  $b_2$  (b) as a function of the maximum wavenumber  $k_{\max}$  considered for the sum defining the Fisher matrix, (185). The analysis corresponds to an ideal geometry survey of volume  $V = 10 \text{ h}^{-3} \text{ Gpc}^3$  and a galaxy number density of  $n_g = 5 \times 10^{-3} \text{ h}^3 \text{ Mpc}^{-3}$ . Dashed (red) lines assume a mean redshift of  $z = 1$ , while continuous (blue) lines assume  $z = 3$ . Both assume that Gaussian initial conditions; that is,  $f_{\text{NL}} = 0$ . The vertical lines correspond to the value of  $k_{\max}$  determined as the inverse of the distance scale  $R$  defined by the condition  $\sigma(R, z) = 0.5$ . (c), (d) Predicted errors on the non-Gaussian parameters  $f_{\text{NL}}^{\text{loc}}$  (c) and  $f_{\text{NL}}^{\text{eq}}$  (d), marginalized over the bias parameters, as a function of  $k_{\max}$ . The effect of shot noise is evident only at very high values of  $k_{\max}$ , while for realistic surveys one can expect a more significant effect at lower  $k$ . The relatively high value  $n_g = 5 \times 10^{-3} \text{ h}^3 \text{ Mpc}^{-3}$  has been chosen here just for illustrative purposes, from the study by Sefusatti and Komatsu in [197].

the choice of  $k_{\max}$  significantly affects the final result. For instance, Sefusatti and Komatsu [197] define  $k_{\max}$ , for a given survey, as the inverse of the scale  $R$  given by the condition  $\sigma(R, z) = 0.5$  to ensure that the tree-level predictions is

applied within the mildly nonlinear range. Notice that the choice of  $k_{\max}$  depends on redshift, since at larger redshift we can expect a larger range of validity of perturbation theory predictions, both for matter and galaxy bispectrums.

The dependence on the survey volume is simply given by  $\sim 1/\sqrt{V}$ .

Scoccimarro et al. [192], from a Fisher matrix analysis as the one described above, have shown that the 2dF and SDSS surveys should be able to probe values of  $f_{\text{NL}}^{\text{loc}} \lesssim 100$ , assuming that  $k_{\text{max}} = 0.3 \text{ h Mpc}^{-1}$ . They suggested as well that an all-sky survey with a galaxy number density of  $n_g \sim 3 \times 10^{-3} \text{ h}^3 \text{ Mpc}^{-3}$  up to redshift  $z \sim 1$  can probe values of  $f_{\text{NL}}^{\text{loc}}$  of order unity.

Sefusatti and Komatsu [197] provided more specific predictions for a choice of planned and proposed high-redshift galaxy surveys, based on a similar Fisher approach, for the errors on non-Gaussian parameters both for the local and equilateral models. It is found that, for equilateral non-Gaussianity, the degeneracy between the non-Gaussian parameter  $f_{\text{NL}}^{\text{eq}}$  and the bias parameters is severe. This is due to the fact that the specific shape dependence of the initial contribution, being somehow complementary to the shape dependence of the bispectrum induced by gravity which is low for nearly equilateral triangles, reduces to a certain extent the overall shape dependence of the total matter bispectrum. In this case, it is then more difficult to distinguish the total matter bispectrum from the component due to nonlinear bias, which at tree-level approximation is a simple constant in the expression for the reduced bispectrum  $Q$ , (177). On the other hand, such degeneracy extends to unphysical regions of the  $b_1$ - $b_2$  plane and it can be significantly reduced by introducing a correlation between linear and quadratic biases as the one predicted by the halo model. The marginalization over bias can be then replaced by a marginalization over the parameters of the Halo Occupation Distribution describing the galaxy population. Sefusatti and Komatsu [197] find that future large-volume surveys ( $V \sim 100 \text{ h}^{-3} \text{ Gpc}^3$  at  $z \sim 1, 2$ ), designed to accurately measure acoustic oscillations in the galaxy correlation function and thus map the late-time expansion of the Universe, should be able to probe  $f_{\text{NL}}^{\text{loc}} \sim 4$  and  $f_{\text{NL}}^{\text{eq}} \sim 20$ , that is, values comparable to those expected from future CMB missions. At that time they constituted the best forecasts for constraints on  $f_{\text{NL}}$  from large-scale structure measurements. These results implied, in particular, that, if Planck will indeed detect primordial non-Gaussianity, a confirmation by large-scale structure observations will be required to firmly establish such an important discovery.

**4.3.4. Effects of Covariance and Current Results.** The simple Fisher matrix analysis described in the previous section makes several approximations, starting with the assumption of an ideal geometry for the survey under consideration, and the Gaussian variance for the galaxy bispectrum configurations. In fact we can expect a proper treatment of the survey selection function *and* of the bispectrum covariance to have a significant impact on the estimation of the non-Gaussian (and bias) parameters. Triangular bispectrum configurations at the largest scales probed by a realistic redshift survey (where the initial component should provide the largest corrections) are indeed highly correlated, because of the limited number of measurable Fourier modes.

The issue of bispectrum covariance has been studied in [165, 184, 185, 192]. For instance, Scoccimarro et al. [192] compare the Fisher matrix results for an ideal survey with a volume and galaxy number density similar to those of the main sample of the SDSS, with the predictions resulting from a likelihood analysis of the same survey, *including* the effects of survey geometry and covariance. Such analysis involves all measurable triangular configurations defined by wavenumbers  $k_1, k_2, k_3 \leq 0.3 \text{ h Mpc}^{-1}$ , with  $\Delta k = 0.015 \text{ h Mpc}^{-1}$ , resulting in a total number of triangle bins,  $N_T = 1015$ . The estimation of the corresponding,  $1015 \times 1015$ , bispectrum covariance matrix clearly represents a challenging computational problem as it cannot be determined from a relatively small number of N-body simulations. This work uses instead a code [165] implementing particle displacements as predicted by second-order Lagrangian perturbation theory (2LPT, see, for instance, [70] and references therein) to produce 6,000 realizations of the density field. Such a large number of realizations are in fact necessary for an accurate determination of the covariance matrix. In addition, the 2LPT results, including particle velocities, allow for an exact redshift mapping. Each mock catalog, in redshift space, is then weighted according to the Feldman-Kaiser-Peacock (FKP) procedure [165, 181, 194] to take into account the SDSS selection function. The same covariance matrix is compared to analytic expressions in the study by Sefusatti et al. in [184].

Given a proper estimate of the covariance matrix, a likelihood function for the reduced bispectrum  $Q_n$  can be defined in terms of the normalized bispectrum eigenmodes  $\hat{q}_n$  that diagonalize it [165]. These can be expressed as

$$\hat{q}_n = \sum_{m=1}^{N_T} \gamma_{mn} \frac{Q_m - \bar{Q}_m}{\Delta Q_m}, \quad (186)$$

where  $\bar{Q}_m \equiv \langle Q_m \rangle$ ,  $\Delta Q_m^2 \equiv \langle (Q_m - \bar{Q}_m)^2 \rangle$  and their signal-to-noise ratio is given by

$$\left( \frac{S}{N} \right)_n = \frac{1}{\lambda_n} \left| \sum_{m=1}^{N_T} \gamma_{mn} \frac{\bar{Q}_m}{\Delta Q_m} \right|, \quad (187)$$

where  $\lambda_n$  represents the eigenvalue for  $\hat{q}_n$ , with  $\langle \hat{q}_n \hat{q}_m \rangle = \lambda_n^2 \delta_{nm}$ . The eigenmodes presenting the largest signal-to-noise ratio can be easily interpreted by considering how they weight different bispectrum configurations. In fact, the largest signal-to-noise ratio corresponds to an eigenmode defined by a nearly equal weighting of *all* triangles, and it therefore represents the overall bispectrum amplitude. The next eigenmode weights instead with opposite sign triangles close to the equilateral shape and nearly collinear triangle. Each eigenmode represents in fact a fraction of the information contained in the bispectrum configurations, and a crucial role in this respect is played by the shape and scale dependence. To illustrate this point, Figure 19 (from [192]) shows the 95% C.L. limits on  $f_{\text{NL}}^{\text{loc}}$  from the likelihood analysis of the IRAS PSCz catalog [198] as a function of the number of eigenmodes included.

Although the diagonalization of the covariance matrix does not ensure the exact independence of the eigenmodes,

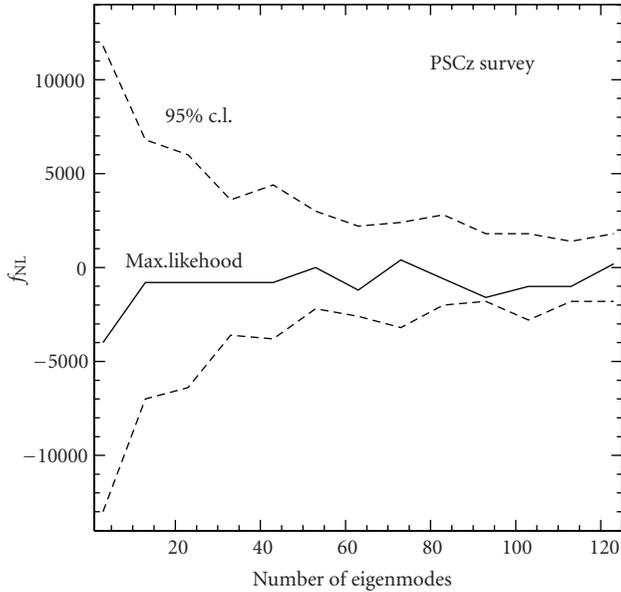


FIGURE 19: 95% confidence limits on  $f_{\text{NL}}$  from the PSCz galaxy bispectrum after marginalization over bias parameters, as a function of the number of eigenmodes included in the likelihood analysis, from the study by Scoccimarro et al. in [192].

which can still present non-vanishing higher-order correlations, it has been shown that this is nevertheless a reasonable assumption in practice [165]. This allows us to write a likelihood function for the non-Gaussian and bias parameters, denoted generically as  $p_\alpha$ , in terms of the product of the probability distribution functions  $P_n(x)$  for each individual eigenmode; that is,

$$\mathcal{L}(\{p_\alpha\}) \propto \prod_{n=1}^{N_T} P_n[\hat{q}_n(\{p_\alpha\})]. \quad (188)$$

The probability distributions  $P_n(x)$ , which can be determined from the mock catalogs, are not expected in general to be Gaussian, although this can be in fact a good first-order approximation in the case of the SDSS main sample [192].

A direct implementation of this kind of analysis, taking into account all measurable bispectrum configurations and their covariance, has been performed by Scoccimarro et al. [199] for different IRAS catalogs [114, 200, 201] and by Feldman et al. [202] for the IRAS PSCz catalog Saunders et al. [198] considering that the case of the  $\chi^2$  model of primordial non-Gaussianity [192, 203] derives the limit  $|f_{\text{NL}}^{\text{loc}}| < 1800$  at 95% C.L. for the bispectrum measured in the PSCz catalog.

Along these lines, Scoccimarro et al. [192] also studied the constraints on  $f_{\text{NL}}^{\text{loc}}$  for local non-Gaussianity that could be obtained from measurements of the galaxy bispectrum in the SDSS main sample, including the effects of the survey geometry and bispectrum covariance, forecasting the  $1\text{-}\sigma$  error  $\Delta f_{\text{NL}}^{\text{loc}} \simeq 150$ , after marginalization over the bias parameters. This work compared this more realistic estimate of the predicted errors on  $f_{\text{NL}}^{\text{loc}}$  from the likelihood analysis of the SDSS bispectrum to the Fisher matrix forecast for an ideal geometry of nearly the same volume and galaxy

density finding a worsening of a factor of 4-5. They point out, however, that the realistic errors, which are an estimate from the north part of SDSS alone, should be taken as an upper bound to the results actually achievable because of the FKP weighting scheme, not optimal at the largest scales where the primordial component is the largest and because of the fact that extra signal can be found as well in open configurations, not considered there, due to the broken translation invariance. We might add, based on the results of Section 4.2.2, that nonlinear corrections present for non-Gaussian initial conditions might increase the overall signal due to a nonzero  $f_{\text{NL}}$ , particularly on small scales where a large number of triangular configurations can be measured.

At this point we should remind the reader that all of the results discussed so far on the galaxy bispectrum and its significance for constraining primordial non-Gaussianity assume the expression (177) to be a reliable prediction. As we will see in the remainder of this section, this is *not* the case, as additional effects of non-Gaussian initial conditions have to be taken into account. Nevertheless, the primordial component, whose direct detection has been the main target of the earlier works discussed above, is still expected to provide a contribution to the galaxy bispectrum, and there are good reasons to believe that these results can be still interpreted as a “conservative estimate” of the possibilities offered by bispectrum measurements in the large-scale structure to test the Gaussianity of the initial conditions.

#### 4.3.5. Primordial Non-Gaussianity and Nonlocal Galaxy Bias.

The constraints and forecasts discussed so far in this section are based on the tree-level expression for the galaxy bispectrum, (177), derived under the assumption of *local* bias, (173). As anticipated, our understanding of galaxy bias in presence of primordial non-Gaussianity *radically changed* in the last two years, after Dalal et al. [132] presented measurements of the halo power spectrum in simulations with non-Gaussian initial conditions of the local kind showing the presence of *large corrections at large scales, not captured by the local bias prescription!* Figure 20 shows the matter-halo cross-power spectrum for different values of  $f_{\text{NL}}$  from these simulations, where the unexpected effect of non-Gaussianity at large scales is evident.

Local bias, (173), in fact, implies a leading contribution to the galaxy (or halo) power spectrum of the simple form

$$P_g(k) = b_1^2 P_0(k), \quad (189)$$

with no dependence on  $f_{\text{NL}}$ , while the simulations results of Dalal et al. [132], later confirmed by Desjacques et al. [133], Grossi et al. [135], and Pillepich et al. [140], are consistent with a *scale-dependent* correction to the linear bias of the form

$$P_g(k) = [b_1 + \Delta b_1(k)]^2 P_0(k), \quad (190)$$

with

$$\Delta b_1(k) = 3 f_{\text{NL}} (b_1 - 1) \delta_c \frac{\Omega_m H_0^2}{k^2 T(k) D(z)}, \quad (191)$$

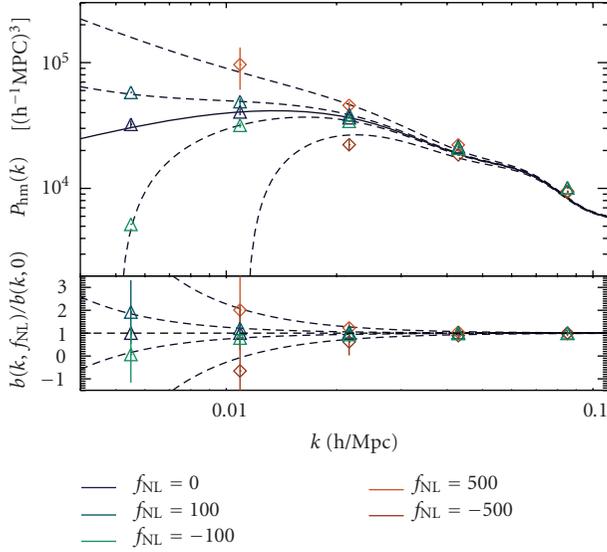


FIGURE 20: Matter-halo cross-power spectrum measured in simulations with *local* non-Gaussian initial conditions for different values of  $f_{\text{NL}}$ , from the study by Dalal et al. in [132].

where  $\delta_c \simeq 1.68$  is the linear critical density for spherical collapse, extrapolated at  $z = 0$ . Such correction therefore increases with the scale, with redshift via the growth factor  $D(z)$  and with the non-Gaussian parameter  $f_{\text{NL}}$ , and vanishes for unbiased populations ( $b_1 = 1$ ).

A theoretical interpretation, based on the peak-background split [204], has been assumed by Afshordi and Tolley [131], Dalal et al. [132], Giannantonio and Porciani [205], Slosar et al. [206], and, with a somehow different derivation, by McDonald [207]. According to these works, the local relation between the galaxy density and the matter density in (173) is modified, in presence of *local* primordial non-Gaussianity, to include an explicit dependence on the primordial curvature perturbation,  $\Phi$ ; that is,

$$\begin{aligned} \delta_g(\mathbf{x}) &= b_1 \delta(\mathbf{x}) + c_1 (f_{\text{NL}}) \Phi(\mathbf{x}) + \frac{1}{2} b_2 \delta^2(\mathbf{x}) + c_2 (f_{\text{NL}}) \delta(\mathbf{x}) \Phi(\mathbf{x}) + \dots, \\ & \quad (192) \end{aligned}$$

where the factors  $c_1$  and  $c_2$  are proportional to  $f_{\text{NL}}$  and depend, in turn, on the linear and quadratic bias parameters  $b_1$  and  $b_2$ . See the study by Giannantonio and Porciani in [205] for a detailed derivation of this expression in the context of the peak-background split. The galaxy two-point function will be given by the following perturbative expansion:

$$\begin{aligned} \langle \delta_g(\mathbf{x}_1) \delta_g(\mathbf{x}_2) \rangle &= b_1^2 \langle \delta(\mathbf{x}_1) \delta(\mathbf{x}_2) \rangle + b_1 c_1 (f_{\text{NL}}) \langle \delta(\mathbf{x}_1) \Phi(\mathbf{x}_2) \rangle + \text{perm.} \dots, \\ & \quad (193) \end{aligned}$$

where the second term can be rewritten as the scale dependence of the linear bias parameter of (191), with

the  $1/k^2$  behavior resulting from the relation between  $\delta_k$  and  $\Phi_k$  given by  $M(k) \sim k^2$ . Giannantonio and Porciani [205] describe, in fact, the galaxy distribution as *multivariate* distribution, although the matter density  $\delta$  and the curvature  $\Phi$  are not two independent random fields, but they are related by the Poisson, (160). It should be noted that no derivation of a similar effect (in the context of the peak-background split) due to a different kind of primordial non-Gaussianity (if feasible) has been, so far, proposed. The derivations presented in the works cited above, in fact, all rely on the relatively simple expression defining the local model (6) while their generalization to a model defined by a generic initial bispectrum is a quite challenging problem.

Following the results of Dalal et al. [132], moreover, an apparently different explanation, resulting in fact in a very similar *but distinct* effect on the galaxy power spectrum, has been proposed by Matarrese and Verde [208] and Taruya et al. [167]. Taruya et al. [167], starting from the *local bias prescription* of (173), point out that the next-to-leading-order correction to the galaxy two-point function in presence of local primordial non-Gaussianity represents, in fact, a large correction, identical up to a constant factor, in the large-scale limit, to the bias correction of (191). The perturbative expression for the galaxy power spectrum is given by

$$P_g(k) \simeq b_1^2 P(k) + b_1 b_2 \int d^3 q B(k, q, |\mathbf{k} - \mathbf{q}|), \quad (194)$$

where the second term, proportional to the quadratic bias parameter  $b_2$  and dependent on the matter bispectrum  $B$ , corresponds to the lowest order, one-loop correction. Remarkably, for local non-Gaussianity, in the limit  $k \rightarrow 0$ , such correction presents the same scale and redshift dependence, and, for massive halos or highly biased populations ( $b_1 \gg 1$ ), even the same amplitude, as the one resulting from (191). The expression, however, can be applied to *any* model of primordial non-Gaussianity, given the appropriate initial matter bispectrum (see, for instance, [209]). In the case of equilateral non-Gaussianity, the correction is almost negligible, while local non-Gaussianity appears to be a limiting case leading to a particularly significant effect. The same correction has been considered already by Scoccimarro [203] in the context of  $\chi^2$  initial conditions, where it leads to a redefinition of the bias parameters, with no additional scale-dependence.

Matarrese and Verde [208] presents a different derivation of an expression similar to the one of (194), based on earlier works on the density peak correlation function [122, 124]. In this case, a specific prediction for the bias parameters, valid however only in the high density threshold limit, is included. It is interesting to notice that the possibility of large-scale effects on the correlations of biased distributions has been explicitly pointed-out by Grinstein and Wise [122], although without further study.

The two distinct corrections to the galaxy power spectrum, one corresponding to the modified bias relation of (192), the other to the perturbative correction due to non-linear bias of (194), have been studied in a comprehensive framework recently by Giannantonio and Porciani [205],

where the authors suggest that the effect measured in N-body simulations is mainly due to the multivariate nature of the galaxy distribution with local primordial non-Gaussianity, rather than the effect of nonlinear bias (194). In addition, Desjacques et al. [133] pointed-out that even the galaxy bias parameters  $b_i$ , related in the framework of the halo model to the halo bias parameters  $b_{h,i}(M)$  for halo populations of mass  $M$ , present a dependence of  $f_{\text{NL}}$  due to the effects of non-Gaussianity on the halo mass function. The picture that has been emerging in the last years is therefore quite complex and it should be stressed that a wide consensus in the community on a well defined model, even for the galaxy power spectrum, is still lacking. For instance, a discrepancy of the order of a 10% between predictions and simulations results, did not find yet a unique interpretation (see discussions in [133, 135, 138, 140, 205]).

This rather surprising effect of local non-Gaussianity on the bias relation leads, remarkably, to the possibility of placing limits on  $f_{\text{NL}}^{\text{loc}}$  from *current* large-scale structure observations, *already comparable to limits from the CMB!* [131, 206]. Specifically, Slosar et al. [206] derived from measurements of the cross-correlation of several large-scale structure datasets and the CMB [210] the  $2\text{-}\sigma$  constraints

$$-29 < f_{\text{NL}}^{\text{loc}} < 70, \quad (195)$$

leading to a marginal improvement of the WMAP results. Encouraging predictions for the constraints that can be derived in future spectroscopic as well as photometric redshift surveys can be found in the study of Carbone et al. in [211]. A fair comparison between these forecasts and those derived for the galaxy bispectrum in the study of Sefusatti and Komatsu in [197] is clearly not possible as the latter do not include the effect on the bias relation discussed above. Two observations, however, are in order. In the first place, these effects on the galaxy power spectrum are specific of the local model of non-Gaussianity, while the galaxy bispectrum is in principle sensitive to *any* initial component  $B_0$ . In the second place, robust results can be obtained from galaxy power spectrum measurements at large scales in *photometric* surveys. The degradation of the information that can be extracted from bispectrum measurements in photometric surveys with respect to spectroscopic ones is still to be properly studied. The impact of photometric errors on the accurate determination of the bispectrum dependence on the triangle shape can in fact be significant.

*4.3.6. The Galaxy Bispectrum after the Paper of Dalal et al. in [132].* First steps in the direction of an extension of the results discussed above to the galaxy bispectrum have been taken by Jeong and Komatsu [212] and Sefusatti [171]. Specifically, Jeong and Komatsu [212] considered an expression for the high-peak three-point function derived by Matarrese et al. in [124], analogous to the one for the two-point function studied by Matarrese and Verde in [208], and applied it to the case of local non-Gaussianity. Sefusatti [171] considered instead the perturbative approach of Taruya et al. [167] based on the local bias expansion of (173), and applied it to local *and* equilateral non-Gaussianity.

These works show that the galaxy bispectrum is expected to be sensitive to both the initial matter bispectrum  $B_0$  as well as to the initial matter *trispectrum*  $T_0$ , by means of a contribution analogous to (194) and given by

$$B_g \simeq b_1^3 B(k_1, k_2, k_3) + \frac{b_1^2 b_2}{2} \int d^3 q T(k_1, k_2, q, |\mathbf{k}_3 - \mathbf{q}|), \quad (196)$$

which represents a large correction at large scales, with an asymptotic behavior characterized by an extra  $1/k^2$  factor with respect to the primordial matter bispectrum component,  $B_0$ , and a dependence on  $f_{\text{NL}}^2$ . In addition, Sefusatti [171] points out that, unlike the power spectrum, large-scale corrections due to nonlinear bias are present as well for *equilateral non-Gaussianity* (and virtually for any nonpathological form of the primordial bispectrum and trispectrum). Figure 21 shows the one-loop corrections to the galaxy bispectrum due to nonlinear bias and primordial non-Gaussianity under the assumption of local bias [171]. Figure 21(a) assumes local non-Gaussianity including a nonzero initial bispectrum and trispectrum, while Figure 21(b) assumes a nonzero initial bispectrum of the equilateral type. Thin lines correspond to Gaussian initial conditions. The black continuous line represents the matter bispectrum and therefore the first term on the r.h.s. of (196), while the blue dashed lines correspond to the second term. Notice that, at next-to-leading order in PT, the matter bispectrum  $B$  depends on the initial trispectrum  $T_0$  as well as the initial bispectrum  $B_0$ , so that an effect is present also for equilateral non-Gaussianity where the figure assumes that  $T_0 = 0$ .

It should be noted, however, that these results ignore, at least for local non-Gaussianity, the modified bias relation of (192) (see, in this respect, some comments in [205]) and do not provide reliable predictions for the constant bias parameters. Furthermore, they have not been properly tested against measurements of the halo bispectrum in numerical simulations. The only work, at the time of writing, in this direction is that of Nishimichi et al. [213] which shows, however, that the dependence of the halo bispectrum on  $f_{\text{NL}}$  is roughly consistent with the functional form resulting from the prediction of (196). The authors attempt as well, using a simple fit to their measurements, a preliminary forecast analysis for a future large-volume ( $100 \text{ h}^{-3} \text{ Gpc}^3$ ), high-redshift survey, finding a detectable value of  $f_{\text{NL}} \simeq 20$ , using a *very limited* number of configurations. Figure 22 from the paper by Nishimichi et al. [213] shows measurements of a set of triangular configurations of the halo bispectrum in simulations with local non-Gaussian initial conditions, as a function of the non-Gaussian parameter  $f_{\text{NL}}$ , where the dependence of the halo bispectrum on  $f_{\text{NL}}^2$  is evident.

A simple but reasonable expectation would be that the inclusion of the effects of primordial non-Gaussianity on galaxy bias will improve the results of Sefusatti and Komatsu [197], which are based on the detectability of the primordial component alone. Our understanding of these phenomena is, however, evolving rapidly in these days, and these notes on recent developments are likely to become outdated relatively soon.

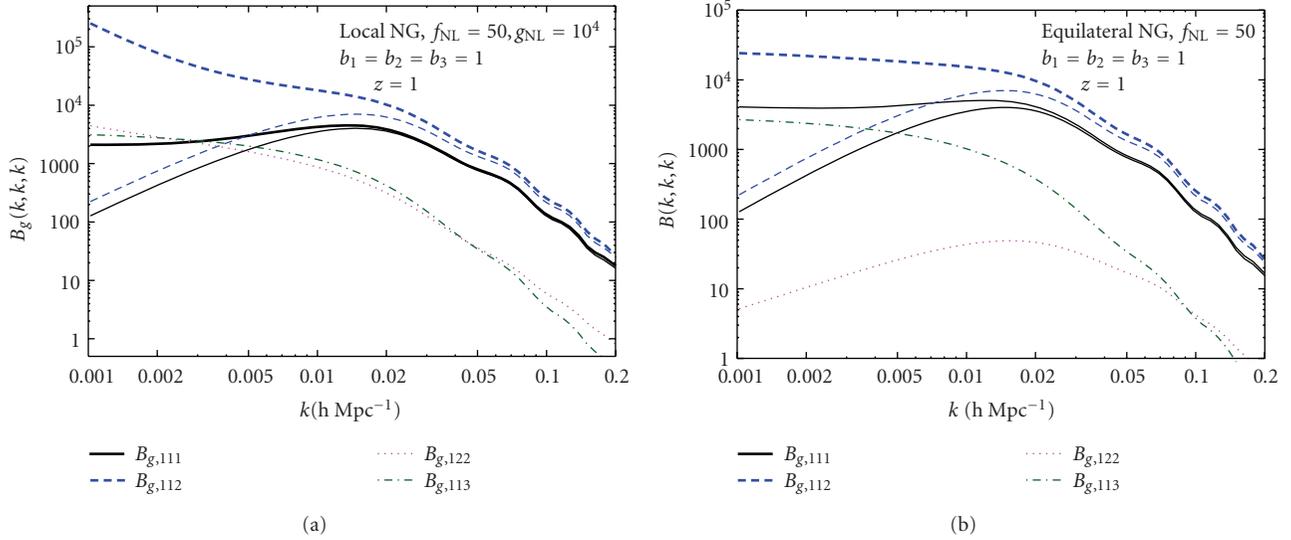


FIGURE 21: Large-scale contributions to the galaxy bispectrum due to primordial non-Gaussianity of the local (a) and equilateral (b) type described as one-loop corrections assuming a local bias prescription. Thin lines correspond to the contributions for Gaussian initial conditions, from the study by Sefusatti in [171]; see the reference for further details.

#### 4.4. Running Non-Gaussianity

**4.4.1. The Case of a Scale-Dependent  $f_{\text{NL}}$ .** DBI models of inflation predict, as we have seen, a primordial curvature bispectrum very close to the equilateral model in its shape dependence. An additional but *quite generic* feature of these models is given by a significant departure from the hierarchical scaling  $B_{\Phi}(k, k, k) \sim P_{\Phi}^2(k)$  [15, 17, 19, 214–216]. More recently, this possibility has been explored as well in models of local non-Gaussianity [31, 217–220].

Under a phenomenological point of view, this extra scale dependence can be described by a *running*  $f_{\text{NL}}(k)$ , or, more properly, in terms of an amplitude parameter  $f_{\text{NL}}$  and a running parameter  $n_{\text{NG}}$ , defined by

$$f_{\text{NL}}(K) \equiv f_{\text{NL}} \left( \frac{K}{k_p} \right)^{n_{\text{NG}}}, \quad (197)$$

where  $k_p$  is a properly chosen pivot scale, while  $K(k_1, k_2, k_3) = (k_1 + k_2 + k_3)/3$  defines an overall scale characteristic of the triangular configuration on which  $B_{\Phi}(k_1, k_2, k_3)$  depends. In other terms, the  $f_{\text{NL}}(K)$  defined above replaces the constant  $f_{\text{NL}}$  in the definitions of the local and equilateral bispectra effectively introducing an extra dependence on scale.

Observational consequences of a running  $f_{\text{NL}}(K)$  have been explored by Lo Verde et al. in [137] and Sefusatti et al. [87], while in the study by Taruya et al. in [167] this effect is included in the prediction for one-loop corrections to the matter and galaxy power spectrum.

Lo Verde et al. [137] provided an analysis of the possibility of constraining the running parameter  $n_{\text{NG}}$  by combining current limits from the CMB on the amplitude parameter  $f_{\text{NL}}$  at the pivot scale  $k_p = 0.04 \text{ Mpc}^{-1}$  with future measurements of cluster abundance. The effect of an  $n_{\text{NG}}$ , significantly different from 1, can result in a much

larger (or smaller) amount of non-Gaussianity on the smaller scales relevant for the cluster mass function. Figure 23(a), from [137] illustrates the difference in the range of scales probed by different observables. Focusing in particular on the equilateral model for the curvature bispectrum, this work assumes the amplitude of  $f_{\text{NL}}(k)$  to be constrained by the CMB bispectrum *at* the pivot point scale  $k_p$  and derives the expected constraints on its running by considering the effective amplitude of  $f_{\text{NL}}(k)$  at the smaller scales ( $k \sim 0.3 - 0.6 \text{ h Mpc}^{-1}$ ) probed by cluster surveys. For an all-sky cluster survey up to redshift  $z_{\text{max}} = 1.3$ , they find the  $1-\sigma$  constraints, marginalized over  $\Omega_m$ ,  $\sigma_8$ , and  $h$ , assuming the fiducial values to be  $f_{\text{NL}} = 38$  and  $n_{\text{NG}} = 0$ ,  $\Delta n_{\text{NG}} \simeq 2$  with a Planck prior  $\Delta f_{\text{NL}}(k = k_p) = 40$ . Their analysis, however, does not include the *simultaneous* limits that measurement of the CMB bispectrum *alone* is expected to provide on both the amplitude  $f_{\text{NL}}$  and running  $n_{\text{NG}}$ .

#### 4.4.2. Running Non-Gaussianity and Bispectrum Measurements.

Sefusatti et al. [87] perform a Fisher matrix analysis of the CMB bispectrum to obtain the sensitivity of this observable to the running of  $f_{\text{NL}}(k)$ . The results in the case of local non-Gaussianity, assuming the same pivot  $k_p = 0.04$  and marginalizing over the amplitude  $f_{\text{NL}}^{\text{loc}}$ , are the  $1-\sigma$  uncertainties of  $\Delta n_{\text{NG}} \simeq 0.68(50/f_{\text{NL}}^{\text{loc}})f_{\text{sky}}^{-1/2}$  for WMAP and  $\Delta n_{\text{NG}} \simeq 0.1(50/f_{\text{NL}}^{\text{loc}})f_{\text{sky}}^{-1/2}$  for Planck, where  $f_{\text{NL}}^{\text{loc}}$  stands for the fiducial value of the amplitude parameter. In the case of equilateral non-Gaussianity, we have  $\Delta n_{\text{NG}} \simeq 1.1(100/f_{\text{NL}}^{\text{eq}})f_{\text{sky}}^{-1/2}$  for WMAP and  $\Delta n_{\text{NG}} \simeq 0.3(100/f_{\text{NL}}^{\text{eq}})f_{\text{sky}}^{-1/2}$  for Planck. Since it is always possible, given the observable of interest (e.g., the CMB bispectrum for a specific experiment) *and* the non-Gaussian model, to choose the pivot point in such a way to remove any degeneracy between the amplitude and the running parameters,

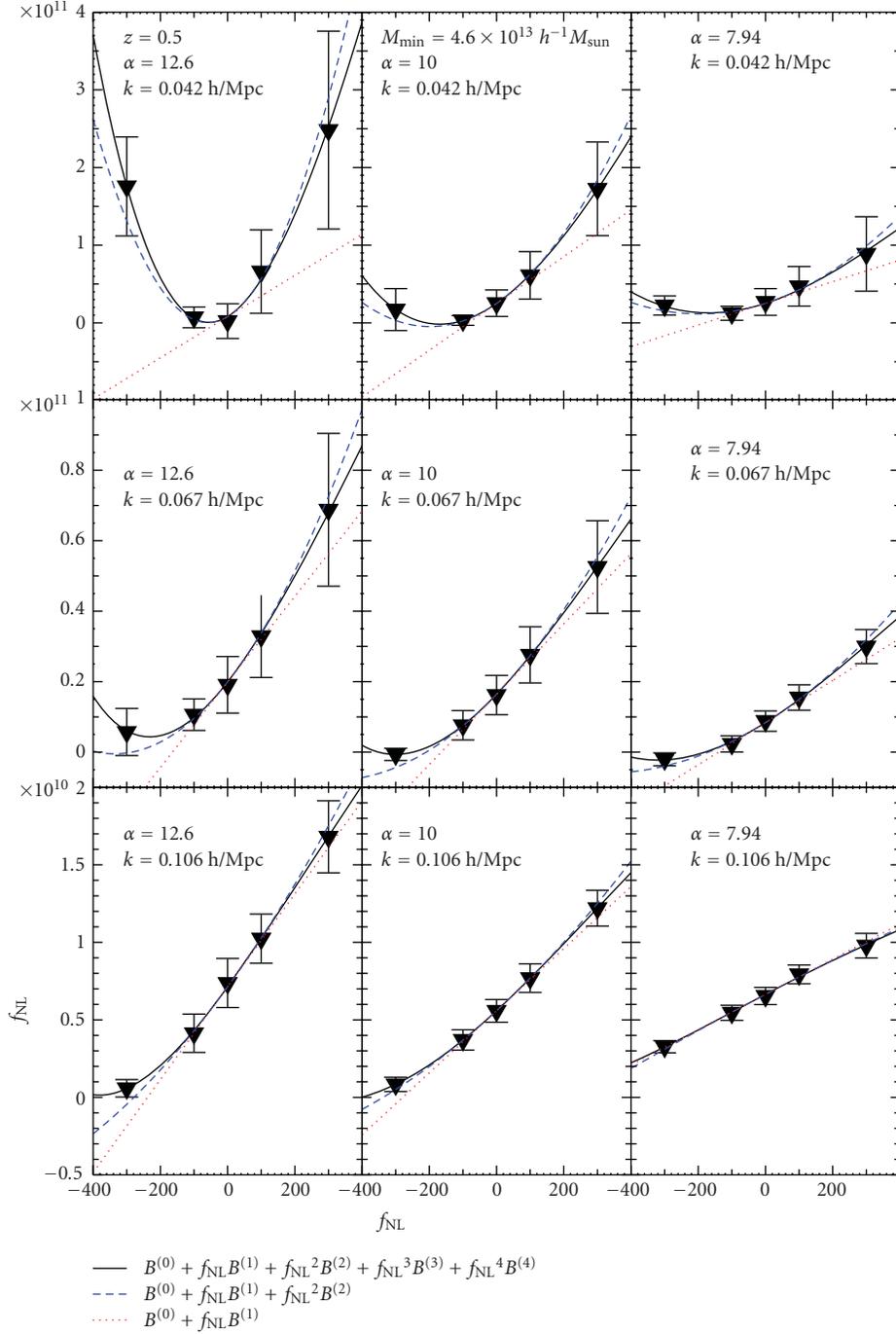


FIGURE 22: Measurements of a set of triangular configurations of the halo bispectrum in N-body simulations with local non-Gaussian initial conditions, as a function of the non-Gaussian parameter  $f_{\text{NL}}$ . Large values of the parameter  $\alpha$  correspond to more squeezed configurations. In the upper panels, the dependence of the halo bispectrum on  $f_{\text{NL}}$  is evident, from the study by Nishimichi et al. in [213].

a measurement of the running parameter comes at *no cost* with respect to the determination of the  $f_{\text{NL}}(k = k_p)$ . Notice, however, that, for reasons related to the numerical implementation of the CMB estimator, Sefusatti et al. [87] assume, for the overall scale representative of a given triangular configuration, the *geometric* mean of the three wavenumbers; that is,  $K \equiv (k_1 k_2 k_3)^{1/3}$ . While the difference

with the more physically motivated definition in terms of the arithmetic mean  $K = (k_1 + k_2 + k_3)/3$  is very small for equilateral non-Gaussianity, in the local model this is not the case.

Sefusatti et al. [171] consider as well the Fisher matrix from large-scale structure information, and specifically the galaxy power spectrum (including the effect on halo bias)

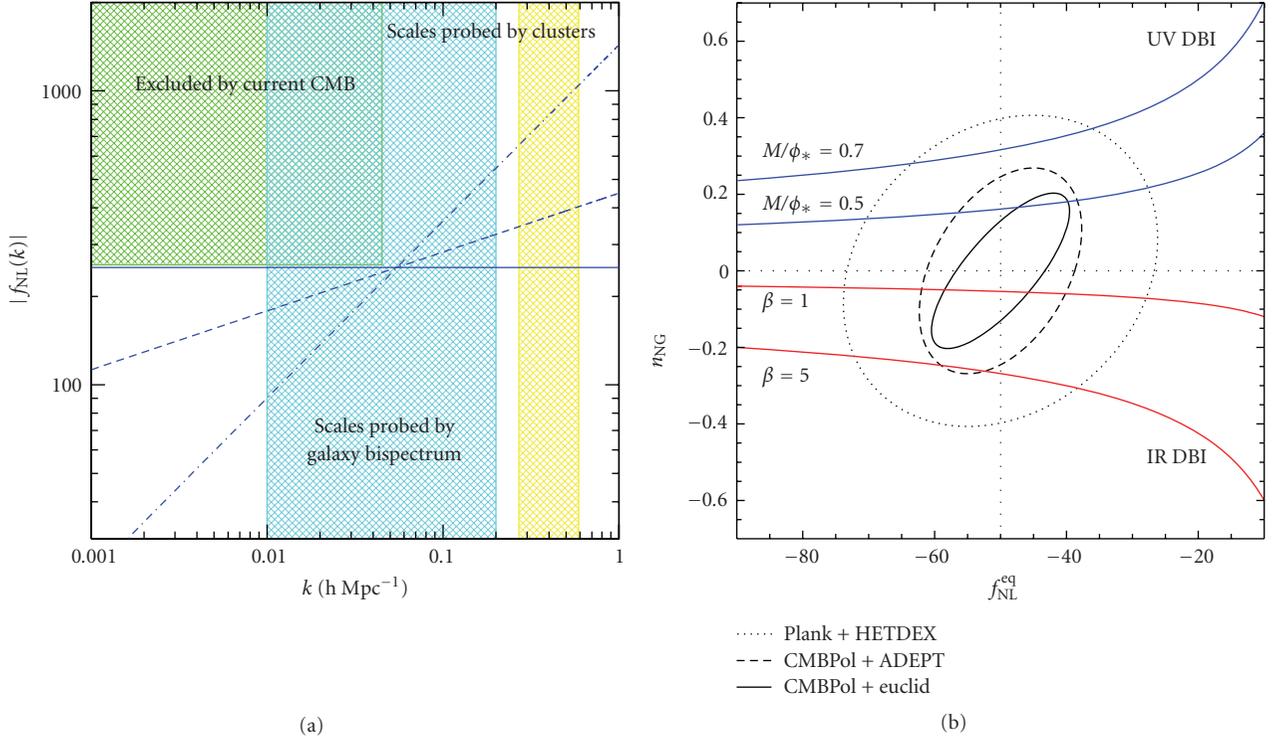


FIGURE 23: (a) Range of scales probed by different observables compared with  $f_{\text{NL}}(k)$  for different values of the running parameter  $n_{\text{NG}}$ , from the study by Lo Verde et al. in [137]. (b) Predictions of DBI models showing the peculiar relation between the amplitude  $f_{\text{NL}}^{\text{eq}}$  and the running  $n_{\text{NG}}$  for different values of the parameters of the inflaton Lagrangian. The figure includes the Fisher matrix forecasts for combined CMB and galaxy bispectrum measurements assuming a fiducial  $f_{\text{NL}}^{\text{eq}} = -50$ , from the study by Sefusatti et al. in [87]; see the reference for further details.

for the local model and the galaxy bispectrum, but in terms of the simple description of (177), therefore excluding halo bias effects. This different choice of observables with respect to the model of primordial non-Gaussianity assumes a negligible effect of equilateral non-Gaussianity on the galaxy power spectrum (still to be confirmed by N-body simulations). It is shown, in particular, that future galaxy redshift surveys can significantly improve CMB results. Figure 23(b) shows the contours plots for the 1- $\sigma$  uncertainties resulting for a joint Fisher matrix analysis of CMB and large-scale structure information. The expected limits are plotted against the predictions for the relation between the amplitude  $f_{\text{NL}}^{\text{loc}}$  and the running  $n_{\text{NG}}$  from DBI inflationary models. It is interesting to notice how these models predict a stronger running for smaller values of the amplitude parameter. In this respect, constraining the value of  $n_{\text{NG}}$  can place additional limits on the parameters of the inflaton Lagrangian.

## 5. Conclusions

Weakly non-Gaussian initial conditions are *defined*, in most of the relevant inflationary models, by a non-vanishing bispectrum for the primordial curvature perturbations. The most direct observables of this primordial density correlator are, naturally, the bispectrum of the temperature fluctuations

in the CMB and the bispectrum of the mass distribution at large scales as probed by galaxy surveys. In this review we presented an overview of the problems, results, and expectations connected with the detection of (or constraints on) primordial non-Gaussianity specifically in bispectrum measurements of the CMB and LSS.

The CMB is an ideal observable for tests of primordial NG because temperature and polarization anisotropies can be described in the linear regime of cosmological perturbations. The statistical properties of the primordial curvature field are thus directly reflected in the pattern of CMB fluctuations. As we have seen, tests of primordial NG are formulated in terms of the estimation of the bispectrum amplitude  $f_{\text{NL}}$  for each of the shapes predicted by different inflationary models. It was originally shown in the literature that a maximum-likelihood estimator of the bispectrum optimally extracts *all* of the  $f_{\text{NL}}$  information from a CMB map. Extracting the primordial nonlinear parameter from the bispectrum has subsequently become the standard way to test primordial NG in the CMB. The best  $f_{\text{NL}}$  measurements to date come from analysis of the WMAP datasets and roughly constrain the primordial bispectrum amplitude to be  $\lesssim 100$  for the local, equilateral, and orthogonal shapes. Despite already being very stringent (the NG part of the CMB temperature anisotropies is constrained at the level of  $10^{-3}$  of the total fluctuation), these bounds are still far from the typical order of magnitude of primordial NG

predicted by most inflationary models. As we have seen, the Fisher matrix forecasts show that future results from the Planck satellite (whose release date is predicted to be in 2012) will improve previous WMAP constraints by roughly one order of magnitude, thus impacting the range of some theoretical predictions. This significant improvement is due to the better sensitivity of Planck to many more bispectrum configurations in the analysis, and the possibility of exploiting both temperature and polarization datasets. Another important limitation on current constraints is that inflationary predictions encompass more shapes than those that have been constrained so far. The reason why many shapes remain to be constrained is that they cannot be written as a *separable* product of one-dimensional functions of a single wavenumber. Separability, as we have seen, is a crucial property since it makes the actual analysis computationally affordable in terms of CPU time. We have reviewed recent work showing that this limitation can also be overcome in future analysis by means of a fully general, and mathematically well-defined, eigenmode expansion of the bispectrum shape. Thanks to this, and in light of the significant improvement in sensitivity provided by Planck, better and more general CMB constraints on primordial NG models will be available in the near future. One caveat is that the high precision of the forthcoming CMB datasets makes them much more sensitive to other spurious (i.e., nonprimordial) sources of NG, which could bias the  $f_{\text{NL}}$  estimate. Achieving an accurate control on these contaminants is clearly a crucial goal for future analysis. As we have seen, much work is being done in order to predict, detect, and isolate nonprimordial NG effects, but some issues still have to be addressed. In particular a complete prediction of the total bispectrum generated by second-order cosmological perturbations is not yet available, although a number of effects have been studied in detail. Accurate characterization of NG from diffuse foreground residuals is another important issue that will require further investigation.

For large-scale structure, many aspects of the general CMB scenario outlined above change, as should be evident from a comparison of the discussions in Sections 3 and 4. In the first place, we cannot rely on a direct relation between the observed galaxy bispectrum and the primordial curvature bispectrum predicted by inflationary models. As we have seen, a small departure from Gaussian initial conditions should result in a correction to the galaxy bispectrum induced by gravitational instability and nonlinear bias, constituting the dominant contributions. The nature of this correction is a complex problem in its own right, since it is due to the linearly evolved initial matter bispectrum as well as to the effects of primordial non-Gaussianity on the galaxy bias relation. Such effects are still under investigations and we do not have, to date, an accurate theoretical model. On the other hand, early results from galaxy power spectrum measurements are very encouraging, albeit restricted at present to the local non-Gaussian model. Current datasets already appear to be able to confirm and improve CMB results. In this respect, it is evident that the ultimate goal is the implementation of a complete large-scale structure

analysis in terms of all measurable correlators, including power spectrum, bispectrum, and beyond, that is, an analysis that fully reflects the non-Gaussian nature of the mass and galaxy distributions even on large scales.

There are several issues which remain to be resolved, for which we can identify three main categories. First, we need to develop a robust *model for the galaxy correlators* accurately accounting for small-scale nonlinearities for both the matter and galaxy density fields, as well as in the presence of non-Gaussian initial conditions; this also must account to describe *nonlocalities* in the bias relation. In this review we have briefly summarized the state of the art, noting that our understanding of these phenomena is evolving rapidly. Secondly, once a reliable model is available, it will be necessary to develop the machinery that will allow us, in the event of a future detection, to properly identify the effects of different models and their bispectrum shapes. In this respect, the CMB results, presented in Section 3, provide an important benchmark. Finally, observational problems connected with redshift surveys, such as the effects of redshift distortions and/or photometric errors, survey selection function, completeness, and so forth, will have to be addressed. We have not discussed these issues here as they are generic to all large-scale structure experiments, but they clearly represent a major challenge for the exploitation of future datasets. Both the first and the last points are crucial for virtually all of the science goals of future ground-based or satellite surveys, particularly dark energy studies. Although only partial results have been obtained so far, there is every indication that characterising non-Gaussianity in future galaxy surveys will result in a significant test of the initial conditions of the Universe.

To summarize, sufficient experimental sensitivity has been reached recently in CMB experiments (namely, WMAP) to allow for meaningful constraints on the non-linear parameter  $f_{\text{NL}}$  for several different families of models. These results are already arguably the most stringent quantitative test of the predictions of standard inflation. However, much tighter constraints on a broader range of models are expected from the future Planck data release. Thus a dramatic confrontation is set to continue between the de facto standard model of inflation and observational datasets from both the CMB and large-scale structure. Tests of primordial non-Gaussianity are rapidly becoming one of the most effective and promising approaches for gleaning important information about the physical processes that generated the primordial cosmological perturbations.

## Appendix

### Basics of Estimation Theory

If a random variable  $\mathbf{x}$  is characterized by a Probability Density Function (PDF)  $p(\mathbf{x} | \lambda)$  dependent on a parameter  $\lambda$ , then an estimator for  $\lambda$  is a function  $\mathcal{E}(\mathbf{x})$  used to infer the value of the parameter. If a given dataset  $\{\mathbf{x}^{\text{obs}}\}$  is drawn from the distribution  $p(\mathbf{x}, \lambda)$ , then  $\hat{\lambda} = \mathcal{E}(\mathbf{x}^{\text{obs}})$  is the estimate of the parameter  $\lambda$  from the given observations. Since  $\mathcal{E}$  is

a function of a random variable, it is itself a random variable. In the literature a random variable obtained as a function of another set of random variables is often referred to as a *statistic*.

A general property usually required when building an estimator is its *unbiasedness*. An estimator for a parameter  $\lambda$  is *unbiased* if its average value is equal to the true value of the parameter:

$$\langle \hat{\lambda} \rangle = \lambda. \quad (\text{A.1})$$

The standard deviation is generally used to determine the error bars on  $\lambda$ ; that is,

$$\sigma_\lambda = \sqrt{\langle (\hat{\lambda} - \lambda)^2 \rangle}, \quad (\text{A.2})$$

where  $\langle \cdot \rangle$  denotes statistical average and  $\sigma^2$  is the variance of the inferred parameter. When we measure a parameter  $\lambda$  from a set of observations drawn from the PDF  $p(\mathbf{x} | \lambda)$ , we clearly would like our estimate not only to be unbiased, but also to have as small error bars as possible. In other words, among all of the possible unbiased estimators of  $\lambda$  that can be built, we look for the one that minimizes  $\sigma_\lambda$  defined in (A.2). If such an estimator exists, then it is called an *optimal estimator*.

In this context a crucial role is played by the *Fisher information matrix*, defined as (Note that for simplicity we work here using a single parameter. The generalization to the multiparameter case is however straightforward.)

$$F_{\lambda\lambda} = \left\langle \left( \frac{\partial^2 (\ln p(\mathbf{x} | \lambda))}{\partial^2 \lambda} \right) \right\rangle. \quad (\text{A.3})$$

The Fisher matrix appears in an important theorem, known as the *Cramer-Rao inequality*, stating that, *for any unbiased estimator of  $\lambda$ ,*

$$\sigma_\lambda \geq \frac{1}{\sqrt{F_{\lambda\lambda}}}. \quad (\text{A.4})$$

This theorem is then placing a *lower bound* on the error bars that can be attained when estimating a given parameter from a given set of observations. No matter which estimator is used, the smallest attainable error bars will be given by the square root of the inverse of the Fisher matrix. For a demonstration of this crucial result see, for example, the paper by Kendall and Stuart in [221] or, in relation to the CMB bispectrum, that of Babich in [69]. It is then clear that the best estimator of a parameter is an unbiased estimator saturating the Rao-Cramer bound. If such an estimator is found, then it is impossible to obtain a better estimate using any other statistic. The question then becomes whether, for a given PDF  $p(\mathbf{x} | \lambda)$ , an estimator saturating the Rao-Cramer bound exists.

It can be shown that a *necessary and sufficient condition* for an estimator  $\mathcal{E}(\mathbf{x})$  of a parameter  $\lambda$  to be optimal is the following:

$$\frac{\partial \ln p(\mathbf{x} | \lambda)}{\partial \lambda} = F_{\lambda\lambda}(\mathcal{E}(\mathbf{x}) - \lambda), \quad (\text{A.5})$$

where  $F$  is the Fisher information matrix just introduced above.

Another crucial quantity in estimation theory is the so-called *maximum-likelihood estimator*. In a maximum-likelihood (ML) approach we take the observed dataset  $\mathbf{x}^{\text{obs}}$  as fixed and we estimate  $\lambda$  as the parameter that maximizes the probability (*likelihood*) to observe the given data. In formulae, the ML estimate of  $\lambda$  is the value  $\hat{\lambda}$  that satisfies

$$\left. \frac{\partial \ln p(\mathbf{x} | \lambda)}{\partial \lambda} \right|_{\lambda=\hat{\lambda}} = 0. \quad (\text{A.6})$$

In this context the PDF  $p(\mathbf{x} | \lambda)$  is often denoted as the *likelihood function* and indicated as  $\mathcal{L}(\mathbf{x}, \lambda)$ . Two powerful theorems involving the likelihood have been proven as follows.

- (1) If there is an optimal unbiased estimator (i.e., an unbiased estimator saturating the Rao-Cramer bound), then it is the maximum-likelihood estimator or a function of it.
- (2) The maximum-likelihood estimator is *asymptotically optimal*; that is, it saturates the Rao-Cramer bound when  $N \rightarrow \infty$ , with  $N$  being the number of repeated observations in our dataset  $\mathbf{x}_{(1)}^{\text{obs}}, \dots, \mathbf{x}_{(N)}^{\text{obs}}$ .

These two theorems answer our initial question about the best estimator choice. The first theorem basically states that, *if a best method exists, then the ML estimator is that method.* Note that this result follows naturally from the optimality condition (A.5) introduced above. The second theorem says that for very large datasets the ML-estimator *is* the best method, that is, the one saturating the Rao-Cramer bound. In other words, when dealing with the practical problem of estimating a parameter from a given dataset, we should in theory always choose an ML approach. However in practice this is not always possible: for example, the PDF  $p(\mathbf{x} | \lambda)$  might be too difficult to calculate or sample numerically, or the ML condition (A.6) (generally a complicated nonlinear equation) too difficult to solve. In this case other approaches and different estimators have to be chosen.

An important role is played by the likelihood of Gaussian random variables. If a given observed variable  $O_\alpha$  is characterized by Gaussianly distributed errors, then it is easy to see that its likelihood is

$$\mathcal{L} = e^{-\chi^2/2}, \quad (\text{A.7})$$

where the  $\chi^2$  statistic is defined as

$$\chi^2 = \sum_\alpha \frac{[O_\alpha(\lambda) - O_\alpha^{\text{obs}}(\lambda)]^2}{(\Delta O_\alpha)^2}, \quad (\text{A.8})$$

where  $O_\alpha^{\text{obs}}$  are the measured values of our observable. In the previous equation we made  $O_\alpha$  dependent on a vector of parameters  $\lambda$  that we want to fit. Our observable could be, for example, the CMB angular power spectrum  $C_\ell$ , the primordial power spectrum  $P(k)$ , or, like in our case, the angular bispectrum  $B_{\ell_1, \ell_2, \ell_3}$ , and we might be

interested in knowing the sensitivity of our observation to any cosmological parameter. Our statistical estimate of  $\lambda$  will be obtained by minimizing  $\chi^2$ . That is clearly equivalent to maximize the likelihood. Let us now for simplicity work in the one-dimensional case (i.e., our observable depends on a single parameter) and expand  $\chi^2$  about its minimum, that is about the best fit value of the parameter  $\bar{\lambda}$ :

$$\chi^2(\lambda) = \chi^2(\bar{\lambda}) + \frac{1}{2} \left. \frac{\partial^2 \chi^2}{\partial \lambda^2} \right|_{\lambda=\bar{\lambda}} (\lambda - \bar{\lambda})^2. \quad (\text{A.9})$$

The linear term vanishes here since we are in the minimum. The quadratic term represents the curvature and defines the error on  $\lambda$ . If  $\chi^2$  moves very quickly away from its minimum, then our determination of  $\lambda$  will be more precise, while the error on  $\lambda$  will be much larger otherwise. If we define

$$F \equiv \frac{1}{2} \left. \frac{\partial^2 \chi^2}{\partial \lambda^2} \right|_{\lambda=\bar{\lambda}}, \quad (\text{A.10})$$

then we can estimate the minimum possible error on  $\lambda$  as  $1/\sqrt{F}$ . It is easy to see that the curvature of the likelihood in the Gaussian case matches exactly the definition of the Fisher matrix given above. The  $1/\sqrt{F}$  lower limit on the error bar then coincides, as it should, with the Rao-Cramer bound. This at the same time validates the choice of  $1/\sqrt{F}$  as the error on the parameter and also shows a simple way to interpret the Rao-Cramer bound. Since the Fisher matrix represents the curvature of the  $\ln$  of the likelihood around its maximum, it also provides an intrinsic minimum error on the measurement of the parameter. A likelihood strongly peaked around its maximum for a given parameter will provide stronger constraints on that parameter and vice versa. We have however to keep in mind that the curvature  $F$  constructed above is the curvature of the likelihood *only* if the distribution of our observable  $O_\alpha$  is Gaussian. This, strictly speaking, is in general not true, but it is a reasonably good approximation in most cases. (A clarifying example is provided by the CMB angular power spectrum. We know that  $C_\ell$  is distributed like a  $\chi^2$  with  $2\ell + 1$  degrees of freedom, which rapidly gets close to a Gaussian as  $\ell$  grows.) The Fisher matrix for any observable is then defined as the second derivative of the  $\chi^2$  statistic (A.8). If we compute it explicitly, then we get

$$F_{\lambda\lambda} = \sum_{\alpha} \frac{1}{(\Delta O_{\alpha})^2} \left[ \left( \frac{\partial O_{\alpha}}{\partial \lambda} \right)^2 + (O_{\alpha} - O_{\alpha}^{\text{obs}}) \frac{\partial^2 O_{\alpha}}{\partial \lambda^2} \right]. \quad (\text{A.11})$$

The second term in the sum above is generally neglected. The idea, as explained by Dodelson in [54] or by Press et al. in [222] is that the observed  $O_{\alpha}$  will oscillate around their real value, making the difference  $(O_{\alpha} - O_{\alpha}^{\text{obs}})$  oscillate around zero, resulting in cancellations. We are then left with the expression generally used in the literature:

$$F_{\lambda\lambda} = \sum_{\alpha} \frac{1}{(\Delta O_{\alpha})^2} \left[ \left( \frac{\partial O_{\alpha}}{\partial \lambda} \right)^2 \right]. \quad (\text{A.12})$$

In this paper, we have applied the basic concepts described in this Appendix to the estimation of the non-Gaussian parameter  $f_{\text{NL}}^{\text{model}}$  from the bispectrum of CMB and

LSS datasets. We would like to stress again that we have just very quickly sketched some essential concepts in estimation theory here. For excellent and much more comprehensive reviews of ideas and applications of estimation theory to cosmology, we refer the reader to the papers by Dodelson in [54], Martinez et al. in [223], and Tegmark et al. in [224]. The brief review provided here was actually largely inspired by those works. A detailed and complete book about statistical methods and estimation theory is, for example, that by Kendall and Stuart in [221].

## Acknowledgments

The first, third, and fourth authors were supported by STFC Grant no. ST/F002998/1 and the Centre for Theoretical Cosmology. The second acknowledges support by the French Agence National de la Recherche under Grant no. BLAN07-1-212615 and by the European Union under the Marie Curie Inter European Fellowship.

## References

- [1] E. Komatsu, J. Dunkley, M. R. Nolte et al., “Five-year *Wilkinson Microwave Anisotropy Probe* observations: cosmological interpretation,” *The Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [2] W. J. Percival, W. Sutherland, J. A. Peacock et al., “Parameter constraints for flat cosmologies from cosmic microwave background and 2dFGRS power spectra,” *Monthly Notices of the Royal Astronomical Society*, vol. 337, no. 3, pp. 1068–1080, 2002.
- [3] M. Tegmark, M. A. Strauss, M. R. Blanton, et al., “Cosmological parameters from SDSS and WMAP,” *Physical Review D*, vol. 69, no. 10, Article ID 103501, 26 pages, 2004.
- [4] A. Refregier, A. Amara, T. D. Kitching, A. Rassat, R. Scaramella, and J. Weller, “Euclid imaging consortium science book,” <http://arxiv.org/abs/1001.0061>.
- [5] D. S. Salopek and J. R. Bond, “Nonlinear evolution of long-wavelength metric fluctuations in inflationary models,” *Physical Review D*, vol. 42, no. 12, pp. 3936–3962, 1990.
- [6] J. R. Fergusson and E. P. S. Shellard, “Primordial non-Gaussianity and the CMB bispectrum,” *Physical Review D*, vol. 76, no. 8, Article ID 083523, 18 pages, 2007.
- [7] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, “Large non-Gaussianity in multiple-field inflation,” *Physical Review D*, vol. 73, no. 8/, Article ID 083522, 19 pages, 2006.
- [8] D. Babich, P. Creminelli, and M. Zaldarriaga, “The shape of non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, no. 8, article 9, 2004.
- [9] J. R. Fergusson and E. P. S. Shellard, “Shape of primordial non-Gaussianity and the CMB bispectrum,” *Physical Review D*, vol. 80, no. 4, Article ID 043510, 26 pages, 2009.
- [10] P. Creminelli, “On non-Gaussianities in single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2003, no. 10, article 3, 2003.
- [11] M. Alishahiha, E. Silverstein, and D. Tong, “DBI in the sky: non-Gaussianity from inflation with a speed limit,” *Physical Review D*, vol. 70, no. 12, Article ID 123505, 2004.
- [12] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant second-order perturbations and non-Gaussianity from inflation,” *Nuclear Physics B*, vol. 667, no. 1-2, pp. 119–148, 2003.

- [13] J. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *Journal of High Energy Physics*, vol. 7, no. 5, pp. 233–264, 2003.
- [14] J. R. Fergusson, M. Liguori, and E. P. S. Shellard, "General CMB and primordial bispectrum estimation: mode expansion, map making, and measures of  $F_{NL}$ ," *Physical Review D*, vol. 82, no. 2, Article ID 023502, 30 pages, 2010.
- [15] X. Chen and Y. Wang, "Large non-Gaussianities with intermediate shapes from quasi-single-field inflation," *Physical Review D*, vol. 81, no. 6, Article ID 063511, 5 pages, 2010.
- [16] F. Vernizzi and D. Wands, "Non-Gaussianities in two-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 19, 2006.
- [17] S. Renaux-Petel, "Combined local and equilateral non-Gaussianities from multifield DBI inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 10, article 12, 2009.
- [18] K. M. Smith, L. Senatore, and M. Zaldarriaga, "Optimal limits on  $f_{NL}^{local}$  from WMAP 5-year data," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article 9, 2009.
- [19] X. Chen, M.-X. Huang, S. Kachru, and G. Shiu, "Observational signatures and non-Gaussianities of general single-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 1, article 2, 2007.
- [20] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, "Ghost inflation," *Journal of Cosmology and Astroparticle Physics*, no. 4, pp. 1–18, 2004.
- [21] D. Seery and J. E. Lidsey, "Primordial non-Gaussianities in single-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 6, pp. 45–70, 2005.
- [22] K. M. Smith and M. Zaldarriaga, "Algorithms for bispectra: forecasting, optimal analysis, and simulation," <http://arxiv.org/abs/astro-ph/0612571>.
- [23] P. D. Meerburg, J. P. van der Schaar, and P. S. Corasaniti, "Signatures of initial state modifications on bispectrum statistics," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 18, 2009.
- [24] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, "Non-Gaussianity from inflation: theory and observations," *Physics Reports*, vol. 402, no. 3–4, pp. 103–266, 2004.
- [25] F. Bernardeau and T. Brunier, "Non-Gaussianities in extended D-term inflation," *Physical Review D*, vol. 76, no. 4, Article ID 043526, 11 pages, 2007.
- [26] F. Bernardeau and J.-P. Uzan, "Non-Gaussianity in multifield inflation," *Physical Review D*, vol. 66, no. 10, Article ID 103506, 14 pages, 2002.
- [27] F. Bernardeau and J.-P. Uzan, "Inflationary models inducing non-Gaussian metric fluctuations," *Physical Review D*, vol. 67, no. 12, Article ID 121301, 5 pages, 2003.
- [28] D. H. Lyth and Y. Rodriguez, "Inflationary prediction for primordial non-Gaussianity," *Physical Review Letters*, vol. 95, no. 12, Article ID 121302, 4 pages, 2005.
- [29] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, "Nonlinear perturbations in multiple-field inflation," *Physical Review D*, vol. 73, no. 8, Article ID 083521, 11 pages, 2006.
- [30] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, "Quantitative bispectra from multifield inflation," *Physical Review D*, vol. 76, no. 8, Article ID 083512, 7 pages, 2007.
- [31] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall, "Conditions for large non-Gaussianity in two-field slow-roll inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 10, article 8, 2008.
- [32] A. Naruko and M. Sasaki, "Large non-Gaussianity from multi-brid inflation," *Progress of Theoretical Physics*, vol. 121, no. 1, pp. 193–210, 2009.
- [33] N. Bartolo, S. Matarrese, and A. Riotto, "Non-Gaussianity in the curvaton scenario," *Physical Review D*, vol. 69, no. 4, Article ID 043503, 7 pages, 2004.
- [34] A. Linde and V. Mukhanov, "The curvaton web," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 4, article 9, 2006.
- [35] D. H. Lyth, C. Ungarelli, and D. Wands, "Primordial density perturbation in the curvaton scenario," *Physical Review D*, vol. 67, no. 2, Article ID 023503, 13 pages, 2003.
- [36] A. Chambers and A. Rajantie, "Non-Gaussianity from massless preheating," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 2, 2008.
- [37] K. Enqvist, A. Jokinen, A. Mazumdar, T. Multamäki, and A. Vähkönen, "Non-Gaussianity from instant and tachyonic preheating," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 3, pp. 133–142, 2005.
- [38] K. Enqvist, A. Jokinen, A. Mazumdar, T. Multamäki, and A. Vähkönen, "Non-Gaussianity from preheating," *Physical Review Letters*, vol. 94, no. 16, Article ID 161301, 2005.
- [39] N. Barnaby and J. M. Cline, "Predictions for non-Gaussianity from non-local inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 6, article 30, 2008.
- [40] J. J. Halliwell, "The interpretation of quantum cosmological models," in *Proceedings of the 13th International Conference on General Relativity and Gravitation*, R. J. Gleiser, C. N. Kozameh, and O. M. Moreschi, Eds., pp. 63–80, IOP Publishers, Bristol, UK, 1993.
- [41] D. Seery, K. A. Malik, and D. H. Lyth, "Non-Gaussianity of inflationary field perturbations from the field equation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 3, article 14, 2008.
- [42] M. Zaldarriaga, "Non-Gaussianities in models with a varying inflaton decay rate," *Physical Review D*, vol. 69, no. 4, Article ID 043508, 7 pages, 2004.
- [43] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, "Non-Gaussianities in new ekpyrotic cosmology," *Physical Review Letters*, vol. 100, no. 17, Article ID 171302, 2008.
- [44] P. Creminelli and L. Senatore, "A smooth bouncing cosmology with scale invariant spectrum," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, article 10, 2007.
- [45] K. Koyama, S. Mizuno, F. Vernizzi, and D. Wands, "Non-Gaussianities from ekpyrotic collapse with multiple fields," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, article 24, 2007.
- [46] J.-L. Lehners and P. J. Steinhardt, "Intuitive understanding of non-Gaussianity in ekpyrotic and cyclic models," *Physical Review D*, vol. 78, no. 12, Article ID 023506, 10 pages, 2009.
- [47] J.-L. Lehners and P. J. Steinhardt, "Non-Gaussian density fluctuations from entropically generated curvature perturbations in ekpyrotic models," *Physical Review D*, vol. 77, Article ID 063533, 5 pages, 2008.
- [48] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, "The three-point correlation function of the cosmic microwave background in inflationary models," *The Astrophysical Journal*, vol. 430, no. 2, pp. 447–457, 1994.
- [49] S. Gupta, A. Berera, A. F. Heavens, and S. Matarrese, "Non-Gaussian signatures in the cosmic background radiation from warm inflation," *Physical Review D*, vol. 66, no. 4, Article ID 043510, 7 pages, 2002.

- [50] R. Holman and A. J. Tolley, “Enhanced non-Gaussianity from excited initial states,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 5, article 1, 2008.
- [51] X. Chen, R. Easther, and E. A. Lim, “Large non-Gaussianities in single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 6, article 23, 2007.
- [52] R. Bean, X. Chen, G. Hailu, S.-H. Tye, and J. Xu, “Duality cascade in brane inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 3, article 26, 2008.
- [53] X. Chen, R. Easther, and E. A. Lim, “Generation and characterization of large non-Gaussianities in single field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article 10, 2008.
- [54] S. Dodelson, *Modern Cosmology*, vol. 13, Academic Press, Amsterdam, The Netherlands, 2003.
- [55] C.-P. Ma and E. Bertschinger, “Cosmological perturbation theory in the synchronous and conformal Newtonian gauges,” *The Astrophysical Journal*, vol. 455, no. 1, pp. 7–25, 1995.
- [56] E. Komatsu, *The pursuit of non-Gaussian fluctuations in the cosmic microwave background*, Ph.D. thesis, Tohoku University, Tokyo, Japan, 2002.
- [57] X. Luo, “The angular bispectrum of the cosmic microwave background,” *The Astrophysical Journal*, vol. 427, no. 2, pp. L71–L74, 1994.
- [58] M. Bucher, B. Van Tent, and C. Sofia Carvalho, “Detecting bispectral acoustic oscillations from inflation using a new flexible estimator,” <http://arxiv.org/abs/0911.1642>.
- [59] D. M. Regan and E. P. S. Shellard, “Cosmic string power spectrum, bispectrum and trispectrum,” <http://arxiv.org/abs/0911.2491>.
- [60] P. G. Ferreira, J. Magueijo, and K. M. Górski, “Evidence for non-Gaussianity in the COBE DMR 4 year sky maps,” *The Astrophysical Journal*, vol. 503, no. 1, pp. L1–L4, 1998.
- [61] B. C. Bromley and M. Tegmark, “Is the cosmic microwave background really non-Gaussian?” *The Astrophysical Journal Letters*, vol. 524, no. 2, pp. L79–L82, 1999.
- [62] J. Magueijo, “Erratum: “New non-Gaussian feature in COBE-DMR 4 year maps”,” *The Astrophysical Journal Letters*, vol. 532, no. 2, p. L157, 2000.
- [63] J. Magueijo, “New non-Gaussian feature in COBE-DMR 4 year maps,” *The Astrophysical Journal Letters*, vol. 528, no. 2, pp. L157–L160, 2000.
- [64] J. Magueijo, P. G. Ferreira, and K. M. Gorski, “Where is COBE maps’ non-Gaussianity?” in *Proceedings of the 2nd International Workshop on Particle Physics and the Early Universe (COSMO ’98)*, D. O. Caldwell, Ed., vol. 478 of *AIP Conference Proceedings*, pp. 176–179, July 1999.
- [65] J. Magueijo, P. G. Ferreira, and K. M. Górski, “Evidence for non-Gaussianity in the CMB,” *Astrophysical Letters and Communications*, vol. 37, no. 3–6, pp. 395–402, 2000.
- [66] A. J. Banday, S. Zaroubi, and K. M. Górski, “On the non-Gaussianity observed in the COBE differential microwave radiometer sky maps,” *The Astrophysical Journal*, vol. 533, no. 2, pp. 575–587, 2000.
- [67] E. Komatsu, B. D. Wandelt, D. N. Spergel, A. J. Banday, and K. M. Górski, “Measurement of the cosmic microwave background bispectrum on the COBE DMR sky maps,” *The Astrophysical Journal*, vol. 566, no. 1, pp. 19–29, 2002.
- [68] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $f_{NL}$  local from WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article 6, 2009.
- [69] D. Babich, “Optimal estimation of non-Gaussianity,” *Physical Review D*, vol. 72, no. 4, Article ID 043003, 13 pages, 2005.
- [70] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccamarro, “Large-scale structure of the Universe and cosmological perturbation theory,” *Physics Report*, vol. 367, no. 1–3, pp. 1–248, 2002.
- [71] F. Bernardeau and L. Kofman, “Properties of the cosmological density distribution function,” *The Astrophysical Journal*, vol. 443, no. 2, pp. 479–498, 1995.
- [72] A. N. Taylor and P. I. R. Watts, “Parameter information from non-linear cosmological fields,” *Monthly Notices of the Royal Astronomical Society*, vol. 328, no. 4, pp. 1027–1038, 2001.
- [73] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, “Limits on non-Gaussianities from WMAP data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 4, 2006.
- [74] E. Komatsu and D. N. Spergel, “Acoustic signatures in the primary microwave background bispectrum,” *Physical Review D*, vol. 63, no. 6, Article ID 063002, 13 pages, 2001.
- [75] P. Creminelli, L. Senatore, M. Zaldarriaga, and M. Tegmark, “Limits on  $f_{NL}$  parameters from Wilkinson Microwave Anisotropy Probe three-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 3, article 5, 2007.
- [76] M. Liguori, A. P. S. Yadav, F. K. Hansen, E. Komatsu, S. Matarrese, and B. D. Wandelt, “Temperature and polarization CMB maps from primordial non-Gaussianities of the local type,” *Physical Review D*, vol. 76, no. 10, Article ID 105016, 10 pages, 2007.
- [77] E. Komatsu, D. N. Spergel, and B. D. Wandelt, “Measuring primordial non-Gaussianity in the cosmic microwave background,” *The Astrophysical Journal*, vol. 634, no. 1, pp. 14–19, 2005.
- [78] E. Komatsu, A. Kogut, M. R. Nolte et al., “First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: tests of Gaussianity,” *The Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 119–134, 2003.
- [79] A. P. S. Yadav and B. D. Wandelt, “Evidence of primordial non-Gaussianity ( $f_{NL}$ ) in the Wilkinson Microwave Anisotropy Probe 3-year data at  $2.8\sigma$ ,” *Physical Review Letters*, vol. 100, no. 18, Article ID 181301, 4 pages, 2008.
- [80] J. R. Shewchuk, “An introduction to the conjugate gradient method without the agonizing pain,” Tech. Rep. CMU-CS-94-125, School of Computer Science, Carnegie Mellon University, Pittsburgh, Pa, USA, 1994.
- [81] K. M. Smith, O. Zahn, and O. Doré, “Detection of gravitational lensing in the cosmic microwave background,” *Physical Review D*, vol. 76, no. 4, Article ID 043510, 28 pages, 2007.
- [82] A. P. S. Yadav, E. Komatsu, B. D. Wandelt, M. Liguori, F. K. Hansen, and S. Matarrese, “Fast estimator of primordial non-Gaussianity from temperature and polarization anisotropies in the cosmic microwave background. II. Partial sky coverage and inhomogeneous noise,” *The Astrophysical Journal*, vol. 678, no. 2, pp. 578–582, 2008.
- [83] E. Komatsu, K. M. Smith, J. Dunkley, et al., “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation,” <http://arxiv.org/abs/1001.4538>.
- [84] D. N. Spergel, R. Bean, O. Doré et al., “Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: implications for cosmology,” *The Astrophysical Journal, Supplement Series*, vol. 170, no. 2, pp. 377–408, 2007.
- [85] T. P. Collaboration, “The scientific programme of planck,” <http://arxiv.org/abs/astro-ph/0604069>.

- [86] D. Babich and M. Zaldarriaga, "Primordial bispectrum information from CMB polarization," *Physical Review D*, vol. 70, no. 8, Article ID 083005, 12 pages, 2004.
- [87] E. Sefusatti, M. Liguori, A. P. S. Yadav, M. G. Jackson, and E. Pajer, "Constraining running non-Gaussianity," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article 22, 2009.
- [88] G. B. Rybicki and A. P. Lightman, *Radiative Processes in Astrophysics*, John Wiley & Sons, New York, NY, USA, 1986.
- [89] A. P. S. Yadav, E. Komatsu, and B. D. Wandelt, "Fast estimator of primordial non-Gaussianity from temperature and polarization anisotropies in the cosmic microwave background," *The Astrophysical Journal*, vol. 664, no. 2, pp. 680–686, 2007.
- [90] D. Baumann, A. Dymarsky, S. Kachru, I. R. Klebanov, and L. McAllister, "Holographic systematics of D-brane inflation," *Journal of High Energy Physics*, vol. 2009, no. 3, article 93, 2009.
- [91] C. L. Bennett, M. Halpern, G. Hinshaw et al., "First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: preliminary maps and basic results," *The Astrophysical Journal, Supplement Series*, vol. 148, no. 197, 117 pages, 2003.
- [92] S. M. Leach, J.-F. Cardoso, C. Baccigalupi et al., "Component separation methods for the PLANCK mission," *Astronomy and Astrophysics*, vol. 491, no. 2, pp. 597–615, 2008.
- [93] D. Munshi and A. Heavens, "A new approach to probing primordial non-Gaussianity," *Monthly Notices of the Royal Astronomical Society*, vol. 401, no. 4, pp. 2406–2418, 2010.
- [94] E. Calabrese, J. Smidt, A. Amblard et al., "Non-Gaussianity in WMAP data due to the correlation of CMB lensing potential with secondary anisotropies," *Physical Review D*, vol. 81, no. 4, Article ID 043529, 2010.
- [95] J. Smidt, A. Amblard, P. Serra, and A. Cooray, "Measurement of primordial non-Gaussianity using the WMAP 5-year temperature skewness power spectrum," *Physical Review D*, vol. 80, no. 12, Article ID 123005, 19 pages, 2009.
- [96] D. Munshi, P. Valageas, A. Cooray, and A. F. Heavens, "Secondary non-Gaussianity and cross-correlation analysis," <http://arxiv4.library.cornell.edu/abs/0907.3229>.
- [97] G. B. Rybicki and W. H. Press, "Interpolation, realization, and reconstruction of noisy, irregularly sampled data," *The Astrophysical Journal*, vol. 398, no. 1, pp. 169–176, 1992.
- [98] P. Cabella, D. Pietrobon, M. Veneziani, et al., "Foreground influence on primordial non-Gaussianity estimates: needlet analysis of WMAP 5-year data," <http://arxiv.org/abs/0910.4362>.
- [99] E. L. Wright, X. Chen, N. Odegard et al., "Five-year Wilkinson Microwave Anisotropy Probe observations: source catalog," *The Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 283–295, 2009.
- [100] N. Bartolo, S. Matarrese, and A. Riotto, "Cosmic microwave background anisotropies at second order: i," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 6, article 24, 2006.
- [101] N. Bartolo, S. Matarrese, and A. Riotto, "The full second-order radiation transfer function for large-scale CMB anisotropies," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 10, 2006.
- [102] C. Pitrou, "The radiative transfer at second order: a full treatment of the Boltzmann equation with polarization," *Classical and Quantum Gravity*, vol. 26, no. 6, Article ID 065006, 2009.
- [103] C. Pitrou, J.-P. Uzan, and F. Bernardeau, "Cosmic microwave background bispectrum on small angular scales," *Physical Review D*, vol. 78, no. 6, Article ID 063526, 20 pages, 2008.
- [104] D. Nitta, E. Komatsu, N. Bartolo, S. Matarrese, and A. Riotto, "CMB anisotropies at second order III: bispectrum from products of the first-order perturbations," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 14, 2009.
- [105] C. Pitrou, J.-P. Uzan, and F. Bernardeau, "The cosmic microwave background bispectrum from the non-linear evolution of the cosmological perturbations," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 7, article 003, 2010.
- [106] P. Serra and A. Cooray, "Impact of secondary non-Gaussianities on the search for primordial non-Gaussianity with CMB maps," *Physical Review D*, vol. 77, no. 10, Article ID 107305, 4 pages, 2008.
- [107] A. Mangilli and L. Verde, "Non-Gaussianity and the CMB bispectrum: confusion between primordial and lensing-Rees-Sciama contribution?" *Physical Review D*, vol. 80, no. 12, Article ID 123007, 11 pages, 2009.
- [108] D. Hanson, K. M. Smith, A. Challinor, and M. Liguori, "CMB lensing and primordial non-Gaussianity," *Physical Review D*, vol. 80, no. 8, Article ID 083004, 9 pages, 2009.
- [109] R. Khatri and B. D. Wandelt, "Crinkles in the last scattering surface: non-Gaussianity from inhomogeneous recombination," *Physical Review D*, vol. 79, no. 2, Article ID 023501, 9 pages, 2009.
- [110] L. Senatore, S. Tassev, and M. Zaldarriaga, "Non-Gaussianities from perturbing recombination," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article 38, 2009.
- [111] D. Babich and E. Pierpaoli, "Point source contamination in CMB non-Gaussianity analyses," *Physical Review D*, vol. 77, no. 12, Article ID 123011, 12 pages, 2008.
- [112] M. Liguori and A. Riotto, "Impact of uncertainties in the cosmological parameters on the measurement of primordial non-Gaussianity," *Physical Review D*, vol. 78, no. 12, Article ID 123004, 9 pages, 2008.
- [113] S. Donzelli, F. K. Hansen, M. Liguori, and D. Maino, "Impact of the  $1/f$  noise and the asymmetric beam on non-Gaussianity searches with planck," *The Astrophysical Journal*, vol. 706, no. 2, pp. 1226–1240, 2009.
- [114] G. P. Efstathiou, N. Kaiser, W. Saunders, et al., "Largescale clustering of IRAS galaxies," *Monthly Notices of the Royal Astronomical Society*, vol. 247, no. 2, 10 pages, 1990.
- [115] D. Maino, C. Burigana, K. M. Górski, N. Mandolesi, and M. Bersanelli, "Removing  $1/f$  noise stripes in cosmic microwave background anisotropy observations," *Astronomy and Astrophysics*, vol. 387, no. 1, pp. 356–365, 2002.
- [116] M. Sandri, F. Villa, N. Mandolesi, M. Bersanelli, and R. Nesti, "Planck low-frequency instrument: a study on the performances of the Planck millimeter space telescope coupled with LFI feed horns," in *Optical, Infrared, and Millimeter Space Telescopes*, J. C. Mather, Ed., vol. 5487 of *Proceedings of SPIE*, Glasgow, UK, June 2004.
- [117] M. Liguori, S. Matarrese, and L. Moscardini, "High-resolution simulations of non-Gaussian cosmic microwave background maps in spherical coordinates," *The Astrophysical Journal*, vol. 597, no. 1, pp. 57–65, 2003.
- [118] M. Liguori, F. K. Hansen, E. Komatsu, S. Matarrese, and A. Riotto, "Testing primordial non-Gaussianity in CMB

- anisotropies,” *Physical Review D*, vol. 73, no. 4, Article ID 043505, 12 pages, 2006.
- [119] F. Elsner and B. D. Wandelt, “Improved simulation of non-gaussian temperature and polarization cosmic microwave background maps,” *The Astrophysical Journal, Supplement Series*, vol. 184, no. 2, pp. 264–270, 2009.
- [120] C. R. Contaldi and J. Magueijo, “Generating non-Gaussian maps with a given power spectrum and bispectrum,” *Physical Review D*, vol. 63, no. 10, Article ID 103512, 10 pages, 2001.
- [121] G. Rocha, M. P. Hobson, S. Smith, P. Ferreira, and A. Challinor, “Simulation of non-Gaussian cosmic microwave background maps,” *Monthly Notices of the Royal Astronomical Society*, vol. 357, no. 1, pp. 1–11, 2005.
- [122] B. Grinstein and M. B. Wise, “Non-Gaussian fluctuations and the correlations of galaxies or rich clusters of galaxies,” *The Astrophysical Journal*, vol. 310, no. 1, pp. 19–22, 1986.
- [123] F. Lucchin and S. Matarrese, “The effect of non-gaussian statistics on the mass multiplicity of cosmic structures,” *The Astrophysical Journal*, vol. 330, pp. 535–544, 1988.
- [124] S. Matarrese, F. Lucchin, and S. A. Bonometto, “A path-integral approach to large-scale matter distribution originated by non-Gaussian fluctuations,” *The Astrophysical Journal*, vol. 310, no. 1, pp. L21–L26, 1986.
- [125] P. Coles, L. Moscardini, F. Lucchin, S. Matarrese, and A. Messina, “Skewness as a test of non-Gaussian primordial density fluctuations,” *Monthly Notices of the Royal Astronomical Society*, vol. 264, no. 3, p. 749, 1993.
- [126] A. Messina, L. Moscardini, F. Lucchin, and S. Matarrese, “Non-Gaussian initial conditions in cosmological N-body simulations. I—space-uncorrelated models,” *Monthly Notices of the Royal Astronomical Society*, vol. 245, pp. 244–254, 1990.
- [127] L. Moscardini, S. Matarrese, F. Lucchin, and A. Messina, “Non-Gaussian initial conditions in cosmological N-body simulations. II—cold dark Matter models,” *Monthly Notices of the Royal Astronomical Society*, vol. 248, pp. 424–438, 1991.
- [128] D. H. Weinberg and S. Cole, “Non-Gaussian fluctuations and the statistics of galaxy clustering,” *Monthly Notices of the Royal Astronomical Society*, vol. 259, no. 4, pp. 652–694, 1992.
- [129] M. White, “Higher order moments of the density field in a parametrized sequence of non-Gaussian theories,” *Monthly Notices of the Royal Astronomical Society*, vol. 310, no. 2, pp. 511–516, 1999.
- [130] E. Sefusatti, C. Vale, K. Kadota, and J. Frieman, “Primordial non-Gaussianity and dark energy constraints from cluster surveys,” *The Astrophysical Journal*, vol. 658, no. 2, pp. 669–679, 2007.
- [131] N. Afshordi and A. J. Tolley, “Primordial non-Gaussianity, statistics of collapsed objects, and the integrated Sachs-Wolfe effect,” *Physical Review D*, vol. 78, no. 12, Article ID 123507, 12 pages, 2008.
- [132] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, “Imprints of primordial non-Gaussianities on large-scale structure: scale-dependent bias and abundance of virialized objects,” *Physical Review D*, vol. 77, no. 12, Article ID 123514, 17 pages, 2008.
- [133] V. Desjacques, U. Seljak, and I. T. Iliev, “Scale-dependent bias induced by local non-Gaussianity: a comparison to N-body simulations,” *Monthly Notices of the Royal Astronomical Society*, vol. 396, no. 1, pp. 85–96, 2009.
- [134] C. Fedeli, L. Moscardini, and S. Matarrese, “The clustering of galaxy clusters in cosmological models with non-Gaussian initial conditions: predictions for future surveys,” *Monthly Notices of the Royal Astronomical Society*, vol. 397, no. 2, pp. 1125–1137, 2009.
- [135] M. Grossi, L. Verde, C. Carbone et al., “Large-scale non-Gaussian mass function and halo bias: tests on N-body simulations,” *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 1, pp. 321–332, 2009.
- [136] T. Y. Lam and R. K. Sheth, “Halo abundances in the fnl model,” *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 4, pp. 2143–2151, 2009.
- [137] M. Lo Verde, A. Miller, S. Shandera, and L. Verde, “Effects of scale-dependent non-Gaussianity on cosmological structures,” *Journal of Cosmology and Astro-Particle Physics*, vol. 2008, no. 4, article 14, 2008.
- [138] M. Maggiore and A. Riotto, “The halo mass function from excursion set theory. III. Non-Gaussian fluctuations,” *The Astrophysical Journal*, vol. 717, no. 1, p. 526, 2010.
- [139] M. Oguri, “Self-calibrated cluster counts as a probe of primordial non-Gaussianity,” *Physical Review Letters*, vol. 102, no. 21, Article ID 211301, 4 pages, 2009.
- [140] A. Pillepich, C. Porciani, and O. Hahn, “Halo mass function and scale-dependent bias from N-body simulations with non-Gaussian initial conditions,” *Monthly Notices of the Royal Astronomical Society*, vol. 402, no. 1, pp. 191–206, 2010.
- [141] P. Valageas, “Mass function and bias of dark matter halos for non-Gaussian initial conditions,” <http://arxiv.org/abs/0906.1042>.
- [142] M. Kamionkowski, L. Verde, and R. Jimenez, “The void abundance with non-Gaussian primordial perturbations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 1, article 10, 2009.
- [143] C. Hikage, P. Coles, M. Grossi et al., “The effect of primordial non-Gaussianity on the topology of large-scale structure,” *Monthly Notices of the Royal Astronomical Society*, vol. 385, no. 3, pp. 1613–1620, 2008.
- [144] C. Hikage, E. Komatsu, and T. Matsubara, “Primordial non-Gaussianity and analytical formula for minkowski functionals of the cosmic microwave background and large-scale structure,” *The Astrophysical Journal*, vol. 653, no. 1, pp. 11–26, 2006.
- [145] D. Crociani, L. Moscardini, M. Viel, and S. Matarrese, “The effects of primordial non-Gaussianity on the cosmological reionization,” *Monthly Notices of the Royal Astronomical Society*, vol. 394, no. 1, pp. 133–141, 2009.
- [146] M. Viel, E. Branchini, K. Dolag, M. Grossi, S. Matarrese, and L. Moscardini, “Primordial non-Gaussianities in the intergalactic medium,” *Monthly Notices of the Royal Astronomical Society*, vol. 393, no. 3, pp. 774–782, 2009.
- [147] A. Cooray, “21-cm background anisotropies can discern primordial non-Gaussianity,” *Physical Review Letters*, vol. 97, no. 26, Article ID 261301, 4 pages, 2006.
- [148] A. Pillepich, C. Porciani, and S. Matarrese, “The bispectrum of redshifted 21 centimeter fluctuations from the dark ages,” *The Astrophysical Journal*, vol. 662, no. 1, pp. 1–14, 2007.
- [149] P. J. E. Peebles, *The Large-Scale Structure of the Universe*, Princeton University Press, Princeton, NJ, USA, 1980.
- [150] F. Bernardeau, “Skewness and kurtosis in large-scale cosmic fields,” *The Astrophysical Journal*, vol. 433, no. 1, pp. 1–18, 1994.
- [151] F. Bernardeau, “The effects of smoothing on the statistical properties of large-scale cosmic fields,” *Astronomy and Astrophysics*, vol. 291, no. 3, pp. 697–712, 1994.
- [152] F. R. Bouchet, R. Juszkiewicz, S. Colombi, and R. Pellat, “Weakly nonlinear gravitational instability for arbitrary  $\Omega$ ,” *The Astrophysical Journal Letters*, vol. 394, no. 1, pp. L5–L8, 1992.

- [153] P. Coles and C. S. Frenk, "Skewness and large-scale structure," *Monthly Notices of the Royal Astronomical Society*, vol. 253, pp. 727–737, 1991.
- [154] J. N. Fry and R. J. Scherrer, "Skewness in large-scale structure and non-Gaussian initial conditions," *The Astrophysical Journal*, vol. 429, no. 1, pp. 36–42, 1994.
- [155] R. Juszkiewicz, F. R. Bouchet, and S. Colombi, "Skewness induced by gravity," *The Astrophysical Journal Letters*, vol. 412, no. 1, pp. L9–L12, 1993.
- [156] O. Lahav, M. Itoh, S. Inagaki, and Y. Suto, "Non-Gaussian signatures from Gaussian initial fluctuations: evolution of skewness and kurtosis from cosmological simulations in the highly nonlinear regime," *The Astrophysical Journal*, vol. 402, no. 2, pp. 387–397, 1993.
- [157] F. Lucchin, S. Matarrese, A. L. Melott, and L. Moscardini, "Higher order moments of the matter distribution in scale-free cosmological simulations with large dynamic range," *The Astrophysical Journal*, vol. 422, no. 2, pp. 430–442, 1994.
- [158] X. Luo and D. N. Schramm, "Kurtosis, skewness, and non-Gaussian cosmological density perturbations," *The Astrophysical Journal*, vol. 408, no. 1, pp. 33–42, 1993.
- [159] J. N. Fry and E. Gaztañaga, "Biasing and hierarchical statistics in large-scale structure," *The Astrophysical Journal*, vol. 413, no. 2, pp. 447–452, 1993.
- [160] M. J. Chodorowski and F. R. Bouchet, "Kurtosis in large-scale structure as a constraint on non-Gaussian initial conditions," *Monthly Notices of the Royal Astronomical Society*, vol. 279, no. 2, pp. 557–563, 1996.
- [161] R. Durrer, R. Juszkiewicz, M. Kunz, and J.-P. Uzan, "Skewness as a probe of non-Gaussian initial conditions," *Physical Review D*, vol. 62, no. 2, Article ID 021301, 5 pages, 2000.
- [162] E. Gaztañaga and P. Fosalba, "Cosmological perturbation theory and the spherical collapse model—II. Non-Gaussian initial conditions," *Monthly Notices of the Royal Astronomical Society*, vol. 301, no. 2, pp. 524–534, 1998.
- [163] E. Gaztanaga and P. Mähönen, "Large-scale clustering from non-Gaussian texture models," *The Astrophysical Journal Letters*, vol. 462, no. 1, pp. L1–L4, 1996.
- [164] E. Gaztanaga, "High-order galaxy correlation functions in the APM galaxy survey," *Monthly Notices of the Royal Astronomical Society*, vol. 268, no. 4, p. 913, 1994.
- [165] R. Scoccimarro, "The bispectrum: from theory to observations," *The Astrophysical Journal*, vol. 544, no. 2, pp. 597–615, 2000.
- [166] J. N. Fry, "The Galaxy correlation hierarchy in perturbation theory," *The Astrophysical Journal*, vol. 279, no. 2, pp. 499–510, 1984.
- [167] A. Taruya, K. Koyama, and T. Matsubara, "Signature of primordial non-Gaussianity on the matter power spectrum," *Physical Review D*, vol. 78, no. 12, Article ID 123534, 13 pages, 2008.
- [168] M. Takada and B. Jain, "Cosmological parameters from lensing power spectrum and bispectrum tomography," *Monthly Notices of the Royal Astronomical Society*, vol. 348, no. 3, pp. 897–915, 2004.
- [169] R. Scoccimarro, "Cosmological perturbations: entering the nonlinear regime," *The Astrophysical Journal*, vol. 487, no. 1, pp. 1–17, 1997.
- [170] R. Scoccimarro, S. Colombi, J. N. Fry, J. A. Frieman, E. Hivon, and A. Melott, "Nonlinear evolution of the bispectrum of cosmological perturbations," *The Astrophysical Journal*, vol. 496, no. 2, pp. 586–604, 1998.
- [171] E. Sefusatti, "One-loop perturbative corrections to the matter and galaxy bispectrum with non-Gaussian initial conditions," *Physical Review D*, vol. 80, no. 12, Article ID 123002, 24 pages, 2009.
- [172] E. Sefusatti, M. Crocce, and V. Desjacques, "The matter bispectrum in N-body simulations with non-Gaussian initial conditions," <http://arxiv.org/abs/1003.0007>.
- [173] F. Bernardeau, M. Crocce, and R. Scoccimarro, "Multipoint propagators in cosmological gravitational instability," *Physical Review D*, vol. 78, no. 10, Article ID 103521, 21 pages, 2008.
- [174] M. Crocce and R. Scoccimarro, "Memory of initial conditions in gravitational clustering," *Physical Review D*, vol. 73, no. 6, Article ID 063520, 13 pages, 2006.
- [175] M. Crocce and R. Scoccimarro, "Renormalized cosmological perturbation theory," *Physical Review D*, vol. 73, no. 6, Article ID 063519, 14 pages, 2006.
- [176] S. Matarrese and M. Pietroni, "Resumming cosmic perturbations," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 6, article 26, 2007.
- [177] M. Pietroni, "On the physical significance of infra-red corrections to inflationary observables," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 10, article 36, 2008.
- [178] N. Bartolo, J. P. Beltrn Almeida, S. Matarrese, M. Pietroni, and A. Riotto, "Signatures of primordial non-Gaussianities in the matter power-spectrum and bispectrum: the time-RG approach," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 3, article 11, 2010.
- [179] J. N. Fry, "Gravity, bias, and the galaxy three-point correlation function," *Physical Review Letters*, vol. 73, no. 3, pp. 215–219, 1994.
- [180] D. Jeong and E. Komatsu, "Perturbation theory reloaded. II. Nonlinear bias, baryon acoustic oscillations, and millennium simulation in real space," *The Astrophysical Journal*, vol. 691, no. 1, pp. 569–595, 2009.
- [181] S. Matarrese, L. Verde, and A. F. Heavens, "Large-scale bias in the universe: bispectrum method," *Monthly Notices of the Royal Astronomical Society*, vol. 290, no. 4, pp. 651–662, 1997.
- [182] T. Nishimichi, A. Taruya, K. Koyama, and C. Sabiu, "Scale dependence of halo bispectrum from non-Gaussian initial conditions in cosmological N-body simulations," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 7, article 002, 2010.
- [183] R. Scoccimarro, H. M. P. Couchman, and J. A. Frieman, "The bispectrum as a signature of gravitational instability in redshift space," *The Astrophysical Journal*, vol. 517, no. 2, pp. 531–540, 1999.
- [184] E. Sefusatti, M. Crocce, S. Pueblas, and R. Scoccimarro, "Cosmology and the bispectrum," *Physical Review D*, vol. 74, no. 2, Article ID 023522, 24 pages, 2006.
- [185] E. Sefusatti and R. Scoccimarro, "Galaxy bias and halo-occupation numbers from large-scale clustering," *Physical Review D*, vol. 71, no. 6, Article ID 063001, 13 pages, 2005.
- [186] H. Guo and Y. P. Jing, "A determination of dark matter bispectrum with a large set of n-body simulations," *The Astrophysical Journal*, vol. 698, no. 1, pp. 479–487, 2009.
- [187] H. Guo and Y. P. Jing, "Determine the galaxy bias factors on large scales using the bispectrum method," *The Astrophysical Journal*, vol. 702, no. 1, pp. 425–432, 2009.
- [188] R. E. Smith, R. K. Sheth, and R. Scoccimarro, "Analytic model for the bispectrum of galaxies in redshift space," *Physical Review D*, vol. 78, no. 2, Article ID 023523, 26 pages, 2008.
- [189] M. Colless, G. Dalton, S. Maddox et al., "The 2dF Galaxy Redshift Survey: spectra and redshifts," *Monthly Notices of the Royal Astronomical Society*, vol. 328, no. 4, pp. 1039–1063, 2001.

- [190] D. G. York, J. Adelman, J. E. Anderson Jr., et al., “The sloan digital sky survey: technical summary,” *The Astrophysical Journal*, vol. 120, no. 3, pp. 1579–1587, 2000.
- [191] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, “Large-scale structure, the cosmic microwave background and primordial non-Gaussianity,” *Monthly Notices of the Royal Astronomical Society*, vol. 313, no. 1, pp. 141–147, 2000.
- [192] R. Scoccimarro, E. Sefusatti, and M. Zaldarriaga, “Probing primordial non-Gaussianity with large-scale structure,” *Physical Review D*, vol. 69, no. 10, Article ID 103513, 12 pages, 2004.
- [193] E. Gaztañaga and R. Scoccimarro, “The three-point function in large-scale structure: redshift distortions and galaxy bias,” *Monthly Notices of the Royal Astronomical Society*, vol. 361, no. 3, pp. 824–836, 2005.
- [194] H. A. Feldman, N. Kaiser, and J. A. Peacock, “Power-spectrum analysis of three-dimensional redshift surveys,” *The Astrophysical Journal*, vol. 426, no. 1, pp. 23–37, 1994.
- [195] A. J. S. Hamilton, “Power spectrum estimation I. Basics,” <http://arxiv.org/abs/astro-ph/0503603>.
- [196] A. J. S. Hamilton, “Power spectrum estimation II. Linear maximum likelihood,” <http://arxiv.org/abs/astro-ph/0503604>.
- [197] E. Sefusatti and E. Komatsu, “Bispectrum of galaxies from high-redshift galaxy surveys: primordial non-Gaussianity and nonlinear galaxy bias,” *Physical Review D*, vol. 76, no. 8, Article ID 083004, 17 pages, 2007.
- [198] W. Saunders, W. J. Sutherland, S. J. Maddox, et al., “The PSCz catalogue,” *Monthly Notices of the Royal Astronomical Society*, vol. 317, no. 1, pp. 55–64, 2002.
- [199] R. Scoccimarro, H. A. Feldman, J. N. Fry, and J. A. Frieman, “The bispectrum of IRAS redshift catalogs,” *The Astrophysical Journal*, vol. 546, no. 2, pp. 652–664, 2001.
- [200] K. B. Fisher, J. P. Huchra, M. A. Strauss, M. Davis, A. Yahil, and D. Schlegel, “The IRAS 1.2 Jy survey: redshift data,” *The Astrophysical Journal, Supplement Series*, vol. 100, no. 1, pp. 69–103, 1995.
- [201] M. A. Strauss, J. P. Huchra, M. Davis, A. Yahil, K. B. Fisher, and J. Tonry, “A redshift survey of IRAS galaxies. VII. The infrared and redshift data for the 1.936 jansky sample,” *The Astrophysical Journal, Supplement Series*, vol. 83, no. 1, pp. 29–63, 1992.
- [202] H. A. Feldman, J. A. Frieman, J. N. Fry, and R. Scoccimarro, “Constraints on galaxy bias, matter density, and primordial non-Gaussianity from the PSCz galaxy redshift survey,” *Physical Review Letters*, vol. 86, no. 8, pp. 1434–1437, 2001.
- [203] R. Scoccimarro, “Gravitational clustering from  $X^2$  initial conditions,” *The Astrophysical Journal*, vol. 542, no. 1, pp. 1–5, 2000.
- [204] S. Cole and N. Kaiser, “Biased clustering in the cold dark matter cosmogony,” *Monthly Notices of the Royal Astronomical Society*, vol. 237, pp. 1127–1146, 1989.
- [205] T. Giannantonio and C. Porciani, “Structure formation from non-Gaussian initial conditions: multivariate biasing, statistics, and comparison with N-body simulations,” *Physical Review D*, vol. 81, no. 6, Article ID 063530, 10 pages, 2010.
- [206] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, “Constraints on local primordial non-Gaussianity from large scale structure,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 31, 2008.
- [207] P. McDonald, “Primordial non-Gaussianity: arge-scale structure signature in the perturbative bias model,” *Physical Review D*, vol. 78, no. 12, Article ID 123519, 9 pages, 2008.
- [208] S. Matarrese and L. Verde, “The effect of primordial non-Gaussianity on halo bias,” *The Astrophysical Journal Letters*, vol. 677, no. 2, pp. L77–L80, 2008.
- [209] L. Verde and S. Matarrese, “Detectability of the effect of inflationary non-Gaussianity on halo bias,” *The Astrophysical Journal Letters*, vol. 706, no. 1, pp. L91–L95, 2009.
- [210] S. Ho, C. Hirata, N. Padmanabhan, U. Seljak, and N. A. Bahcall, “Correlation of CMB with large-scale structure. I. Integrated Sachs-Wolfe tomography and cosmological implications,” *Physical Review D*, vol. 78, no. 4, Article ID 043519, 35 pages, 2008.
- [211] C. Carbone, L. Verde, and S. Matarrese, “Non-Gaussian halo bias and future galaxy surveys,” *The Astrophysical Journal Letters*, vol. 684, no. 1, pp. L1–L4, 2008.
- [212] D. Jeong and E. Komatsu, “Primordial non-Gaussianity, scale-dependent bias, and the bispectrum of galaxies,” *The Astrophysical Journal*, vol. 703, no. 2, pp. 1230–1248, 2009.
- [213] T. Nishimichi, I. Kayo, C. Hikage et al., “Bispectrum and non-linear biasing of galaxies: perturbation analysis, numerical simulation, and SDSS galaxy clustering,” *Publications of the Astronomical Society of Japan*, vol. 59, no. 1, pp. 93–106, 2007.
- [214] X. Chen, “Running non-Gaussianities in Dirac-Born-Infeld inflation,” *Physical Review D*, vol. 72, no. 12, Article ID 123518, 7 pages, 2005.
- [215] J. Khoury and F. Piazza, “Rapidly-varying speed of sound, scale invariance and non-Gaussian signatures,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 7, article 26, 2009.
- [216] S. E. Shandera and S.-H. H. Tye, “Observing brane inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 7, 2006.
- [217] C. T. Byrnes, K. Y. Choi, and L. M.H. Hall, “Large non-Gaussianity from two-component hybrid inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 2, article 17, 2009.
- [218] C. T. Byrnes, S. Nurmi, G. Tasinato, and D. Wands, “Scale dependence of local  $F_{NL}$ ,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 2, article 034, 2010.
- [219] C. T. Byrnes and G. Tasinato, “Non-Gaussianity beyond slow roll in multi-field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 16, 2009.
- [220] J. Kumar, L. Leblond, and A. Rajaraman, “Scale dependent local non-Gaussianity from loops,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 4, article 24, 2010.
- [221] M. Kendall and A. Stuart, *The Advanced Theory of Statistics. Vol.2: Inference and Relationship*, Griffin, London, UK, 4th edition, 1979.
- [222] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN. The Art of Scientific Computing*, Cambridge University Press, Cambridge, UK, 1992.
- [223] V. J. Martinez, E. Saar, E. M. Gonzales, and M. J. Pons-Borderia, Eds., *Data Analysis in Cosmology*, vol. 665 of *Lecture Notes in Physics*, Springer, Berlin, Germany, 2009.
- [224] M. Tegmark, A. N. Taylor, and A. F. Heavens, “Karhunen-Loève eigenvalue problems in cosmology: how should we tackle large data sets?” *The Astrophysical Journal*, vol. 480, no. 1, pp. 22–35, 1997.

## Review Article

# Large-Angle Anomalies in the CMB

**Craig J. Copi,<sup>1</sup> Dragan Huterer,<sup>2</sup> Dominik J. Schwarz,<sup>3</sup> and Glenn D. Starkman<sup>4</sup>**

<sup>1</sup> CERCA, Department of Physics, Case Western Reserve University, Cleveland, OH 44106-7079, USA

<sup>2</sup> Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, MI 48109-1040, USA

<sup>3</sup> Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany

<sup>4</sup> CERCA/ISO, Department of Physics, Case Western Reserve University, Cleveland, OH 44106-7079, USA

Correspondence should be addressed to Dragan Huterer, [huterer@umich.edu](mailto:huterer@umich.edu)

Received 30 April 2010; Revised 7 August 2010; Accepted 19 August 2010

Academic Editor: Eiichiro Komatsu

Copyright © 2010 Craig J. Copi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We review the recently found large-scale anomalies in the maps of temperature anisotropies in the cosmic microwave background. These include alignments of the largest modes of CMB anisotropy with each other and with geometry and direction of motion of the solar system, and the unusually low power at these largest scales. We discuss these findings in relation to expectation from standard inflationary cosmology, their statistical significance, the tools to study them, and the various attempts to explain them.

## 1. Introduction: Why Large Scales are Interesting?

The Copernican principle states that the Earth does not occupy a special place in the universe and that observations made from Earth can be taken to be broadly characteristic of what would be seen from any other point in the universe at the same epoch. The microwave sky is isotropic, apart from a Doppler dipole and a microwave foreground from the Milky Way. Together with the Copernican principle and some technical assumptions, an oft-inferred consequence is the so-called cosmological principle. It states that the distributions of matter and light in the Universe are homogeneous and isotropic at any epoch and thus also defines what we mean by cosmic time.

This set of assumptions is a crucial, implicit ingredient in obtaining most important results in quantitative cosmology, for example, it allows us to treat cosmic microwave background (CMB) temperature fluctuations in different directions in the sky as multiple probes of a single statistical ensemble, leading to the precision determinations of cosmological parameters that we have today.

Although we have some observational evidence that homogeneity and isotropy are reasonably good approximations to reality, neither of these are actual logical consequences of the Copernican principle, for example, the

geometry of space could be homogeneous but anisotropic—like the surface of a sharp mountain ridge, with a gentle path ahead but the ground dropping steeply away to the sides. Indeed, three-dimensional space admits not just the three well known homogeneous isotropic geometries (Euclidean, spherical and hyperbolic— $E^3$ ,  $S^3$ , and  $H^3$ ) but five others which are homogeneous but anisotropic. The two simplest are  $S^2 \times E^1$  and  $H^2 \times E^1$ . These spaces support the cosmological principle but have preferred directions.

Similarly, although the Earth might not occupy a privileged place in the universe, it is not necessarily true that all points of observation are equivalent, for example, the topology of space may not be simply connected—we could live in a three dimensional generalization of a torus so that if you travel far enough in certain directions you come back to where you started. While such three-spaces generically admit locally homogeneous and isotropic geometries, certain directions or points might be singled out when nonlocal measurements are considered, for example the length of the shortest closed nontrivial geodesic through a point depends on the location of that point within the fundamental domain. Similarly, the inhomogeneity and anisotropy of eigenmodes of differential operators on such spaces are likely to translate into statistically inhomogeneous and anisotropic large scale structure, in the manner of Chladni figures on vibrating plates.

The existence of nontrivial cosmic topology and of anisotropic geometry are questions that can only be answered observationally. In this regard, it is worth noting that our record at predicting the gross properties of the universe on large scales from first principles has been rather poor. According to the standard concordance model of cosmology, over 95% of the energy content of the universe is extraordinary—dark matter or dark energy whose existence has been inferred from the failure of the Standard Model of particle physics plus General Relativity to describe the behavior of astrophysical systems larger than a stellar cluster—while the very homogeneity and isotropy (and inhomogeneity) of the universe owe to the influence of an inflaton field whose particle-physics-identity is completely mysterious even after three decades of theorizing.

The stakes are set even higher with the recent discovery of dark energy that makes the universe undergo accelerated expansion. It is known that dark energy can affect the largest scales of the universe, for example, the clustering scale of dark energy may be about the horizon size today. Similarly, inflationary models can induce observable effects on the largest scales via either explicit or spontaneous violations of statistical isotropy. It is reasonable to suggest that statistical isotropy and homogeneity should be substantiated observationally, not just assumed. More generally, testing the cosmological principle should be one of the key goals of modern observational cosmology.

With the advent of high signal-to-noise maps of the cosmic microwave background anisotropies and with the conduct of nearly-full-sky deep galaxy surveys, statistical isotropy *has* begun to be precisely tested. Extraordinary full-sky temperature maps produced by the Wilkinson Microwave Anisotropy Probe (WMAP), in particular, are revolutionizing our ability to probe the universe on its largest scales [1–6]. In the near future, these will be joined by higher resolution temperature maps and high-resolution polarization maps and, eventually, by deep all-sky surveys, and perhaps by tomographic 21-cm line observations that will extend our detailed knowledge of the universe’s background geometry and fluctuations into the interior of the sphere of last scattering.

In this brief paper, we describe the large-scale anomalies in the CMB data, some of which were first reported on by the Cosmic Background Explorer (COBE) Differential Microwave Radiometer (DMR) collaboration in the mid 1990s. In particular, we report on alignments of the largest modes of CMB anisotropy with each other, and with geometry and direction of motion of the Solar System, as well as on unusually low angular correlations at the largest angular scales. We discuss these findings and, as this is not meant to be a comprehensive review and we emphasize results based on our own work in the area, we refer the reader to literature for all developments in the field. This paper extends an earlier review on the subject by Huterer [7], and complements another review on statistical isotropy in this special issue [8].

The paper is organized as follows. In Section 2 we describe the statistical quantities that describe the CMB and the expectations for their values in the currently favored

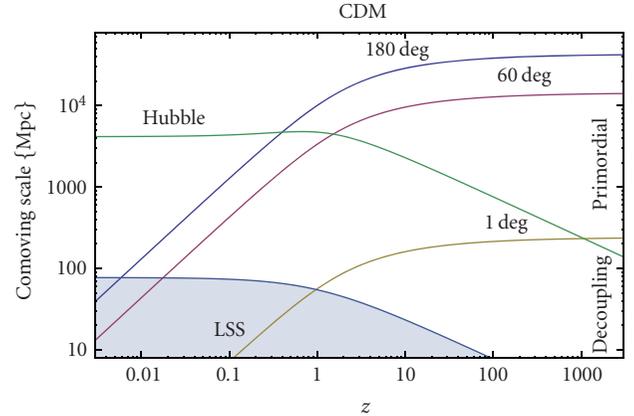


FIGURE 1: The comoving length within the context of the concordance model of an arc seen at a fixed angle and the comoving Hubble length as functions of redshift. Linear perturbation theory is expected to work well outside the shaded region, in which the large scale structure (LSS) forms.

$\Lambda$ CDM cosmological model. In Section 3, we describe the alignments at the largest scales, as well as multipole vectors, which is a tool to study them. In Section 4, we describe findings of low power at largest scales in the CMB. Section 5 categorizes and covers the variety of possible explanations for these anomalies. We conclude in Section 6.

## 2. Expectations from Cosmological Inflation

A fixed angular scale on the sky probes the physics of the universe at a range of physical distances corresponding to the range of observable redshifts. This is illustrated in Figure 1, where the comoving lengths of arcs at fixed angle are shown as a function of redshift, together with the comoving Hubble scale. Angles of 1 degree and less probe events that were in causal contact at all epochs between the redshift of decoupling and today; this redshift range includes physical processes such as the secondary CMB anisotropies. The situation is different for angles  $>60$  degrees, which subtend arcs that enter our Hubble patch only at  $z \lesssim 1$ . Therefore, the primordial CMB signal on such large angular scales could only be modified by the physics of local foregrounds and cosmology in the relatively recent past ( $z \lesssim 1$ ). Because they correspond to such large physical scales, the largest observable angular scales provide the most direct probe of the primordial fluctuations—whether generated during the epoch of cosmological inflation or preceding it.

**2.1. Statistical Isotropy.** What do we expect for the large angular scales of the CMB? A crucial ingredient of cosmology’s concordance model is cosmological inflation—a period of accelerating cosmic expansion in the early universe. If we assume that inflationary expansion persisted for sufficiently many e-folds, then we expect to live in a homogeneous and isotropic universe within a domain larger than our Hubble volume. This homogeneity and isotropy will not be exact but should characterize both the background and

the statistical distributions of matter and metric fluctuations around that background. These fluctuations are made visible as anisotropies of the CMB temperature and polarization, which are expected to inherit the underlying statistical isotropy. The temperature  $T$  seen in direction  $\hat{e}$  is predicted to be described by a Gaussian random field on the sky (i.e., the 2-sphere  $S^2$ ), which implies that we can expand it in terms of spherical harmonics  $Y_{\ell m}(\hat{e})$  multiplied by independent Gaussian random coefficients  $a_{\ell m}$  of zero mean.

Statistical isotropy implies that the expectation values of all  $n$ -point correlation functions (of the temperature or polarization) are invariant under arbitrary rotations of the sky. As a consequence the expectation of the temperature coefficients is zero,  $\langle a_{\ell m} \rangle = 0$ , for all  $\ell > 0$  and  $m = -\ell, -\ell + 1, \dots, +\ell$ . The two-point correlation becomes a function of  $\cos \theta \equiv \hat{e}_1 \cdot \hat{e}_2$  only and can be expanded in terms of Legendre polynomials:

$$\langle T(\hat{e}_1)T(\hat{e}_2) \rangle \equiv C(\theta) = \frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_{\ell} P_{\ell}(\cos \theta). \quad (1)$$

Statistical independence implies that expectations of  $a_{\ell m}$  with different  $\ell$  and  $m$  vanish. In particular, the two-point correlation function is diagonal in  $\ell$  and  $m$ :

$$\langle a_{\ell m}^* a_{\ell' m'} \rangle \propto \delta_{\ell \ell'} \delta_{m m'}. \quad (2)$$

Statistical isotropy adds that the constant of proportionality depends only on  $\ell$ , not  $m$ :

$$\langle a_{\ell m}^* a_{\ell' m'} \rangle = \delta_{\ell \ell'} \delta_{m m'} C_{\ell}. \quad (3)$$

The variance  $C_{\ell}$  is called the angular power of the multipole  $\ell$ . The higher  $n$ -point correlation functions are constrained in similar ways but, as we will see below, are not expected to provide independent information if a simple inflationary scenario was realized by nature.

**2.2. Gaussianity.** If inflation is driven by a single dynamically relevant degree of freedom with appropriate properties (minimal coupling, Minkowski vacuum in UV limit, etc.), then we can reduce the quantization of matter and space-time fluctuations during inflation to the problem of quantizing free scalar fields. For free fields the only nontrivial object is the two-point correlation (the propagator), and all higher correlation functions either vanish or are just some trivial combination of the two-point function. This property is mapped onto the temperature field of the CMB. A classical random field with these properties is a Gaussian with mean  $T_0$  and variance  $C(\theta)$ . Thus the brightness of the primordial CMB sky is completely characterized by  $T_0$  and  $C(\theta)$  (or  $C_{\ell}$ ). Note that evolution of perturbations leads to deviations from Gaussianity that would mostly be evident at very small scales ( $\ell \gg 100$ ). Moreover, many inflationary models predict small deviations from Gaussianity; these are described in other contributions to this volume [9, 10].

**2.3. Scale Invariance.** Another generic feature of inflation is the almost scale invariance of the power spectrum of

fluctuations. This can be understood easily, as the Hubble scale is approximately constant during inflation as the wavelengths of observable modes are redshifted beyond the horizon. Given that fluctuations of modes on horizon exit are related to the Hubble parameter,  $\delta\phi = H/2\pi$ , these modes have similar amplitudes. However, scale invariance is not exact. In canonical slow-roll inflation models, the deviation from exact scale invariance is due to the evolution of the Hubble parameter during inflation, which is measured by the so-called first slow-roll function  $\epsilon_1 \equiv \dot{d}_H$  where  $d_H \equiv H^{-1}$  is the Hubble distance. From the weak energy condition  $\epsilon_1 > 0$ , while  $\epsilon_1 \ll 1$  during slow-roll inflation.

At the level of the angular power spectrum, exact scale invariance implies the Sachs-Wolfe ‘‘plateau’’ (i.e., constancy of  $l(l+1)C_{\ell}$  at low  $\ell$ ) [11]

$$C_{\ell} = \frac{2\pi A}{\ell(\ell+1)}. \quad (4)$$

Here, again in the slow-roll parameterization,  $A \sim (H_{\text{infl}}/M_{\text{P}})^2 T_0^2 / \epsilon_1$ . This neglects secondary anisotropies like the late time, integrated Sachs-Wolfe effect (particularly important at very low  $\ell$ ) and the contribution from gravitational waves. Furthermore, inflation predicts a small departure from scale invariance, which has recently been detected (e.g., [6]), and which also contributes to a tilt in the aforementioned plateau.

**2.4. Cosmic Variance.** As we can measure only one sky, it is important to find the best estimators of  $C_{\ell}$  and  $C(\theta)$ . Let us for the moment assume that we are able to measure the primordial CMB of the full-sky, without any instrumental noise. We also restrict ourselves to  $\ell \geq 2$ , as the variance of the monopole cannot be defined and the measured dipole is dominated by our motion through the universe rather than by primordial physics. (Separation of the Doppler dipole from the intrinsic dipole is possible in principle [12, 13], but not with existing data.) Statistical isotropy suggests to estimate the angular power by

$$\hat{C}_{\ell} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{+\ell} |a_{\ell m}|^2, \quad (5)$$

which satisfies  $\langle \hat{C}_{\ell} \rangle = C_{\ell}$  and is thus unbiased. The variance of this estimator can be calculated assuming Gaussianity:

$$\text{Var}(\hat{C}_{\ell}) = \frac{2}{2\ell+1} \hat{C}_{\ell}^2. \quad (6)$$

It can be shown that, assuming statistical isotropy and Gaussianity,  $\hat{C}_{\ell}$  is the best estimator in the sense that it has minimal variance and is unbiased. However, we emphasize that these qualities depend intrinsically on the correctness of the underlying assumptions.

With these same assumptions, the variance of the two-point correlation function is easily shown to be

$$\text{Var}[\hat{C}(\theta)] = \frac{1}{8\pi^2} \sum_{\ell} (2\ell + 1) C_{\ell}^2 P_{\ell}^2(\cos \theta), \quad (7)$$

where  $\hat{C}(\theta)$  is calculated from  $\hat{C}_{\ell}$  following (1).

Putting the results of this section together allows us to come up with a generic prediction of inflationary cosmology for  $C(\theta)$  on the largest angular scales; see Figure 2.

### 3. Alignments

In brief, the upshot of the previous section is that the twin assumptions of statistical isotropy and Gaussianity are the starting point of *any* CMB analysis. The measurements of the CMB monopole, dipole, and  $(\Delta T)_{\text{rms}}$  tell us that isotropy is observationally established at the percent level without any cosmological assumption, and at a level  $10^{-4}$  if we attribute the dominant contribution to the dipole to our peculiar motion. For the purpose of cosmological parameter estimation, the task is to test the statistical isotropy of the CMB brightness fluctuations. At the largest angular scales, this can only be done by means of full-sky maps.

Let us assume that the various methods that have been developed to get rid of the Galactic foreground in single frequency band maps of the microwave sky are reliable (though we argue below that this might not be the case). Our review of alignments will be based on the internal linear combination (ILC) map produced by the WMAP team, which is based on a minimal variance combination of the WMAP frequency bands. The weights for the five frequency band maps are adjusted in 12 regions of the sky, one region lying outside the Milky Way and 11 regions along the Galactic plane.

**3.1. Multipole Vectors.** To study the orientation and alignment of CMB multipoles, Copi et al. [15] introduced to cosmology the use of multipole vectors; an alternative representation of data on the sphere. The multipole vectors contain information about the “directions” associated with each multipole. In this new basis the temperature fluctuation multipole,  $\ell$ , may be expanded as

$$T_{\ell} \equiv \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m} \equiv A^{(\ell)} \left[ \prod_{i=1}^{\ell} (\hat{v}^{(\ell,i)} \cdot \hat{e}) - \mathcal{T}_{\ell} \right], \quad (8)$$

where  $\hat{v}^{(\ell,i)}$  is the  $i$ th vector for the  $\ell$ th multipole,  $\hat{e}$  is the usual radial unit vector,  $\mathcal{T}_{\ell}$  is the trace of the preceding product of multipole vector terms, and  $A^{(\ell)}$  is the “power” in the multipole. By construction, we immediately see that  $T_{\ell}$  is a symmetric traceless, rank  $\ell$  tensor. Subtracting  $\mathcal{T}_{\ell}$  ensures that  $T_{\ell}$  is traceless and the dot products explicitly show this is a rotationally invariant quantity (a scalar under rotations). This form makes the symmetry properties obvious. As an example the quadrupole is written as

$$T_2 = A^{(2)} \left[ (\hat{v}^{(2,1)} \cdot \hat{e}) (\hat{v}^{(2,2)} \cdot \hat{e}) - \frac{1}{3} \hat{v}^{(2,1)} \cdot \hat{v}^{(2,2)} \right]. \quad (9)$$

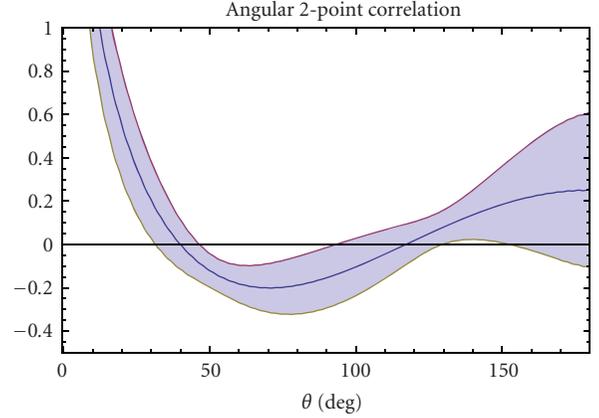


FIGURE 2: Mean and (cosmic) variance of the angular two-point correlation function as expected from cosmological inflation (arbitrary normalization). Only statistical isotropy, Gaussianity and scale invariance are assumed. Tensors, spectral tilt, reionization and the integrated Sachs-Wolfe effect are neglected for the purpose of this plot. Comparison to the prediction from the best-fit  $\Lambda$ CDM model (Figure 5) reveals that these corrections are subdominant. Note that cosmic variance errors at different values of  $\theta$  are very highly correlated.

These two forms for representing  $T_{\ell}$ , harmonic, and multipole-vector, both contain the same information. At the same time, they are fundamentally different from each other. Each unit vector  $\hat{v}^{(\ell,i)}$  has two degrees of freedom while the scalar,  $A^{(\ell)}$  has one; thus the multipole vector representation contains the full  $2\ell + 1$  degrees of freedom. Note that we call each  $\hat{v}^{(\ell,i)}$  a multipole vector but it is only defined up to a sign. We can always reflect the vector through the origin by absorbing a negative sign into the scalar  $A^{(\ell)}$ ; thus these vectors actually are headless. Regardless, we will continue to refer to them as multipole vectors and not use the overall sign of the vector in our analysis. This issue is equivalent to choosing a phase convention, such as the Condon-Shortley phase for the spherical harmonics. For the work reviewed here the overall phase is not relevant and thus will not be specified.

An efficient algorithm to compute the multipole vectors for low- $\ell$  has been presented in [15] and is publicly available [16]; other algorithms have been proposed as well in [17–19]. Interestingly, after the publication of the CHS paper [15], Weeks [18] pointed out that multipole vectors were actually first used by Maxwell [20] more than 100 years ago in his study of multipole moments in electrodynamics. They remain in use in geometrology, nuclear physics, and other fields.

The relation between multipole vectors and the usual harmonic basis is very much the same as that between Cartesian and spherical coordinates of standard geometry: both are complete bases, but specific problems are much more easily addressed in one basis than the other. In particular, we and others have found that multipole vectors are particularly well suited for tests of planarity and alignment of the CMB anisotropy pattern. Moreover, a number of

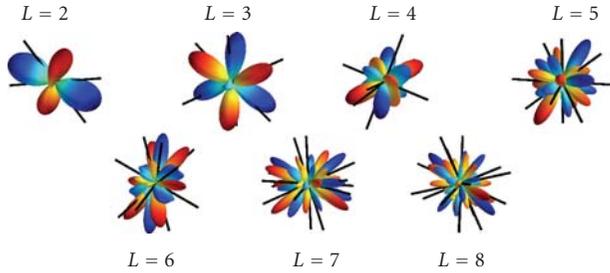


FIGURE 3: Multipole vectors of our sky, based on WMAP five-year full-sky ILC map and with galactic plane coinciding with the plane of the page. The temperature pattern at each multipole  $\ell$  ( $2 \leq \ell \leq 8$ ) can either be described by an angular temperature pattern (colored lobes in this figure), or alternatively by precisely  $\ell$  multipole vectors (black “sticks”). While the multipole vectors contain all information about the directionality of the CMB temperature pattern, they are not simply related to the hot and cold spots and, for example, do not correspond to the temperature minima/maxima [14]. Notice that  $\ell = 2$  and 3 temperature patterns are rather planar with the same plane and that their vectors lie approximately in this plane. Adopted from [15].

interesting theoretical results have been found; for example, Dennis and Phys [21] analytically computed the two-point correlation function of multipole vectors for a Gaussian random, isotropic underlying field. Numerous quantities have been proposed for assigning directions to multipoles and statistics on these quantities have been studied. In the work of Copi et al. [14] we have summarized these attempts and have shown their connections to the multipole vectors.

**3.2. Planarity and Alignments.** Tegmark et al. [22] and de Oliveira-Costa et al. [23] first argued that the octopole is planar and that the quadrupole and octopole planes are aligned. In the work of Schwarz et al. [24], followed up by Copi et al. [14, 25], we investigated the quadrupole-octopole shape and orientation using the multipole vectors. The quadrupole is fully described by two multipole vectors, which define a plane. This plane can be described by the “oriented area” vector

$$\vec{w}^{(\ell;i,j)} \equiv \hat{v}^{(\ell,i)} \times \hat{v}^{(\ell,j)}. \quad (10)$$

(Note that the oriented area vector *does not* fully characterize the quadrupole, as pairs of quadrupole multipole vectors related by a rotation about the oriented area vector lead to the same oriented area vector.) The octopole is defined by three multipole vectors which determine (but again are not fully determined by) three area vectors. Hence there are a total of four planes determined by the quadrupole and octopole.

In the work of Copi et al. [25] we found that (see Figure 4)

- (i) the four area vectors of the quadrupole and octopole are mutually close (i.e., the quadrupole and octopole planes are aligned) at the 99.6% C.L.;
- (ii) the quadrupole and octopole planes are orthogonal to the ecliptic at the 95.9% C.L.; this alignment was at

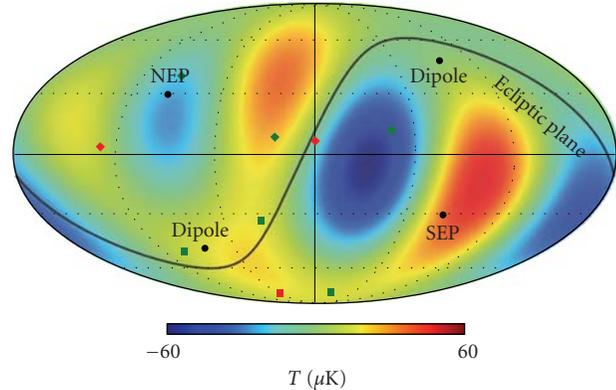


FIGURE 4: Quadrupole and octopole ( $\ell = 2$  and 3) temperature anisotropy of the WMAP sky map in galactic coordinates, shown with the ecliptic plane and the cosmological dipole. Included are the multipole vectors (solid diamonds): two for the quadrupole (red diamonds) and three for the octopole (green diamonds). We also show the four normals (solid squares) to the planes defined by vectors that describe the quadrupole and octopole temperature anisotropy; one normal is defined by the quadrupole (red square) and three by the octopole (green squares). Note that three out of four normals lie very close to the dipole direction. The probability of this alignment being accidental is about one part in a thousand. Moreover, the ecliptic plane traces out a locus of zero of the combined quadrupole and octopole over a broad swath of the sky—neatly separating a hot spot in the northern sky from a cold spot in the south. These apparent correlations with the solar system geometry are puzzling and currently unexplained.

98.5% C.L. in our analysis of the WMAP 1 year maps. The reduction of alignment was due to WMAP’s adaption of a new radiometer gain model for the 3 year data analysis, that took seasonal variations of the receiver box temperature into account—a systematic that is indeed correlated with the ecliptic plane. We regard that as clear evidence that multipole vectors are a sensitive probe of alignments;

- (iii) the normals to these four planes are aligned with the direction of the cosmological dipole (and with the equinoxes) at a level inconsistent with Gaussian random, statistically isotropic skies at 99.7% C.L.;
- (iv) the ecliptic threads between a hot and a cold spot of the combined quadrupole and octopole map, following a node line across about 1/3 of the sky and separating the three strong extrema from the three weak extrema of the map; this is unlikely at about the 95% C.L.

These numbers refer to the WMAP ILC map from three years of data; other maps give similar results. Moreover, correction for the kinematic quadrupole—slight modification of the quadrupole due to our motion through the CMB rest frame—must be made and increases significance of the alignments. See [25, Table 3] for the illustration of both of these points.

While not all of these alignments are statistically independent, their combined statistical significance is certainly

greater than their individual significances; for example, given their mutual alignments, the conditional probability of the four normals lying so close to the ecliptic is less than 2%; the combined probability of the four normals being both so aligned with each other and so close to the ecliptic is less than  $0.4\% \times 2\% = 0.008\%$ . These are therefore clearly surprising, highly statistically significant anomalies—unexpected in the standard inflationary theory and the accepted cosmological model.

Particularly puzzling are the alignments with solar system features. CMB anisotropy should clearly not be correlated with our local habitat. While the observed correlations seem to hint that there is contamination by a foreground or perhaps by the scanning strategy of the telescope, closer inspection reveals that there is no obvious way to explain the observed correlations. Moreover, if their explanation is that they are a foreground, then that will likely exacerbate other anomalies that we will discuss in Section 4.2.

Our studies (see [14]) indicate that the observed alignments are with the ecliptic plane, with the equinox, or with the CMB dipole, and *not* with the Galactic plane: the alignments of the quadrupole and octopole planes with the equinox/ecliptic/dipole directions are much more significant than those for the Galactic plane. Moreover, it is remarkably curious that it is precisely the ecliptic alignment that has been found on somewhat smaller scales using the power spectrum analyses of statistical isotropy [26–29].

Finally, it is important to make sure that the results are unbiased by unfairly chosen statistics. We have studied this issue extensively in [14], and here we briefly describe the principal statistics used to quantify the probability of alignments quoted just above.

To define statistics we first compute the three dot-products between the quadrupole area vector and the three octopole area vectors:

$$A_k \equiv \left| \vec{w}^{(2;1,2)} \cdot \vec{w}^{(3;i,j)} \right|. \quad (11)$$

The absolute value is included since the multipole vectors are headless; thus each  $A_k$  lies in the interval  $[0, 1]$ . Two natural choices of statistics that are independent of the ordering of  $A_k$  are

$$S \equiv \frac{1}{3}(A_1 + A_2 + A_3), \quad (12)$$

$$T \equiv 1 - \frac{1}{3}[(1 - A_1)^2 + (1 - A_2)^2 + (1 - A_3)^2].$$

Both  $S$  and  $T$  statistics can be viewed as the suitably defined “distance” to the vertex  $(A_1, A_2, A_3) = (0, 0, 0)$ . A third obvious choice,  $(A_1^2 + A_2^2 + A_3^2)/3$ , is just  $2S - T$ . To test alignment of the quadrupole and octopole planes with one another we quoted the  $S$  statistic numbers;  $T$  gives similar results.

Alternatively, generalizing the definition in [22], one can find, for each  $\ell$ , the choice,  $n_\ell$ , of  $z$  axis that maximizes the angular momentum dispersion

$$\hat{L}_\ell^2 \equiv \frac{\sum_{m=-\ell}^{\ell} m^2 |a_{\ell m}|^2}{\ell^2 \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2}. \quad (13)$$

One can then compare the maximized value with that from simulated isotropic skies [14]. Because  $\ell = 2, 3$  are both planar (the quadrupole trivially so, the octopole because the three planes of the octopole are nearly parallel), the direction that maximizes the angular momentum dispersion of each is nearly the same as the (average) direction of that multipole’s planes. Thus, the alignment of the octopole and quadrupole can be seen either from the  $\mathcal{S}$  statistic, or by looking at the alignment of  $n_2$  with  $n_3$ .

To test alignments of multipole planes with physical directions, we find the plane whose normal,  $\hat{n}$ , has the largest dot product with the sum of the four quadrupole and octopole area vectors [14]. Again, since  $\vec{w}_i \cdot \hat{n}$  is defined only up to a sign, we take the absolute value of each dot product. Therefore, we find the direction  $\hat{n}$  that maximizes

$$\mathcal{S} \equiv \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} |\vec{w}_i \cdot \hat{n}|. \quad (14)$$

**3.3. Summary.** The study of alignments in the low  $\ell$  CMB has found a number of peculiarities. We have shown that the alignment of the quadrupole and octopole planes is inconsistent with Gaussian, statistically isotropic skies at least at the 99% confidence level. Further a, number of (possibly related) alignments occur at 95% confidence levels or greater. Putting together these provides a strong indication that the full-sky CMB WMAP maps are inconsistent with the standard cosmological model at the large-angles. Even more peculiar is the alignment of the quadrupole and octopole with solar system features (the ecliptic plane and the dipole).

This is strongly suggestive of an unknown systematic in the data reduction; however, careful scrutiny has revealed no such systematic (except the mentioned modification of the radiometer gain model, that leads to a reduction of ecliptic alignment); see Sections 5.3 and 5.4 for further discussion of the data analysis and instrumental explanations. We again stress that these results hold for full-sky maps; maps that are produced through combination of the individual frequency maps in such a way as to remove foregrounds. An alternative approach that removes the need for full-sky maps is presented in the next section.

## 4. Two-point Angular Correlation Function

The usual CMB analysis solely involves the spherical harmonic decomposition and the two-point angular power spectrum. There are many reasons for this. Firstly, when working with a statistically isotropic universe the angular power spectrum contains all of the physical information. Secondly, the standard theory predicts the  $a_{\ell m}$  and their statistical properties, through the  $C_\ell$ , thus the spherical harmonic basis is a natural one to employ. Finally, as measured today the angular size of the horizon at the time of last scattering is approximately 1 degree. Since  $\theta(\text{deg})200/\ell$ , the causal physics at the surface of last scattering leaves its imprint on the CMB on small scales,  $\theta \lesssim 1^\circ$  or  $\ell \gtrsim 100$ . The two-point angular power spectrum focuses on these small scales, making it a good means of exploring the physics of

the last scattering surface. The tremendous success of the standard model of cosmology has been the agreement of the theory and observations on these small scales, allowing for the precise determination of the fundamental cosmological parameters [30].

The two-point angular correlation function provides another means of analyzing CMB observations and should not be ignored even if, in principle, it contains the same information as the angular power spectrum. Thus, even in the case of full-sky observations and/or statistical isotropy there are benefits in looking at the data in different ways. The situation is similar to a function in one dimension where it is widely appreciated that features easily found in the real space analysis can be very difficult to find in the Fourier transform, and vice versa. Furthermore, the two-point angular correlation function highlights behavior at large-angles (small  $\ell$ ); the opposite of the two-point angular power spectrum. Thus the angular correlation function allows for easier study of the temperature fluctuation modes that are super-horizon sized at the time of last scattering. Finally, the angular correlation function in its simplest form is a direct pixel-based measure (see below). Thus it does not rely on the reconstruction of contaminated regions of the sky to employ. This makes it a simple, robust measure even for partial sky coverage.

**4.1. Definition.** Care should be taken when discussing statistical quantities of the CMB and their estimators. It rarely is in the literature. Here we follow the notation of Copi et al. [25], also see [31]. The two-point angular correlation function,

$$\tilde{\mathcal{C}}(\hat{e}_1, \hat{e}_2) \equiv \langle T(\hat{e}_1)T(\hat{e}_2) \rangle, \quad (15)$$

is the *ensemble average* (represented by the angle brackets,  $\langle \cdot \rangle$ ) of the product of the temperatures in the directions  $\hat{e}_1$  and  $\hat{e}_2$ . Unfortunately we only have one universe to observe so this ensemble average cannot be calculated. Instead we average over the sky so that what we mean by the two-point angular correlation function is a *sky average*,

$$\mathcal{C}(\theta) \equiv \overline{T(\hat{e}_1)T(\hat{e}_2)}, \quad (16)$$

where the average is over all pairs of pixels with  $\hat{e}_1 \cdot \hat{e}_2 = \cos \theta$ . This is a pixel-based quantity and can be calculated for any region of the sky (of course not all separations  $\theta$  may be represented on a given patch of the sky, depending on its geometry).

**4.2. Missing Angular Power at Large Scales.** Spergel et al. [2] found that the two-point correlation function nearly vanishes on scales greater than about 60 degrees, contrary to what the standard  $\Lambda$ CDM theory predicts, and in agreement with the same finding obtained from COBE data about a decade earlier [32].

We have revisited the angular two-point function in the 3-yr WMAP data in [25] and the 5-yr WMAP data in [31]; see Figure 5. From this figure we qualitatively see the following.

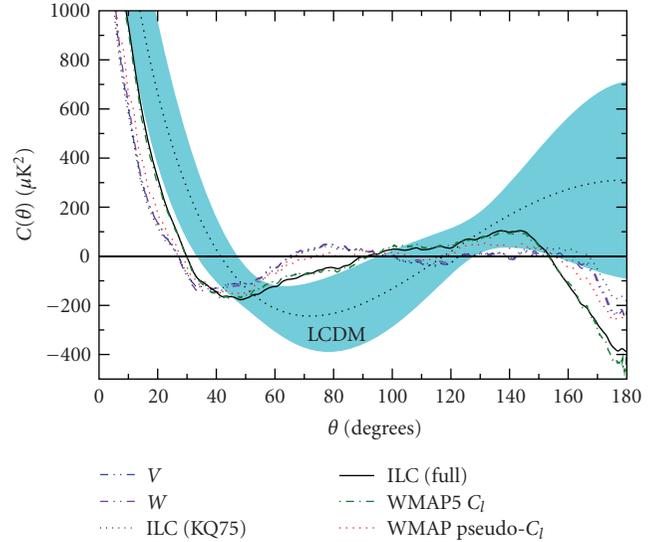


FIGURE 5: Two-point angular correlation function,  $\mathcal{C}(\theta)$ , computed in pixel space, for three different bands masked with the KQ75 mask (from WMAP 5 year data). Also shown are the correlation function for the ILC map with and without the mask, and the value expected for a statistically isotropic sky with best-fit  $\Lambda$ CDM cosmology together with 68% cosmic variance error bars. Even by eye, it is apparent that masked maps have  $\mathcal{C}(\theta)$  that is consistent with zero at  $\theta \gtrsim 60$  deg. We also show the  $\mathcal{C}(\theta)$  computed from the “official” published maximum likelihood estimator-based  $C_\ell$ . Clearly, the MLE-based  $C_\ell$ , as well as  $\mathcal{C}(\theta)$  computed from the full-sky ILC maps, is in significant disagreement with the angular correlation function computed from cut-sky maps. Adopted from [31].

- (i) All of the cut-sky map curves are very similar to each other, and they are also very similar to the Legendre transform of the pseudo- $C_\ell$  estimate of the angular power spectrum, which is not surprising given that the two are formally equivalent [33]. Meanwhile the full-sky ILC  $\mathcal{C}(\theta)$  and the Legendre transform of the maximum likelihood estimator (MLE) of the  $C_\ell$  agree well with each other, but not with any of the others.
- (ii) The most striking feature of the cut-sky (and pseudo- $C_\ell$ )  $\mathcal{C}(\theta)$ , is that all of them are very nearly zero above about 60°, except for some anticorrelation near 180°. This is also true for the full-sky curves, but less so.

In order to be more quantitative about these observations we adopt the  $S_{1/2}$  statistic introduced by the WMAP team [2] which quantifies the deviation of the two-point correlation function from zero:

$$S_{1/2} \equiv \int_{-1}^{1/2} [\mathcal{C}(\theta)]^2 d(\cos \theta). \quad (17)$$

Spergel et al. [2] found that only 0.15% of the parameter sets in their Markov chain of  $\Lambda$ CDM model CMB skies had lower values of  $S_{1/2}$  than the observed one-year WMAP sky.

Applying this statistic we have found that the two-point function computed from the various cut-sky maps shows an even stronger lack of power, for WMAP 5 year maps

significant at the 0.037%–0.025% level depending on the map used; see Figure (5). However, we also found that, while  $\mathcal{C}(\theta)$  computed in pixel space over the masked sky agrees with the harmonic space calculation that uses the pseudo- $C_\ell$  estimator, it disagrees with the  $C_\ell$  obtained using the MLE (advocated in the 3rd year WMAP release [4]). The MLE-based  $C_\ell$  leads to  $C(\theta)$  that is low (according to the  $S_{1/2}$  statistic) only at the 4.6% level.

There are actually two interesting questions one can ask.

- (i) Is the correlation function measured on the cut-sky compatible with cut-sky expectation from the Gaussian random, isotropic underlying model?
- (ii) Is the reconstruction of the full-sky correlation function from partial information compatible with the expectation from the Gaussian random, isotropic underlying model?

Our results refer to the first question above. The second question, while also extremely interesting, is more difficult to be robustly resolved because the reconstruction uses assumptions about statistical isotropy (see the next subsection).

The little large-angle correlation that does appear in the full-sky maps (e.g., the solid, black line in Figure 5) is associated with points inside the masked region. Shown in Figure 6 are the normalized contributions to  $\mathcal{C}(\theta)$  from different parts of the map. In particular, we see that almost all of the contribution to the full-sky two-point angular correlation function comes from correlations with at least one point inside the masked region. Conversely, there is essentially no large-angle correlation for points outside the masked region and even very little among the points completely inside the mask. We also see that all the curves cross zero at nearly the same angle,  $\theta \sim 90^\circ$ . We have no explanation for these results though they may point to systematics in the data.

**4.3. Alternative Statistics.** The two-point angular correlation function,  $\mathcal{C}(\theta)$ , as defined above in (15) is a simple pixel-based measure of correlations. It makes no assumptions about the underlying theory, which can be taken as a feature or as a flaw. On the positive side, (15) does not assume that the standard model is correct and tries to “force” it on the data. On the negative side we are not utilizing the full information available when comparing to the standard model.

Various approaches have been taken to incorporate the standard model in the analysis, for example; Hajian [34] defined a statistic that explicitly takes into account the covariance in the quantity  $C(\theta)$ :

$$A(x) \equiv \int_{-1}^x \int_{-1}^x C(\theta) F^{-1}(\theta, \theta') C(\theta') d(\cos \theta) d(\cos \theta'), \quad (18)$$

where  $F$  is the aforementioned covariance

$$F(\theta, \theta') \equiv \langle [C(\theta) - \langle C(\theta) \rangle][C(\theta') - \langle C(\theta') \rangle] \rangle, \quad (19)$$

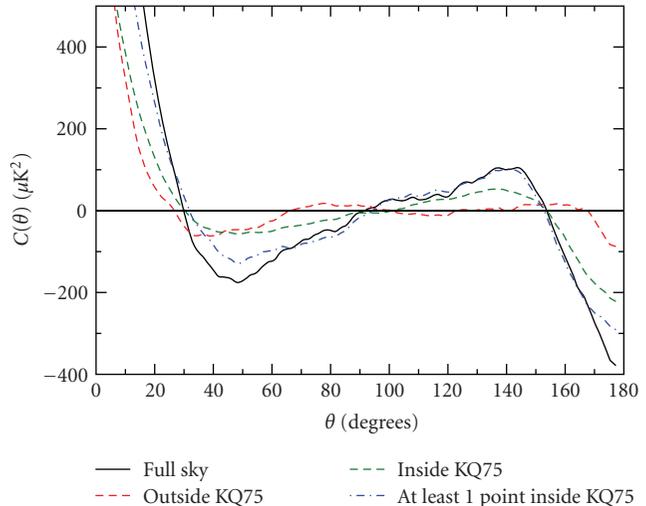


FIGURE 6: The two-point angular correlation function from the WMAP 5-year results. Plotted are  $\mathcal{C}(\theta)$  for the ILC calculated separately on the part of the sky outside the KQ75 cut (dashed line), inside the KQ75 cut (dotted line), and on the part of the sky with at least one point inside the KQ75 cut (dotted-dashed line). For better comparison to the full-sky  $\mathcal{C}(\theta)$  (solid line), the partial sky  $\mathcal{C}(\theta)$  has been scaled by the fraction of the sky over which they are calculated. Adopted from [31].

and as usual the angle brackets denote an ensemble average. Note that in the limit that  $C(\theta)$  is uncorrelated then  $A(1/2) = S_{1/2}$ . Clearly this statistic relies on a model to calculate  $F(\theta, \theta')$ . With this statistic and assuming the concordance model; it is found that less than 1% of realizations of the standard model have an  $A(0.53)$  less than those found for the masked skies. For the full-sky ILC map approximately 8% of realizations have a smaller value. Though less constraining than making no assumptions about the theory through the use of the  $S_{1/2}$  statistic, these results are consistent with those we previously found.

Another approach advocated by Efstathiou et al. [35] is to reconstruct the full-sky  $C(\theta)$  from the partial sky and compare the reconstructed full-sky  $C(\theta)$  (using, say  $S_{1/2}$ ) to the predictions of the model. This approach employs the usual map making algorithm on the low- $\ell$  spherical harmonic coefficients,  $a_{\ell m}$ . In this approach it is assumed that the statistical properties of the  $a_{\ell m}$  above some  $\ell_{\max}$  are known. In particular it assumes there are no correlations between the  $a_{\ell m}$  with  $\ell < \ell_{\max}$  and those with  $\ell > \ell_{\max}$ . The method is very similar to a maximum likelihood analysis (see e.g., [36]). As we have seen above (e.g., in Figure 5) it is not surprising that this approach will be consistent with the full-sky ILC results, as these authors have verified. As mentioned in the previous subsection, however, this procedure poses a *different* question of the data than the one that has been addressed by the  $S_{1/2}$  statistic applied to a masked sky. With the map-making technique a full-sky map is constructed that is consistent with the sky outside the mask but relies on assumptions to fill in the masked region. As is clear from Figure 6 it is precisely the region inside the mask that is

introducing correlations with the region outside the mask. Thus, the assumptions required to allow filling in the masked region also produce two-point angular correlations.

Whether or not reconstructing the full-sky is a “more optimal” approach than direct calculation of the cut-sky  $C(\theta)$  is moot, but likely depends on the actual (rather than the assumed) statistical properties of the underlying fluctuations as well as on the particular realization of those distributions. What *is* important is to make an “apples to apples” comparison between the observed sky and simulated realizations of the ensemble of possible skies. As we have shown, the pixel-based two-point correlation function on the region of the sky outside a conservative galactic mask is inconsistent with the predictions of the standard  $\Lambda$ CDM model for the identical pixel-based two-point correlation function on the identically masked sky. The fact that the full-sky analysis shows less statistical significance is not in contradiction with the cut-sky result, although it may eventually help in pointing to a cause of this anomaly.

**4.4. Summary.** The striking feature of the two-point angular correlation function as seen in Figure 5 is not that it disagrees with  $\Lambda$ CDM (though it does at  $> 90\%$  C.L.) but that at large-angles it is nearly zero. This lack of large-angle correlations is unexpected in inflationary models. The  $S_{1/2}$  statistic quantifies the deviation from zero and shows that a discrepancy exists at more than 99.9% C.L. Equally striking is the fact that the little correlation that does exist in the full-sky ILC map, or equally in the MLE estimated  $a_{\ell m}$ , comes from correlations between the masked foreground region and the expectedly cleaner CMB regions of the sky. Thus this residual correlation, which still is discrepant with generic inflationary predictions at about 95% C.L., comes from the reconstruction procedure. This surprising lack of large-angle correlation outside the masked region remains an open problem.

We also note that the vanishing of power is much more apparent in real space (as in  $\mathcal{C}(\theta)$ ) than in multipole space (as in  $C_\ell$ ). The harmonic-space quadrupole and octopole are only moderately low (e.g., [37]), and it is really a range of low multipoles that conspires to make up the vanishing  $\mathcal{C}(\theta)$ . Specifically, as discussed in [31], there is a cancellation between the combined contributions of  $C_2, \dots, C_5$  and the contributions of  $C_\ell$  with  $\ell \geq 6$ . It is this conspiracy that is most disturbing, since it violates the independence of the  $C_\ell$  of different  $\ell$  that defines statistical isotropy.

In [31] we therefore explored the possibility that the vanishing of  $C(\theta)$  could be explained simply by changing the values of the theoretical low- $\ell C_\ell$ , as might be the result, say, of a modified inflaton potential. In particular, we replaced  $C_2$  through  $C_{20}$  in the best-fit  $\Lambda$ CDM model with the values extracted from the cut-sky ILC five-year map. From these  $C_\ell$ 's, 200000 random maps were created, masked, and  $S_{1/2}$  computed. Under the assumptions of Gaussianity and statistical isotropy of these  $C_\ell$ 's only 3 percent of the generated maps had  $S_{1/2}$  less than the cut-sky ILC5 value. Thus, even if the  $C_\ell$  are set to the specific values that produce such a low  $S_{1/2}$ , a Gaussian random, statistically isotropic

realization is unlikely to produce the observed lack of large-angle correlations at the 97% C.L. Moreover, in work in progress we show that almost all of those 3% are skies with several anomalously low  $C_\ell$ —not at all the sky we see. Only a tiny fraction of the 3% represent skies in which most of the individual  $C_\ell$ 's were close to the  $\Lambda$ CDM prediction but they conspire to cancel one another in the large-angle  $C(\theta)$ . This shows that either (i) the low- $\ell C_\ell$ 's are correlated, contrary to the assumption of statistical isotropy or (ii) our Universe is an extremely unlikely realization of whatever statistically isotropic model one devises (Though this appears to be an unlikely explanation since correlations between the  $C_\ell$  generically increase the variance of the  $S_{1/2}$  statistic [33]).

It is for this reason that theoretical efforts to explain “low power on large scales” must focus on explaining the low  $\mathcal{C}(\theta)$  at  $\theta \gtrsim 60$  deg, rather than the low quadrupole.

Finally, one might ask if the observed lack of correlation and the alignment of quadrupole and octopole are correlated. This issue was studied by Rakić & Schwarz [38] for the full-sky and by Sarkar et al. [39] for the cut-sky case. In both cases, it was shown that low power and alignments are uncorrelated, that is, that having one does not imply a larger or smaller probability of having the other. This was shown by applying a Monte-Carlo analysis to sky realizations with the underlying standard Gaussian random, statistically isotropic cosmological model, without any further constraints. Thus one might view the 99.6% C.L. of quadrupole-octopole alignment presented in the previous section and the 95% C.L. for lack of correlation in full-sky maps reported in this section as statistically independent.

## 5. Quest for an Explanation

Understanding the origin of CMB anomalies is clearly important. Both the observed alignments of the low- $\ell$  full-sky multipoles, and the absence of large-angle correlations (especially on the galaxy cut-sky) are severely inconsistent with predictions of standard cosmological theory. There are four classes of possible explanations:

- (1) astrophysical foregrounds,
- (2) artifacts of faulty data analysis,
- (3) instrumental systematics,
- (4) theoretical/cosmological.

In this section, we review these four classes of explanations, giving examples from each. First, however, we discuss two generic ways to break statistical isotropy and affect the intrinsic (true) CMB signal—additive and multiplicative modulations—and illustrate in general terms why it has been so difficult to *explain* the anomalies.

**5.1. Additive versus Multiplicative Effects.** Why is it difficult to explaining the observed CMB anomalies? There are three basic reasons:

- (i) Most explanations work by *adding* power to the large-angle CMB, while the observed anisotropies actually

have *less* large-scale power and particularly less large-angle correlation, than the  $\Lambda$ CDM cosmological model predicts.

- (ii) Unaccounting for sources of CMB fluctuations in the foreground, even if possessing/causing aligned low- $\ell$  multipoles of their own, cannot bring unaligned statistically isotropic cosmological perturbations into alignment. Therefore, aligned foregrounds as an explanation for alignment work only if the cosmological signal is subdominant, thus exacerbating the lack of large-angle correlations.
- (iii) The alignments of the quadrupole and octopole are with respect to the ecliptic plane and near the dipole direction. It is generally difficult to have these directions naturally be picked out by any class of explanations (though there are exceptions to this—see the instrumental example below).

Gordon et al. [40], Rakić and Schwarz [38] and Bunn and Bourdon [41] explored generic “additive models” where the temperature modification that causes the alignment is added to

$$T_{\text{observed}}(\hat{e}) = T_{\text{intrinsic}}(\hat{e}) + T_{\text{add}}(\hat{e}). \quad (20)$$

Here  $T_{\text{add}}(\hat{e})$  is the additive term—perhaps contamination by a foreground and perhaps an additive instrumental or cosmological effect. They showed that additive modulations of the CMB sky that ameliorate the alignment problems tend to *worsen* the overall likelihood at large scales (though they may pick up offsetting positive likelihood contribution from higher multipoles). The intuitive reason for this is that there are two penalties incurred by the additive modulation. First, since the power spectrum at low  $\ell$  is lower than expected, one typically needs to arrange for an accidental cancellation between  $T_{\text{intrinsic}}$  and  $T_{\text{add}}$ ; the cancellation moreover must leave aligned quadrupole and octopole even though the quadrupole and octopole of  $T_{\text{intrinsic}}$  are not aligned. (Very similar reasoning argues against additive explanations of the suppression of large-angle correlations.) Second, the simplest models for the additive contribution that are based on an azimuthally symmetric modulation of a gradient field can only affect  $m = 0$  multipoles around the preferred axis, while as we mentioned earlier the observed quadrupole and octopole as seen in the preferred (dipole) frame are dominated by the  $m = \ell$  components.

In contrast to the additive models, the multiplicative mechanisms, where the intrinsic temperature is multiplied by a spatially varying modulation, are phenomenologically more promising. As a proof of principle, a toy-model modulation [40]

$$T_{\text{observed}}(\hat{e}) = f[1 + w_2 Y_{20}(\hat{e})] T_{\text{intrinsic}}(\hat{e}) \quad (21)$$

(where the modulation is a pure  $Y_{20}$  along the dipole axis) can increase the likelihood of the WMAP data by a factor of  $\exp(16/2)$  and, at the same time, increase the probability of obtaining a sky with more alignment (e.g., higher angular momentum statistic) 200 times, to 45%; see Figure 7.

Indeed, Groeneboom et al. [42], building on the work of Groeneboom and Eriksen [43] and Hanson and Lewis [44] and motivated by a model due to Ackerman et al. [45], have identified a  $9\sigma$  quadrupolar power asymmetry, recently confirmed by the WMAP team [46]; this anomaly can, however, be fully explained by accounting for asymmetric beams [47]. Recently, Hoftuft et al. [48] found a greater than  $3\text{-}\sigma$  evidence for nonzero dipolar modulation of the power.

*5.2. Astrophysical Explanations.* One fairly obvious possibility is that there is a pernicious foreground that contaminates the primordial CMB and leads to the observed anomalies. Such foregrounds are, of course, additive mechanisms, in the sense of the preceding section, and so suffer from the shortcomings described therein. Moreover, most such foregrounds are Galactic, while the observed alignments are with respect to the ecliptic plane. One would expect that Galactic foregrounds should lead to Galactic and not ecliptic foregrounds. This simple expectation was confirmed in [14], where we showed that, by artificially adding a large admixture of Galactic foregrounds to WMAP CMB maps, the quadrupole vectors move near the  $z$ -axis and the normal into the Galactic plane, while for the octopole all three normals become close to the Galactic disk at  $90^\circ$  from the Galactic center. Therefore, as expected Galactic foregrounds lead to Galactic, and not ecliptic, correlations of the quadrupole and octopole (see also studies by [49, 50]).

Moreover, in [14], we have shown that the known Galactic foregrounds possess a multipole vector structure very different from that of the observed quadrupole and octopole. The quadrupole is nearly pure  $Y_{22}$  in the frame where the  $z$ -axis is parallel to the dipole (or  $\hat{w}^{(2,1,2)}$  or any nearly equivalent direction), while the octopole is dominantly  $Y_{33}$  in the same frame. Mechanisms which produce an alteration of the microwave signal from a relatively small patch of sky—and all of the recent proposals fall into this class—are most likely to produce aligned  $Y_{20}$  and  $Y_{30}$ . This is essentially because the low- $\ell$  multipole vectors will all be parallel to each other, leading to a  $Y_{\ell 0}$  in this frame.

A number of authors have attempted to explain the observed quadrupole-octopole correlations in terms of a *new* foreground—for example the Rees-Sciama effect [38, 51], interstellar dust [52], local voids [53], or the Sunyaev-Zeldovich effect [54]. Most if not all of these proposals have a difficult time explaining the anomalies without severe fine-tuning. For example, Vale [55] cleverly suggested that the moving lens effect, with the Great Attractor as a source, might be responsible for the extra anisotropy; however, Cooray and Seto [56] have argued that the lensing effect is far too small and requires too large a mass of the Attractor.

It is also interesting to ask if any known or unknown Solar system physics could lead to the observed alignments. Dikarev et al. [57, 58] studied the question of whether solar system dust could give rise to sizable levels of microwave emission or absorption. Surprisingly, very little is known about dust grains of mm to cm size in the Solar system, and their absorption/emission properties strongly depend on their chemical composition. While iron and ice particles

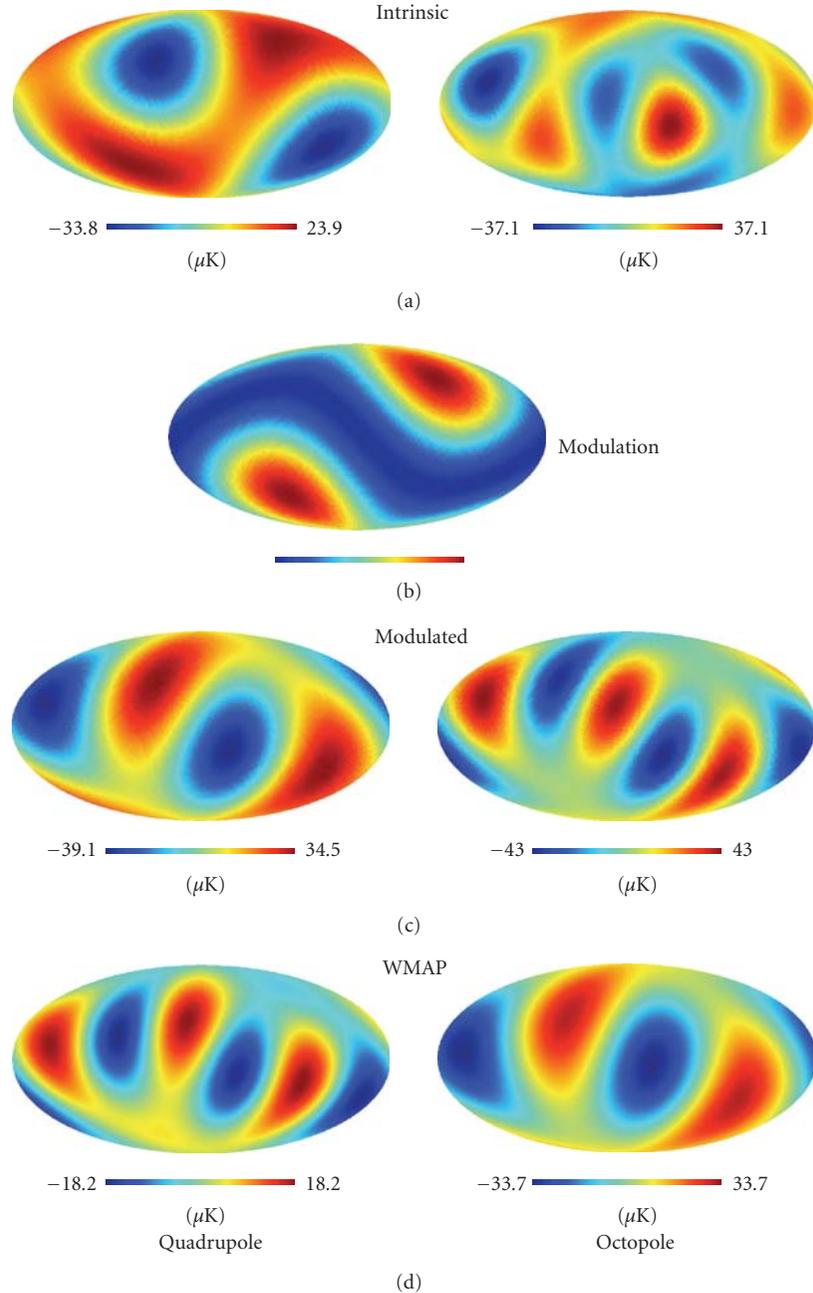


FIGURE 7: A realization of the multiplicative model where the quadrupole (left column) and octopole (right column) exhibit an alignment similar to WMAP. (a): intrinsic (unmodulated) sky from a Gaussian random isotropic realization. (b) (single column): the quadrupolar modulation with  $f = -1$  and  $w_2 = -7$  (see (21)) in the dipole direction. (c): the modulated sky of the observed CMB. (d): WMAP full-sky quadrupole and octopole. Adopted from [40].

can definitely be excluded to contribute at significant levels, carbonaceous and silicate dust grains might contribute up to a few  $\mu\text{K}$  close to the ecliptic plane, for example, due to the trans-Neptunian object belt. Such an extra contribution along the ecliptic could give rise to CMB structures aligned with the ecliptic, but those would look very different from the observed ones. On top of that, Solar system dust would be a new additive foreground and could not explain the lack of large-angle correlations. Thus it seems unlikely

that Solar system dust grains cause the reported large-angle anomalies, nevertheless they are sources of microwave absorption and emission and may become important to precision measurements in the future.

Finally, it has often been suggested to some of us in private communications that the anomalies may not reflect an unknown foreground that has been neglected, but rather the “misubtraction” of a known foreground. However, it has never quite been clear to us how this leads to the

observed alignments or lack of large-angle correlations, and we are unaware of any literature that realizes this suggestion successfully.

**5.3. Data Analysis Explanations.** Most of the results discussed so far have been obtained using reconstructed full-sky maps of the WMAP observations [1, 22, 59]. In the presence of the sky cut of even just a few degrees, the errors in the reconstructed anisotropy pattern, and the directions of multipole vectors, are too large to allow drawing quantitative conclusions about the observed alignments [15]. These large errors are expected: while the *power* in the CMB (represented, say, by the angular power spectrum  $C_\ell$ ) can be accurately recovered since there are  $2\ell + 1$  modes available for each  $\ell$ , there are only 2 modes available for each multipole vector; hence the cut-sky reconstruction is noisier. However the cut-sky alignment probabilities, while very uncertain, are consistent with the full-sky values [14, 50]; more generally, the alignments appear to be rather robust to Galactic cuts and foreground contamination [60].

A different kind of explanation of missing large-scale power, or missing large-angle correlations, has been taken by Efstathiou et al. [35] who argued that maximum likelihood estimators can be applied to the cut-sky maps to reliably and optimally reconstruct the CMB anisotropy of the whole sky; for a recent work that extends these ideas, see [33]. This approach yields more two-point correlations on large scales ( $S_{1/2} \sim 8000 (\mu K)^4$ , which is  $\sim 5\%$  likely) than the direct cut-sky pixel-based calculation which gives  $S_{1/2} \sim 1000 (\mu K)^4$  result and is  $\sim 0.03\%$  likely. These authors then argue that the extremely low  $S_{1/2}$  obtained by the pixel-based approach applied to the cut-sky is essentially a fluke, and the more reliable result comes from their maximum-likelihood reconstruction of the full-sky. It may indeed be true that the pixel-based calculation is a suboptimal estimate of the *full-sky*  $C(\theta)$  for a *statistically isotropic cosmology*. However, quantities calculated on the cut-sky are clearly insensitive to assumptions about what lies behind the cut. We can only observe reliably the  $\sim 75\%$  of the sky that was not masked, and that is where the large-angle two-point-correlation is near-vanishing. Any attempt to reconstruct the full-sky must make assumptions about the statistical properties of the CMB sky, and would clearly be affected by the coupling of small-scale and large-scale modes—exactly what is necessary to have a sky in which  $S_{1/2}$  is anomalously low, while the  $C_\ell$ 's are individually approximately consistent with the standard cosmology.

**5.4. Instrumental Explanations.** Are instrumental artifacts the cause of the observed alignments (and/or the low large-scale power)? One possible scenario would go as follows. WMAP avoids making observations near the Sun, therefore covering regions away from the ecliptic more than those near the ecliptic. While the corresponding variations in the noise per pixel are well known (e.g., as the number of observations per pixel,  $N_{\text{obs}}$ ; see [1, Figure 3]), and its effects on the large-scale anomalies are ostensibly small, they could, in principle, be amplified and create the observed ecliptic anomalies.

However a successful proposal for such an amplification has not yet been put forward.

Another possibility is that an imperfect instrument couples with dominant signals from the sky to create anomalies. Let us review an example given in [40]: suppose that the instrumental response  $T_{\text{instr}}(\hat{e})$  to the true sky signal  $T(\hat{e})$  is nonlinear

$$T_{\text{instr}}(\hat{e}) = f \sum_i \alpha_i \left[ \frac{T(\hat{e})}{f} \right]^i. \quad (22)$$

Here  $f$  is an arbitrary normalization scale for the nonlinearity of the response, and  $\alpha_i$  are arbitrary coefficients with  $\alpha_1 = 1$ . If  $\alpha_{i>1} \neq 0$  then  $T_{\text{instr}} \neq T$  and the observed temperature is a nonlinear modulation of the true temperature. The dominant temperature signal for a differencing experiment such as WMAP is the dipole arising from our peculiar motion,  $T(\hat{e}) = T_{\text{dip}} \cos \theta$ , with  $T_{\text{dip}} = 3.35$  mK and  $\theta$  the polar angle in the dipole frame. Taking  $f = T_{\text{dip}}$ ,

$$\begin{aligned} \frac{T_{\text{instr}}(\hat{e})}{T_{\text{dip}}} &= \alpha_1 P_1(\cos \theta) + \alpha_2 \left[ \frac{2}{3P_2(\cos \theta)} + 1 \right] \\ &+ \alpha_3 \left[ \frac{2}{5P_3(\cos \theta)} + \frac{3}{5P_1(\cos \theta)} \right] + \dots, \end{aligned} \quad (23)$$

where  $P_\ell$  are the Legendre polynomials. Note that with  $\alpha_2 \sim \alpha_3 \sim O(10^{-2})$ , the  $10^{-3}$  dipole anisotropy is modulated into a  $10^{-5}$  quadrupole and octopole anisotropy which are aligned *in the dipole frame* with the  $m = 0$  multipole structure. Unfortunately (or fortunately!), WMAP detectors are known to be linear to much better than 1%, so this particular realization of the instrumental explanation does not work. As an aside, note that this type of explanation needs to assume that the higher multipoles are not aligned with the dipole/ecliptic and, moreover, requires essentially no intrinsic power at large scales (that is, even less than what is observed).

To summarize, even though the ecliptic alignments (and the north-south power asymmetry) hint at a systematic effect due to some kind of coupling of an observational strategy and the instrument, to date no plausible proposal of this sort has been put forth.

**5.5. Cosmological Explanations.** The most exciting possibility is that the observed anomalies have primordial origin, and potentially inform us about the conditions in the early universe. One expects that in this case the alignments with the dipole, or with the solar system, would be statistical flukes.

The breaking of statistical isotropy implies that the usual relation  $\langle a_{\ell m}^* a_{\ell' m'} \rangle = C_\ell \delta_{\ell\ell'} \delta_{mm'}$  does not hold any more; instead

$$\langle a_{\ell m}^* a_{\ell' m'} \rangle = C_{\ell\ell' mm'}, \quad (24)$$

where the detailed form of the quantity on the right-hand side is model dependent (see the more detailed discussions in [61–63]).

There are many possibilities for how the absence of statistical isotropy might arise; for example, in a nontrivial spatial topology, the fundamental domain would not be rotationally invariant, and so the spherical harmonics (times an appropriate radial function) would not be a basis of independent eigenmodes of the fluctuations. This would certainly lead to a correlation among  $a_{\ell m}$  of different  $m$  and also  $\ell$ , although not necessarily to aligned multipoles. The hope would be that the shape of the fundamental domain would lead to these alignments, while a lower density of states at long wavelength (compared to the covering space) would lead to the absence of large-angle correlations. No specific model has been suggested to accomplish all of those, and the matter is complicated by known bounds on cosmic topology [64–68] which force the fundamental domain to be relatively large.

Alternately, in the early universe, asymmetry in the stress-energy tensor of dark energy [69] or a long-wavelength dark energy density mode with a gradient in the desired direction [40], could both imprint the observed alignments via the integrated Sachs-Wolfe mechanism; but it is hard to see how they would explain the lack of large-angle correlations. The authors of [70] have put forward a model where the Sachs-Wolfe contribution to low- $\ell$  multipoles is partly cancelled by the Integrated Sachs-Wolfe contribution, but which still fails to explain the lowness of  $S_{1/2}$  or the alignments of low- $\ell$  multipoles. Models where the anomalies are caused by breaking of the statistical isotropy [71, 72] are especially well studied; see [73] for the equally interesting possibility that the *translational* invariance is broken.

A commonly used mechanism to explain such anomalies is inflationary models that contain implicit breaking of isotropy [40, 74–80]. Pontzen [81] helpfully shows temperature and polarization patterns caused by various classes of Bianchi models that explicitly break the statistical isotropy. However, outside of explaining the anomalies, the motivation for these anisotropic models is not compelling and they seem somewhat contrived. Moreover, the authors have not investigated whether the low  $S_{1/2}$  is also observed. Nevertheless, given the large-scale CMB observations, as well as the lack of fundamental theory that would explain inflation, investigating such models is well worthwhile.

A very reasonable approach is to describe breaking of the isotropy with a phenomenological model, measure the parameters of the model, and then try to draw inferences about the underlying physical mechanism. For example, a convenient approach is to describe the breaking of isotropy via the direction-dependent power spectrum of dark matter perturbations [82]

$$P(\vec{k}) = A(k) \left[ 1 + \sum_{\ell m} g_{\ell m}(k) Y_{\ell m}(\hat{k}) \right], \quad (25)$$

where  $k = |\vec{k}|$ ,  $g_{\ell m}(k)$  quantifies the departure from statistical isotropy as a function of wavenumber  $k$ , and  $A(k)$  refers to the statistically isotropic part of the power spectrum. In this model, the power spectrum, normally considered to depend only on scale  $k$ , now depends on direction in a

parametric way. Statistically significant finding that  $g_{\ell m} \neq 0$  for any  $(\ell, m)$  would signal a violation of statistical isotropy.

As with the other attempts to explain the anomalies, we conclude that, while there have been some interesting and even promising suggestions, no cosmological explanation to date has been compelling.

*5.6. Alignment Explanations: What Next.* While future WMAP data is not expected to change any of the observed results, our understanding and analysis techniques are likely to improve. The most interesting test will come from the Planck satellite, whose temperature maps, obtained with a completely different instrument and observational technique than WMAP, could shed significant new light on the alignments. Moreover, polarization information could be extremely useful in distinguishing between different models and classes of explanations in general; for example, Dvorkin et al. [83] explicitly show how polarization information expected from Planck can help identify the cause of the alignments. Finally, one could use the large-scale structure (i.e., galaxy distribution) data on the largest observable scales from surveys such as Dark Energy Survey (DES) and Large Synoptic Survey Telescope (LSST) to test cosmological explanations (see, e.g. [84, 85]).

## 6. Explanations from the WMAP Team

In their seven year data release the WMAP team explicitly discusses several CMB anomalies [46] including the two main ones described in this paper. For the first major issue—the alignment of low multipoles with each other—the WMAP team agrees that the alignment is observed and argue, based on work by Francis and Peacock [86], that the integrated Sachs-Wolfe (ISW) contribution of structures at small redshifts ( $z \ll 1$ ) could be held responsible. There are serious problems with this argument. Firstly, the ordinary Sachs-Wolfe (SW) effect typically dominates at these  $\ell$  over the ISW. Thus, only if the ordinary SW effect on the last scattering surface is anomalously low will the ISW contribution dominate. Secondly, though the ISW may lead to alignment of the quadrupole and octopole it is not an explanation for the observed Solar system alignments. This alignment would need to be an additional statistical fluke. Finally, this explanation does nothing whatsoever to mitigate the lack of large scale angular correlation because the ISW effect acts as an additive component and should be statistically uncorrelated from the primordial CMB. Therefore, even if the ISW reconstruction is taken as reliable, this argument would imply

- (1) an accidental downward fluctuation of the SW sufficient for the ISW of local structure to dominate and cause an alignment,
- (2) an accidental cancellation in angular correlation between the SW and ISW temperature patterns.

Neither the WMAP team nor Francis and Peacock estimate the likelihood of these two newly created puzzles.

Regarding the second major issue—the lack of angular correlation—the WMAP team refers to a recent work by Efstathiou et al. [35] who argue that quadratic estimators are better estimates of the full-sky from cut-sky data and are in better agreement with the concordance model. While these estimators have been shown to be optimal under the assumption of statistical isotropy, it is unclear why they should be employed when this assumption is to be tested. (For a contrary view, see [33].) The pixel-based estimator applied to the cut-sky in our analysis does not rely on statistical isotropy and it is more conservative as it does not try to reconstruct temperature anisotropies inside the cut. Finally, our claims that the pixel-based cut-sky two-point correlation function is highly anomalous rely on comparisons to the identical correlation function calculated on simulated cut skies. Whether or not it is a good estimate of the full-sky correlation function is answering a different question, (see Section 5.2). Conversely, the estimator suggested in [35] assumes that whatever is within the cut can be reconstructed reliably by truncating the number of multipole moments considered. The latter logic is equivalent to the assumption of the statistical independence of low and high multipoles, which is exactly a consequence of statistical isotropy.

These arguments from the WMAP team offer neither new nor convincing explanations of the observed anomalies discussed in this paper. At best they replace one set of anomalies by another.

## 7. Conclusions

The CMB is widely regarded as offering strong substantiating evidence for the concordance model of cosmology. Indeed the agreement between theory and data is remarkable—the patterns in the two-point correlation functions (TT, TE, and EE) of Doppler peaks and troughs are reproduced in detail by fitting with only six (or so) cosmological parameters. This agreement should not be taken lightly; it shows our precise understanding of the causal physics on the last scattering surface. Even so, the cosmological model we arrive at is baroque, requiring the introduction at different scales and epochs of three sources of energy density that are only detected gravitationally—dark matter, dark energy and the inflaton. This alone should encourage us to continuously challenge the model and probe the observations particularly on scales larger than the horizon at the time of last scattering.

At the very least, probes of the large-angle (low- $\ell$ ) properties of the CMB reveal that we do not live in a typical realization of the concordance model of inflationary  $\Lambda$ CDM. We have reviewed a number of the ways in which that is true: the peculiar geometry of the  $\ell = 2$  and 3 multipoles—their planarity, their mutual alignment, their alignment perpendicular to the ecliptic and to the dipole; the north-south asymmetry; and the near absence of two-point correlations for points separated by more than  $60^\circ$ .

If indeed the observed  $\ell = 2$  and 3 CMB fluctuations are not cosmological, one must reconsider all CMB results that rely on the low  $\ell$ , for example, the measurement of the optical depth from CMB polarization at low  $\ell$  or the spectral

index of scalar perturbations and its running. Moreover, the CMB galaxy cross-correlation, which has been used to provide evidence for the Integrated Sachs-Wolfe effect and hence the existence of dark energy, also gets contributions from the lowest multipoles (though the main contribution comes from slightly smaller scales,  $\ell \sim 10$ ). Indeed, it is quite possible that the underlying physical mechanism does not cut off abruptly at the octopole, but rather affects the higher multipoles. Indeed, several pieces of evidence have been presented for anomalies at  $l > 3$  (e.g., [87, 88]). One of these is the parity of the microwave sky. While the observational fact that the octopole is larger than the quadrupole ( $C_3 > C_2$ ) is not remarkable on its own, including higher multipoles (up to  $\ell \sim 20$ ) the microwave sky appears to be parity odd at a statistically significant level (since WMAP 5yr) [89–91]. It is hard to imagine a cosmological explanation for a parity odd universe, but the same holds true for unidentified systematics or unaccounted astrophysical foregrounds, especially as this recently noticed puzzle shows up in the very well studied angular power spectrum.

While the further WMAP data is not expected to change any of the observed results, our understanding and analysis techniques are likely to improve. Much work remains to study the large-scale correlations using improved foreground treatment, accounting even for the subtle systematics and in particular studying the time-ordered data from the spacecraft. The Planck experiment will be of great importance, as it will provide maps of the largest scales obtained using a very different experimental approach than WMAP—measuring the absolute temperature rather than temperature differences. Polarization maps, when available at high enough signal-to-noise at large scales (which may not be soon), will be a fantastic independent test of the alignments, as each explanation for the alignments, in principle, also predicts the statistics of the polarization pattern on the sky.

## Acknowledgment

D. Huterer is supported by DOE OJI grant under Contract DE-FG02-95ER40899 and NSF under contract AST-0807564. D. Huterer and C. J. Copi are supported by NASA under Contract NNX09AC89G; D. J. Schwarz is supported by Deutsche Forschungsgemeinschaft (DFG); G. D. Starkman is supported by a grant from the US Department of Energy; both G. D. Starkman and C. J. Copi are supported by NASA under Cooperative Agreement NNX07AG89G.

## References

- [1] C. L. Bennett et al., “First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: preliminary maps and basic results,” *The Astrophysical Journal Supplement*, vol. 148, no. 1, 2003.
- [2] D. N. Spergel, L. Verde, H. V. Peiris et al., “First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: determination of cosmological parameters,” *Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 175–194, 2003.

- [3] G. Hinshaw, D. N. Spergel, L. Verde et al., “First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: the angular power spectrum,” *Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 135–159, 2003.
- [4] D. N. Spergel, R. Bean, O. Doré et al., “Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: implications for cosmology,” *Astrophysical Journal, Supplement Series*, vol. 170, no. 2, pp. 377–408, 2007.
- [5] E. Komatsu, J. Dunkley, M. R. Nolta et al., “Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [6] E. Komatsu et al., “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation,” <http://arxiv.org/abs/1001.4538>.
- [7] D. Huterer, “Mysteries on universe’s largest observable scales,” *New Astronomy Reviews*, vol. 50, no. 11–12, pp. 868–874, 2006.
- [8] L. R. Abramo and T. S. Pereira, “Testing gaussianity, homogeneity and isotropy with the cosmic microwave background,” <http://arxiv.org/abs/1002.3173>.
- [9] C. T. Byrnes and K.-Y. Choi, “Review of local non-Gaussianity from multi-field inflation,” <http://arxiv.org/abs/1002.3110>.
- [10] X. Chen, “Primordial non-Gaussianities from inflation models,” <http://arxiv.org/abs/1002.1416>.
- [11] R. K. Sachs and A. M. Wolfe, *Astrophysical Journal*, vol. 147, p. 73, 1967.
- [12] P. J. E. Peebles and D. T. Wilkinson, “Comment on the anisotropy of the primeval fireball,” *Physical Review*, vol. 174, no. 5, p. 2168, 1968.
- [13] M. Kamionkowski and L. Knox, “Aspects of the cosmic microwave background dipole,” *Physical Review D*, vol. 67, no. 6, Article ID 063001, 2003.
- [14] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, “On the large-angle anomalies of the microwave sky,” *Monthly Notices of the Royal Astronomical Society*, vol. 367, no. 1, pp. 79–102, 2006.
- [15] C. J. Copi, D. Huterer, and G. D. Starkman, “Multipole vectors: a new representation of the CMB sky and evidence for statistical anisotropy or non-Gaussianity at  $2 \leq \ell \leq 8$ ,” *Physical Review D*, vol. 70, no. 4, Article ID 043515, 13 pages, 2004.
- [16] C. J. Copi, “Freely downloadable multipole vector code,” <http://www.phys.cwru.edu/projects/mpvectors/#code>.
- [17] G. Katz and J. Weeks, “Polynomial interpretation of multipole vectors,” *Physical Review D*, vol. 70, no. 6, Article ID 063527, 2004.
- [18] J. R. Weeks, “Maxwell’s multipole vectors and the CMB,” <http://arxiv.org/abs/astro-ph/0412231>.
- [19] R. C. Helling, P. Schupp, and T. Tesileanu, “CMB statistical anisotropy, multipole vectors, and the influence of the dipole,” *Physical Review D*, vol. 74, no. 6, Article ID 063004, 2006.
- [20] J. C. Maxwell, *A Treatise on Electricity and Magnetism, Vol. I*, Clarendon Press, London, UK, 3rd edition, 1891.
- [21] M. R. Dennis, “Correlations between Maxwell’s multipoles for Gaussian random functions on the sphere,” *Journal of Physics A*, vol. 38, no. 8, pp. 1653–1658, 2005.
- [22] M. Tegmark, A. de Oliveira-Costa, and A. J. S. Hamilton, “High resolution foreground cleaned CMB map from WMAP,” *Physical Review D*, vol. 68, no. 12, Article ID 123523, 2003.
- [23] A. de Oliveira-Costa, M. Tegmark, M. Zaldarriaga, and A. Hamilton, “Significance of the largest scale CMB fluctuations in WMAP,” *Physical Review D*, vol. 69, no. 6, Article ID 063516, 2004.
- [24] D. J. Schwarz, G. D. Starkman, D. Huterer, and C. J. Copi, “Is the low- $\ell$  microwave background cosmic?” *Physical Review Letters*, vol. 93, no. 22, Article ID 221301, 2004.
- [25] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, “Uncorrelated universe: statistical anisotropy and the vanishing angular correlation function in WMAP years 1–3,” *Physical Review D*, vol. 75, no. 2, Article ID 023507, 2007.
- [26] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Górski, and P. B. Lilje, “Asymmetries in the cosmic microwave background anisotropy field,” *Astrophysical Journal*, vol. 605, no. 1, pp. 14–20, 2004.
- [27] F. K. Hansen, P. Cabella, D. Marinucci, and N. Vittorio, “Asymmetries in the local curvature of the wilkinson microwave anisotropy probe data,” *Astrophysical Journal*, vol. 607, no. 2, pp. L67–L70, 2004.
- [28] A. Bernui, T. Villela, C. A. Wuensche, R. Leonardi, and I. Ferreira, “On the cosmic microwave background large-scale angular correlations,” *Astronomy and Astrophysics*, vol. 454, no. 2, pp. 409–414, 2006.
- [29] A. Bernui, B. Mota, M. J. Rebouças, and R. Tavakol, “Mapping the large-scale anisotropy in the WMAP data,” *Astronomy and Astrophysics*, vol. 464, no. 2, pp. 479–485, 2007.
- [30] M. R. Nolta, J. Dunkley, R. S. Hill et al., “Five-year wilkinson microwave anisotropy probe observations: angular power spectra,” *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 296–305, 2009.
- [31] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, “No large-angle correlations on the non-Galactic microwave sky,” *Monthly Notices of the Royal Astronomical Society*, vol. 399, no. 1, pp. 295–303, 2009.
- [32] G. Hinshaw, A. J. Banday, C. L. Bennett et al., “Two-point correlations in the COBE 1 DMR four-year anisotropy maps,” *Astrophysical Journal*, vol. 464, no. 1, pp. L25–L28, 1996.
- [33] A. Pontzen and H. V. Peiris, “The cut-sky cosmic microwave background is not anomalous,” *Physical Review D*, vol. 81, no. 10, Article ID 103008, 2010.
- [34] A. Hajian, “Analysis of the apparent lack of power in the cosmic microwave background anisotropy at large angular scales,” <http://arxiv.org/abs/astro-ph/0702723>.
- [35] G. Efstathiou, Y.-Z. Ma, and D. Hanson, “Large-angle correlations in the cosmic microwave background,” *Monthly Notices of the Royal Astronomical Society*, vol. 407, no. 3, pp. 2530–2542, 2010.
- [36] G. Efstathiou, “A maximum likelihood analysis of the low cosmic microwave background multipoles from the Wilkinson Microwave Anisotropy Probe,” *Monthly Notices of the Royal Astronomical Society*, vol. 348, no. 3, pp. 885–896, 2004.
- [37] I. J. O’Dwyer, H. K. Eriksen, B. D. Wandelt et al., “Bayesian power spectrum analysis of the first-year Wilkinson Microwave Anisotropy Probe data,” *Astrophysical Journal*, vol. 617, no. 2, pp. L99–L102, 2004.
- [38] A. Rakić and D. J. Schwarz, “Correlating anomalies of the microwave sky,” *Physical Review D*, vol. 75, no. 10, Article ID 103002, 2007.
- [39] D. Sarkar, D. Huterer, C. J. Copi, G. D. Starkman, and D. J. Schwarz, “Missing power vs low- $l$  alignments in the cosmic microwave background: no correlation in the standard cosmological model,” <http://arxiv.org/abs/1004.3784>.
- [40] C. Gordon, W. Hu, D. Huterer, and T. Crawford, “Spontaneous isotropy breaking: a mechanism for CMB multipole alignments,” *Physical Review D*, vol. 72, no. 10, Article ID 103002, 13 pages, 2005.
- [41] E. F. Bunn and A. Bourdon, “Contamination cannot explain the lack of large-scale power in the cosmic microwave

- background radiation,” *Physical Review D*, vol. 78, no. 12, Article ID 123509, 2008.
- [42] N. E. Groeneboom, H. K. Eriksen, K. Gorski, G. Huey, J. Jewell, and B. Wandelt, “Bayesian analysis of white noise levels in the five-year WMAP data,” *Astrophysical Journal*, vol. 702, no. 1, pp. L87–L90, 2009.
- [43] N. E. Groeneboom and H. K. Eriksen, “Bayesian analysis of sparse anisotropic universe models and application to the five-year WMAP data,” *Astrophysical Journal*, vol. 690, no. 2, pp. 1807–1819, 2009.
- [44] D. Hanson and A. Lewis, “Estimators for CMB statistical anisotropy,” *Physical Review D*, vol. 80, no. 6, Article ID 063004, 2009.
- [45] L. Ackerman, S. M. Carroll, and M. B. Wise, “Imprints of a primordial preferred direction on the microwave background,” *Physical Review D*, vol. 75, no. 8, Article ID 083502, 2007.
- [46] C. L. Bennett et al., “Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: are there cosmic microwave background anomalies?” <http://arxiv.org/abs/1001.4758>.
- [47] D. Hanson, A. Lewis, and A. Challinor, “Asymmetric beams and CMB statistical anisotropy,” *Physical Review D*, vol. 81, no. 10, Article ID 103003, 2010.
- [48] J. Hoftuft, H. K. Eriksen, A. J. Banday, K. M. Górski, F. K. Hansen, and P. B. Lilje, “Increasing evidence for hemispherical power asymmetry in the five-year wmap data,” *Astrophysical Journal*, vol. 699, no. 2, pp. 985–989, 2009.
- [49] S. Bashinsky and U. Seljak, “Signatures of relativistic neutrinos in CMB anisotropy and matter clustering,” *Physical Review D*, vol. 69, no. 8, Article ID 083002, 35 pages, 2004.
- [50] P. Bielewicz, H. K. Eriksen, A. J. Banday, K. M. Górski, and P. B. Lilje, “Multipole vector anomalies in the first-year WMAP data: a cut-sky analysis,” *Astrophysical Journal*, vol. 635, no. 2, pp. 750–760, 2005.
- [51] A. Rakić, S. Räsänen, and D. J. Schwarz, *Monthly Notices of the Royal Astronomical Society*, vol. 369, pp. L27–L31, 2006.
- [52] P. C. Frisch, “Tentative identification of interstellar dust in the magnetic wall of the heliosphere,” *Astrophysical Journal*, vol. 632, no. 2, pp. L143–L146, 2005.
- [53] K. T. Indue and J. Silk, “Local voids as the origin of large-angle cosmic microwave background anomalies. I,” *Astrophysical Journal*, vol. 648, no. 1, pp. 23–30, 2006.
- [54] H. V. Peiris and T. L. Smith, “CMB isotropy anomalies and the local kinetic Sunyaev-Zel’dovich effect,” *Physical Review D*, vol. 81, no. 12, Article ID 123517, 2010.
- [55] C. Vale, “Local pancake defeats axis of evil,” <http://arxiv.org/abs/astro-ph/0509039>.
- [56] A. Cooray and N. Seto, “Did the Wilkinson Microwave Anisotropy Probe see moving local structures?” *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 12, article 4, pp. 41–49, 2005.
- [57] V. Dikarev, O. Preuß, S. Solanki, H. Krüger, and A. Krivov, “Understanding the WMAP results: low-order multipoles and dust in the vicinity of the solar system,” *Earth, Moon and Planets*, vol. 102, no. 1–4, pp. 555–561, 2008.
- [58] V. Dikarev, O. Preuß, S. Solanki, H. Krüger, and A. Krivov, “The local dust foregrounds in the microwave sky. I. thermal emission spectra,” *Astrophysical Journal*, vol. 705, no. 1, pp. 670–682, 2009.
- [59] H. K. Eriksen, A. J. Banday, K. M. Górski, and P. B. Lilje, “On foreground removal from the wilkinson microwave anisotropy probe data by an internal linear combination method: limitations and implications,” *Astrophysical Journal*, vol. 612, no. 2, pp. 633–646, 2004.
- [60] A. de Oliveira-Costa and M. Tegmark, “CMB multipole measurements in the presence of foregrounds,” *Physical Review D*, vol. 74, no. 2, Article ID 023005, 2006.
- [61] A. Hajian and T. Souradeep, “Measuring the statistical isotropy of the cosmic microwave background anisotropy,” *Astrophysical Journal*, vol. 597, no. 1, pp. L5–L8, 2003.
- [62] A. Hajian, T. Souradeep, and N. Cornish, “Statistical isotropy of the wilkinson microwave anisotropy probe data: a bipolar power spectrum analysis,” *Astrophysical Journal*, vol. 618, no. 2, pp. L63–L66, 2005.
- [63] T. S. Pereira and L. R. Abramo, “Angular-planar CMB power spectrum,” *Physical Review D*, vol. 80, no. 6, Article ID 063525, 2009.
- [64] N. J. Cornish, D. N. Spergel, and G. D. Starkman, “Measuring the topology of the universe,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 1, pp. 82–84, 1998.
- [65] N. J. Cornish, D. Spergel, and G. Starkman, “Can COBE see the shape of the universe?” *Physical Review D*, vol. 57, no. 10, pp. 5982–5996, 1998.
- [66] N. J. Cornish, D. N. Spergel, and G. D. Starkman, “Circles in the sky: finding topology with the microwave background radiation,” *Classical and Quantum Gravity*, vol. 15, no. 9, pp. 2657–2670, 1998.
- [67] N. J. Cornish, D. N. Spergel, G. D. Starkman, and E. Komatsu, “Constraining the topology of the Universe,” *Physical Review Letters*, vol. 92, no. 20, Article ID 201302, 2004.
- [68] J. S. Key, N. J. Cornish, D. N. Spergel, and G. D. Starkman, “Extending the WMAP bound on the size of the Universe,” *Physical Review D*, vol. 75, no. 8, Article ID 084034, 2007.
- [69] R. Battye and A. Moss, “Anisotropic dark energy and CMB anomalies,” *Physical Review D*, vol. 80, no. 2, Article ID 023531, 2009.
- [70] N. Afshordi, G. Geshnizjani, and J. Khoury, “Do observations offer evidence for cosmological-scale extra dimensions?” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 30, 2009.
- [71] T. R. Jaffe, A. J. Banday, H. K. Eriksen, K. M. Górski, and F. K. Hansen, “Evidence of vorticity and shear at large angular scales in the WMAP data: a violation of cosmological isotropy?” *Astrophysical Journal*, vol. 629, no. 1, pp. L1–L4, 2005.
- [72] T. Ghosh, A. Hajian, and T. Souradeep, “Unveiling hidden patterns in CMB anisotropy maps,” *Physical Review D*, vol. 75, no. 8, Article ID 083007, 2007.
- [73] S. M. Carroll, C.-Y. Tseng, and M. B. Wise, “Translational invariance and the anisotropy of the cosmic microwave background,” *Physical Review D*, vol. 81, no. 8, Article ID 083501, 2010.
- [74] L. Ackerman, S. M. Carroll, and M. B. Wise, “Imprints of a primordial preferred direction on the microwave background,” *Physical Review D*, vol. 75, no. 8, Article ID 083502, 2007.
- [75] C. Armendariz-Picon, “Creating statistically anisotropic and inhomogeneous perturbations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 9, article 14, 2007.
- [76] A. E. Gümrükçüoğlu, C. R. Contaldi, and M. Peloso, “Inflationary perturbations in anisotropic backgrounds and their imprint on the cosmic microwave background,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, article 5, 2007.
- [77] D. C. Rodrigues, “Anisotropic cosmological constant and the CMB quadrupole anomaly,” *Physical Review D*, vol. 77, no. 2, Article ID 023534, 2008.

- [78] C. Pitrou, T. S. Pereira, and J.-P. Uzan, “Predictions from an anisotropic inflationary era,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article 4, 2008.
- [79] A. L. Erickcek, M. Kamionkowski, and S. M. Carroll, “A hemispherical power asymmetry from inflation,” *Physical Review D*, vol. 78, no. 12, Article ID 123520, 2008.
- [80] A. L. Erickcek, C. M. Hirata, and M. Kamionkowski, “A scale-dependent power asymmetry from isocurvature perturbations,” *Physical Review D*, vol. 80, no. 8, Article ID 083507, 2009.
- [81] A. Pontzen, “Rogues’ gallery: the full freedom of the Bianchi CMB anomalies,” *Physical Review D*, vol. 79, no. 10, Article ID 103518, 2009.
- [82] A. R. Pullen and M. Kamionkowski, “Cosmic microwave background statistics for a direction-dependent primordial power spectrum,” *Physical Review D*, vol. 76, no. 10, Article ID 103529, 2007.
- [83] C. Dvorkin, H. V. Peiris, and W. Hu, “Testable polarization predictions for models of CMB isotropy anomalies,” *Physical Review D*, vol. 77, no. 6, Article ID 063008, 2008.
- [84] H. Zhan, L. Knox, J. A. Tyson, and V. Margoniner, “Exploring large-scale structure with billions of galaxies,” *Astrophysical Journal*, vol. 640, no. 1, pp. 8–17, 2006.
- [85] A. R. Pullen and C. M. Hirata, “Non-detection of a statistically anisotropic power spectrum in large-scale structure,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 5, article 27, 2010.
- [86] C. L. Francis and J. A. Peacock, “An estimate of the local integrated Sachs-Wolfe signal and its impact on cosmic microwave background anomalies,” *Monthly Notices of the Royal Astronomical Society*, vol. 406, no. 1, pp. 14–21, 2010.
- [87] K. Land and J. Magueijo, “Examination of evidence for a preferred axis in the cosmic radiation anisotropy,” *Physical Review Letters*, vol. 95, no. 7, Article ID 071301, 4 pages, 2005.
- [88] K. Land and J. Magueijo, “The axis of evil revisited,” *Monthly Notices of the Royal Astronomical Society*, vol. 378, no. 1, pp. 153–158, 2007.
- [89] K. Land and J. Magueijo, “Is the Universe odd?” *Physical Review D*, vol. 72, no. 10, Article ID 101302, 5 pages, 2005.
- [90] J. Kim and P. Naselsky, “Anomalous parity asymmetry of the Wilkinson microwave anisotropy probe power spectrum data at low multipoles,” *Astrophysical Journal*, vol. 714, no. 2, pp. L265–L267, 2010.
- [91] J. Kim and P. Naselsky, “Anomalous parity asymmetry of WMAP power spectrum data at low multipoles: is it cosmological or systematics?” <http://arxiv.org/abs/1002.0148>.

## Review Article

# Testing Gaussianity, Homogeneity, and Isotropy with the Cosmic Microwave Background

L. Raul Abramo<sup>1</sup> and Thiago S. Pereira<sup>2</sup>

<sup>1</sup>*Instituto de Física, Universidade de São Paulo, CP 66318, CEP 05315-970 São Paulo, Brazil*

<sup>2</sup>*Instituto de Física Teórica, Universidade Estadual Paulista, CP 70532-2, CEP 01156-970 São Paulo, Brazil*

Correspondence should be addressed to L. Raul Abramo, lrwabramo@gmail.com

Received 3 February 2010; Accepted 12 May 2010

Academic Editor: Dragan Huterer

Copyright © 2010 L. R. Abramo and T. S. Pereira. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We review the basic hypotheses which motivate the statistical framework used to analyze the cosmic microwave background, and how that framework can be enlarged as we relax those hypotheses. In particular, we try to separate as much as possible the questions of gaussianity, homogeneity, and isotropy from each other. We focus both on isotropic estimators of nongaussianity as well as statistically anisotropic estimators of gaussianity, giving particular emphasis on their signatures and the enhanced “cosmic variances” that become increasingly important as our putative Universe becomes less symmetric. After reviewing the formalism behind some simple model-independent tests, we discuss how these tests can be applied to CMB data when searching for large-scale “anomalies”.

## 1. Introduction

According to our current understanding of the Universe, the morphology of the cosmic microwave background (CMB) temperature field, as well as all cosmological structures that are now visible, like galaxies, clusters of galaxies, and the whole web of large-scale structure, are probably the descendants of quantum process that took place some  $10^{-35}$  seconds after the Big Bang. In the standard lore, the machinery responsible for these processes is termed cosmic inflation and, in general terms, what it means is that microscopic quantum fluctuations pervading the primordial Universe are stretched to what correspond, today, to cosmological scales (see [1–3] for comprehensive introductions to inflation.) These primordial perturbations serve as initial conditions for the process of structure formation, which enhance these initial perturbations through gravitational instability. The subsequent (classical) evolution of these instabilities preserves the main statistical features of these fluctuations that were inherited from their inflationary origin—provided, of course, that we restrain ourselves to linear perturbation theory.

However, given that matter has a natural tendency to cluster, and this inevitably leads to nonlinearities (not to mention the sorts of complications that come with baryonic physics), the structures which are visible today are far from ideal probes of those statistical properties. CMB photons, on the other hand, to an excellent approximation experience free streaming since the time of decoupling ( $z \approx 1100$ ) and are therefore exempt from these non-linearities (except, of course, for secondary anisotropies such as the Rees-Sciama effect or the Sunyaev-Zel’dovich effect), which implies that they constitute an ideal window to the physics of the early Universe—see, for example, [4–6]. In fact, we can determine the primary CMB anisotropies as well as most of the secondary anisotropies on large scales, such as the Integrated Sachs-Wolfe effect, completely in terms of the initial conditions by means of a *linear kernel* as follows:

$$\Theta(\hat{n}) \equiv \frac{\Delta T(\hat{n}; \eta_0)}{T(\eta_0)} = \int d^3x' \int_0^{\eta_0} d\eta' \sum_i K_i(\vec{x}', \eta'; \hat{n}) S^i(\vec{x}', \eta'), \quad (1)$$

where  $\eta'$  is conformal time, and  $\mathcal{S}^i$  denote the initial conditions of all matter and metric fields (as well as their time derivatives, if the initial conditions are nonadiabatic). Here  $K_i$  is a linear kernel, or a retarded Green's function, that propagates the radiation field to the time and place of its detection, here on Earth. Since that kernel is insensitive to the statistical nature of the initial conditions (which can be thought of as constants which multiply the source terms), those properties are precisely transferred to the CMB temperature field  $\Theta$ .

The statistical properties of the primordial fluctuations are, to lowest order in perturbation theory, quite simple; because the quantum fluctuations that get stretched and enhanced by inflation are basically harmonic oscillators in their ground state, the distribution of those fluctuations is Gaussian, with each mode an independent random variable. The Fourier modes of these fluctuations are characterized by random phases (corresponding to the random initial values of the oscillators), with zero mean, and variances which are given simply by the field mass and the mode's wavenumber  $k = 2\pi/\lambda$ . The presence of higher-order interactions (which exist even for free fields, because of gravity) changes this simple picture, introducing higher-order correlations which destroy gaussianity—even in the simplest scenario of inflation [7–9]. However, since these interactions are typically suppressed by powers of the factor  $GH^2 \simeq 10^{-12}$ , where  $G$  is Newton's constant and  $H$  the Hubble parameter during inflation, the corrections are small—but, at least in principle, detectable [10–12].

Since these statistical properties are a generic prediction of (essentially) all inflationary models, they can also be inferred from two ingredients that are usually assumed as a first approximation to our Universe. First, since inflation was designed to stretch our Universe until it became spatially homogeneous and isotropic, it is reasonable to expect that all statistical momenta of the CMB should be spatially homogeneous and rotationally invariant, regardless of their general form. Second, in linear perturbation theory [13] where we have a large number of cosmological fluctuations evolving independently, we can expect, based on the *central limit theorem*, that the Universe will obey a Gaussian distribution.

The power of this program lies, therefore, in its simplicity: if the Universe is indeed Gaussian, homogeneous, and statistically isotropic (SI), then essentially all the information about inflation and the linear (low redshift) evolution of the Universe is encoded in the variance, or two-point correlation function, of large-scale cosmological structures and/or the CMB. As it turns out, the five-year dataset from the Wilkinson Microwave anisotropy probe (WMAP) strongly supports these predictions [11, 14]. Moreover, the measurements of the CMB temperature power spectrum by the WMAP team, alongside measurements of the matter power spectrum from existing survey of galaxies [15, 16] and data from type Ia supernovae [17–19], have shown remarkable consistency with a *concordance model* ( $\Lambda$ CDM), in which the cosmos figures as a Gaussian, spatially flat, approximately homogeneous, and statistically isotropic web

of structures composed mainly of baryons, dark matter, and dark energy.

However, while the detection of a nearly scale-invariant and Gaussian spectrum is a powerful boost to the idea of inflation, just knowing the variance of the primordial fluctuations is not sufficient to single out which particular inflationary model was realized in our Universe. For that, we will need not only the 2-point function, but the higher momenta of the distribution as well. Therefore, in order to break this model degeneracy, we must go beyond the framework of the  $\Lambda$ CDM, Gaussian, spatially homogeneous, and statistically isotropic Universe.

Reconstructing our cosmic history, however, is not the only reason to explore further the statistical properties of the CMB. The full-sky temperature maps by WMAP [11, 20] have revealed the existence of a series of large-angle anomalies—which, incidentally, were (on hindsight) already visible in the lower-resolution COBE data [21]. These anomalies suggest that at least one of our cherished hypotheses underlying the standard cosmological model might be wrong—even as a first-order approximation. Perhaps the most intriguing anomalies (described in more detail in other review papers in this volume) are the low value of the quadrupole and its alignment of the quadrupole ( $\ell = 2$ ) with the octupole ( $\ell = 3$ ) [22–27], the sphericity [26] (or lack of planarity [28]), of the multipole  $\ell = 5$ , and the north-south asymmetry [29–33]. In the framework of the standard cosmological model, these are very unlikely statistical events, and yet the evidence that they exist in the real data (and are not artifacts of poorly subtracted extended foregrounds—e.g., [34]) is strong.

Concerning theoretical explanations, even though we have by now an arsenal of *ad hoc* models designed to account for the existence of these anomalies, none has yet quite succeeded in explaining their origin. Nevertheless, they all share the point of view that the detected anomalies might be related to a deviation of gaussianity and/or statistical isotropy.

In this paper, we will describe, first, how to characterize, from the point of view of the underlying spacetime symmetries, both non-gaussianity and statistical anisotropy. We will adopt two guiding principles. The first is that gaussianity and SI, being completely different properties of a random variable, should be treated separately, whenever possible or practical. Second, since there is only one type of gaussianity and SI but virtually infinite ways away from them, it is important to try to measure these deviations without a particular model or anomaly in mind—although we may eventually appeal to particular models as illustrations or as a means of comparison. This approach is not new and, although not usually mentioned explicitly, it has been adopted in a number of recent papers [35, 36].

One of the main motivations for this model-independent approach is the difficult issue of *aprioristic* statistics; one can only test the random nature of a process if it can be repeated a very large (formally, infinite) number of times. Since the CMB only changes on a timescale of tens of millions of years, waiting for our surface of last scattering to probe a different region of the Universe is not a practical proposition. Instead,

we are stuck with one dataset (a sequence of apparently random numbers), which we can subject to any number of tests. Clearly, by sheer chance, about 30% of the tests will give a positive detection with 70% confidence level (CL), 10% will give a positive detection with 90% CL, and so on. With enough time, anyone can come up with detections of arbitrarily high significance—and ingenuity will surely accelerate this process. Hence, it would be useful to have a few guiding principles to inform and motivate our statistical tests, so that we do not end up shooting blindly at a finite number of fish in a small wheelbarrow.

This paper is divided into two parts. We start Part I by reviewing the basic statistical framework behind linear perturbation theory (Section 2). This serves as a motivation for Section 3, where we discuss the formal aspects of non-Gaussian and statistically isotropic models (Section 2.1), as well as Gaussian models of statistical anisotropy (Section 2.2). Part II is devoted to a discussion on model-independent cosmological tests of non-gaussianity and statistical anisotropy and their application to CMB data. We focus on two particular tests, namely, the multipole vectors statistics (Section 3) and functional modifications of the two-point correlation function (Section 4). After discussing how such tests are usually carried out when searching for anomalies in CMB data (Section 6.1), we present a new formalism which generalizes the standard procedure by including the ergodicity of cosmological data as a possible source of errors (Section 6.2). This formalism is illustrated in Section 7, where we carry a search of planar-type deviations of isotropy in CMB data. We then conclude in Section 8.

## Part I: The Linearized Universe

### 2. General Structure

We start by defining the temperature fluctuation field. Since the background radiation is known to have an average temperature of 2.725 K, we are interested only in deviations from this value at a given direction  $\hat{n}$  in the CMB sky. So let us consider the dimensionless function on  $S^2$  as follows:

$$\Theta(\hat{n}) \equiv \frac{T(\hat{n}) - T_0}{T_0}, \quad (2)$$

where  $T_0 = 2.725$  K is the blackbody temperature of the mean photon energy distribution—which, if homogeneity holds, is also equal to the ensemble average of the temperature.

In full generality, the fluctuation field is not only a function of the position vector  $\vec{n}$ , but also of the time in which our measurements are taken. In practice, the time and displacement of measurements vary so slowly that we can ignore these dependences altogether. Therefore, we can equally well consider this function as one defined only on the unit radius sphere  $S^2$ , for which the following decomposition holds:

$$\Theta(\hat{n}) = \sum_{\ell, m} a_{\ell m} Y_{\ell m}(\hat{n}). \quad (3)$$

Since the spherical harmonics  $Y_{\ell m}(\hat{n})$  obey the symmetry  $Y_{\ell m}^*(\hat{n}) = (-1)^m Y_{\ell, -m}(\hat{n})$ , the fact that the temperature field is a real function implies the identity  $a_{\ell m}^* = (-1)^m a_{\ell, -m}$ . This means that each temperature multipole  $\ell$  is completely characterized by  $2\ell + 1$  real degrees of freedom.

#### 2.1. From Inhomogeneities to Anisotropies: Linear Theory.

The ultimate source of anisotropies in the Universe is the inhomogeneities in the baryon-photon fluid, as well as their associated spacetime metric fluctuations. If the photons were in perfect equilibrium with the baryons up to a sharply defined moment in time (the so-called instant recombination approximation), their distribution would have only one parameter (the equilibrium temperature at each point), so that photons flying off in any direction would have exactly the same energies. In that case, the photons we see today coming from a line-of-sight  $\hat{n}$  would reflect simply the density and gravitational potentials (the “sources”) at the position  $R \hat{n}$ , where  $R$  is the radius to that (instantaneous) last scattering surface. Evidently, multiple scatterings at the epoch of recombination, combined with the fact that anisotropies themselves act as sources for more anisotropies, complicate this picture, and in general the relationship of the sources with the anisotropies must be calculated from either a set of Einstein-Boltzmann equations or, equivalently, from the line-of-sight integral equations coupled with the Einstein, continuity, and Euler equations [6].

Assuming for simplicity that recombination was instantaneous, at a time  $\eta_R$ , the linear kernels of (1) reduce to  $K_i(\vec{x}', \eta'; \hat{n}) \rightarrow \beta_i \delta(\eta' - \eta_R) \delta(\vec{x}' - \hat{n}R)$ , where  $R = \eta_0 - \eta_R$  and  $\beta_i$  are constant coefficients. The photon distribution that we measure on Earth would therefore be given by

$$\Theta(\hat{n}) \approx \sum_i \beta_i S^i(\vec{x}' = \hat{n}R, \eta' = \eta_R). \quad (4)$$

We can also express this result in terms of the Fourier spectrum of the sources as follows:

$$\Theta(\hat{n}) \approx \sum_i \beta_i \int \frac{d^3 k}{(2\pi)^3} e^{i\vec{k} \cdot \hat{n}R} S^i(\vec{k}, \eta_R). \quad (5)$$

Now we can use what is usually referred to as “Rayleigh’s expansion” (though Watson, in his classic book on Bessel functions, attributes this to Bauer, J. f. Math. LVI, 1859) as follows:

$$e^{i\vec{k} \cdot \vec{x}} = 4\pi \sum_{\ell m} i^\ell j_\ell(kx) Y_{\ell m}^*(\hat{k}) Y_{\ell m}(\hat{x}), \quad (6)$$

where  $j_\ell(z)$  are the spherical Bessel functions. Substituting (6) into (5) we obtain that

$$\begin{aligned} a_{\ell m} &= \int d^2 \hat{n} Y_{\ell m}^*(\hat{n}) \Theta(\hat{n}) \\ &\approx \int \frac{d^3 k}{(2\pi)^3} \Theta(\vec{k}) \times 4\pi i^\ell j_\ell(kR) Y_{\ell m}^*(\hat{k}), \end{aligned} \quad (7)$$

where we have loosely collected the sources into the term  $\Theta(\vec{k}) \equiv \sum_i \beta_i S^i(\vec{k}, \eta_R)$ . This expression conveys well the

simple relation between the Fourier modes and the spherical harmonic modes. Therefore, up to coefficients which are known given some background cosmology, the statistical properties of the harmonic coefficients  $a_{\ell m}$  are inherited from those of the Fourier modes  $\Theta(\vec{k})$  of the underlying matter and metric fields. Notice that the properties of the  $a_{\ell m}$ s under rotations, on the other hand, have nothing to do with the statistical properties of the fluctuations; they come directly from the spherical harmonic functions  $Y_{\ell m}$ .

*2.2. Statistics in Fourier Space.* The characterization of the statistics of random variables is most commonly expressed in terms of the correlation functions. The two-point correlation function is the ensemble expectation value,

$$C(\vec{k}, \vec{k}') \equiv \langle \Theta(\vec{k}) \Theta(\vec{k}') \rangle. \quad (8)$$

In the absence of any symmetries, this would be a generic function of the arguments  $\vec{k}$  and  $\vec{k}'$ , with only two constraints: first, because  $\Theta(\vec{x})$  is a real function,  $\Theta^*(\vec{k}) = \Theta(-\vec{k})$ , hence, in our definition  $C^*(\vec{k}, \vec{k}') = C(-\vec{k}, -\vec{k}')$ ; second, due to the associative nature of the expectation value,  $C(\vec{k}, \vec{k}') = C(\vec{k}', \vec{k})$ . It is obvious how to generalize this definition to 3, 4, or an arbitrary number of fields at different  $\vec{k}$ s (or “points”).

Let us first discuss the issue of gaussianity. If we say that the variables  $\Theta(\vec{k})$  are Gaussian random numbers, then all the information that characterizes their distribution is contained in their two-point function  $C(\vec{k}, \vec{k}')$ . The probability distribution function (pdf) is then formally given by

$$P[\Theta(\vec{k}), \Theta(\vec{k}')] \sim \exp\left[-\frac{\Theta(\vec{k})\Theta(\vec{k}')}{2C(\vec{k}, \vec{k}')}\right]. \quad (9)$$

In this case, all higher-order correlation functions are either zero (for odd numbers of points) or they are simply connected to the two-point function by means of Wick's Theorem as follows:

$$\begin{aligned} & \langle \Theta(\vec{k}_1) \Theta(\vec{k}_2) \cdots \Theta(\vec{k}_{2N}) \rangle_G \\ &= \sum_{i,j} \prod_{\alpha=1}^N B_{i,j}^\alpha \langle \Theta(\vec{k}_i) \Theta(\vec{k}_j) \rangle, \end{aligned} \quad (10)$$

where the sum runs over all permutations of the pairs of wave vectors and  $B_{i,j}$  are weights.

Second, let us consider the issue of homogeneity. A field is homogeneous if its expectation values (or averages) do not depend on the spatial points where they are evaluated. In terms of the  $N$ -point functions in real space, we should have the following

$$\begin{aligned} & \langle \Theta(\vec{x}_1) \Theta(\vec{x}_2) \cdots \Theta(\vec{x}_N) \rangle \\ & \xrightarrow{\text{Homog.}} C_N(\vec{x}_1 - \vec{x}_2, \dots, \vec{x}_{N-1} - \vec{x}_N). \end{aligned} \quad (11)$$

Writing this expression in terms of the Fourier modes, we get the following

$$\begin{aligned} & \langle \Theta(\vec{x}_1) \Theta(\vec{x}_2) \cdots \Theta(\vec{x}_N) \rangle \\ &= \int \frac{d^3 k_1 d^3 k_2 \cdots d^3 k_N}{(2\pi)^{3N}} e^{-i\vec{k}_1 \cdot \vec{x}_1} e^{-i\vec{k}_2 \cdot \vec{x}_2} \cdots e^{-i\vec{k}_N \cdot \vec{x}_N} \\ & \quad \times \langle \Theta(\vec{k}_1) \Theta(\vec{k}_2) \cdots \Theta(\vec{k}_N) \rangle. \end{aligned} \quad (12)$$

Homogeneity demands that the expression in (12) is a function of the *distances* between spatial points only, not of the points themselves. Hence, the expectation value in Fourier space on the right-hand side of this expression must be proportional to  $\delta(\vec{k}_1 + \vec{k}_2 + \cdots + \vec{k}_N)$ . In other words, the hypothesis of homogeneity constrains the  $N$ -point function in Fourier space to be of the following form:

$$\begin{aligned} & \langle \Theta(\vec{k}_1) \Theta(\vec{k}_2) \cdots \Theta(\vec{k}_N) \rangle_H \\ &= (2\pi)^3 \tilde{N}(\vec{k}_1, \vec{k}_2, \dots, \vec{k}_N) \\ & \quad \times \delta(\vec{k}_1 + \vec{k}_2 + \cdots + \vec{k}_N). \end{aligned} \quad (13)$$

Notice that the “ $(N-1)$ -spectrum” in Fourier space,  $\tilde{N}$ , can still be a function of the directions of the wavenumbers  $\vec{k}_i$  (it will be, in fact, a function of  $N-1$  such vectors, due to the global momentum conservation expressed by the  $\delta$ -function.) Models which realize the general idea of (13) correspond to homogeneous but anisotropic universes [37–40].

There is a useful diagrammatic illustration for the  $N$ -point functions in Fourier space that enforce homogeneity. Notice that we could use the  $\delta$ -function in (13) to integrate out any one of the momenta  $\vec{k}_i$  in (12). Let us instead rewrite the  $\delta$ -functions in terms of triangles, so for the 4-point function we have

$$\begin{aligned} & \delta(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4) \\ &= \int d^3 q \delta(\vec{k}_1 + \vec{k}_2 - \vec{q}) \delta(\vec{k}_3 + \vec{k}_4 + \vec{q}), \end{aligned} \quad (14)$$

whereas for the 5-point function we have

$$\begin{aligned} & \delta(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4 + \vec{k}_5) \\ &= \int d^3 q d^3 q' \delta(\vec{k}_1 + \vec{k}_2 - \vec{q}) \delta(\vec{q} + \vec{k}_3 - \vec{q}') \\ & \quad \times \delta(\vec{q}' + \vec{k}_4 + \vec{k}_5), \end{aligned} \quad (15)$$

and so on, so that the  $N$ -point  $\delta$ -function is reduced to  $N-2$  triangles with  $N-3$  “internal momenta” (the idea is nicely illustrated in Figure 1.) Substituting the expression for the  $N$ -point  $\delta$ -function into (12) and integrating out all external

momenta but the first ( $\vec{k}_1$ ) and last ( $\vec{k}_N$ ), the result is as follows

$$\begin{aligned} & \langle \Theta(\vec{x}_1)\Theta(\vec{x}_2)\cdots\Theta(\vec{x}_N) \rangle \\ &= \frac{1}{(2\pi)^{3N}} \int d^3k_1 d^3q_1 \cdots d^3q_{N-3} d^3k_N \\ & \quad \times e^{i\vec{k}_1 \cdot (\vec{x}_1 - \vec{x}_2)} e^{i\vec{q}_1 \cdot (\vec{x}_2 - \vec{x}_3)} \cdots e^{i\vec{q}_{N-3} \cdot (\vec{x}_{N-2} - \vec{x}_{N-1})} \\ & \quad \times e^{i\vec{k}_N \cdot (\vec{x}_{N-1} - \vec{x}_N)} \langle \Theta(\vec{k}_1)\Theta(\vec{q}_1 - \vec{k}_1) \cdots \Theta(\vec{k}_N) \rangle. \end{aligned} \quad (16)$$

This expression shows explicitly that the real-space  $N$ -point function above does not depend on any particular spatial point, only on the intervals between points.

Finally, what are the constraints imposed on the  $N$ -point functions that come from isotropy alone? Clearly, no dependence on the *directions* defined by the points,  $\vec{x}_i - \vec{x}_j$ , can arise in the final expression for the  $N$ -point functions in real space, so from (12) we see that the  $N$ -point function in Fourier space should depend only on the moduli of the wavenumbers—up to some momentum-conservation  $\delta$ -functions, which naturally carry vector degrees of freedom.

In this paper, we will mostly be concerned with tests of isotropy given homogeneity (but not necessarily Gaussianity), so in our case we will usually assume that the  $N$ -point function in Fourier space assumes the form given in (13).

**2.3. Statistics in Harmonic Space.** In the previous Section, we characterized the statistics of our field in Fourier space, which in most cases is most easily related to fundamental models such as inflation. Now we will change to harmonic representation, because that is what is most directly related to the observations of the CMB,  $\Theta(\hat{n})$ , which are taken over the unit sphere  $S^2$ . We will discuss mostly the two-point function here, and we defer a fuller discussion of  $N$ -point functions in harmonic space to Section 3.

From (7), we can start by taking the two-point function in harmonic space, and computing it in terms of the two-point function in Fourier space as follows

$$\begin{aligned} \langle a_{\ell m} a_{\ell' m'}^* \rangle &= \int \frac{d^3k d^3k'}{(2\pi)^6} (4\pi)^2 i^{\ell} (-i)^{\ell'} j_{\ell}(kR) j_{\ell'}(k'R) Y_{\ell m}(\hat{k}) \\ & \quad \times Y_{\ell' m'}^*(\hat{k}') \langle \Theta(\vec{k}) \Theta^*(\vec{k}') \rangle. \end{aligned} \quad (17)$$

Under the hypothesis of homogeneity, this expression simplifies considerably, leading to

$$\begin{aligned} \langle a_{\ell m} a_{\ell' m'}^* \rangle_{\text{H}} &= \int d^3k \frac{2}{\pi} i^{\ell} (-i)^{\ell'} j_{\ell}(kR) j_{\ell'}(kR) Y_{\ell m}(\hat{k}) Y_{\ell' m'}^*(\hat{k}) \times \tilde{N}_2(\vec{k}). \end{aligned} \quad (18)$$

If, in addition to homogeneity, we also assume isotropy, then  $\tilde{N}_2 \rightarrow P(k)$ , and the integration over angles factors

out, leading to the orthogonality condition for spherical harmonics as follows

$$\int d^2\hat{k} Y_{\ell m}(\hat{k}) Y_{\ell' m'}^*(\hat{k}) = \delta_{\ell\ell'} \delta_{mm'}, \quad (19)$$

and as a result the covariance of the  $a_{\ell m}$ s becomes diagonal as follows

$$\begin{aligned} \langle a_{\ell m} a_{\ell' m'}^* \rangle_{\text{H,I}} &= \delta_{\ell\ell'} \delta_{mm'} \int \frac{dk}{k} j_{\ell}^2(kR) \frac{2}{\pi} k^3 P(k) \\ &= 4\pi \delta_{\ell\ell'} \delta_{mm'} \int d \log k j_{\ell}^2(kR) \Delta_T^2(k) \\ &\equiv C_{\ell} \delta_{\ell\ell'} \delta_{mm'}, \end{aligned} \quad (20)$$

where we have defined the usual *temperature power spectrum*  $\Delta_T(k) = k^3 P(k)/2\pi^2$  in the middle line, and the angular power spectrum  $C_{\ell}$  in the last line of (20). As a pedagogical note, let us recall that the power spectrum basically expresses how much power the two-point correlation function has per unit  $\log k$  as follows:

$$\langle \Theta(\vec{x}) \Theta(\vec{x}') \rangle_{\text{H,I}} = \int d \log k \frac{\sin\left(k \left| \frac{\vec{x} - \vec{x}'}{|\vec{x} - \vec{x}'|} \right| \right)}{k \left| \frac{\vec{x} - \vec{x}'}{|\vec{x} - \vec{x}'|} \right|} \Delta_T^2(k). \quad (21)$$

In an analogous manner to what was done above, we can also construct the *angular* two-point correlation function in harmonic space as follows:

$$\langle \Theta(\hat{n}) \Theta(\hat{n}') \rangle = \sum_{\ell m} \sum_{\ell' m'} \langle a_{\ell m} a_{\ell' m'}^* \rangle Y_{\ell m}(\hat{n}) Y_{\ell' m'}^*(\hat{n}'). \quad (22)$$

The hypothesis of homogeneity by itself does not lead to significant simplifications, but isotropy leads to a very intuitive expression for the angular two-point function as follows:

$$\begin{aligned} \langle \Theta(\hat{n}) \Theta(\hat{n}') \rangle_{\text{H,I}} &= \sum_{\ell m} \sum_{\ell' m'} C_{\ell} \delta_{\ell\ell'} \delta_{mm'} Y_{\ell m}(\hat{n}) Y_{\ell' m'}^*(\hat{n}') \\ &= \sum_{\ell} C_{\ell} \frac{2\ell + 1}{4\pi} P_{\ell}(\hat{n} \cdot \hat{n}'). \end{aligned} \quad (23)$$

Clearly, not only is this expression the analogous in  $S^2$  of (21), but in fact the Fourier power spectrum  $\Delta_T^2(k)$  and the angular power spectrum  $C_{\ell}$  are defined in terms of each other as indicated in (23) as follows:

$$C_{\ell} = 4\pi \int d \log k j_{\ell}^2(kR) \Delta_T^2(k). \quad (24)$$

Now, using the facts that the spherical Bessel function of order  $\ell$  peaks when its argument is approximately given by  $\ell$ , and that  $\int d \log z j_{\ell}^2(z) = 1/(2\ell(\ell + 1))$ , we obtain the following (this is one type of what has become known in the literature as *Limber's approximations*)

$$C_{\ell} \approx \frac{2\pi}{\ell(\ell + 1)} \Delta_T^2\left(k = \frac{\ell}{R}\right). \quad (25)$$

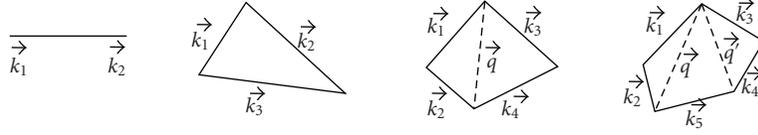


FIGURE 1: Diagrammatic representation of the 2, 3, 4, and 5-point correlation functions in Fourier space. The dashed lines represent internal momenta.

Incidentally, from this expression it is clear why it is customary to define

$$C_\ell \equiv \frac{\ell(\ell+1)}{2\pi} C_\ell \approx \Delta_T^2 \left( k = \frac{\ell}{R} \right). \quad (26)$$

Using (12), we can easily generalize the results of this subsection to  $N$ -point functions in  $S^2$  and in harmonic space, however, the assumption of isotropy alone does very little to simplify our life. The hypothesis of homogeneity, on the other hand, greatly simplifies the angular  $N$ -point functions, and most of the work in statistical anisotropy of the CMB that goes beyond the two-point function assumes that homogeneity holds. Notice that the issue of gaussianity is, as always, confined to the question of whether or not the two-point function holds all information about the distribution of the relevant variables and is therefore completely separated from questions about homogeneity and/or isotropy.

Also notice that the separable nature of the definition (22) implies here as well, like in Fourier space, a reciprocity relation for the correlation function

$$C(\hat{n}_1, \hat{n}_2) = C(\hat{n}_2, \hat{n}_1). \quad (27)$$

This symmetry must hold regardless of underlying models and is important in order to analyze the symmetries of the correlation function, as we will see later.

Before we move on, it is perhaps important to mention that the decomposition (22) is not unique. In fact, instead of the angular momenta of the parts,  $(\ell_1, m_1; \ell_2, m_2)$ , we could equally well have used the basis of total angular momentum  $(L, M; \ell_1, \ell_2)$  and decomposed that expression as

$$C(\hat{n}_1, \hat{n}_2) = \sum_{L, M} \sum_{\ell_1, \ell_2} \mathcal{A}_{\ell_1 \ell_2}^{LM} \mathcal{Y}_{\ell_1 \ell_2}^{LM}(\hat{n}_1, \hat{n}_2), \quad (28)$$

where  $\mathcal{Y}_{\ell_1 \ell_2}^{LM}$  are known as the bipolar spherical harmonics, defined by [41]

$$\mathcal{Y}_{\ell_1 \ell_2}^{LM}(\hat{n}_1, \hat{n}_2) = [Y_{\ell_1}(\hat{n}_1) \otimes Y_{\ell_2}(\hat{n}_2)]_{LM}, \quad (29)$$

where  $L$  and  $M = m_1 + m_2$  are the eigenvalues of the total and azimuthal angular momentum operators, respectively. This decomposition is completely equivalent to (22), and we can exchange from one decomposition to another by using the relation

$$\mathcal{A}_{\ell_1 \ell_2}^{LM} = \sum_{m_1 m_2} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle (-1)^{M+\ell_1-\ell_2} \sqrt{2L+1} \begin{pmatrix} \ell_1 & \ell_2 & L \\ m_1 & m_2 & -M \end{pmatrix}, \quad (30)$$

where the  $3 \times 2$  matrices above are the well-known 3-j coefficients. At this point, it is only a matter of mathematical convenience whether we choose to decompose the correlation function as in (22) or as in (28). Although the bipolar harmonics behave similarly to the usual spherical harmonics in many aspects, the modulations of the correlation function as described in this basis have a peculiar interpretation. We will not go further into detail about this decomposition here, as it is discussed at length in another review article in this volume.

**2.4. Estimators and Cosmic Variance.** Returning to the covariance matrix (20), we see that, if we assume gaussianity of the  $a_{\ell m}$ s, then the angular power spectrum suffices to describe statistically how much the temperature fluctuates in any given angular scale; all we have to do is to calculate the average (20). This can be a problem, though, since we have only one Universe to measure, and therefore only one set of  $a_{\ell m}$ s. In other words, the average in (20) is poorly determined.

At this point, the hypothesis that our Universe is spatially homogeneous and isotropic at cosmological scales comes not only as simplifying assumption about the spacetime symmetries, but also as a remedy to this unavoidable smallness of the working cosmologist's sample space. If isotropy holds, different cosmological scales are statistically independent, which means that we can take advantage of the ergodic hypothesis and trade averaging over an ensemble for averaging over space. In other words, for a given  $\ell$  we can consider each of the  $2\ell + 1$  real numbers in  $a_{\ell m}$  as statistically independent Gaussian random variables, and define a *statistical estimator* for their variances as the average

$$\hat{C}_\ell \equiv \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2. \quad (31)$$

The smaller the angular scales ( $\ell$  bigger), the larger the number of independent patches that the CMB sky can be divided into. Therefore, in this limit we should have

$$\lim_{\ell \rightarrow \infty} \hat{C}_\ell = C_\ell. \quad (32)$$

On the other hand, for large angular scales (small  $\ell$ 's), the number of independent patches of our Universe becomes smaller, and (31) becomes a weak estimation of the  $C_\ell$ s. This means that any statistical analysis of the Universe on large scales will be plagued by this intrinsic *cosmic sample variance*. Notice that this is an unavoidable limit as long as we have only one observable Universe.

Finally, it is important to keep in mind the clear distinction between the angular power spectrum  $C_\ell$  and its estimator (31). The former is a theoretical variable which can be calculated from first principles, as we have shown in Section 2.1. The latter, being a function of the data, is itself a random variable. In fact, if the  $a_{\ell m}$ s are Gaussian, then we can rewrite expression (31) as

$$\frac{(2\ell+1)}{C_\ell} \hat{C}_\ell = X_\ell, \quad X_\ell = \sum_{m=-\ell}^{\ell} \frac{|a_{\ell m}|^2}{C_\ell}, \quad (33)$$

where  $X_\ell$  is a chi-square random variable with  $2\ell+1$  degrees of freedom. According to the central limit theorem, when  $\ell \rightarrow \infty$ ,  $X_\ell$  approaches a standard normal variable (A standard normal variable is a Gaussian variable  $X$  with zero mean and unit variance. Any other Gaussian variable  $Y$  with mean  $\mu$  and variance  $\sigma$  can be obtained from  $X$  through  $Y = \sigma X + \mu$ .) which implies that  $\hat{C}_\ell$  will itself follow a Gaussian distribution. Its mean can be easily calculated using (20) and (31) and is of course given by

$$\langle \hat{C}_\ell \rangle = C_\ell, \quad (34)$$

which shows that the  $\hat{C}_\ell$ s are *unbiased* estimators of the  $C_\ell$ s. It is also straightforward to calculate its variance (valid for any  $\ell$ )

$$\langle (\hat{C}_\ell - C_\ell)(\hat{C}_{\ell'} - C_{\ell'}) \rangle = \frac{2}{2\ell+1} C_\ell^2 \delta_{\ell\ell'}. \quad (35)$$

Because this estimator does not couple different cosmological scales, it has the minimum *cosmic variance* we can expect from an estimator due to the finiteness of our sample—so it is *optimal* in that sense.  $\hat{C}_\ell$  is therefore the best estimator we can build to measure the statistical properties of the multipolar coefficients  $a_{\ell m}$  when both statistical isotropy and gaussianity hold.

In later Sections, we will explore angular or harmonic  $N$ -point functions for which the assumption of isotropy does not hold. However, it is important to remember at all times that we have only one map, which means one set of  $a_{\ell m}$ s. The estimator for the angular power spectrum,  $\hat{C}_\ell$ , takes into account *all* the  $a_{\ell m}$ s by dividing them into the different  $\ell$ s and summing over all  $m \in (-\ell, \ell)$ . Clearly, it will inherit a sample variance for small  $\ell$ s, when the  $a_{\ell m}$ s can only be divided into a few “independent parts”. As we try to estimate higher-order objects such as the  $N$ -point functions, we will have to subdivide the  $a_{\ell m}$ ’s into smaller and smaller subsamples, which are not necessarily independent (in the statistical sense) of each other. So, the price to pay for aiming at higher-order statistics is a worsening of the cosmic sample variance.

**2.5. Correlation and Statistical Independence.** The covariance given in (20) has two distinct, important properties. First, note that its diagonal entries, the  $C_\ell$ ’s, are  $m$ -independent coefficients; this is crucial for having statistical isotropy, as we will show latter. Second, statistical isotropy at the Gaussian level implies that different cosmological “scales” (understood

here as meaning the modes with total angular momentum  $\ell$  and azimuthal momentum  $m$ ) should be *statistically independent* of each other—and this is represented by the Kronecker deltas in (20).

In fact, statistical independence of cosmological scales is a particular property of Gaussian and statistically isotropic random fields and is not guaranteed to hold when gaussianity is relaxed. We will see in the next Section that the rotationally invariant 3-point correlation function (and in general any  $N > 2$  correlation function) couples to the three scales involved. In particular, if it happens that the Gaussian contribution of the temperature field is given by (20), but at least one of its non-Gaussian moments are nonzero, then the fact that a particular correlation is zero, like for example  $\langle a_{2m_1} a_{3m_2}^* \rangle$ , does not imply that the scales  $\ell = 2$  and  $\ell = 3$  are (statistically) independent. This is just a restatement of the fact that, while statistical independence implies null correlation, the opposite is not necessarily true. This can be illustrated by the following example: consider a random variable  $\alpha$  distributed as

$$P(\alpha) = \begin{cases} 1, & \alpha \in [0, 1], \\ 0, & \text{otherwise.} \end{cases} \quad (36)$$

Let us now define two other variables  $x = \cos(2\pi\alpha)$  and  $y = \sin(2\pi\alpha)$ . From these definitions, it follows that  $x$  and  $y$  are statistically dependent variables, since knowledge of the mean/variance of  $x$  automatically gives the mean/variance of  $y$ . However, these variables are clearly uncorrelated

$$\langle xy \rangle = \frac{1}{2\pi} \int_0^{2\pi} \cos \eta \sin \eta d\eta = 0. \quad (37)$$

Although correlations are among cosmologist’s most popular tools when analyzing CMB properties, statistical independence may turn out to be an important property as well, specially at large angular scales, where cosmic variance is more of a critical issue.

### 3. Beyond the Standard Statistical Model

Until now we have been analyzing the properties of Gaussian and statistically isotropic random temperature fluctuations. This gives us a fairly good statistical description of the Universe in its linear regime, as confirmed by the astonishing success of the  $\Lambda$ CDM model. This picture is incomplete though, and we have good reasons to search for deviations of either gaussianity and/or statistical isotropy. For example, the observed clustering of matter in galactic environments certainly goes beyond the linear regime where the central limit theorem can be applied, therefore leading to large deviations of gaussianity in the matter power spectrum statistics. Besides, deviations of the cosmological principle may leave an imprint in the statistical moments of cosmological observables, which can be tested by searching for spatial inhomogeneities [42] or directionalities [43].

But how do we plan to go beyond the standard model, given that there is only one Gaussian and statistically isotropic description of the Universe, but infinite possibilities

otherwise? This is in fact an ambitious endeavor, which may strongly depend on observational and theoretical hints on the type of signatures we are looking for. In the absence of extra input, it is important to classify these signatures in a general scheme, differentiating those which are non-Gaussian from those which are anisotropic. Furthermore, given that the signatures of non-gaussianity may in principle be quite different from that of statistical anisotropy, such a classification is crucial for data analysis, which requires sophisticated tools capable of separating these two issues. (Although gaussianity and homogeneity/isotropy are mathematically distinct properties, it is possible for a Gaussian but inhomogeneous/anisotropic model to look like an isotropic and homogeneous non-gaussian model. See, e.g., [44].)

We therefore start Section 3.1 by analyzing deviations of gaussianity when statistical isotropy holds. In Section 3.2, we keep the hypothesis of gaussianity and analyze the consequences of breaking statistical rotational invariance.

### 3.1. Non-Gaussian and SI Models

**3.1.1. Rotational Invariance of  $N$ -Point Correlation Functions.** We turn now to the question of non-Gaussian but statistically isotropic probabilities distributions. We will keep working with the  $N$ -point correlation function defined in harmonic space,

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} \dots a_{\ell_N m_N} \rangle, \quad (38)$$

since knowledge of these functions enables one to fully reconstruct the CMB temperature probability distribution. Specifically, we would like to know the form of any  $N$ -point correlation function which is invariant under arbitrary 3-dimensional spatial rotations. When rotated to a new (primed) coordinate system, the  $N$ -point correlation function transforms as

$$\begin{aligned} & \langle a_{\ell_1 m'_1} a_{\ell_2 m'_2} \dots a_{\ell_N m'_N} \rangle \\ &= \sum_{\text{all } m} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} \dots a_{\ell_N m_N} \rangle D_{m'_1 m_1}^{\ell_1} D_{m'_2 m_2}^{\ell_2} \dots D_{m'_N m_N}^{\ell_N}, \end{aligned} \quad (39)$$

where the  $D_{m'_i m_i}^{\ell_i}(\alpha, \beta, \gamma)$ s are the coefficients of the Wigner rotation-matrix, which depend on the three Euler-angles  $\alpha$ ,  $\beta$ , and  $\gamma$  characterizing the rotation. Notice that in this notation the primed (rotated) system is indicated by the primed  $m$ s. For the 2-point correlation function, we have already seen that the well-known expression

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle = (-1)^{m_2} C_{\ell_1} \delta_{\ell_1 \ell_2} \delta_{m_1, -m_2} \quad (40)$$

does the job

$$\begin{aligned} \langle a_{\ell_1 m'_1} a_{\ell_2 m'_2} \rangle &= C_{\ell_1} \left( \sum_{m_1} (-1)^{m_1} D_{m'_1 m_1}^{\ell_1} D_{m'_2 -m_1}^{\ell_2} \right) \delta_{\ell_1 \ell_2} \\ &= (-1)^{m'_2} C_{\ell_1} \delta_{\ell_1 \ell_2} \delta_{m'_1, -m'_2}. \end{aligned} \quad (41)$$

Note the importance of the *angular spectrum*,  $C_\ell$ , being an  $m$ -independent function.

What about the 3-point function? In this case, the invariant combination is found to be

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = B_{\ell_1 \ell_2 \ell_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \quad (42)$$

which can be verified by straightforward calculations. Again, the nontrivial physical content of this statistical moment is contained in an arbitrary but otherwise  $m$ -independent function: the *bispectrum*  $B_{\ell_1 \ell_2 \ell_3}$  [45–47]. As we anticipated in Section 2, rotational invariance of the 3-point correlation is not enough to guarantee statistical independence of the three cosmological scales involved in the bispectrum, although in principle a particular model could be formulated to ensure that  $B_{\ell_1 \ell_2 \ell_3} \propto \delta_{\ell_1 \ell_2} \delta_{\ell_2 \ell_3}$ , at least for some subset of a general geometric configuration of the 3-point correlation function.

These general properties hold for all the  $N$ -point correlation function. For the 4-point correlation function, for example, Hu [48] has found the following rotationally invariant combination

$$\begin{aligned} & \langle a_{\ell_1 m_1} \dots a_{\ell_4 m_4} \rangle \\ &= \sum_{LM} Q_{\ell_3 \ell_4}^{\ell_1 \ell_2}(L) (-1)^M \begin{pmatrix} \ell_1 & \ell_2 & L \\ m_1 & m_2 & -M \end{pmatrix} \begin{pmatrix} \ell_3 & \ell_4 & L \\ m_3 & m_4 & M \end{pmatrix}, \end{aligned} \quad (43)$$

where the  $Q_{\ell_3 \ell_4}^{\ell_1 \ell_2}(L)$  function is known as the *trispectrum*, and  $L$  is an internal angular momentum needed to ensure parity invariance. In a likewise manner, it can be verified that the following expression

$$\begin{aligned} & \langle a_{\ell_1 m_1} \dots a_{\ell_5 m_5} \rangle \\ &= \sum_{LM L'M'} P_{\ell_3 \ell_4 \ell_5}^{\ell_1 \ell_2}(L, L') (-1)^{M+M'} \begin{pmatrix} \ell_1 & \ell_2 & L \\ m_1 & m_2 & -M \end{pmatrix} \\ & \times \begin{pmatrix} \ell_3 & \ell_4 & L' \\ m_3 & m_4 & -M' \end{pmatrix} \begin{pmatrix} \ell_5 & L & L' \\ m_5 & M & M' \end{pmatrix} \end{aligned} \quad (44)$$

gives the rotationally invariant quadrispectrum  $P_{\ell_3 \ell_4 \ell_5}^{\ell_1 \ell_2}(L, L')$ .

The examples above should be enough to show how the general structure of these functions emerges under SI; apart from an  $m$ -independent function, every pair of momenta  $\ell_i$  in these functions are connected by a triangle, which in turn connects itself to other triangles through internal momenta when more than 3 scales are present. In Figure 2, we show some diagrams representing the functions above.

Although we have always shown  $N$ -point functions which are rotationally invariant, the procedure used for obtaining them was rather intuitive, and therefore does not offer a recipe for constructing general invariant correlation functions. Furthermore, it does not guarantee that this procedure can be extended for arbitrary  $N$ s. Here we will present a recipe for doing that, which also guarantees the uniqueness of the solution.

The general recipe for obtaining the rotationally invariant  $N$ -point function is as follows: from the expression

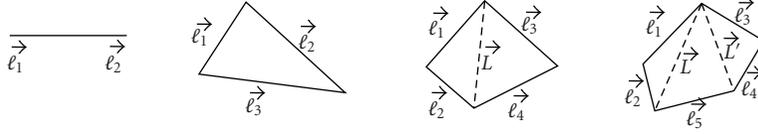


FIGURE 2: Diagrammatic representation of the 2, 3, 4, and 5-point correlation functions in harmonic space. Here  $\vec{\ell}$  actually represents the pair  $(\ell, m)$ .

(39) above, we start by contracting every pairs of Wigner functions, where by “contracting” we mean using the identity

$$D_{m'_1 m_1}^{\ell_1}(\omega) D_{m'_2 m_2}^{\ell_2}(\omega) = \sum_{L, M, M'} \begin{pmatrix} \ell_1 & \ell_2 & L \\ m'_1 & m'_2 & -M \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & L \\ m_1 & m_2 & -M' \end{pmatrix} \times (2L+1)(-1)^{M+M'} D_{M'M}^L(\omega), \quad (45)$$

and where  $\omega = \{\alpha, \beta, \gamma\}$  is a shortcut notation for the three Euler angles. Once this contraction is done, there will remain  $N/2D$ -functions, which can again be contracted in pairs. This procedure should be repeated until there is only one Wigner function left, in which case we will have an expression of the following form:

$$\langle a_{\ell_1 m_1} \cdots a_{\ell_N m_N} \rangle = \sum_{\text{all } m'} \langle a_{\ell_1 m'_1} \cdots a_{\ell_N m'_N} \rangle \times \sum \text{geometrical factors} \times D_{MM'}^L(\omega). \quad (46)$$

Now, we see that the only way for this combination to be rotationally invariant is when the remaining  $D_{MM'}^L$  function above does not depend on  $\omega$ , that is,  $D_{MM'}^L(\omega) = \delta_{L0} \delta_{M0} \delta_{M'0}$ . Once this identity is applied to the geometrical factors, we are done, and the remaining terms inside the primed  $m$ -summation will give the rotationally invariant  $(N-1)$ -spectrum.

As an illustration of this algorithm, let us construct the rotationally invariant spectrum and bispectrum. For the 2-point function, there is only one contraction to be done, and after we simplify the last Wigner function, we arrive at

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle = \left[ \sum_{m'_1} \frac{\langle |a_{\ell_1 m'_1}|^2 \rangle}{2\ell_1 + 1} \right] (-1)^{m_2} \delta_{\ell_1 \ell_2} \delta_{m_1, -m_2}, \quad (47)$$

where, of course

$$C_\ell \equiv \frac{1}{2\ell + 1} \sum_m \langle |a_{\ell m}|^2 \rangle \quad (48)$$

is the well-known definition of the temperature angular spectrum. For the 3-point function, there are two contractions, and the simplification of the last Wigner function gives

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle = \left[ \sum_{m'_1, m'_2, m'_3} \langle a_{\ell_1 m'_1} a_{\ell_2 m'_2} a_{\ell_3 m'_3} \rangle \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m'_1 & m'_2 & m'_3 \end{pmatrix} \right] \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \quad (49)$$

From this expression and the orthogonality of the 3-j symbols (see the Appendix), we can immediately identify the definition of the bispectrum as follows:

$$B_{\ell_1 \ell_2 \ell_3} \equiv \sum_{m_1, m_2, m_3} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \quad (50)$$

It should be mentioned that this recipe not only enables us to establish the rotational invariance of any  $N$ -point correlation function, but it also furnishes a straightforward definition of unbiased estimators for the  $N$ -point functions. All we have to do is to drop the ensemble average of the primed  $a_{\ell m}$ 's. So, for example, for the 2- and 3-point functions above, the unbiased estimators are given, respectively, by

$$\hat{C}_\ell = \frac{1}{2\ell + 1} \sum_m a_{\ell m} a_{\ell m}^*, \quad (51)$$

$$\hat{B}_{\ell_1 \ell_2 \ell_3} = \sum_{m_1, m_2, m_3} a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}.$$

Notice that isotropy plays the same role, in  $S^2$ , that homogeneity plays in  $\mathbb{R}^3$ . What enforces homogeneity in  $\mathbb{R}^3$  is the Fourier-space  $\delta$ -functions, as in the discussion around Figure 1. However, in  $S^2$ , the equivalents of the Fourier modes are the harmonic modes, for which there is only a discrete notion of orthogonality—and no Dirac  $\delta$ -function. What we found above is that the Wigner 3-j symbols play the same role as the Fourier space  $\delta$ -functions; they are the enforcers of isotropy (rotational invariance) for the  $N$ -point angular correlation function. Hence, the diagrammatic representations of the constituents of the  $N$ -point functions in Fourier (Figure 1) and in harmonic space (Figure 2) really do convey the same physical idea—one in  $\mathbb{R}^3$ , the other in  $S^2$ .

**3.2. Gaussian and Statistically Anisotropic Models.** In the last Section, we have developed an algorithm which enables one to establish the rotational invariance of any  $N$ -point correlation function. As we have shown, this is also an algorithm for building unbiased estimators of non-Gaussian correlations. In this Section, we will change the perspective and analyze the case of Gaussian but statistically anisotropic models of the Universe.

There are many ways in which statistical anisotropy may be manifested in CMB. From a fundamental perspective, a short phase of inflation which produces just enough e-folds to solve the standard Big Bang problems may leave imprints on the largest scales of the Universe, provided that the

spacetime is sufficiently anisotropic at the onset of inflation [39]. Another source of anisotropy may result from our inability to efficiently clean foreground contaminations from temperature maps. Usually, the cleaning procedure involves the application of a mask function in order to eliminate contaminations of the galactic plane from raw data. As a consequence, this procedure may either induce, as well as hide some anomalies in CMB maps [28].

It is important to mention that these two examples can be perfectly treated as Gaussian: in the first case, the anisotropy of the spacetime can be established in the linear regime of perturbation theory and therefore will not destroy gaussianity of the quantum modes, provided that they are initially Gaussian. In the second case, the mask acts linearly over the temperature maps, therefore preserving its probability distribution [49].

**3.2.1. Primordial Anisotropy.** Recently, there have been many attempts to test the isotropy of the primordial Universe through the signatures of an anisotropic inflationary phase [38–40, 50, 51]. A generic prediction of such models is the linear coupling of the scalar, vector, and tensor modes through the spatial shear, which is in turn induced by anisotropy of the spacetime [38]. Whenever that happens, the matter power spectrum, defined in a similar way as in (13), will acquire a directionality dependence due to this type of see-saw mechanism. This dependence can be accommodated in a harmonic expansion of the form

$$P(\vec{k}) = \sum_{\ell, m} r_{\ell m}(k) Y_{\ell m}(\hat{k}), \quad (52)$$

where the reality of  $P(\vec{k})$  requires that  $r_{\ell m}(k) = (-1)^m r_{\ell, -m}^*(k)$ . Given that temperature perturbations  $\Theta(\vec{x})$  are real, their Fourier components must satisfy the relation  $\Theta(\vec{k}) = \Theta^*(-\vec{k})$ . This property taken together with the definition (13) implies that

$$P(\vec{k}) = P(-\vec{k}), \quad (53)$$

which in turn restricts the  $\ell$ s in (52) to even values. Also, note that by relaxing the assumption of spatial isotropy, we are only breaking a continuous spacetime symmetry, but discrete symmetries such as parity should still be present in this class of models. Indeed, by imposing invariance of the spectrum under the transformation  $z \rightarrow -z$ , we find that  $(-1)^{\ell-m} = 1$ . Similarly, invariance under the transformations  $x \rightarrow -x$  and  $y \rightarrow -y$  implies the conditions  $r_{\ell m} = (-1)^m r_{\ell, -m}$  and  $r_{\ell m} = r_{\ell, -m}$ , respectively. Gathering all these constraints with the parity of  $\ell$ , we conclude that

$$r_{\ell m} \in \mathbb{R}, \quad \ell, m \in 2\mathbb{N}. \quad (54)$$

That is, from the initial  $2\ell + 1$  degrees of freedom, only  $\ell/2 + 1$  of them contribute to the anisotropic spectrum [39].

**3.2.2. Signatures of Statistical Anisotropies.** The selection rules (54) are the most generic predictions we can expect

from models with global break of anisotropy. We will now work out the consequences of these rules to the temperature power spectrum and check whether they can say something about the CMB large-scale anomalies.

From the expressions (17) and (52), we can immediately calculate the most general anisotropic covariance matrix [37] as follows

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2}^* \rangle = \sum_{\ell_3, m_3} \mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} H_{\ell_1 \ell_2}^{\ell_3 m_3}, \quad (55)$$

where

$$\begin{aligned} \mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} &= (-1)^{m_1} \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \\ &\times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ -m_1 & m_2 & m_3 \end{pmatrix} \end{aligned} \quad (56)$$

are the Gaunt coefficients resulting from the integral of three spherical harmonics (see the Appendix). These coefficients are zero unless the following conditions are met:

$$\begin{aligned} \ell_1 + \ell_2 + \ell_3 &\in 2\mathbb{N}, \\ m_1 + m_2 + m_3 &= 0, \end{aligned} \quad (57)$$

$$|\ell_i - \ell_j| \leq \ell_k \leq \ell_i + \ell_j, \quad \forall i, j, k \in \{1, 2, 3\}.$$

The remaining coefficients in (55) are given by

$$H_{\ell_1 \ell_2}^{\ell_3 m_3} = 4\pi i^{\ell_1 - \ell_2} \int_0^\infty d \log k r_{\ell_3 m_3}(k) j_{\ell_1}(kR) j_{\ell_2}(kR) \quad (58)$$

and correspond to the anisotropic generalization of temperature power spectrum (20).

The selection rules (57), taken together with (54), lead to important signatures in the CMB. In particular, since  $\ell_3$  is even, the quantity  $\ell_1 \pm \ell_2$  must also be even, that is, multipoles with different parity do not couple to each other in this class of models

$$\langle a_{\ell_1 m_1} a_{\ell_2 m_2}^* \rangle = 0, \quad \ell_1 + \ell_2 = \text{odd}. \quad (59)$$

This result is in fact expected on theoretical grounds, because by breaking a continuous symmetry (isotropy) we cannot expect to generate a discrete signature (parity) in CMB. However, notice that the absence of correlations between, say, the quadrupole and the octupole, does not imply that there will be no alignment between them. One example of this would be a covariance matrix of the form  $C_{\ell m} \delta_{\ell \ell'} \delta_{m m'}$ . If the  $C_{\ell, 0}$  happens to be zero, for example, then all multipoles will present a preferred direction (in this case, the  $z$ -axis).

**3.2.3. Isotropic Signature of Statistical Anisotropy.** We have just shown that a generic consequence of an early anisotropic phase of the Universe is the generation of even-parity signatures in CMB maps. Interestingly, these signatures may be present even in the isotropic angular spectrum, since the  $C_\ell$ s acquire some additional modulations in the presence of statistical anisotropies. In principle, these modulations could

be constrained by measuring an *effective* angular spectrum of the form

$$\langle a_{\ell m} a_{\ell m}^* \rangle = C_\ell + \epsilon \sum_{\ell' > 0} g_{m, -m, 0}^{\ell \ell \ell'} H_{\ell \ell}^{\ell' 0}, \quad (60)$$

where we have introduced a small  $\epsilon$  parameter to quantify the amount of primordial anisotropy.

In order to constrain these modulations, we have to build a statistical estimator for  $\langle a_{\ell m} a_{\ell m}^* \rangle$ . Since we are looking for the diagonal entries of the matrix (55), a first guess would be

$$\widehat{C}_\ell^{\text{eff}} = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} a_{\ell m} a_{\ell m}^*. \quad (61)$$

To check whether this is an unbiased estimator of the effective angular spectrum, we apply it to (55) and take its average

$$\langle \widehat{C}_\ell^{\text{eff}} \rangle = C_\ell + \epsilon \sum_{\ell' > 0, m} g_{m, -m, 0}^{\ell \ell \ell'} H_{\ell \ell}^{\ell' 0}. \quad (62)$$

Using the definition (56) of the Gaunt coefficients, the  $m$ -summation in the expression above becomes [41]

$$\sum_m (-1)^{\ell-m} \begin{pmatrix} \ell & \ell & \ell' \\ m & -m & 0 \end{pmatrix} = \sqrt{2\ell + 1} \delta_{\ell', 0} = 0, \quad (63)$$

where the last equality follows because  $\ell' > 0$ . Consequently, we conclude that

$$\langle \widehat{C}_\ell^{\text{eff}} \rangle = C_\ell \neq \langle a_{\ell m} a_{\ell m}^* \rangle. \quad (64)$$

At first sight, this result may seem innocuous, showing only that this is not an appropriate estimator for (60). Note, however, that (61) is in fact the estimator of the angular spectrum usually applied to CMB data under the assumption of statistical isotropy. In other words, by means of the usual procedure we may be neglecting important information about statistical anisotropy. Moreover, the cosmic variance induced by the application of this estimator on anisotropic CMB maps is small, because, as it can easily be checked,

$$\langle (\widehat{C}_\ell^{\text{eff}} - C_\ell) (\widehat{C}_{\ell'}^{\text{eff}} - C_{\ell'}) \rangle = \frac{2}{2\ell + 1} C_\ell^2 \delta_{\ell \ell'} + \mathcal{O}(\epsilon^2). \quad (65)$$

This result shows that the construction of statistical estimators strongly depends on our prejudices about what non-gaussianity and statistical anisotropy should look like. Consequently, an estimator built to measure one particular property of the CMB may equally well hide other important signatures. One possible solution to this problem is to let the construction of our estimators be based on what the observations seem to tell us, as we will do in the next section.

## Part II: Cosmological Tests

So much for mathematical formalism. We will now turn to the question of how the hypotheses of gaussianity and statistical isotropy of the Universe can be tested. Though we are primarily interested in applying these tests to CMB temperature maps, most of the tools we will be dealing with

can be applied to polarization ( $E$ - and  $B$ -mode maps) and to catalogs of large-scale structures as well.

Testing the gaussianity and SI of our Universe is a difficult task. Specially because, as we have seen, there is only one Gaussian and SI Universe, but infinitely many universes which are neither Gaussian nor isotropic. So what type of non-gaussianity and statistical anisotropy should we test for? In order to attack this problem, we can follow two different routes. In the *bottom-up* approach, models for the cosmic evolution are formulated in such a way as to account for some specific deviations from gaussianity and SI. These physical principles range from nontrivial cosmic topologies [52–54], primordial magnetic fields [55–59], and local [43, 60–62] and global [38–40, 50] manifestation of anisotropy, to nonminimal inflationary models [51, 63–67]. The main advantage of the bottom-up approach is that we know exactly what feature of the CMB is being measured. One of its drawbacks is the plethora of different models and mechanisms that can be tested.

The second possibility is the *top-down*, or model-independent approach. Here, we are not concerned with the mechanisms responsible for deviations of gaussianity or SI, but rather with the qualitative features of any such deviation. Once these features are understood, we can use them as a guide for model building. Examples here include constructs of *a posteriori* statistics [29, 35, 36] and functional modifications of the two-point correlation function [28, 37, 68–71].

In the next Section, we will explore two different model-independent tests: one based on functional modifications of the two-point correlation function, and another one based on the so-called Maxwell's multipole vectors.

## 4. Multipole Vectors

Multipole vectors were first introduced in cosmology by Copi et al. [23] as a new mathematical representation for the primordial temperature field, where each of its multipoles  $\ell$  are represented by  $\ell$  unit real vectors. Later, it was realized that this idea is in fact much older [72], being proposed originally by J. C. Maxwell in his *Treatise on Electricity and Magnetism*.

The power of this approach is that the multipole vectors can be entirely calculated in terms of a temperature map, without any reference to external reference frames. This makes them ideal tools to test the morphology of CMB maps, like the quadrupole-octupole alignment.

The purpose of the following presentation is only comprehensiveness. A mathematically rigorous introduction to the subject can be found in references [72–74], as well as other papers in this review volume.

*4.1. Maxwell's Representation of Harmonic Functions.* We start our presentation of the multipole vectors by recalling some terminology. A harmonic function in three dimensions is any (twice differentiable) function  $h$  that satisfies Laplace's equation

$$\nabla^2 h = 0, \quad (66)$$

where  $\nabla^2$  is the Laplace operator. In spherical coordinates, the formal solution to Laplace's equation which is regular at the origin ( $r = 0$ ) is

$$h = \sum_{\ell=0}^{\infty} h_{\ell}(r, \theta, \varphi), \quad h_{\ell} = \sum_{m=-\ell}^{\ell} a_{\ell m} r^{\ell} Y_{\ell m}(\theta, \varphi). \quad (67)$$

The functions  $r^{\ell} Y_{\ell m}$  are known as the *solid spherical harmonics* [74]. Since they agree with the usual spherical harmonics on the unit sphere, it is sometimes stated in the literature that the latter form a set of harmonic functions. This is an abuse of nomenclature though, and the reader should be careful.

Given the scalar nature of Laplace's operator, it is possible to find solutions to (66) in terms of Cartesian coordinates. Such solutions can be constructed by combining homogeneous polynomials (i.e., sums of monomials of the same order) of order  $\ell$ :

$$h = \sum_{\ell=0}^{\infty} h_{\ell}(x, y, z), \quad h_{\ell} = \sum_{abc}^{\ell} \lambda_{abc} x^a y^b z^c, \quad (a + b + c = \ell). \quad (68)$$

In three dimensions, the most general homogeneous polynomial of order  $\ell$  contains  $(\ell + 2)/(2\ell!)$  independent coefficients. However, since each polynomial must independently satisfy (66), precisely  $\ell!/(2!(\ell - 2)!)$  of these coefficients will depend on each other. This constraint leaves us with  $(\ell + 1)(\ell + 2)/2 - \ell(\ell - 1)/2 = 2\ell + 1$  independent degrees of freedom in each multipole  $\ell$ —which is, of course, the same number of independent degrees of freedom appearing in (67).

Based on this analysis, Maxwell introduced his own representation of harmonic functions. He noticed that by successively applying directional derivatives of the form  $\vec{v} \cdot \nabla \equiv \nabla_{\vec{v}}$  over the monopole potential  $1/r$ , where  $r = \sqrt{x^2 + y^2 + z^2}$  and  $\vec{v}$  is a unit vector, he could construct solutions of the form (68). That is,

$$f_{\ell}(x, y, z) = \lambda_{\ell} \nabla_{\vec{v}_\ell} \cdots \nabla_{\vec{v}_2} \nabla_{\vec{v}_1} \frac{1}{r} \Big|_{r=1}, \quad (69)$$

where  $\lambda_{\ell}$  are real constants. We are now going to show that this construction does indeed lead to solutions of the form (68). First, note that there is a pattern which emerges from successive application of directional derivatives over the monopole function

$$\begin{aligned} f_0 &= \frac{1}{r}, \\ f_1 &= \nabla_{\vec{v}_1} f_0 = \frac{-\vec{v}_1 \cdot \vec{r}}{r^3}, \\ f_2 &= \nabla_{\vec{v}_2} f_1 = \frac{3(\vec{v}_2 \cdot \vec{r})(\vec{v}_1 \cdot \vec{r}) - r^2(\vec{v}_1 \cdot \vec{v}_2)}{r^3}, \end{aligned} \quad (70)$$

and so on. By induction, one can show that the general expression will be given by

$$f_{\ell} = \frac{(-1)^{\ell} (2\ell - 1)!! \prod_{i=1}^{\ell} \vec{v}_i \cdot \vec{r} + r^2 Q_{\ell-2}}{r^{2\ell+1}}, \quad (71)$$

where  $Q_{\ell-2}$  is a homogeneous polynomial of order  $\ell - 2$  which only involves combinations of the vectors  $\vec{v}_i$  and  $\vec{r}$ .

The numerator of the function  $f_{\ell}$  given by (71) is clearly a homogeneous polynomial of order  $\ell$  (as one can easily check for some  $\ell$ s and also prove by mathematical induction). A not so obvious result is that this polynomial is also harmonic. To prove that, let us define

$$g_{\ell} = (-1)^{\ell} (2\ell - 1)!! \prod_{i=1}^{\ell} \vec{v}_i \cdot \vec{r} + r^2 Q_{\ell-2} \quad (72)$$

and consider the application of the operator  $\nabla^2$  over the combination  $r^{\alpha} g_{\ell}$ . For the  $x$ -component we get

$$\partial_x^2 (r^{\alpha} g_{\ell}) = r^{\alpha} \partial_x^2 g_{\ell} + 2\alpha r^{\alpha-2} x \partial_x g_{\ell} + [\alpha r^{\alpha-2} + \alpha(\alpha-2)x^2 r^{\alpha-4}] g_{\ell}. \quad (73)$$

Repeating this process for the  $y$ - and  $z$ -components and then adding the results, we find

$$\begin{aligned} \nabla^2 (r^{\alpha} g_{\ell}) &= r^{\alpha} \nabla^2 g_{\ell} + 2\alpha r^{\alpha-2} (x \partial_x g_{\ell} + y \partial_y g_{\ell} + z \partial_z g_{\ell}) \\ &\quad + \alpha(\alpha+1) r^{\alpha-2} g_{\ell} \\ &= r^{\alpha} \nabla^2 g_{\ell} + \alpha(\alpha+2\ell+1) (r^{\alpha-2} g_{\ell}), \end{aligned} \quad (74)$$

where in the last step we have used Euler's theorem on homogeneous functions, that is,  $\vec{r} \cdot \nabla g_{\ell} = \ell g_{\ell}$ . If we now choose  $\alpha = -(2\ell + 1)$ , we find immediately that

$$\nabla^2 \left( \frac{g_{\ell}}{r^{2\ell+1}} \right) = \frac{\nabla^2 g_{\ell}}{r^{2\ell+1}}. \quad (75)$$

By construction, the left-hand side of the above expression is equal to  $\nabla^2 f_{\ell}$ . But according to the definition (69), this quantity is also zero, since Laplace's operator commutes with directional derivatives and  $\nabla^2(1/r) = 0$  for  $r > 0$ . Therefore,  $\nabla^2 g_{\ell} = 0$  and  $g_{\ell}$  is harmonic, which completes our proof.

In conclusion, Maxwell's construction of harmonic functions, (69), is completely equivalent to the standard representation in terms of spherical harmonics. More importantly, this gives a one-to-one relationship between temperature maps (given by the  $a_{\ell m}$ 's) and  $\ell$  unit vectors  $\vec{v}_i$ . This means that the multipole vectors can be directly calculated from a CMB map, without any reference to external reference frames or additional geometrical constructs. The reader interested in algorithms to construct the vectors from CMB maps may check references [23, 72], as well as the other papers in this volume that review this approach.

**4.2. Multipole Vectors Statistics.** It should be clear from the discussion above that the multipole vectors give an intuitive way to discover, interpret, and visualize phase correlations between different multipoles in the CMB maps. But they also can reveal intramultipole features, such as planarity (when a given multipole presents a preferred plane). Some of the most conspicuous hints of statistical anisotropy in the CMB, like the quadrupole-octupole alignment, have indeed been first found with less mathematically elegant methods

[22, 27], but are best described in terms of the multipole vector formalism [23, 24, 26, 72]. However, this case should also sound an alarm, because some feature of a (presumably) random realization was found, then further scrutinized with a certain test which was, to some extent (intentionally or not), tailored to single out that very feature.

The pitfalls of aprioristic approaches are sometimes unavoidable, and all we can do is to take a second look at our sample with a more generic set of tools, to try to assess how significant our result really is in the context of a larger set of statistical tests. Multipole vectors are, in fact, ideally suited to this, since they can be found for all multipoles, and it is relatively easy to construct scalar combinations of these vectors with the usual methods of linear algebra. Also convenient is the fact that simulating these vectors from maps, or even directly, is also relatively easy, so the standard model of a Gaussian random field can be easily translated into the pdfs for the tests constructed with the multipole vectors. An important drawback of the multipole vectors is that, because they are computed in terms of equations which are nonlinear in the temperature fields, the distribution functions of statistical tests involving these vectors are highly non-Gaussian [75]. This means that these statistics usually have to be estimated by means of simulations (usually assuming that the underlying temperature field is itself Gaussian). Another delicate issue is how stable the multipole vectors are to instrumental noise and sky cut—and here, again, we must rely on numerical simulations to compare the observations with theoretical models.

In order to see how one should go about constructing a general set of tests, it should be noted first that the multipole vectors define *directions*, but they have no sense (they are “headless vectors”). This follows from the fact that the sign of the constants  $\lambda_\ell$  is degenerated with the sign of the vectors  $\vec{v}_{\ell,p}$  ( $p = 1, \dots, \ell$ ). Hence, the first requirement is that our tests should be independent of this sign ambiguity. Notice that this ambiguity extends to ancillary constructs such as the normal vectors, defined as the vector product:

$$\vec{w}_{\ell,p,p'} \equiv \vec{v}_{\ell,p} \otimes \vec{v}_{\ell,p'}. \quad (76)$$

Mathematically, the sign ambiguity implies that the multipole vectors belong to the quotient space  $S^2/\mathbb{Z}_2$  (also known as  $\mathbb{R}P^2$ ) while the normal vectors belong to  $\mathbb{R}^3/\mathbb{Z}_2$ .

In principle, any positive-definite scalar constructed through the multipole or the normal vectors is “fair game”, but there are some guidelines; for example, one should avoid double-counting the same degrees of freedom. Below we review some of these tests, based on the work of [35].

**4.2.1. The R Statistic.** The first example of a test involving the multipole vectors would be a scalar product, such as  $\vec{v} \cdot \vec{v}'$ . The most natural test would involve asking whether the  $\ell$  multipole vectors of the given multipole  $\ell$  are especially aligned or not. This means computing

$$R_{\ell\ell} = \frac{2}{\ell(\ell-1)} \sum_{p,p'>p}^{\ell} \left| \vec{v}_{\ell,p} \cdot \vec{v}_{\ell,p'} \right|, \quad (77)$$

where the normalization was introduced to make  $0 \leq R_{\ell\ell} \leq 1$ .

This idea could be generalized to test alignments between multipole vectors at different multipoles:

$$R_{\ell\ell'} = \frac{1}{\ell\ell'} \sum_{p,p'} \left| \vec{v}_{\ell,p} \cdot \vec{v}_{\ell',p'} \right|. \quad (78)$$

In fact, the quadrupole-octupole alignment can already be seen with this simple test: for essentially all CMB maps the significance of the alignment as measured by  $R_{23}$  is of the order of 90%–95% CL [35].

**4.2.2. The S Statistic.** The second most natural test does not involve directly the multipole vectors themselves, but the normal vectors that can be produced by taking the vector product between the multipole vectors. So, we take

$$\vec{w}_{\ell,p,p'} = \vec{v}_{\ell,p} \otimes \vec{v}_{\ell',p'}. \quad (79)$$

Notice that the number of normal vectors for a given multipole  $\ell$  is  $l = \ell(\ell-1)/2$ —so the number of normal vectors grows rapidly for larger multipoles, making it harder to use and meaning that the same degrees of freedom may be overcounted, at least for  $\ell > 3$ .

Again, the best strategy is simplicity: we can ask whether the normal vectors are aligned, within and between multipoles. This means computing

$$S_{\ell\ell} = \frac{2}{l(l-1)} \sum_{p,p'>p}^l \left| \vec{w}_{\ell,p} \cdot \vec{w}_{\ell,p'} \right|, \quad (80)$$

where the normalization was introduced again to make  $0 \leq S_{\ell\ell} \leq 1$ .

And yet again this idea is easily generalized to test alignments between normal vectors at different multipoles:

$$S_{\ell\ell'} = \frac{1}{ll'} \sum_{p,p'} \left| \vec{w}_{\ell,p} \cdot \vec{w}_{\ell',p'} \right|. \quad (81)$$

With this test, the quadrupole-octupole alignment is much more significant: for essentially all CMB maps the test  $S_{23}$  deviates from the expected range with 98% CL [35].

**4.2.3. Other Tests with Multipole Vectors.** One can go on and expand the types of tests using both multipole and normal vectors. One idea would be, for example, to disregard the moduli of the normal vectors:

$$D_{\ell\ell'} = \frac{1}{ll'} \sum_{p,p'} \left| \hat{w}_{\ell,p} \cdot \hat{w}_{\ell',p'} \right|. \quad (82)$$

This test is therefore insensitive to the relative angle between the multipole vectors that produce any given normal vectors. This idea is similar to the planar modulations that will be discussed in the next section. With this test, the quadrupole-octupole alignment is significant to about 95%–98% CL. This test can also be generalized to a self-alignment test ( $\ell = \ell'$ ), with just an adjustment to the normalization.

Another possibility would be to measure the alignment between multipole vectors and normal vectors:

$$B_{\ell\ell'} = \frac{1}{\ell'} \sum_{p,p'} \left| \vec{v}_{\ell,p} \cdot \vec{w}_{\ell',p'} \right|. \quad (83)$$

Of course, this test cannot be easily generalized to a self-alignment. With the  $B_{23}$  test, the quadrupole-octupole alignment is significant to about 95%–98% CL

We could go on here, but it should be clear that all the information (the  $2\ell + 1$  real degrees of freedom for each multipole  $\ell$ ) has already been exhausted in the tests above.

## 5. Temperature Correlation Function

Despite their strong cosmological appeal, the multipole vectors have some limitations. Not only their directions in the CMB sky are sometimes difficult to interpret physically [26], but they also have the additional drawback of mixing, in a nontrivial manner, information on both gaussianity and SI of the map being analyzed [26, 75].

Another way of quantifying deviations in the standard statistical framework of cosmology is through functional modifications of the two-point correlation function [37, 68]. Although this approach does not offer an optimal separation between gaussianity and SI (which is, by the way, an open problem in this field), working with the two-point correlation function makes it easier to test Gaussian models of statistical anisotropy [28, 71].

The most general 2-point correlation function (2pcf) of two independent unit vectors is a function  $C$  of the form

$$C : S^2 \times S^2 \longrightarrow \mathbb{R}. \quad (84)$$

We have seen in Section 2.3 that if we choose spherical coordinates  $(\theta_i, \varphi_i)$  to describe each vector  $\hat{n}_i$ , the function above can be decomposed either in terms of two spherical harmonics  $Y_{\ell,m_i}$  or in terms of the bipolar spherical harmonic  $\mathcal{Y}_{\ell_i, \ell_j}^{LM}$ . In any case, the 2pcf will have the following functional dependence:

$$C = C(\theta_1, \varphi_1, \theta_2, \varphi_2). \quad (85)$$

This function is absolutely general. If the Universe has any cosmological deviation of isotropy, whatever it is, it can be described by the function above (see also Figure 5).

Unfortunately, this function will be of limited theoretical interest unless we have some hints on how to select its relevant degrees of freedom. This difficulty is in fact a general characteristic of model-independent tools, which at some stages forces us to rely on our theoretical prejudices about the statistical nature of the Universe in order to construct estimators of non-Gaussianity and/or statistical anisotropy. Nonetheless, it is still possible to construct statistical estimators of anisotropy based on (85). For example, Hajian and Souradeep [68] have constructed an unbiased estimator  $\kappa_\ell$  for this function based solely on the requirement that this estimator should be rotationally invariant. Although it is true that any statistically significant  $\kappa_\ell > 0$  will point towards

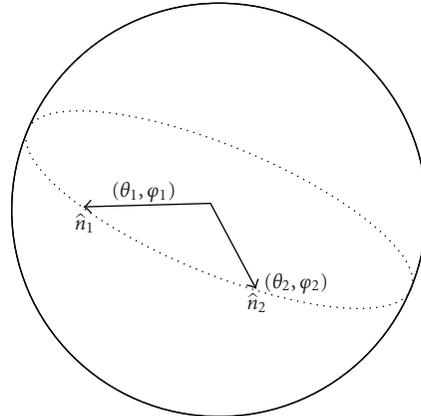


FIGURE 3: Geometrical representation of the 2pcf in terms of two unit vectors.

anisotropy, it is not clear what type of anisotropy is being detected by this estimator.

We can still use (85) to search for deviations of SI if we restrict its domain to a smaller and nontrivial subdomain. For example, we can take the vectors  $\hat{n}_1$  and  $\hat{n}_2$  to be the same and expand a function of the form

$$C = C(\theta_1, \varphi_1), \quad (86)$$

which is equivalent to  $C : S^2 \rightarrow \mathbb{R}$ . This form of the 2pcf makes it ideal for searching for power multipole moments in CMB, once a suitable estimator is defined [37]. Unfortunately, when we take  $\hat{n}_1 = \hat{n}_2$ , we are in fact considering a one-point correlation function, which by construction does not allow us to measure correlations between different points in the sky.

It seems in principle that the functions (86) and (85) are the only possibilities besides the isotropic 2pcf (24). If not, what other combinations of the vectors  $\hat{n}_1$  and  $\hat{n}_2$  can we consider? As a matter of fact, these two vectors are geometrical quantities intuitively bound to our notion of two-point correlation functions on the sphere. From this perspective they are not fundamental quantities. In fact, we can equally well represent the 2pcf by a disc living inside the unit sphere, as shown in Figure 4.

In this representation,  $\theta$  is the angle between the vectors  $\hat{n}_1$  and  $\hat{n}_2$ , as usual. The normal to the plane,  $\hat{n}$ , is represented by two spherical angles  $(\Theta, \Phi)$ . Finally, there is an overall orientation  $\varphi$  of the disc around its unit vector which completes the four degrees of freedom contained in (85). We have found therefore another valid geometrical representation of the most general 2pcf:

$$C = C(\Theta, \Phi, \theta, \varphi). \quad (87)$$

The main advantage of the above representation when compared to (85) is its straightforward geometrical interpretation. First, note that the angular separation  $\theta$  of the isotropic 2pcf is trivially included in this definition and does not need to be obtained as a consequence of rotational invariance (see the discussion in Section 3.2.1.) Second,

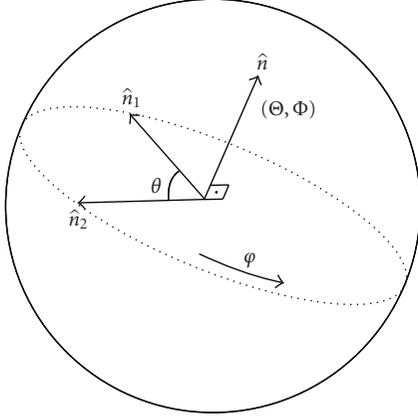


FIGURE 4: Geometrical representation of the 2pcf in terms of a unit disc (or plane).

by characterizing the correlation function in terms of the geometrical components of a disc, we know exactly what are the degrees of freedom involved. This makes it easier to construct estimators of statistical anisotropy, alleviating the drawbacks of model-independent approaches mentioned above.

**5.1. Anisotropy through Planarity.** An immediate application of the representation (87) is its use in the search for planar deviations of isotropy in CMB [28, 71]. Planar modulations of astrophysical origin may play an important role to the CMB morphology. One example is the role played by the galactic and ecliptic plane in the quadrupole-octupole/north-south anomalies [26]. Also, it is well known that our galactic plane is sensible source of foreground contamination in the construction of cleaned CMB maps. These hints indicate that CMB modulations induced by the disc in Figure 4 are not only a mathematical possibility but perhaps also a symmetry of cosmological relevance.

Since we are primarily interested in measuring planar modulations of CMB, but including the usual angular modulation as the isotropic limit of the 2pcf, we can consider only the azimuthal average of (87):

$$C \longrightarrow \frac{1}{2\pi} \int_0^{2\pi} C(\Theta, \Phi, \theta, \varphi) d\varphi. \quad (88)$$

The resulting function can be easily expand in terms of simple special functions as

$$C(\Theta, \Phi, \theta) = \sum_{\ell} \sum_{l,m} \frac{2\ell+1}{\sqrt{4\pi}} \mathcal{C}_{\ell}^{lm} P_{\ell}(\cos \theta) Y_{lm}(\Theta, \Phi), \quad l \in 2\mathbb{N}, \quad (89)$$

where the restriction on the  $l$ -mode results from the symmetry  $\hat{n}_1 \leftrightarrow \hat{n}_2$ . (See [71] for more details. Incidentally, the functional dependence of (89) differs from that originally shown in [71], which does not correctly take the planar dependence into account. We thank Yuri Shtanov for pointing this out.) The multipolar coefficients  $\mathcal{C}_{\ell}^{lm}$  correspond to

a generalization of the usual angular power spectrum  $C_{\ell}$ 's. In fact, they can be seen as the coefficients of a spherical harmonic decomposition of the function  $C_{\ell}(\hat{n})$ , provided that this function suffers modulations as we sweep planes on the sphere.

**5.1.1. Angular-Planar Power Spectrum.** Since we are restricting our analysis to the Gaussian framework, the set of coefficients  $\mathcal{C}_{\ell}^{lm}$  is all we need to characterize the two-point correlation function. However, the final product of CMB observations is temperature maps, and not correlation maps. What we need then is an algebraic relation between the multipolar coefficients  $\mathcal{C}_{\ell}^{lm}$  and the temperature coefficients  $a_{\ell m}$  defined in (3). At first sight, this relation could be obtained by equating expression (89) to its standard definition in (22) and then using the orthogonality of the special functions to isolate the  $\mathcal{C}_{\ell}^{lm}$ 's in terms of the  $a_{\ell m}$ 's. But for that to work we need the relation between the set of angles  $(\Theta, \Phi, \theta)$  and  $(\theta_1, \varphi_1, \theta_2, \varphi_2)$  which, depending on the reference frame we choose, is extremely complicated. Fortunately, all we need is the following relation:

$$\hat{n}_1 \cdot \hat{n}_2 = \cos \theta = \cos \theta_1 \cos \theta_2 + \sin \theta_1 \sin \theta_2 \cos(\varphi_1 - \varphi_2), \quad (90)$$

together with a suitable choice of our coordinate system. For example, we can use the invariance of the scalar product  $\hat{n}_1 \cdot \hat{n}_2$  and choose our coordinate system such that the discs of Figure 4 lies in the  $xy$  plane. With this choice we will have

$$(\Theta, \Phi) = (0, 0), \quad \cos \theta = \cos(\varphi_1 - \varphi_2), \quad (91)$$

and the integration over  $\theta$  becomes simple. Once this is done, we make a passive rotation of the coordinate system and then we integrate over the remaining angles  $\Theta$  and  $\Phi$ , which will then be given precisely by the Euler angles used in the rotation. The details are rather technical and can be found in the Appendix. The final expression is

$$\frac{(-1)^m \mathcal{C}_{\ell}^{lm}}{\sqrt{2l+1}} = 2\pi \sum_{\ell_1, m_1} \sum_{\ell_2, m_2} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle \begin{pmatrix} l & \ell_1 & \ell_2 \\ -m & m_1 & m_2 \end{pmatrix} I_{\ell_1 \ell_2}^{l \ell}, \quad (92)$$

where

$$I_{\ell_1 \ell_2}^{l \ell} \equiv \sum_m (-1)^m \lambda_{\ell_1 m} \lambda_{\ell_2 m} \begin{pmatrix} l & \ell_1 & \ell_2 \\ 0 & m & -m \end{pmatrix} \int_0^{\pi} d(-\cos \theta) P_{\ell}(\cos \theta) e^{im\theta}, \quad (93)$$

and where the  $\lambda_{\ell, m}$ 's form a set of coefficients resulting from the  $\theta$  integration, which are zero unless  $\ell_i + m = \text{even}$  (see the Appendix for more details).

Expression (92) is what we were looking for. With this relation, the angular-planar power spectrum  $\mathcal{C}_{\ell}^{lm}$  can be calculated from first principles for any model predicting a specific covariance matrix. Moreover, since the angular-planar function (89) is, after all, a correlation function,

it should be possible to relate the angular-planar power spectrum  $\mathcal{C}_\ell^{lm}$  to the bipolar power spectrum  $\mathcal{A}_{\ell_1\ell_2}^{LM}$  of Hajian and Souradeep. In fact, by inverting expression (30) and plugging the result in (92), we find a linear relation between these two set of coefficients:

$$\mathcal{C}_\ell^{lm} = 2\pi \sum_{\ell_1, \ell_2} A_{\ell_1\ell_2}^{lm} (-1)^{\ell_1 - \ell_2} (2l + 1) I_{\ell_1\ell_2}^{l,\ell}. \quad (94)$$

Here, the set of geometrical coefficients  $I_{\ell_1\ell_2}^{l,\ell}$  plays a similar role to the 3-j symbols in expression (30). Note also that the angular-planar power spectrum has only three free indices while the bipolar power spectrum has four. This is a consequence of the azimuthal average we took in (88), which further constrains the degrees of freedom of the correlation function.

*5.1.2. Statistical Estimators and  $\chi^2$  Analysis.* We have shown that the angular-planar power spectrum (92) is given in terms of an ensemble average of temperature maps. Evidently, we cannot calculate it directly from data, for we have only one CMB map (the one taken from our own Universe.) The best we can do is to estimate the statistical properties of (92), like its mean and variance, and see whether these quantities agree, in the statistical sense, with what we would expect to obtain from a particular model of the Universe.

The reader should note that this procedure is not new—its limitation is due to the same cosmic variance which lead us to construct an estimator for the angular power spectrum  $C_\ell$  (see the discussion of Section 2.4). For the same reason, we will need to construct an *estimator* for the angular-planar power spectrum. An obvious choice is

$$\hat{\mathcal{C}}_\ell^{lm} \equiv 2\pi\sqrt{2l+1} \sum_{\ell_1, m_1, \ell_2, m_2} a_{\ell_1 m_1} a_{\ell_2 m_2} \begin{pmatrix} l & \ell_1 & \ell_2 \\ m & m_1 & m_2 \end{pmatrix} I_{\ell_1\ell_2}^{l,\ell}, \quad (95)$$

for in this case we have an unbiased estimator:

$$\langle \hat{\mathcal{C}}_\ell^{lm} \rangle = \mathcal{C}_\ell^{lm}, \quad (96)$$

regardless of the underlying model.

If we now have a model predicting the angular-planar power spectrum, we can ask how good this model fits the observational data once it is calculated using (95). All we need is a simple chi-square ( $\chi^2$ ) goodness-of-fit test, which in our case can be written in the following generalized form:

$$(\chi_v^2)_\ell^l \equiv \frac{1}{2l+1} \sum_{m=-l}^l \frac{|\hat{\mathcal{C}}_\ell^{lm} - \mathcal{C}_\ell^{lm}|^2}{(\sigma_\ell^{lm})^2}, \quad (97)$$

in which  $\ell$  and  $l$  are the angular and planar degrees of freedom, respectively, and where  $\sigma_\ell^{lm}$  is just the standard deviation of the estimator  $\hat{\mathcal{C}}_\ell^{lm}$ . The  $(2l+1)^{-1}$  factor accounts for the  $2l+1$  planar degrees of freedom and was introduced for latter convenience.

In Section 7 we will apply this test to the 5-year WMAP temperature maps in order to check the robustness of the  $\Lambda$ CDM model against the hypothesis of SI. Before that, we will stop and digress a little about how observational uncertainties should be included in our analysis.

## 6. Theory v. Observations: Cosmic and Statistical Variances

Until now we have been concerned with the formal aspects of non-Gaussian and statistically anisotropic universes, and how model-independent tests might be designed to detect deviations of either gaussianity or statistical isotropy. We will now discuss how such tests can be carried out and interpreted once we possess cosmological data.

In a great variety of tests, statistical tools are designed to detect particular deviations of gaussianity or SI from cosmological data like, for example, the CMB. The final outcome of these tests is usually a probability (a pure number), which should be interpreted as the chance that a Universe like ours might result from an ensemble of “equally prepared” Gaussian and SI universes. An anomaly in CMB is usually understood as a measure of how unlikely a particular feature of our Universe is according to this specific test. The multipole vector statistics, for example, when applied to a large number of simulated (Gaussian and SI) CMB maps, show that only  $\sim 0.01\%$  of these maps have a quadrupole-octupole alignment as strong as WMAP maps [23, 26].

There are two points to keep in mind when carrying this type of analysis. The first is that a particular “detection” may always turn out to be a statistical fluctuation revealed by one specific tool. The robustness of an anomaly then depends on the number of independent tests pointing to the same result. The second is that the implementation of statistical tools is sensitive to the way we extract information from data, requiring an accurate separation between cosmological signal and astrophysical/instrumental noise.

In this Section we present a critical review of the standard procedure used to implement cosmological tests. We show that it does not account for the intrinsic uncertainties of cosmological observations, which may possibly lead to an under/overestimation of anomalies. We then present a generalization of this process which naturally accounts for these uncertainties.

*6.1. Standard Calculations.* Suppose  $x$  is a random variable predicted by a particular model and that cosmological observations of this quantity returned the value  $x_0$ . (Rigorously,  $x$  is only *one realization* of a random variable  $X$ , which is a real-valued function defined on a sample space. By the same reason we should not call the  $a_{\ell m}$ ’s in (3) random variables, though we will stick to this nomenclature throughout this text.) We would like to calculate the probability, *according to this model*, that in a random Universe we would have  $x \leq x_0$ . Assuming that  $\mathcal{P}_{\text{th}}$  is the (normalized) probability density function (pdf) of  $x$ , this probability is commonly defined to be

$$P_{\leq} \equiv \int_{-\infty}^{x_0} \mathcal{P}_{\text{th}}(x) dx. \quad (98)$$

The probability of having  $x > x_0$  is then simply given by  $P_{>} = 1 - P_{\leq}$ . See Figure 5.

If  $P_{\leq}$  is found to be too small or, equivalently, too high, we might be tempted to interpret  $x_0$  as “anomalous” according to this model. However, this definition of probability

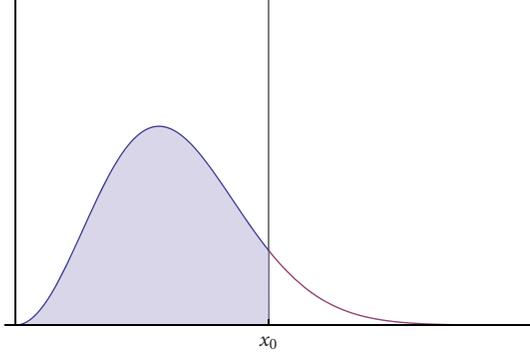


FIGURE 5: Probability density function for the theoretical variable  $x$ . The shaded area is the probability (98).

assumes that  $x_0$  was measured with infinite precision, and so it says nothing about an important question we must deal with: typically, the measurement of  $x_0$  has an uncertainty which needs to be folded into the final probabilities that the observations match the theoretical expectations.

Moreover, since no two equally prepared experiments will ever return the same value  $x_0$ , our measurements should also be regarded as random events. In a more rigorous approach, we would have to consider  $x_0$  itself as one realization of a random variable, conditioned to the distribution of the signal. In the case of CMB, however, this would be only part of the whole picture, since the randomness of the measurements of  $x_0$  should also be related to the way this data is reduced to its final form. This happens because different map cleaning procedures will lead to slightly different values for  $x_0$ . This difference induces a variance in the data which reflects the remaining foreground contamination of the temperature map. We will elaborate more on this point through a concrete example, after we show how (98) may be changed in order to include the indeterminacy of cosmological measurements.

**6.2. Convolution Probabilities.** The question we want to answer is how to calculate the probability of  $x$  being smaller than our measurements when the latter are also random events. Let us suppose that our measurement is described by the random variable  $y$  and that  $x_0$  is its most probable value. The probability of having  $x \leq y$  is simply the probability of having

$$z \equiv x - y \leq 0 \quad (99)$$

for some particular realization of the variable  $y$ . It should be clear by now that if we know the pdf of  $z$ , the probability we are looking for is simply the area under this distribution for  $-\infty \leq z \leq 0$ . The probability  $P_{\leq}$  of  $z$  being smaller or equal to zero can be calculated as

$$P_{\leq} \equiv P\{(x, y) \mid x - y \leq 0\} = \iint_{x-y \leq 0} \mathcal{P}(x, y) dx dy. \quad (100)$$

Now, under the hypothesis of independence of  $x$  and  $y$  we have  $\mathcal{P}(x, y) = \mathcal{P}_{\text{th}}(x)\mathcal{P}_{\text{obs}}(y)$ , where  $\mathcal{P}_{\text{obs}}$  is the (normalized) probability distribution function of the variable  $y$ . We can therefore rewrite the last expression as

$$P_{\leq} = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^y \mathcal{P}_{\text{th}}(x) dx \right] \mathcal{P}_{\text{obs}}(y) dy. \quad (101)$$

If we now hold  $y$  fixed and do the change of variables  $x = u + y$ , we get

$$P_{\leq} = \int_{-\infty}^0 \left[ \int_{-\infty}^{\infty} \mathcal{P}_{\text{th}}(u + y) \mathcal{P}_{\text{obs}}(y) dy \right] du, \quad (102)$$

where we have changed the position of the integrals. The reader will now notice that term inside brackets is precisely the pdf we were looking for. Since there is nothing special about the variable  $y$ , we can equally well hold  $x$  fixed and repeat the calculus, obtaining the symmetric version of this result. In fact, the final pdf for  $z$  is nothing else than the convolution of the pdfs of each variable [76]:

$$\begin{aligned} \mathcal{P}(z) &\equiv (\mathcal{P}_{\text{obs}} * \mathcal{P}_{\text{th}})(z) = \int_{-\infty}^{\infty} \mathcal{P}_{\text{obs}}(y) \mathcal{P}_{\text{th}}(z \pm y) dy \\ &= \int_{-\infty}^{\infty} \mathcal{P}_{\text{obs}}(x \mp z) \mathcal{P}_{\text{th}}(x) dx, \end{aligned} \quad (103)$$

where the plus (minus) sign refers to the difference (sum) of  $x$  and  $y$ . Integrating this pdf from  $(-\infty, 0]$  we get our answer

$$P_{\leq} \equiv \int_{-\infty}^0 \mathcal{P}(z) dz. \quad (104)$$

As a consistency check, notice that in the limit where observations are made with infinite precision,  $\mathcal{P}_{\text{obs}}(y)$  becomes a delta function and we have

$$P_{\leq} = \int_{-\infty}^0 \int_{-\infty}^{+\infty} \delta(y - x_0) \mathcal{P}_{\text{th}}(z \pm y) dy dz = \int_{-\infty}^{\pm x_0} \mathcal{P}_{\text{th}}(x) dx \quad (105)$$

which agrees with our previous definition. The reader must be careful, though, not to think of (98) as some lower bound to (104). Since none of the pdfs appearing in (103) are necessarily symmetric, a large distance from  $x_0$  to the most probable (not the mean!) value  $x$  would not, by itself, constitute sufficient grounds to claim that the measured value of this observable is “unusual” in any sense, simply because a large overlap between the two pdfs can render the result usual according to (104).

To illustrate this point, let us calculate  $P_{\leq}$  using (98) and (104) for the pdf’s which appear in Figure 6. For pedagogical reasons, we have chosen  $\mathcal{P}_{\text{th}}$  and  $\mathcal{P}_{\text{obs}}$  as positively (Maxwell distribution) and negatively (Gumbel distribution) skewed pdf’s, respectively. The convolved distribution appears as the solid (black) line. For these pdf’s, (98) gives 99.2%, while (104) gives 93.6% of chance of  $x$  being smaller than the observed  $x_0$ ; all pdf’s were normalized to one.

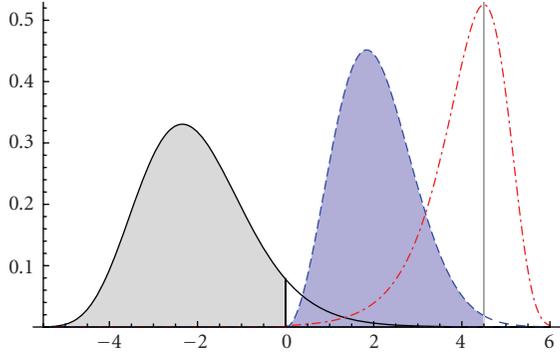


FIGURE 6: Probability density functions for  $x$  (blue, dashed line),  $y$  (red, dot-dashed line), and  $z$  (solid line). The shaded area gives the probability of  $x$  being smaller than  $x_0$  (dashed vertical line). See the text for more details.

## 7. A $\chi^2$ Test of Statistical Isotropy

Although the last example was constructed to emphasize an important feature of the formalism developed in Section 6, cosmological observables designed to measure deviations of either gaussianity or statistical isotropy will often follow asymmetric distributions. The intrinsic uncertainties of cosmological measurements, specially the ones originating from map cleaning procedures, may be crucial when searching for any map's anomalies. We will now make a concrete application of this formalism using the angular-planar chi-square test developed in Section 5.1.

**7.1.  $\Lambda$ CDM Model.** In the simplest realization of the  $\Lambda$ CDM model, the covariance matrix of temperature maps is determined by  $\langle a_{\ell_1 m_1} a_{\ell_2 m_2} \rangle = (-1)^{m_2} C_{\ell_1} \delta_{\ell_1 \ell_2} \delta_{m_1, -m_2}$ . Using this expression in (92), we find that

$$\mathcal{C}_\ell^{lm} \stackrel{(SI)}{=} C_\ell \delta_{l0} \delta_{m0}. \quad (106)$$

On the other hand, if the only nonzero  $\mathcal{C}_\ell^{lm}$ 's are given by  $l = m = 0$ , then  $\mathcal{C}_\ell^{00} = C_\ell$ . Therefore, statistical isotropy is achieved if and only if the angular-planar power spectrum is of the form (106). Since we are only interested in nontrivial planar modulations, we will restrict our analysis to the cases where  $l \neq 0$ , that is,

$$\mathcal{C}_\ell^{lm} = 0, \quad (l \geq 2), \quad (107)$$

where, we remind the reader, the parity of  $\ell$  comes from the symmetry (27).

For this particular model, we can also calculate the covariance matrix of the estimator (95) explicitly. Using the null hypothesis above, we find after some algebra that

$$\langle (\hat{\mathcal{C}}_\ell^{lm})^* \hat{\mathcal{C}}_\ell^{l'm'} \rangle = 8\pi^2 \sum_{\ell_1, \ell_2} C_{\ell_1} C_{\ell_2} (I_{\ell_1 \ell_2}^{l \ell})^2 \delta_{l'l'} \delta_{mm'}. \quad (108)$$

This matrix has some interesting properties. First, note that the planar degrees of freedom are independent in this case (which justifies the  $(2l+1)^{-1}$  factor introduced in (97)).

Second, its diagonal elements are given by the variance  $(\sigma_\ell^{lm})^2 = \langle (\hat{\mathcal{C}}_\ell^{lm})^* \hat{\mathcal{C}}_\ell^{lm} \rangle$ , which now becomes  $m$ -independent:

$$(\sigma_\ell^{lm})^2 \rightarrow (\sigma_\ell^l)^2 = 8\pi^2 \sum_{\ell_1, \ell_2} C_{\ell_1} C_{\ell_2} (I_{\ell_1 \ell_2}^{l \ell})^2. \quad (109)$$

Therefore, for the particular case of the  $\Lambda$ CDM model, the chi-square test (97) gets even simpler. Using (107) and (109), we find

$$(\chi_v^2)_\ell^l = \frac{1}{2l+1} \sum_{m=-l}^l \frac{|\hat{\mathcal{C}}_\ell^{lm}|^2}{(\sigma_\ell^l)^2}. \quad (110)$$

It is now clear that if the data under analysis are really described by this model, then it must be true that

$$\langle (\chi_v^2)_\ell^l \rangle = \frac{1}{2l+1} \sum_{m=-l}^l \frac{\langle (\hat{\mathcal{C}}_\ell^{lm})^* \hat{\mathcal{C}}_\ell^{lm} \rangle}{(\sigma_\ell^l)^2} = 1, \quad (111)$$

where we have used (109). This shows that any large deviation of this test from unity will be an indication of planar modulation in temperature maps, up, of course, to error bars. For convenience, let us define a new quantity as

$$\bar{\chi}_\ell^l \equiv (\chi_v^2)_\ell^l - 1 \quad (112)$$

which will quantify anisotropies whenever  $\bar{\chi}_\ell^l$  is significantly positive or negative.

This generalized chi-square test furnishes a complete prescription when searching for planar modulations of temperature in CMB maps. We emphasize, though, that for a given CMB map, the chi-square analysis must be done entirely in terms of that map's data. Since we are performing a model-independent test, we are not allowed to introduce fiducial biases in the analysis (e.g., by calculating  $\sigma_\ell^l$  using  $C_\ell^{\text{model}}$ ), which would only include our *a priori* prejudices about what the map's anisotropies should look like. Since the  $C_\ell$ 's are, by construction, a measure of statistical isotropy, consequently, an "anomalous" detection of  $C_\ell$ 's is by no means a measure of statistical anisotropy, and it is this particular value that should be used in (112) if we want to find deviations of isotropy, regardless of how high/low it is.

**7.2. Searching for Planar Signatures in WMAP.** In order to apply the test (112) to the 5-year WMAP data [11, 14], we will define two new variables:

$$x \equiv (\bar{\chi}^2)_{\ell(\text{th})}^l, \quad y \equiv (\bar{\chi}^2)_{\ell(\text{obs})}^l, \quad (113)$$

which will be jointly analyzed using the formalism of Section 6. Still, there remains the question of how to obtain their pdfs. These functions can be obtained numerically provided that the number of realizations of each variable is large enough, since in this case their histograms can be considered as piecewise constant functions which approximate the real pdfs. For the case of the (theoretical) variable  $x$  defined above

TABLE 1: Full-sky foreground-cleaned CMB maps from WMAP data used in our analysis to estimate the variable  $y$  (see the text for more details). Note that the reference [78] includes the analysis of maps from the three- and five-year WMAP releases.

Full sky maps	References
Hinshaw et al.	[14, 79]
de Oliveira-Costa et al.	[80]
Kim et al.	[78]
Park et al.	[81]
Delabrouille et al.	[82]

we have run  $2 \times 10^4$  Monte Carlo simulations of Gaussian and statistically isotropic CMB maps using the  $\Lambda$ CDM best-fit  $C_\ell$ 's provided by the WMAP team [77]. With these maps we have then constructed  $2 \times 10^4$  realizations of the variable  $x$ .

The simulation of the (observational) variable  $y$  is more difficult and depends on the way we estimate contamination from residual foregrounds in CMB maps. As is well known, not only instrumental noise, but also systematic errors (e.g., in the map-making process), the inhomogeneous scanning of the sky (i.e., the exposure function of the probe), or unremoved foreground emissions (even after applying a cut-sky mask) could corrupt—at distinct levels—the CMB data.

Foreground contamination, on the other hand, may have several different sources, many of which are far beyond our present scopes. However, since different teams apply distinct procedures on the raw data in order to produce a final map, we will make the hypothesis that maps cleaned by different teams represent—to a good extent—“independent” CMB maps. Therefore, we can estimate the residual foreground contaminations by comparing these different foreground-cleaned maps.

In fact, the WMAP science team has made substantial efforts to improve the data products by minimizing the contaminating effects caused by diffuse galactic foregrounds, astrophysical point sources, artifacts from the instruments and measurement process, and systematic errors [83, 84]. As a result, multifrequency foreground-cleaned full-sky CMB maps were produced, named Internal Linear Combination maps, corresponding to three- and five-year WMAP data [14, 79]. To account for the mentioned randomness, systematic, and contaminating effects of the CMB data, we will use in our analyses several full-sky foreground-cleaned CMB maps, listed in Table 1, which were produced using both the three- and five-year WMAP data.

The prescription we adopt to determine the distribution of the observational variable  $y$  is as follows: we simulate Gaussian random  $a_{\ell m}$ 's in such a way that *their central values are given by the five-year ILC5 data* [14, 79], and with a variance which is estimated from the *sample standard deviation* of all the maps listed in Table 1. So, for example, suppose we have  $n$  different full-sky temperature maps at hand and we want to estimate the randomness inherent in the determination of, let's say,  $a_{32}$ ; therefore, we take

$$\mathcal{N}(a_{32}^{\text{ILC5}}, \sigma_{32}) \rightarrow a_{32}, \quad (114)$$

with

$$\sigma_{32} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (a_{32}^i - \bar{a}_{32})^2}, \quad \bar{a}_{32} = \frac{1}{n} \sum_{i=1}^n a_{32}^i, \quad (115)$$

where  $\mathcal{N}(\mu, \sigma)$  represents a Gaussian distribution with mean  $\mu$  and standard deviation  $\sigma$ . Note that if the residual contamination is indeed weak, then the sample variance above will be small, and our procedure will reduce to the standard way of calculating probabilities. As for the use of a Gaussian in (114), this choice was dictated not only by simplicity, but rather by the fact that the propagation of uncertainties in physical experiments is usually assumed to follow a normal distribution. However; note that there are some instrumental uncertainties, such as beam or gain uncertainties, which will not in general follow normal distributions. In fact, some of them may even fail to be additive, meaning that our convolution formula will be inapplicable in these cases. In our analysis, we have focused only on foreground residuals, where the normality hypothesis is reasonable. (We thank the referee for pointing out this important aspect of the formalism.)

**7.3. Full-Sky Maps.** Following this procedure, we have used the full-sky maps shown in Table 1 to construct  $10^4$  Gaussian random  $a_{\ell m}$ 's, which were then used to calculate  $10^4$  realizations of  $y = (\bar{\chi}^2)_{\ell(\text{obs})}^l$ . With those variables we constructed histograms which, together with the histograms for the (full-sky) variable  $x$ , were used to calculate the final probability (104). We have restricted our analysis to the range of values  $(\ell, l) \in [2, 10]$ , since the low multipolar sector (i.e., large angular scales) is where most of the anomalies were reported. The resulting histograms and pdf's are shown in Figure 7, and the final probabilities we obtained are shown in Table 2.

Overall, our results show no significant planar deviations of anisotropy in WMAP data. The most unlikely individual values in Table 2 are in the sectors  $(l, \ell)$  given by (2, 5), (10, 5), (4, 7), and (6, 8) and are all above a relative chance of 5% of either being too negative [(2, 5), (10, 5)] or too positive [(4, 7), (6, 8)]. However, it is perhaps worth mentioning that not only the individual values of  $(\bar{\chi}^2)_\ell^l$  are relevant, but also their coherence over a range of angular or planar momenta carries interesting information. So, for example, a set of  $(\bar{\chi}^2)_\ell^l$ 's which are all individually within the cosmic variance bounds, but which are all positive (or negative), can be an indication of an excess (or lack) of planar modulation. This type of coherent behavior appears in the following cases:  $(\bar{\chi}^2)_2^l$ ,  $(\bar{\chi}^2)_3^l$  and, to a lesser extent,  $(\bar{\chi}^2)_\ell^4$  (see Table 2). The angular quadrupole  $\ell = 2$ , as well as the angular octupole  $\ell = 3$ , have all positive planar spectra (for all values of  $l$  which we were able to compute), indicated by probabilities larger than 50%. The planar hexadecupole  $l = 4$  also has 8 out of 9 angular spectra assuming positive values (only  $\ell = 5$  is negative).

The data analyzed in this Section relates to the full-sky maps, which are certainly still affected by residual galactic foregrounds. The reader interested in the complete analysis,

TABLE 2: Final probabilities of obtaining, in a random  $\Lambda$ CDM Universe, a chi-square value smaller or equal to  $(\bar{\chi}^2)_{\ell(\text{obs})}^l$ , as given by full-sky temperature maps.

$l \setminus \ell$	2	3	4	5	6	7	8	9	10
2	81.1%	73.7%	54.1%	<b>6.1%</b>	80.7%	46.4%	36.9%	47.5%	81.8%
4	74.0%	72.6%	55.0%	39.6%	74.2%	<b>93.2%</b>	51.6%	55.4%	56.3%
6	78.1%	80.7%	69.3%	52.3%	33.6%	80.0%	<b>95.0%</b>	50.3%	82.2%
8	63.5%	87.7%	18.8%	51.5%	21.4%	66.4%	31.6%	27.5%	82.3%
10	67.9%	50.0%	61.0%	<b>8.7%</b>	37.7%	59.5%	36.6%	29.2%	35.7%

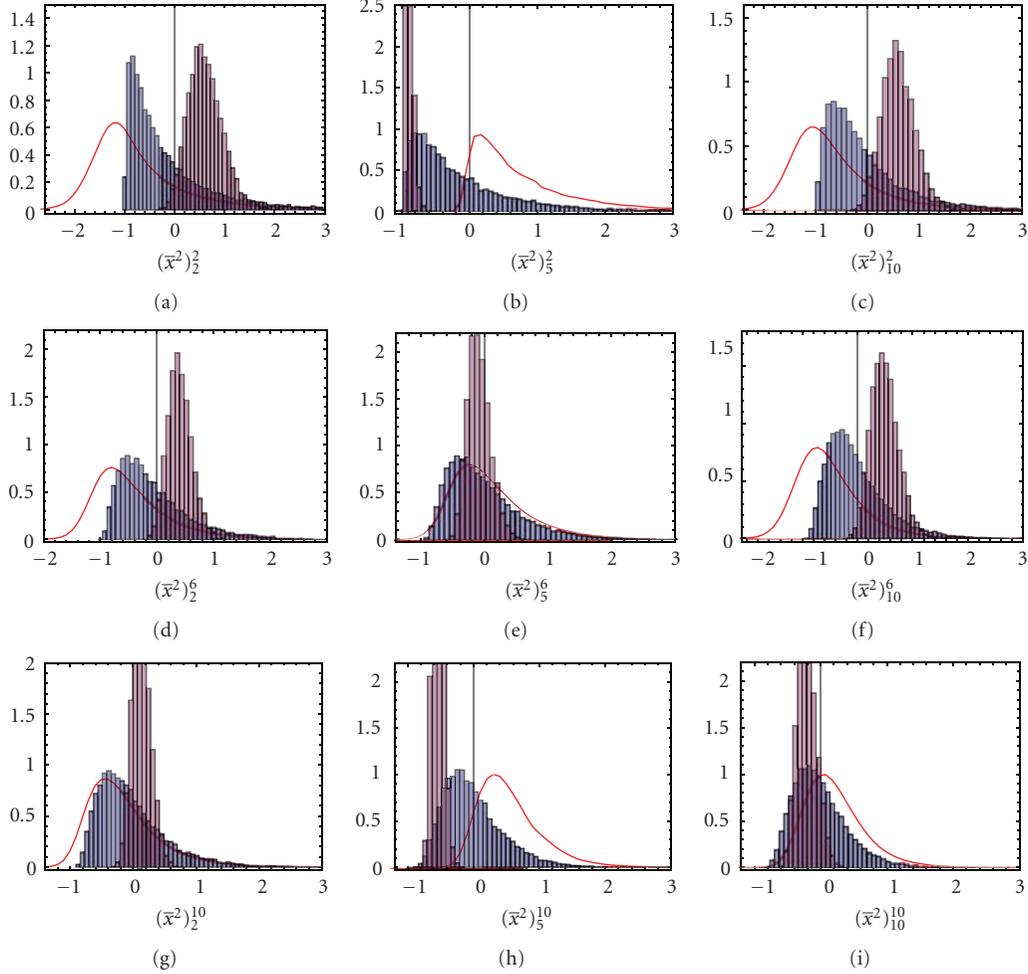


FIGURE 7: Histograms for  $(\bar{\chi}^2)_{\ell(\text{th})}^l$  (blue),  $(\bar{\chi}^2)_{\ell(\text{obs})}^l$  (purple) and for the difference  $(\bar{\chi}^2)_{\ell(\text{th})}^l - (\bar{\chi}^2)_{\ell(\text{obs})}^l$  (solid, red line). We show only a few representative figures. The final probabilities are shown in Table 2 and correspond to the area under the solid curve from  $-\infty$  to 0. All pdfs are normalized to 1.

including data from masked CMB maps, can check reference [28].

## 8. Conclusions

We know our Universe is not perfectly Gaussian, homogeneous, or isotropic. The deviations from an idealized picture (or the lack thereof), whether predictably small or surprisingly large, can tell us a great deal about the Universe we live in. Since the types of physical mechanisms

behind deviations from perfect gaussianity, homogeneity, or isotropy are typically very different, we should try to measure these individual features separately—whenever possible or practical.

Recently, it has been suggested that some of the most discussed anomalies in the CMB can be explained away [85], or that the evidence for them is statistically weak [86]. But even if it turns out that our Universe is a plain vanilla kind of place, where everything goes according to the inflationary theorist's dreams, we would still need to analyze it with tools

that allow us to check the standard picture against the data. In addition, local physics (related to the solar system, or our galaxy), as well as instrumental quirks, tend to leave imprints on the CMB which are clearly anisotropic, but have a certain coherence which can be detected, and possibly corrected for, with the help of these checks.

However, in an era where at least the large-scale maps of the CMB are likely to remain basically unchanged, we should be careful not to over analyze the data with the benefit of an ever greater hindsight (put another way, *a posteriori* conundrums only get worse with time.) This can only be achieved if we find natural and generally agreed-upon classifications of the types of deviations that may occur, without too much guidance from what the data is telling us. We believe that focusing on the possible underlying symmetries, with perhaps some guidance from group-theoretic arguments, is one way to settle these issues. We have presented a few methods along these lines, one using multipole vectors, the other using a natural generalization of the two-point correlation function (also other methods have been presented in this paper).

Perhaps the best indication that we are on the right track is the fact that most of these methods are applicable in other areas of physics and astronomy, and that in some cases we have adapted tests of anisotropy from other areas, such as scattering theory and the theory of angular momentum in quantum mechanics. So, even if these anomalies eventually perish, they will be survived by the powerful methods that have been devised to test them.

## Appendix

### A. Geometrical Identities and Derivations

*A.1. Wigner D-Functions.* From the unitarity of the rotation operators  $D(R)$ , we have

$$\sum_m D_{m'm}^\ell(\omega_1) D_{m''m}^\ell(\omega_2) = D_{m''m'}^\ell(\omega_1\omega_2). \quad (\text{A.1})$$

If  $\omega_2 = \omega_1^{-1}$ , then using the identity  $D_{m''m}^\ell(\omega^{-1}) = D_{m'm''}^{*\ell}(\omega)$ , we find

$$\sum_m D_{m'm}^\ell(\omega_1) D_{m''m}^{*\ell}(\omega_1) = \delta_{m''m'}. \quad (\text{A.2})$$

*A.2. Gaunt Integral.* The definition of the Gaunt integral used in this paper is

$$\mathcal{G}_{m_1 m_2 m_3}^{\ell_1 \ell_2 \ell_3} = (-1)^{m_1} \int d^2\hat{n} Y_{\ell_1, -m_1}(\hat{n}) Y_{\ell_2, m_2}(\hat{n}) Y_{\ell_2, m_2}(\hat{n}). \quad (\text{A.3})$$

*A.3. 3-j Symbols.* We present here some useful identities related to the 3-j symbols

(i) Isotropic limit

$$\begin{pmatrix} l_1 & l_2 & 0 \\ m_1 & m_2 & 0 \end{pmatrix} = \frac{(-1)^{l_2 - m_1}}{\sqrt{2l_1 + 1}} \delta_{l_1 l_2} \delta_{m_1, -m_2}. \quad (\text{A.4})$$

(ii) Parity and permutations

$$\begin{aligned} \begin{pmatrix} l_1 & l_2 & l \\ m_1 & m_2 & m \end{pmatrix} &= \begin{pmatrix} l & l_1 & l_2 \\ m & m_1 & m_2 \end{pmatrix} \\ &= (-1)^{l_1 + l_2 + l} \begin{pmatrix} l_2 & l_1 & l \\ m_2 & m_1 & m \end{pmatrix} \\ &= (-1)^{l_1 + l_2 + l} \begin{pmatrix} l_1 & l_2 & l \\ -m_1 & -m_2 & -m \end{pmatrix}. \end{aligned} \quad (\text{A.5})$$

(iii) Orthogonality

$$\begin{aligned} \sum_{m_1 = -l_1}^{l_1} \sum_{m_2 = -l_2}^{l_2} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l'_3 \\ m_1 & m_2 & m'_3 \end{pmatrix} \\ &= \frac{\delta_{l_3 l'_3} \delta_{m_3 m'_3}}{2l_3 + 1}, \\ \sum_{l_1 = |l_2 - l_3|}^{l_2 + l_3} \sum_{m_1 = -l_1}^{l_1} (2l + 1) \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_4 \\ m_1 & m'_2 & m'_3 \end{pmatrix} \\ &= \delta_{m_2 m'_2} \delta_{m_3 m'_3}, \\ \sum_{m = -l}^l (-1)^{l-m} \begin{pmatrix} l & l & \ell \\ m & -m & 0 \end{pmatrix} \\ &= \sqrt{2l + 1} \delta_{\ell, 0}. \end{aligned} \quad (\text{A.6})$$

The last expression is particularly useful in the derivation of (106).

*A.4. Derivation of (92).* We start by equating expressions (89) and (22)

$$\begin{aligned} \sum_{\ell} \sum_{l, m} \frac{2\ell + 1}{\sqrt{4\pi}} \mathcal{C}_\ell^{lm} P_\ell(\cos \vartheta) Y_{lm}(\hat{n}) \\ = \sum_{\ell_1, m_1} \sum_{\ell_2, m_2} \langle a_{\ell_1, m_1} a_{\ell_2, m_2}^* \rangle Y_{\ell_1, m_1}(\hat{n}_1) Y_{\ell_2, m_2}(\hat{n}_2). \end{aligned} \quad (\text{A.7})$$

As mentioned in the main text, the inversion of  $\mathcal{C}_\ell^{lm}$  as a function of the  $a_{\ell m}$ 's is not a trivial task, since the angles  $(\Theta, \Phi, \theta)$  depend nonlinearly on the angles  $(\theta_1, \varphi_1, \theta_2, \varphi_2)$ . The easiest way to achieve this goal is to pick up a coordinate system where only the  $\theta$  dependence is present. After integrating it out, we rotate our coordinate system using three Euler angles to recover back the  $(\Theta, \Phi)$  dependence, which can then be integrated with the help of some Wigner matrices identities. We start by positioning the vectors  $\hat{n}_1$  and  $\hat{n}_2$  in the  $xy$  plane, that is; we chose  $\hat{n}_1 = (\pi/2, \phi_1)$ ,

$\hat{n}_2 = (\pi/2, \phi_2)$ . By (90), we then have  $\cos \vartheta = \cos(\phi_1 - \phi_2)$ . Using the relation [41]

$$Y_{\ell m}\left(\frac{\pi}{2}, \phi\right) = \lambda_{\ell m} e^{im\phi},$$

$$\lambda_{\ell m} = \begin{cases} (-1)^{(\ell+m)/2} \sqrt{\frac{2\ell+1}{4\pi} \frac{\mathfrak{A}!!}{(\ell+m)!!} \frac{\mathfrak{B}!!}{(\ell-m)!!}}, & \text{if } \ell+m \in 2\mathbb{N}, \\ 0, & \text{otherwise} \end{cases} \quad (\text{A.8})$$

where  $\mathfrak{A}$  denotes  $(\ell+m-1)!!$  and  $\mathfrak{B}$  denotes  $(\ell-m-1)!!$ , we can integrate the  $\theta$  dependence on both sides of (A.7). This gives us

$$\frac{1}{\sqrt{\pi}} \sum_{l,m} \mathcal{C}_\ell^{lm} Y_{lm}(0,0) = \sum_{\ell_1, m_1} \sum_{\ell_2, m_2} \langle a_{\ell_1 m_1} a_{\ell_2 m_2}^* \rangle I_{\ell_1 m_1 \ell_2 m_2}^\ell, \quad (\text{A.9})$$

where we have introduced the following definition:

$$\begin{aligned} I_{\ell_1 m_1 \ell_2 m_2}^\ell &\equiv -\lambda_{\ell_1 m_1} \lambda_{\ell_2 m_2} \\ &\times \int_0^\pi P_\ell(\cos(\varphi_1 - \varphi_2)) e^{i(m_1 \varphi_1 - m_2 \varphi_2)} d(\cos(\varphi_1 - \varphi_2)). \end{aligned} \quad (\text{A.10})$$

We need now to integrate out the  $\Theta$  and  $\Phi$  dependence in the right-hand side of (A.9) which was hidden due to our choice of a particular coordinate system. In order to do that, we keep the vectors  $\hat{n}_1$  and  $\hat{n}_2$  fixed and make a rotation of our coordinate system using three Euler angles  $\omega = \{\alpha, \beta, \gamma\}$ . This rotation changes the coefficients  $\mathcal{C}_\ell^{lm}$ 's and  $a_{\ell m}$ 's according to

$$a_{\ell m} = \sum_{m'} D_{mm'}^\ell(\omega) \tilde{a}_{\ell m'}, \quad \mathcal{C}_\ell^{lm} = \sum_{m'} D_{mm'}^l(\omega) \tilde{\mathcal{C}}_\ell^{lm'}, \quad (\text{A.11})$$

where  $\tilde{\mathcal{C}}_\ell^{lm}$  and  $\tilde{a}_{\ell m}$  are the multipolar coefficients in the new coordinate system, and where  $D_{mm'}^l(\omega)$  are the elements of the Wigner rotation matrix. The advantage of positioning the vectors  $\hat{n}_1$  and  $\hat{n}_2$  in the plane  $xy$  are that now the angles  $\Theta$  and  $\Phi$  are given precisely by the Euler angles  $\beta$  and  $\gamma$ , regardless of the value of  $\alpha$

$$\begin{aligned} \sum_{l,m} \mathcal{C}_\ell^{lm} Y_{lm}(0,0) &= \sum_{l,m'} \tilde{\mathcal{C}}_\ell^{lm'} \left( \sum_m D_{mm'}^l(\alpha, \beta, \gamma) Y_{lm}(0,0) \right) \\ &= \sum_{l,m'} \tilde{\mathcal{C}}_\ell^{lm'} Y_{l,-m}(\beta, \gamma), \end{aligned} \quad (\text{A.12})$$

where in the last step we have used  $Y_{lm}(0,0) = \sqrt{(2l+1)/4\pi} \delta_{m0}$ . Therefore, in our new coordinate system we have (dropping the “ $\sim$ ” in our notation)

$$\begin{aligned} &\frac{1}{2\pi} \sum_{l,m} \mathcal{C}_\ell^{lm} D_{0m}^l(\omega) \sqrt{2l+1} \\ &= \sum_{\ell_1, m_1} \sum_{\ell_2, m_2} \langle a_{\ell_1 m_1} a_{\ell_2 m_2}^* \rangle \sum_{m'_1, m'_2} I_{\ell_1 m'_1 \ell_2 m'_2}^\ell D_{m'_1 m_1}^{\ell_1}(\omega) D_{m'_2 m_2}^{\ell_2}(\omega). \end{aligned} \quad (\text{A.13})$$

We may now isolate  $\mathcal{C}_\ell^{lm}$  using the identities [41]

$$\begin{aligned} &\int d\omega D_{m'_1 m_1}^{\ell_1}(\omega) D_{m'_2 m_2}^{\ell_2}(\omega) \\ &= \frac{8\pi^2}{2l_1+1} \delta_{l_1 l_2} \delta_{m_1 m_2} \delta_{m'_1 m'_2}, \\ &\int d\omega D_{m'_1 m_1}^{\ell_1}(\omega) D_{m'_2 m_2}^{\ell_2}(\omega) D_{m'_3 m_3}^{\ell_3}(\omega) \\ &= 8\pi^2 \begin{pmatrix} l_1 & l_2 & l_3 \\ m'_1 & m'_2 & m'_3 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \end{aligned} \quad (\text{A.14})$$

where  $d\omega = \sin \beta d\beta d\alpha d\gamma$ , to obtain

$$\begin{aligned} &\frac{1}{\sqrt{2l+1}} \mathcal{C}_\ell^{lm} \\ &= 2\pi \sum_{\ell_1, m_1} \sum_{\ell_2, m_2} \langle a_{\ell_1 m_1} a_{\ell_2 m_2}^* \rangle \sum_{m'_1, m'_2} I_{\ell_1 m'_1 \ell_2 m'_2}^\ell (-1)^{m_2+m'_2+m} \\ &\quad \times \begin{pmatrix} \ell_1 & \ell_2 & l \\ m'_1 & -m'_2 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & l \\ m_1 & -m_2 & -m \end{pmatrix}. \end{aligned} \quad (\text{A.15})$$

If we now redefine  $-m_2 \rightarrow m_2$ , and note that the first 3-j symbol above is identically zero unless  $m'_1 = m'_2$ , we finally obtain (92).

*A.5. Some Properties of the Integral (93).* The geometrical coefficients  $I_{\ell_1 \ell_2}^{l, \ell}$  defined in (93) have many interesting properties which can be explored in order to speed up numerical computation of (92). First, we note that it is symmetric under permutation of  $\ell_1$  and  $\ell_2$

$$\begin{aligned} I_{\ell_1 \ell_2}^{l, \ell} &= \sum_m I_{\ell_1 m \ell_2 m}^\ell (-1)^m \begin{pmatrix} \ell_1 & \ell_2 & l \\ m & -m & 0 \end{pmatrix} \\ &= \sum_m I_{\ell_2 m \ell_1 m}^\ell (-1)^{m+\ell_1+\ell_2+l} \begin{pmatrix} \ell_2 & \ell_1 & l \\ -m & m & 0 \end{pmatrix} \\ &= \sum_m I_{\ell_2 m \ell_1 m}^\ell (-1)^{m+2(\ell_1+\ell_2+l)} \begin{pmatrix} \ell_2 & \ell_1 & l \\ m & -m & 0 \end{pmatrix} \\ &= I_{\ell_2 \ell_1}^{l, \ell}. \end{aligned} \quad (\text{A.16})$$

Some of the other properties are a consequence of the integral  $I_{\ell_1 m \ell_2 m}^\ell$  defined in (A.10). We may note for example

that, due to the symmetry of the  $\lambda_{\ell m}$  coefficient defined in (A.8), we will have

$$I_{\ell_1 \ell_2}^{\ell} = 0, \quad \text{for any } \{(\ell_1, \ell_2) \in \mathbb{N} \mid \ell_1 + \ell_2 = \text{odd}\}. \quad (\text{A.17})$$

Furthermore, the  $\lambda_{\ell m}$  coefficients restrict the  $m$  summation above to their values which obey  $m + \ell_1 + \ell_2 = \text{even}$ . If we further notice that (A.10) is proportional to the integral of an integral of the form  $\int_{-1}^1 P_\ell(\cos \theta) \cos m\theta d\theta$ , and that this integral is zero unless  $\ell + m = \text{even}$ , we conclude that

$$I_{\ell_1 \ell_2}^{\ell} = 0, \quad \text{for any } \{(\ell_1, \ell_2, \ell) \in \mathbb{N} \mid \ell_1 + \ell_2 + \ell = \text{odd}\}. \quad (\text{A.18})$$

Besides, using the fact that the integral  $\int_{-1}^1 P_\ell(\cos \theta) \cdot \cos m\theta d\theta$  is zero for any  $m < l$ , we find

$$I_{\ell_1 \ell_2}^{\ell} = 0, \quad \text{for any } \{(\ell_1, \ell_2, \ell) \in \mathbb{N} \mid \ell_1 < l, \ell_2 < l\}. \quad (\text{A.19})$$

We finally comment on the special case where  $l = 0$ , for which we have

$$I_{\ell_1 \ell_2}^{0, \ell} = \frac{(-1)^{\ell_1}}{\sqrt{2\ell_1 + 1}} \left( \sum_m I_{\ell_1 m \ell_2}^{\ell} \right) \delta_{\ell_1 \ell_2}. \quad (\text{A.20})$$

However,

$$\begin{aligned} & \sum_{m=-\ell'}^{\ell'} I_{\ell' m \ell'}^{\ell} \\ &= \int_0^\pi P_\ell(\cos \vartheta) \\ & \times \left( \sum_{m=-\ell'}^{\ell'} \frac{(2\ell' + 1)}{4\pi} \frac{\mathfrak{D}!!}{(\ell' + m)!!} \frac{\mathfrak{G}!!}{(\ell' - m)!!} e^{im\vartheta} \right) d(-\cos \vartheta) \\ &= \frac{2\ell' + 1}{4\pi} \int_{-1}^1 P_\ell(x) P_{\ell'}(x) dx \\ &= \frac{1}{2\pi} \delta_{\ell \ell'}, \end{aligned} \quad (\text{A.21})$$

where  $\mathfrak{D}$  denotes  $(\ell' + m - 1)!$ , and  $\mathfrak{G}$  denotes  $(\ell' - m - 1)!$ , and in the derivation above we have made use of the Fourier series expansion of the Legendre polynomial. So we conclude that

$$I_{\ell_1 \ell_2}^{0, \ell} = \frac{(-1)^{\ell_1}}{2\pi \sqrt{2\ell_1 + 1}} \delta_{\ell \ell_1} \delta_{\ell_1 \ell_2}, \quad (\text{A.22})$$

which is needed in the derivation of (106).

## Acknowledgments

The authors would like to thank Glenn Starkman and Yuri Shtanov for their enlightening discussions during the development of this paper. This paper was supported by FAPESP and CNPq.

## References

- [1] V. Mukhanov, *Physical Foundations of Cosmology*, Cambridge University Press, Cambridge, UK, 2005.
- [2] S. Weinberg, *Cosmology*, Oxford University Press, Oxford, UK, 2008.
- [3] P. Peter and J.-P. Uzan, *Primordial Cosmology*, Oxford University Press, Oxford, UK, 2009.
- [4] W. Hu, U. Seljak, M. White, and M. Zaldarriaga, "Complete treatment of CMB anisotropies in a FRW universe," *Physical Review D*, vol. 57, no. 6, pp. 3290–3301, 1998.
- [5] W. Hu, "CMB temperature and polarization anisotropy fundamentals," *Annals of Physics*, vol. 303, no. 1, pp. 203–225, 2003.
- [6] S. Dodelson, *Modern Cosmology*, Scottish Academic Press, Edinburgh, UK, 2003.
- [7] J. M. Maldacena, "Non-gaussian features of primordial fluctuations in single field inflationary models," *Journal of High Energy Physics*, vol. 2003, no. 05, article 013, 2003.
- [8] S. Weinberg, "Quantum contributions to cosmological correlations," *Physical Review D*, vol. 72, no. 4, Article ID 043514, 19 pages, 2005.
- [9] S. Weinberg, "Quantum contributions to cosmological correlations. II. Can these corrections become large?" *Physical Review D*, vol. 74, no. 2, Article ID 023508, 5 pages, 2006.
- [10] K. M. Huffenberger and U. Seljak, "Prospects for ACT: simulations, power spectrum, and non-Gaussian analysis," *New Astronomy*, vol. 10, no. 6, pp. 491–515, 2005.
- [11] E. Komatsu, J. Dunkley, M. R. Nolte et al., "Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation," *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [12] A. Bernui and M. J. Rebouças, "Searching for non-Gaussianity in the WMAP data," *Physical Review D*, vol. 79, no. 6, Article ID 063528, 11 pages, 2009.
- [13] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, "Theory of cosmological perturbations," *Physics Report*, vol. 215, no. 5-6, pp. 203–333, 1992.
- [14] G. Hinshaw, J. L. Weiland, R. S. Hill et al., "Five-year wilkinson microwave anisotropy probe observations: data processing, sky maps and basic results," *The Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 225–245, 2009.
- [15] M. Tegmark, M. R. Blanton, M. A. Strauss et al., "The three-dimensional power spectrum of galaxies from the sloan digital sky survey," *Astrophysical Journal*, vol. 606, no. 2, pp. 702–740, 2004.
- [16] S. Cole, W. J. Percival, J. A. Peacock, et al., "The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications," *Monthly Notices of the Royal Astronomical Society*, vol. 362, no. 2, pp. 505–534, 2005.
- [17] S. Perlmutter, G. Aldering, G. Goldhaber et al., "Measurements of  $\Omega$  and  $\Lambda$  from 42 high-redshift supernovae," *Astrophysical Journal*, vol. 517, no. 2, pp. 565–586, 1999.
- [18] A. G. Riess, A. V. Filippenko, P. Challis et al., "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astronomical Journal*, vol. 116, no. 3, pp. 1009–1038, 1998.
- [19] W. M. Wood-Vasey, G. Miknaitis, C. W. Stubbs et al., "Observational constraints on the nature of dark energy: first cosmological results from the essence supernova survey," *Astrophysical Journal*, vol. 666, no. 2, pp. 694–715, 2007.

- [20] C. L. Bennett, M. Halpern, G. Hinshaw et al., “First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: preliminary maps and basic results,” *Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 1–27, 2003.
- [21] G. F. Smoot, C. L. Bennett, A. Kogut, et al., “Structure in the COBE differential microwave radiometer first-year maps,” *Astrophysical Journal*, vol. 396, no. 1, pp. 1–5, 1992.
- [22] A. de Oliveira-Costa, M. Tegmark, M. Zaldarriaga, and A. Hamilton, “Significance of the largest scale CMB fluctuations in WMAP,” *Physical Review D*, vol. 69, no. 6, Article ID 063516, 12 pages, 2004.
- [23] C. J. Copi, D. Huterer, and G. D. Starkman, “Multipole vectors: a new representation of the CMB sky and evidence for statistical anisotropy or non-Gaussianity at  $2 \leq \ell \leq 8$ ,” *Physical Review D*, vol. 70, no. 4, Article ID 043515, 13 pages, 2004.
- [24] D. J. Schwarz, G. D. Starkman, D. Huterer, and C. J. Copi, “Is the low- $\ell$  microwave background cosmic?” *Physical Review Letters*, vol. 93, no. 22, Article ID 221301, 4 pages, 2004.
- [25] K. Land and J. Magueijo, “Examination of evidence for a preferred axis in the cosmic radiation anisotropy,” *Physical Review Letters*, vol. 95, no. 37, Article ID 071301, 4 pages, 2005.
- [26] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, “On the large-angle anomalies of the microwave sky,” *Monthly Notices of the Royal Astronomical Society*, vol. 367, no. 1, pp. 79–102, 2006.
- [27] M. Tegmark, A. de Oliveira-Costa, and A. J. S. Hamilton, “High resolution foreground cleaned CMB map from WMAP,” *Physical Review D*, vol. 68, no. 12, Article ID 123523, 14 pages, 2003.
- [28] L. R. Abramo, A. Bernui, and T. S. Pereira, “Searching for planar signatures in WMAP,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article 013, 2009.
- [29] A. Bernui, B. Mota, M. J. Rebouças, and R. Tavakol, “Mapping the large-scale anisotropy in the WMAP data,” *Astronomy and Astrophysics*, vol. 464, no. 2, pp. 479–485, 2007.
- [30] A. Bernui, “Anomalous CMB north-south asymmetry,” *Physical Review D*, vol. 78, no. 6, Article ID 063531, 10 pages, 2008.
- [31] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Górski, and P. B. Lilje, “Asymmetries in the cosmic microwave background anisotropy field,” *Astrophysical Journal*, vol. 605, no. 1 I, pp. 14–20, 2004.
- [32] H. K. Eriksen, A. J. Banday, K. M. Górski, F. K. Hansen, and P. B. Lilje, “Hemispherical power asymmetry in the third-year Wilkinson Microwave Anisotropy Probe sky maps,” *Astrophysical Journal*, vol. 660, no. 2, pp. 81–84, 2007.
- [33] D. Pietrobon, P. Cabella, A. Balbi, R. Crittenden, G. de Gasperis, and N. Vittorio, “Needlet bispectrum asymmetries in the WMAP 5-year data,” *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 402, no. 1, pp. L34–L38, 2010.
- [34] L. R. Abramo, L. Sodré Jr., and C. A. Wuensche, “Anomalies in the low CMB multipoles and extended foregrounds,” *Physical Review D*, vol. 74, no. 8, Article ID 083515, 12 pages, 2006.
- [35] L. R. Abramo, A. Bernui, I. S. Ferreira, T. Villela, and C. A. Wuensche, “Alignment tests for low CMB multipoles,” *Physical Review D*, vol. 74, no. 6, Article ID 063506, 10 pages, 2006.
- [36] D. Hanson and A. Lewis, “Estimators for CMB statistical anisotropy,” *Physical Review D*, vol. 80, no. 6, Article ID 063004, 15 pages, 2009.
- [37] A. R. Pullen and M. Kamionkowski, “Cosmic microwave background statistics for a direction-dependent primordial power spectrum,” *Physical Review D*, vol. 76, no. 10, Article ID 103529, 8 pages, 2007.
- [38] T. S. Pereira, C. Pitrou, and J.-P. Uzan, “Theory of cosmological perturbations in an anisotropic universe,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 09, article 006, 2007.
- [39] C. Pitrou, T. S. Pereira, and J.-P. Uzan, “Predictions from an anisotropic inflationary era,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 04, article 004, 2008.
- [40] A. E. Gümrükçüoğlu, C. R. Contaldi, and M. Peloso, “Inflationary perturbations in anisotropic backgrounds and their imprint on the cosmic microwave background,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, pp. 1–35, 2007.
- [41] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonsky, *Quantum Theory of Angular Momentum*, World Scientific, Singapore, 1988.
- [42] J.-P. Uzan, C. Clarkson, and G. F. R. Ellis, “Time drift of cosmological redshifts as a test of the copernican principle,” *Physical Review Letters*, vol. 100, no. 19, Article ID 191303, 4 pages, 2008.
- [43] S. Prunet, J.-P. Uzan, F. Bernardeau, and T. Brunier, “Constraints on mode couplings and modulation of the CMB with WMAP data,” *Physical Review D*, vol. 71, no. 8, pp. 1–12, 2005.
- [44] P. G. Ferreira and J. Magueijo, “Closet non-Gaussianity of anisotropic Gaussian fluctuations,” *Physical Review D*, vol. 56, no. 8, pp. 4578–4591, 1997.
- [45] J. N. Fry and A. L. Melott, “Statistical comparison of galaxy formation models: the bispectrum,” NSF-ITP-84-143.
- [46] X. Luo, “The angular bispectrum of the cosmic microwave background,” *Astrophysical Journal*, vol. 427, no. 2, pp. 71–74, 1994.
- [47] D. N. Spergel and D. M. Goldberg, “Microwave background bispectrum. I. Basic formalism,” *Physical Review D*, vol. 59, no. 10, Article ID 103001, 8 pages, 1999.
- [48] W. Hu, “Angular trispectrum of the cosmic microwave background,” *Physical Review D*, vol. 64, no. 8, Article ID 083005, 15 pages, 2001.
- [49] E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen, “Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets,” *Astrophysical Journal*, vol. 567, no. 1, pp. 2–17, 2002.
- [50] L. Ackerman, S. M. Carroll, and M. B. Wise, “Imprints of a primordial preferred direction on the microwave background,” *Physical Review D*, vol. 75, no. 8, Article ID 083502, 7 pages, 2007.
- [51] Y. Shtanov and H. Pyatkovska, “Statistical anisotropy in the inflationary universe,” *Physical Review D*, vol. 80, no. 2, Article ID 023521, 19 pages, 2009.
- [52] J.-P. Luminet, J. R. Weeks, A. Riazuelo, R. Lehoucq, and J.-P. Uzan, “Dodecahedral space topology as an explanation for weak wide-angle temperature correlations in the cosmic microwave background,” *Nature*, vol. 425, no. 6958, pp. 593–595, 2003.
- [53] A. Riazuelo, J. Weeks, J.-P. Uzan, R. Lehoucq, and J.-P. Luminet, “Cosmic microwave background anisotropies in multi-connected flat spaces,” *Physical Review D*, vol. 69, no. 10, Article ID 103518, 25 pages, 2004.
- [54] W. S. Hipólito-Ricaldi and G. I. Gomero, “Topological signatures in CMB temperature anisotropy maps,” *Physical Review D*, vol. 72, no. 10, Article ID 103008, 24 pages, 2005.

- [55] T. Kahniashvili, G. Lavrelashvili, and B. Ratra, "CMB temperature anisotropy from broken spatial isotropy due to a homogeneous cosmological magnetic field," *Physical Review D*, vol. 78, no. 6, Article ID 063012, 15 pages, 2008.
- [56] T. Kahniashvili, Y. Maravin, and A. Kosowsky, "Faraday rotation limits on a primordial magnetic field from Wilkinson Microwave Anisotropy Probe five-year data," *Physical Review D*, vol. 80, no. 2, Article ID 023009, 5 pages, 2009.
- [57] A. Bernui and W. S. Hipólito-Ricaldi, "Can a primordial magnetic field originate large-scale anomalies in WMAP data?" *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 3, pp. 1453–1460, 2008.
- [58] C. Caprini, F. Finelli, D. Paoletti, and A. Riotto, "The cosmic microwave background temperature bispectrum from scalar perturbations induced by primordial magnetic fields," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 6, article 021, 2009.
- [59] T. R. Seshadri and K. Subramanian, "Cosmic microwave background bispectrum from primordial magnetic fields on large angular scales," *Physical Review Letters*, vol. 103, no. 8, Article ID 081303, 4 pages, 2009.
- [60] C. Gordon, W. Hu, D. Huterer, and T. Crawford, "Spontaneous isotropy breaking: a mechanism for CMB multipole alignments," *Physical Review D*, vol. 72, no. 10, Article ID 103002, 13 pages, 2005.
- [61] L. Campanelli, P. Cea, and L. Tedesco, "Cosmic microwave background quadrupole and ellipsoidal universe," *Physical Review D*, vol. 76, no. 6, Article ID 063007, 10 pages, 2007.
- [62] L. Campanelli, "Model of universe anisotropization," *Physical Review D*, vol. 80, no. 6, Article ID 063006, 12 pages, 2009.
- [63] A. L. Erickcek, M. Kamionkowski, and S. M. Carroll, "A hemispherical power asymmetry from inflation," *Physical Review D*, vol. 78, no. 12, Article ID 123520, 5 pages, 2008.
- [64] A. L. Erickcek, C. M. Hirata, and M. Kamionkowski, "A scale-dependent power asymmetry from isocurvature perturbations," *Physical Review D*, vol. 80, no. 8, Article ID 083507, 17 pages, 2009.
- [65] J. F. Donoghue, K. Dutta, and A. Ross, "Nonisotropy in the CMB power spectrum in single field inflation," *Physical Review D*, vol. 80, no. 2, Article ID 023526, 13 pages, 2009.
- [66] M. Kawasaki, K. Nakayama, T. Sekiguchi, T. Suyama, and F. Takahashi, "Non-Gaussianity from isocurvature perturbations," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 11, article 019, 2008.
- [67] M. Kawasaki, K. Nakayama, T. Sekiguchi, T. Suyama, and F. Takahashi, "A general analysis of non-gaussianity from isocurvature perturbations," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 1, article 042, 2009.
- [68] A. Hajian and T. Souradeep, "Measuring the statistical isotropy of the cosmic microwave background anisotropy," *Astrophysical Journal*, vol. 597, no. 1, pp. 5–8, 2003.
- [69] A. Hajian, T. Souradeep, and N. Cornish, "Statistical isotropy of the wilkinson microwave anisotropy probe data: a bipolar power spectrum analysis," *Astrophysical Journal*, vol. 618, no. 2, pp. 63–66, 2005.
- [70] A. Hajian and T. Souradeep, "The Cosmic Microwave Background bipolar power spectrum: basic formalism and applications," <http://arxiv.org/abs/astro-ph/0501001>.
- [71] T. S. Pereira and L. R. Abramo, "Angular-planar CMB power spectrum," *Physical Review D*, vol. 80, no. 6, Article ID 063525, 12 pages, 2009.
- [72] J. R. Weeks, "Maxwell's multipole vectors and the CMB," <http://arxiv.org/abs/astro-ph/0412231>.
- [73] G. Katz and J. Weeks, "Polynomial interpretation of multipole vectors," *Physical Review D*, vol. 70, no. 6, Article ID 063527, 6 pages, 2004.
- [74] E. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics*, Chelsea Publishing, London, UK, 1955.
- [75] M. R. Dennis and K. Land, "Probability density of the multipole vectors for a Gaussian cosmic microwave background," *Monthly Notices of the Royal Astronomical Society*, vol. 383, no. 2, pp. 424–434, 2008.
- [76] R. B. Ash, *Basic Probability Theory*, Dover, New York, NY, USA, 2008.
- [77] <http://lambda.gsfc.nasa.gov/>.
- [78] J. Kim, P. Naselsky, and P. R. Christensen, "CMB map derived from the WMAP data through harmonic internal linear combination," *Physical Review D*, vol. 77, no. 10, Article ID 103002, 14 pages, 2008.
- [79] G. Hinshaw, M. R. Nolte, C. L. Bennett et al., "Three-year Wilkinson Microwave Anisotropyprobe (WMAP1) observations: temperature analysis," *Astrophysical Journal, Supplement Series*, vol. 170, no. 2, pp. 288–334, 2007.
- [80] A. de Oliveira-Costa and M. Tegmark, "CMB multipole measurements in the presence of foregrounds," *Physical Review D*, vol. 74, no. 2, Article ID 023005, 13 pages, 2006.
- [81] C.-G. Park, C. Park, and J. R. Gott III, "Cleaned 3 year Wilkinson Microwave anistropy probe cosmic microwave background map: magnitude of the quadrupole and alignment of large-scale modes," *Astrophysical Journal*, vol. 660, no. 2 I, pp. 959–975, 2007.
- [82] J. Delabrouille, J.-F. Cardoso, M. Le Jeune, M. Betoule, G. Fay, and F. Guilloux, "A full sky, low foreground, high resolution CMB map from WMAP," *Astronomy and Astrophysics*, vol. 493, no. 3, pp. 835–837, 2009.
- [83] N. Jarosik, C. Barnes, M. R. Greason et al., "Three-year Wilkinson Microwave Anisotropyprobe (WMAP)1 observations: beam profiles, data processing, radiometer characterization, and systematic error limits," *Astrophysical Journal, Supplement Series*, vol. 170, no. 2, pp. 263–287, 2007.
- [84] B. Gold, C. L. Bennett, R. S. Hill et al., "Five-year wilkinson microwave anisotropy probe observations: galactic foreground emission," *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 265–282, 2009.
- [85] C. L. Francis and J. A. Peacock, "An estimate of the local ISW signal, and its impact on CMB anomalies," *Monthly Notices of the Royal Astronomical Society*, vol. 406, no. 1, pp. 14–21, 2010.
- [86] C. L. Bennett, R. S. Hill, and G. Hinshaw, "Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: are there cosmic microwave background anomalies?" <http://arxiv.org/abs/1001.4758>.

## Review Article

# Cosmic Strings and Their Induced Non-Gaussianities in the Cosmic Microwave Background

**Christophe Ringeval**

*Institute of Mathematics and Physics, Centre for Cosmology, Particle Physics and Phenomenology, Louvain University,  
2 Chemin du Cyclotron, 348 Louvain-la-Neuve, Belgium*

Correspondence should be addressed to Christophe Ringeval, ringeval@fyma.ucl.ac.be

Received 15 January 2010; Accepted 27 May 2010

Academic Editor: Eiichiro Komatsu

Copyright © 2010 Christophe Ringeval. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Motivated by the fact that cosmological perturbations of inflationary quantum origin were born Gaussian, the search for non-Gaussianities in the cosmic microwave background (CMB) anisotropies is considered as the privileged probe of nonlinear physics in the early universe. Cosmic strings are active sources of gravitational perturbations and incessantly produce non-Gaussian distortions in the CMB. Even if, on the currently observed angular scales, they can only contribute a small fraction of the CMB angular power spectrum, cosmic strings could actually be the main source of its non-Gaussianities. In this paper, after having reviewed the basic cosmological properties of a string network, we present the signatures Nambu-Goto cosmic strings would induce in various observables ranging from the one-point function of the temperature anisotropies to the bispectrum and trispectrum. It is shown that string imprints are significantly different than those expected from the primordial type of non-Gaussianity and could therefore be easily distinguished.

## 1. Motivations

The origin of cosmic strings dates back to the discovery that cosmological phase transitions triggered by the spontaneous breakdown of the fundamental interaction symmetries may form topological defects [1–3]. Cosmic strings belong to the class of line-like topological defects, as opposed to point-like monopoles and the membrane shaped domain walls. As shown by Kibble, the appearance of defects in any field theory is related to the topology of the vacuum manifold [3]. If the ground state of a field theory experiences a spontaneous breakdown from a symmetry group  $G$  to a subgroup  $H$ , Kibble showed that cosmic strings will be formed if the first homotopy group  $\pi_1(G/H) \neq I$  is nontrivial. In other words, if noncontractile loops can be found in the manifold  $\mathcal{M} = G/H$  of equivalent vacua. Similarly, the other homotopy groups  $\pi_0$  and  $\pi_2$  determine the formation of domain walls and monopoles, respectively. Once formed and cooled, these defects cannot be unfolded, precisely due to their nontrivial topological configuration over the vacuum manifold of the theory. This simple statement suggests that cosmic strings, and topological defects in general, are a natural outcome of

the unification of the fundamental interactions in the context of Cosmology. As remnants of unified forces, their discovery would be an incredible opportunity to probe extremely high-energy physics with “a telescope”.

In the last thirty years, many works have been devoted to the cosmological consequences, signatures, and searches for topological defects [4–7]. They have pushed cosmic strings to the privileged place to be generically compatible with observations. Indeed, domain walls and monopoles are prone to suffer from the cosmological catastrophe problem; their formation is sufficiently efficient (or their annihilation sufficiently inefficient) to either overclose the universe or spoil the Big-Bang Nucleosynthesis (BBN) predictions [8, 9]. For domain walls, this implies that either they should be extremely light, that is, formed at an energy scale less than a few MeV, or no discrete symmetry should have been broken during the cooling of the universe. There is not so much choice for the monopoles; if interactions were unified, monopoles would have been formed. The homotopy group of  $\pi_2(G/H)$  with  $H$  containing the  $U(1)$  of electroweak interactions is indeed nontrivial. (As often with topological defects, sensitivity to the underlying model is such that one

can often find a counter-example of any result. Both of these statements, on walls and monopoles, can be evaded in some particular models or with some amount of fine-tuning, as for instance if cosmic strings can be attached to them and catalyze annihilations [10, 11].) Cosmic inflation was originally designed to solve the monopole problem. If a phase of accelerated expansion of the universe occurs, then any defects will be diluted enough to no longer have any (dramatic) consequences on cosmology [12–15]. Meanwhile, Inflationary Cosmology solves the flatness and homogeneity problem of the standard Big-Bang model, explains the origin and spectrum of the cosmic microwave background (CMB) anisotropies, as the formation of the large-scale structures [16–19]. Inflation provides a priori an easy solution to the topological defects problem by diluting them to at most one per Hubble radius. However, one has to keep in mind that this mechanism works only if the defects were formed before inflation, and even in that case some may survive [20]. This has to be the case for monopoles and heavy walls, but not for local strings. On the contrary, exhaustive analysis of particle physics motivated inflationary models, embedding the Standard Model  $SU(3) \times SU(2) \times U(1)$ , has shown that strings are generically produced at the end of inflation [21]. In this picture, our universe should contain cosmic strings whose properties are closely related to those of the inflation [22–24]. String Theory provides an alternative framework to Field Theories: brane inflationary models propose that the accelerated expansion of the universe is induced by the motion of branes in warped and compact extradimensions [25–28]. Inflation ends when two branes collide and such a mechanism again triggers the formation of one-dimensional cosmological extended objects, dubbed cosmic superstrings [29–32]. These objects may be cosmologically stretched fundamental strings or one-dimensional D-brane [33, 34]. Although cosmic superstrings are of a different nature than their topological analogue, they produce the same gravitational effects and share similar cosmological signatures [35, 36].

Among the expected signatures, cosmic strings induce temperature anisotropies in the CMB with an amplitude typically given by  $GU$ , where  $U$  is the string energy per unit length and  $G$  the Newton constant [37]. (To avoid any confusion with Greek tensor indices, we will use the Carter's notations  $U$  and  $T$  for the string energy density and tension [38].) For the Grand Unified Theory (GUT) energy scale, one has  $GU \approx 10^{-5}$ , which precisely corresponds to the observed amplitude of the CMB temperature fluctuations [39]. However, the power spectra do not match; topological defects are active sources of gravitational perturbations, that is, they produce perturbations all along the universe history, and cannot produce the characteristic coherent patterns of the acoustic peaks [40–44]. Current CMB data analyses including a string contribution suggest that they can only contribute to at most 10% of the overall anisotropies on the observed angular scales [45, 46]. For Abelian cosmic strings (see Section 2), numerical simulations in Friedmann-Lemaître-Robertson-Walker (FLRW) spacetimes show that this corresponds to an upper two-sigma bound  $GU < 7 \times 10^{-7}$  [47]. Direct detection searches provide less stringent

limits but are applicable to all cosmic string models:  $GU < 4 \times 10^{-6}$  [48–50]. Detecting cosmic strings in the CMB certainly requires one to go further than the power spectrum [51, 52] (see, however, Section 4.5). In fact, strings induce line-like discontinuities in the CMB temperature through the so-called Gott-Kaiser-Stebbins effect, which are intrinsically of non-Gaussian nature [53, 54]. In the inflationary picture, cosmological perturbations find their origin in the quantum fluctuations of the field-metric system, and therefore were born generically Gaussian. Non-Gaussianities can nevertheless appear from non-linear effects during inflation or from couplings to other fields (see the other articles in this issue). These non-Gaussianities are of the primordial type, that is, they exist before the cosmological perturbations reenter the Hubble radius. On the other hand, cosmic strings are a source of non-Gaussianity at all times and, as we will see, produce different signals from the CMB point of view. (Notice that second-order perturbations, being non-linear, actively generate non-Gaussianities but at a relatively small amplitude [55–58].)

In this paper, we review the non-Gaussian features a cosmological network of cosmic strings produce in the CMB anisotropies. In a first section, we briefly scan various cosmic string models and emphasize their similarities and differences for cosmology. Making observable predictions for cosmic strings faces the problem of understanding their cosmological evolution. Not only one has to solve the local dynamics in curved space, but as extended objects, cosmic strings follow a globally nonlocal evolution: the fate of one string depends on its interactions with the others. The cosmological evolution of a network of cosmic strings is a nontrivial problem which can be overcome by means of numerical simulations. These simulations permit an estimation of the various statistical properties affecting the observational signatures, such as the number of strings per Hubble radius, their shapes, velocities, or the loop density distribution. Latest results in this area, for the Nambu-Goto (NG) type of cosmic strings, are presented in Section 3. Once the statistical properties of a cosmological cosmic strings network are known, it is possible to extract meaningful observables depending only on the unique model parameter  $U$ . (If no currents are flowing along the string, Lorentz invariance implies that the string tension  $T$  equals the energy density  $U$ .) In Section 4, we recap the expected CMB temperature anisotropies induced by cosmic strings, derived from various methods. Particular attention is paid to small angle CMB maps which preserve all of the projected statistical information. We then derive the cosmic string signals expected in various non-Gaussian estimators ranging from the one-point function of the CMB temperature fluctuations to the bispectrum and trispectrum. We conclude in Section 5 and discuss various non-Gaussian aspects which still have to be explored.

## 2. Cosmic Strings of Various Origins

Cosmic strings of cosmological interest can be of various kinds depending on the microscopic model they stem from.

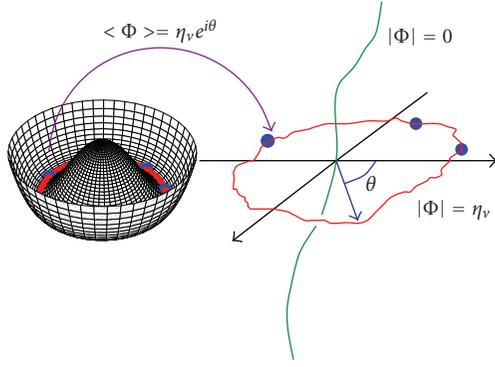


FIGURE 1: The Abelian Higgs potential in the complex plane  $[\Re(\Phi), \Im(\Phi)]$ . The nontrivial phase mapping from the internal space to the physical space (right) leads to the formation of a cosmic string. The old vacuum  $|\Phi| = 0$  becomes trapped inside the new one  $|\Phi| = \eta_v$ .

As mentioned in the introduction, they can either be nontrivial stable, or metastable, field configurations or more fundamental objects in String Theory. From a gravitational point of view, they all are, however, line-like energy density and pressure distributions. In the following, we briefly review the different kinds of string having a cosmological interest and we emphasize their similarities and differences.

**2.1. Abelian Vortices.** The simplest example of cosmic string illustrating the Kibble mechanism is the Abelian Higgs model. The theory is invariant under a local gauge group  $U(1)$  and the Higgs potential assumes its standard Mexican hat renormalisable form

$$V(\Phi) = \frac{\lambda}{8} (|\Phi|^2 - \eta_v^2)^2, \quad (1)$$

where  $\lambda$  is the self-coupling constant and  $\eta_v$  the vacuum expectation value of the Higgs field  $\Phi$ . In Minkowski space, the Lagrangian reads

$$\mathcal{L}_h = \frac{1}{2} (D_\mu \Phi)^\dagger (D^\mu \Phi) - \frac{1}{4} H_{\mu\nu} H^{\mu\nu} - V(\Phi), \quad (2)$$

where  $H_{\mu\nu}$  is the field strength tensor associated with the vector gauge boson  $B_\mu$  and

$$D_\mu = \partial_\mu + igB_\mu. \quad (3)$$

At high enough temperature, loop corrections from the thermal bath restore the  $U(1)$  symmetry and the effective potential has an overall minimum at  $|\Phi| = 0$  [1, 59]. Starting from high enough temperature, one therefore expects the  $U(1)$  symmetry to be spontaneously broken during the expansion and cooling of the universe. During the phase transition, the Higgs field reaches its new vacuum expectation value  $\Phi = \eta_v e^{i\theta}$ . At each spacetime location, the phase  $\theta(x^\mu)$  will have a given value, all of them being uncorrelated on distances larger than the typical correlation length of the phase transition. As pointed by Kibble, this is at most the horizon size  $d_h \propto t$  although one expects

it to be much smaller [3, 60–63]. As a result, there exists closed paths in space along which  $\theta$  varies from 0 to  $2\pi$  (or a multiple of  $2\pi$ ). Such phase configurations necessarily encompass a point at which  $|\Phi| = 0$  (see Figure 1); the old vacuum has been trapped into a nontrivial configuration of the new vacuum, and this prevents its decay. Such a structure is invariant by translations along the third spatial dimension and is string shaped.

Solitonic solutions of the field equations describing a static straight Abelian string can easily be computed under the Nielsen–Olesen ansatz. The transverse profile of the Higgs and gauge field are assumed to be [64]

$$\Phi = \eta_v H(\varrho) e^{in\theta}, \quad B_\mu = \frac{Q(\varrho) - n}{g} \delta_{\mu\theta}, \quad (4)$$

where  $(r, \theta)$  stands for a polar coordinate system aligned along the string. The dimensionless radial coordinate has been defined by  $\varrho = m_h r$  where,  $m_h = \sqrt{\lambda} \eta_v$  is the mass of the Higgs boson. The integer  $n$  is the “winding number” and gives the number of times the Higgs winds the potential for one rotation around the string. From (2), the dimensionless equations of motion read

$$\begin{aligned} \frac{d^2 H}{d\varrho^2} + \frac{1}{\varrho} \frac{dH}{d\varrho} &= \frac{HQ^2}{\varrho^2} + \frac{1}{2} H(H^2 - 1), \\ \frac{d^2 Q}{d\varrho^2} - \frac{1}{\varrho} \frac{dQ}{d\varrho} &= \frac{m_b^2}{m_h^2} H^2 Q, \end{aligned} \quad (5)$$

where  $m_b = g\eta_v$  is the mass of the vector gauge boson. In Figure 2, we have represented the string solution to these equations in Minkowski spacetime [65, 66]. The boundary conditions are such that the Higgs field vanishes at the center of the string to reach its vacuum expectation value (vev) asymptotically. This typically happens after a length scale given by its Compton wavelength  $1/m_h$ . Similarly, the gauge field boundary conditions are such that it has vanishing derivative in the core and remains finite far from the string. As shown in Figure 2, it actually condenses inside the string with a spatial extension roughly equal to  $1/m_b$ .

The energy content of such a string is given by the stress tensor stemming from the Lagrangian of (2). Along the string worldsheet,

$$\begin{aligned} T^{tt} &= -T^{zz} \\ &= \frac{\lambda \eta_v^4}{2} \left[ (\partial_\varrho H)^2 + \frac{Q^2 H^2}{\varrho^2} + \frac{(H^2 - 1)^2}{4} + \frac{\lambda}{g^2} \frac{(\partial_\varrho Q)^2}{\varrho^2} \right], \end{aligned} \quad (6)$$

which are the only two components which do not vanish after an integration over the transverse coordinates. Integrating the temporal part gives the string energy per unit length  $U$ , whereas the longitudinal component gives  $-T$ . One finally gets

$$U = T = C \left( \frac{\lambda}{g^2} \right) \eta_v^2, \quad (7)$$

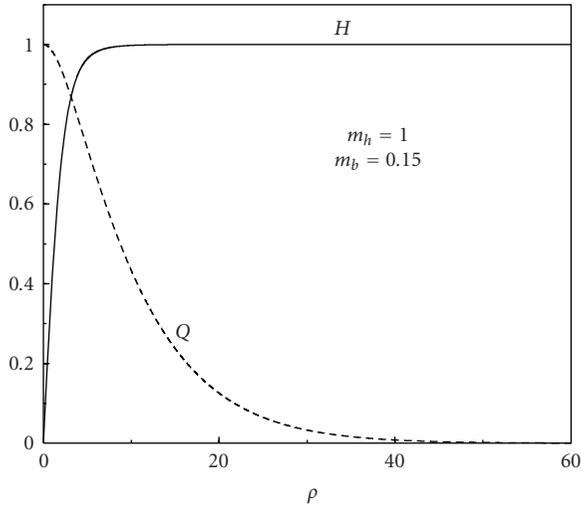


FIGURE 2: String forming field profiles in the Abelian Higgs model with unity winding number. The Higgs field  $H$  vanishes in the vortex core and reaches its vacuum expectation value within distances given by the Higgs Compton wavelength  $1/m_h$ . The gauge field  $Q$  condenses inside the vortex over distances given by  $1/m_b$ .

where  $C(\lambda/g^2)$  is an order unity function at fixed winding number. Increasing the winding number centrifuges the energy density around the core such that  $U$  is changed in a more complex way [5]. This immediately shows that cosmic strings generically carry an energy density and tension of the order of the symmetry breaking energy scale  $U \simeq \eta_v^2$ . Notice that along the string direction the pressure  $P_z = -T = -U$  is negative, and we are in presence of a “cosmological constant wire”, as one may expect from a Lorentz invariant vacuum object. Consequently, the trace of the stress tensor  $\eta_{\mu\nu} T^{\mu\nu}$  vanishes and cosmic strings do not induce any Newtonian gravitational potential. Together with the so-called cosmological scaling behaviour (see below), this is the mere reason why they remain cosmologically acceptable. They do however induce dynamical gravitational effects, the metric far from the string core being Minkowski with a missing angle [67] (see Section 4).

The Abelian string model is intensively used in the literature to explore the string forming phase transition and string interactions [68, 69]. From the Kibble’s argument, one expects the phase of the Higgs field to be random and the resulting string path should be a self-avoiding random walk with a given correlation length [70]. Performing lattice simulations allows to probe in more details the string forming mechanism and gives a more accurate picture of a cosmic string network just after its formation [71–74]. Abelian Higgs simulations are also used to compute the cosmological evolution of such a network [75–77] (see Section 3).

## 2.2. Other Flux Tubes

*Global String.* The Abelian string provides an explicit example of the formation of line-like topological defects by the

spontaneous breakdown of a gauged symmetry. Breaking a  $U(1)$  global symmetry can also produce topological defects, the so-called global strings. However, in the absence of gauge fields, one can show that global cosmic strings exhibit long-range interactions and Goldstone radiation [78]. Their dynamics can however mimic local strings and being cosmologically acceptable in some regime [79–81].

*Non-Abelian String.* If the broken symmetry group  $G$  is non-Abelian, the cosmic strings formed during the phase transition exhibit new properties compared to the  $U(1)$  kind [5]. In particular, the mapping of the Higgs field to the real space can be made along different broken generators  $\Phi_a = \eta_v \exp(iT_a \theta)$ . This implies that different type of non-Abelian strings may be formed and will interact with each others according to their respective windings. The classic example being the appearance of a  $[T_a, T_b]$ -string from the crossing between a  $T_a$ -string and  $T_b$ -string [82]. In the cosmological framework, new strings can potentially be formed at each interaction leading to a frustrated intricate configuration [83–86]. Such an outcome depends on the underlying non-Abelian gauge group and Abelian string-like evolution can also be recovered, as for instance in the  $U(N)$  models [87–91].

*Semilocal String.* String-shaped energy density distribution can also appear even if the vacuum manifold is simply connected. The nontrivial topology argument is indeed only a sufficient condition of defect appearance. The electroweak symmetry breaking scheme enters this class, although the first homotopy group is trivial, semilocal strings can be formed [92, 93]. A simple description of these strings can be obtained by replacing the Higgs field in the Abelian model by a doublet in a  $SU(2)$  global representation [94]. These flux tube configurations are stabilised because they can be energetically favoured for some values of the model parameters, typically for  $m_b > m_h$  [95]. Let us notice that the currently measured electroweak model parameters do not support stable vortices [94]. Similar vortices could also be formed during a chiral symmetry breaking phase transition [96–103].

*K- and DBI-String.* These are another extensions of the Abelian Higgs string for which the scalar and gauge field kinetic terms are noncanonical, or of the Dirac-Born-Infeld form [104–107]. These strings essentially differ from their Abelian counterparts when the gradient terms are non-vanishing, that is, in the core.

*Current Carrying String.* In minimal extensions of the Abelian Higgs model, one may couple extra-scalar fields to the string forming Higgs field. As shown by Witten, this can lead to the condensation of the extrascalar over the string core [108]. The resulting string is carrying a current that breaks the longitudinal Lorentz invariance;  $U$  and  $T$  are no longer degenerated and the string dynamics is affected [38, 65, 109]. One of the most important consequence of these currents is the potential appearance of centrifugally

supported loops. If stable, these so-called vortons could efficiently populate the universe and avoiding the overclosure gives strong constraints on the cosmic string energy scale [110, 111]. A similar mechanism works for the fermionic fields which are Yukawa coupled to the string forming Higgs field. They generically produce currents along the string with a discrete mass spectrum, in a way similar to the photon propagation in waveguides [112]. Unless the massive propagation modes are not excited, the resulting loops are however expected to be unstable [113, 114].

**2.3. Cosmic Superstrings.** Cosmic superstrings are fundamental line-shaped objects that can be formed at the end of brane-inflation (see [34–36, 115, 116] for reviews). The idea that fundamental quantum strings can be stretched to cosmological distances has been mentioned by Witten [117]. If stable, one would expect fundamental strings to be at an energy scale close to the String Theory scale, that is, close to the Planck mass, and this is trivially ruled out by observations. In addition, current CMB constraints tell us that the energy scale of inflation is at most the GUT scale [118], implying that strings formed at a higher energy would have been diluted anyway. The situation changed with the discovery that inflation within String Theory could be a geometrical phenomena induced by the motion of a brane moving in a warped throat, somewhere in the compact manifold of the extra-dimensions [26]. In the KKLT model [28, 119], the inflaton is a scalar degree of freedom associated with the position  $r$  of a D3 brane in a warped throat. Within a ten-dimensional super-gravity ansatz for the metric, in the type IIB String Theory,

$$ds^2 = \frac{1}{\sqrt{h(r)}} g_{\mu\nu} dx^\mu dx^\nu + \sqrt{h(r)} (dr^2 + r^2 d_5s^2), \quad (8)$$

and the throat is described by the warping function  $h(r)$  (explicitly, it can be the Klebanov-Strassler conifold [120]). In this system, accelerated expansion of our universe comes from the interaction of this brane with an anti-D3 brane sitting at the bottom of the throat  $r_0$ . Current CMB data suggest that inflation preferentially ends by violation of the slow-roll conditions, whereas the system continues to evolve till the two branes collide [121]. The brane interactions at that stage require String Theory calculations and are expected to trigger a reheating era accompanied by a copious production of various D1-branes and fundamental F-strings [29–32]. Since the brane annihilation takes place at the bottom of the throat, due to the warped metric, the cosmic superstring tensions measured by an exterior four-dimensional observer are redshifted by a factor  $h^{1/2}(r_0)$ . The resulting effect is to significantly lower the string tension down to acceptable values. In fact, the stability of the produced F-strings and D-strings require additional constraints on the model parameters and the spectrum of superstring tensions depending on the underlying scenario [32]. For instance, in the KKLT model, one expects  $10^{-10} < GU < 10^{-7}$  [122].

Cosmic superstrings differ from the Abelian strings in various aspects. In addition to the coexistence of two

different types, they can form bound states of  $p$  F-strings and  $q$  D-strings. The tension of these  $(p, q)$ -strings depends on  $p, q$ , the binding energy but also on their configuration in the throat [123, 124]. In fact, many of  $(p, q)$ -string properties mimic the non-Abelian type of topological vortex, as the existence of bound states and Y-junctions [125–128]. Such similarities have actually been used to probe the properties of the cosmic superstrings through the more tractable framework of field theory [129–132].

**2.4. Infinitely Thin Strings.** These are the one-dimensional version of the relativistic point particles. Following Carter macroscopic covariant approach [38, 109, 133, 134], string events can be localised in the four-dimensional spacetime by the so-called embedding functions  $x^\mu = X^\mu(\xi^a)$ , where  $\xi^0$  and  $\xi^1$  are a timelike and spacelike internal coordinate of the string worldsheet. Denoting by  $g$  the four-dimensional metric tensor, one can define the two-dimensional induced metric

$$\gamma_{ab} = g_{\mu\nu} \frac{\partial X^\mu}{\partial \xi^a} \frac{\partial X^\nu}{\partial \xi^b}, \quad (9)$$

such that the infinitesimal interval between two events reduces to  $ds^2 = \gamma_{ab} d\xi^a d\xi^b$ . From its inverse, one can define the first fundamental tensor

$$q^{\mu\nu} = \gamma^{ab} \frac{\partial X^\mu}{\partial \xi^a} \frac{\partial X^\nu}{\partial \xi^b}, \quad (10)$$

which is nothing but a projector over the string worldsheet. Similarly,  $\perp_\nu^\mu \equiv g_\nu^\mu - q_\nu^\mu$  is an orthogonal projector and they verify

$$q_\rho^\mu q_\nu^\rho = q_\nu^\mu, \quad \perp_\rho^\mu \perp_\nu^\rho = \perp_\nu^\mu, \quad q_\rho^\mu \perp_\nu^\rho = 0. \quad (11)$$

Variations of the first fundamental form are encoded in the second fundamental tensor

$$K_{\mu\nu}^\rho \equiv q_\nu^\alpha \bar{\nabla}_\mu q_\alpha^\rho, \quad (12)$$

where a bar quantity stands for the projection of its four-dimensional analogue over the worldsheet, that is,  $\bar{\nabla}_\mu \equiv q_\mu^\alpha \nabla_\alpha$ . Integrability imposes  $K_{[\mu\nu]}^\rho = 0$  and, by construction, the second fundamental form is, respectively, tangent and orthogonal to the worldsheet on its first and last indices. As a result, contracting the first two tangential indices gives a purely orthogonal vector which measures the string extrinsic curvature [134]

$$K^\rho \equiv K_\mu^{\mu\rho} = \bar{\nabla}_\mu q^{\mu\rho}. \quad (13)$$

The energy content of a spacetime two-dimensional surface can be characterised by its internal stress energy tensor. Similarly to the cosmological perfect fluid, one may consider a string whose stress-energy tensor is diagonal in a preferred basis. Positivity of the energy conditions ensures that the timelike eigenvalue  $U > 0$ , while the spacelike eigenvalue  $T$  should verify  $|T| \leq U$  [135]. In this frame,  $U$  represents the energy per unit length of the string and  $T$  the

string tension. Denoting by  $u^\mu$  and  $v^\nu$  the respective timelike and spacelike orthonormal eigenvectors, one has

$$\bar{T}^{\mu\nu} = Uu^\mu u^\nu - Tv^\mu v^\nu = (U - T)u^\mu u^\nu - Tq^{\mu\nu}, \quad (14)$$

where

$$\begin{aligned} u^\alpha u_\alpha &= -1, & v^\alpha v_\alpha &= 1, & u^\alpha v_\alpha &= 0, \\ q^{\mu\nu} &= -u^\mu u^\nu + v^\mu v^\nu. \end{aligned} \quad (15)$$

In the absence of external forces, reparametrisation invariance of the string worldsheet ensures the stress-energy pseudo-conservation from Noether's theorem [136]

$$\bar{\nabla}_\rho \bar{T}^{\rho\sigma} = 0. \quad (16)$$

As for a cosmological fluid, these equations are not sufficient to close the equations of motion for the string. One has to supplement them by an equation of state of the fluid under scrutiny. The simplest case is the so-called barotropic model for which the equation of state is the relation  $U(T)$ . One can then introduce the two Legendre conjugated parameters

$$\ln \bar{\nu} = \int \frac{dU}{U - T}, \quad \ln \bar{\mu} = \int \frac{dT}{T - U}, \quad (17)$$

such that  $U - T = \bar{\mu}\bar{\nu}$ . Clearly,  $\bar{\nu}$  plays the role of a number density and its Legendre conjugated parameter  $\bar{\mu}$  will therefore be a chemical potential, that is, an effective mass carried per unit number density. Defining their respective worldsheet current density by

$$\bar{\mu}^\rho \equiv \bar{\mu}u^\rho, \quad \bar{\nu}^\rho \equiv \bar{\nu}v^\rho, \quad (18)$$

one can rewrite (16) as

$$\bar{\nabla}_\rho \bar{T}_\sigma^\rho = \bar{\mu}_\sigma \bar{\nabla}_\rho \bar{\nu}^\rho + \bar{\nu}^\rho \bar{\nabla}_\rho \bar{\mu}_\sigma - \bar{\nu}^\rho \bar{\nabla}_\sigma \bar{\mu}_\rho - T K_\sigma = 0. \quad (19)$$

Contracting (19) with  $\bar{\nu}^\sigma$  ensures the current conservation along the string

$$\bar{\nabla}_\rho \bar{\nu}^\rho = 0, \quad (20)$$

while its projection onto the worldsheet gives the momentum transport law

$$q_\alpha^\sigma u^\rho \bar{\nabla}_{[\rho} \bar{\mu}_{\sigma]} = 0. \quad (21)$$

Finally, the orthogonal projection of (19) reduces to

$$K^\rho = \perp_\sigma^\rho \frac{\bar{\nu}^\alpha \bar{\nabla}_\alpha \bar{\mu}^\sigma}{T} = \perp_\sigma^\rho \left( \frac{U}{T} - 1 \right) \dot{u}^\sigma, \quad (22)$$

where the string acceleration  $\dot{u}^\sigma$  stands for

$$\dot{u}^\sigma \equiv u^\alpha \bar{\nabla}_\alpha u^\sigma. \quad (23)$$

As should be clear from (14), the barotropic equation of state breaks Lorentz invariance along the string for  $U \neq T$ . In fact, it describes a wide class of elastic string models [137–139], and as suggested by (20), the scalar current carrying cosmic strings [140]. Conversely, imposing Lorentz

invariance along the worldsheet reduces the equation of state to the trivial form  $U = T$  (see (14)), which is also the relation found for the Abelian Higgs string. This infinitely thin string is the Nambu-Goto (NG) string and does not possess any internal structure [141]. The associated equations of motion are purely geometrical and do not depend on  $U$ . From (22), they reduce to the vanishing of the extrinsic curvature vector, that is,

$$K^\rho = 0, \quad (24)$$

which can be rewritten in a coordinate-dependant way by using(13)

$$K^\mu = \frac{1}{\sqrt{-\gamma}} \partial_a (\sqrt{-\gamma} \gamma^{ab} \partial_b X^\mu) + \Gamma_{\nu\rho}^\mu \gamma^{ab} \partial_a X^\nu \partial_b X^\rho. \quad (25)$$

The connections  $\Gamma_{\nu\rho}^\mu$  are for the background spacetime of metric  $g_{\mu\nu}$  while  $\gamma$  is the determinant of the induced metric. These equations can also be recovered from the usual NG action with an explicit coordinate system [5]

$$S = -U \int d^2\xi \sqrt{-\gamma}. \quad (26)$$

### 3. Cosmological Evolution of Nambu-Goto Strings

The previous section shows that the equations of motion of an isolated string depend on the underlying microscopic model. The type of string is more determinant when two strings interact; cosmic superstrings may form bound states, while non-Abelian vortices may weave new vortices from each of their interaction points. Understanding the cosmological evolution of a string network requires one to solve both the local equations of motion for each string and the outcome of their interactions when they meet. Moreover, the evolution of a system of strings starts from an initial configuration which should describe the network configuration just after its formation. Numerical simulations have been used to overcome some of the above-mentioned difficulties and, up to now, Friedmann-Lemaître-Robertson-Walker (FLRW) network simulations have only been performed with Nambu-Goto strings, Abelian Higgs strings and semi-local strings [75–77, 142–149], up to some variations [80, 81, 131, 150]. As a result, extrapolating the following results to other types of string should be made with caution. On the bright side, (22) suggests that as long as the string acceleration remains small compared to  $T/(U - T)$ , one expects the equations of motion of the string to be close to the NG case (up to the eventual vortons appearance). In the following, we describe the results obtained for NG strings. Some differences exist with the results obtained in the Abelian Higgs string simulations.

Before entering into details, let us summarize two fundamental properties these simulations have revealed. The first is that a cosmic string network avoids cosmological domination by evacuating most of its excess energy through some complex mechanisms, which typically result in transferring energy between the horizon-sized distances

and the smaller length scales. For NG simulations, this is the formation of cosmic string loops whereas in Abelian Higgs simulation boson radiation is involved. The second property is that the influence of the initial conditions is expected to disappear on the length scales of astrophysical interests. A network of cosmic strings relaxes towards a cosmological attractor which depends only on the expansion rate; this is the so-called scaling regime.

**3.1. Dynamics.** The equations of motion for NG strings are the vanishing of the extrinsic curvature vector  $K^\mu = 0$ . In a flat FLRW background,

$$ds^2 = a^2(\eta) \left( -d\eta^2 + \delta_{ij} dx^i dx^j \right), \quad (27)$$

Equation (25) can be simplified with the transverse gauge fixing conditions

$$g_{\mu\nu} \frac{\partial X^\mu}{\partial \tau} \frac{\partial X^\nu}{\partial \sigma} = 0, \quad (28)$$

with the notation  $\tau = \xi^0$  and  $\sigma = \xi^1$  for the timelike and spacelike string coordinates. Such a choice of coordinates reflects the property that an NG string is Lorentz invariant along the worldsheet; there is no physical longitudinal component of the string velocity. In this gauge, the equations of motion read

$$\begin{aligned} \ddot{X}^\mu + \left( \frac{\dot{\varepsilon}}{\varepsilon} + \frac{2}{a} \frac{da}{dX^0} \dot{X}^0 \right) \dot{X}^\mu - \frac{1}{\varepsilon} \left( \frac{\dot{X}^\mu}{\varepsilon} \right)' \\ - \frac{2}{a} \frac{da}{dX^0} \frac{\dot{X}^0}{\varepsilon} \frac{\dot{X}^\mu}{\varepsilon} + \delta_0^\mu \frac{2}{a} \frac{da}{dX^0} \dot{X}^2 = 0, \end{aligned} \quad (29)$$

where a ‘‘dot’’ and a ‘‘prime’’ stand, respectively, for differentiation with respect to  $\tau$  and  $\sigma$ . We have also defined the quantity

$$\varepsilon \equiv \sqrt{-\frac{\dot{X}^2}{\dot{X}^2}}. \quad (30)$$

The conditions in (28) do not completely fix the coordinate degrees of freedom and one can supplement them with the so-called temporal gauge fixing which identifies the timelike coordinate with the background time at the string event:  $\tau = X^0 = \eta$ . In the transverse temporal gauge, (28) reads  $\dot{X} \cdot \dot{X} = 0$ , while (29) simplifies to

$$\ddot{X} + 2\mathcal{H} \left( 1 - \dot{X}^2 \right) - \frac{1}{\varepsilon} \left( \frac{\dot{X}}{\varepsilon} \right)' = 0, \quad \dot{\varepsilon} + 2\mathcal{H}\varepsilon\dot{X}^2 = 0, \quad (31)$$

with

$$\varepsilon = \sqrt{\frac{\dot{X}^2}{1 - \dot{X}^2}}, \quad (32)$$

and  $\mathcal{H}$  is the conformal Hubble parameter. The vector symbols being understood as three-dimensional spatial vectors. Numerically, it is much more convenient to solve an equivalent set of equations found by Bennett and Bouchet [144]. Defining the new vectors  $\vec{p}$  and  $\vec{q}$  as

$$\vec{p}(\tau, u) \equiv \frac{\dot{X}}{\varepsilon} - \dot{X}, \quad \vec{q}(\tau, v) \equiv \frac{\dot{X}}{\varepsilon} + \dot{X}, \quad (33)$$

evaluated at the new coordinates  $u = \int \varepsilon d\sigma - \tau$  and  $v = \int \varepsilon d\sigma + \tau$ , the equations of motion (31) can be recast into

$$\frac{\partial \vec{p}}{\partial \tau} = -\mathcal{H} \left[ \vec{q} - \vec{p} (\vec{p} \cdot \vec{q}) \right], \quad (34)$$

$$\frac{\partial \vec{q}}{\partial \tau} = -\mathcal{H} \left[ \vec{p} - \vec{q} (\vec{p} \cdot \vec{q}) \right],$$

$$\frac{\dot{\varepsilon}}{\varepsilon} = -\mathcal{H} (1 - \vec{p} \cdot \vec{q}). \quad (35)$$

As an illustrative example, these equations have an exact solution in Minkowski space. Taking  $\mathcal{H} = 0$ , one immediately gets  $\varepsilon = 1$  (up to a normalisation constant),  $\vec{p}(u)$  and  $\vec{q}(v)$  are constant over the characteristics  $u = \sigma - \tau$  and  $v = \sigma + \tau$ . Inverting (33) gives

$$\frac{\dot{X}}{\varepsilon}(\tau, \sigma) = \frac{1}{2} \left[ \vec{p}(\sigma + \tau) + \vec{q}(\sigma - \tau) \right], \quad (36)$$

which describes the propagation of left and right moving string deformations at the speed of light. In the FLRW background, these modes are no longer free moving, but interact through the Hubble term (see (34)). Solving these equations gives the  $X^\mu(\tau, \sigma)$  for each strings but does not predict what happens when two strings collide.

**3.2. Collisions.** In the infinitely thin approach, the outcome of an NG string intersection event cannot be predicted. Stress tensor conservation equations require either that the two strings pass through each others, or they intercommute as sketched in Figure 3. The outcome of a string collision process can only be addressed within the framework of a microscopic model. Numerical simulations of interactions have been performed for a variety of models, and in particular for the Abelian Higgs string as represented in Figure 4. In this case, unless the relative string velocity is close to unity [151, 152], or the strings are almost parallel, string intercommutation generically occurs [153, 154]. Let us notice that for type I Abelian strings (i.e., having  $m_b > m_h$ ), bound states of the two strings can also be formed at low velocity [155–157].

The situation is not necessarily the same for the other types of string. As already mentioned, non-Abelian strings can weave new strings from their intersection points while current carrying cosmic strings intercommute for bosonic carriers [158]. In the case of cosmic superstrings, it has been shown that they intercommute with a probability depending on the fundamental string coupling, a quantity

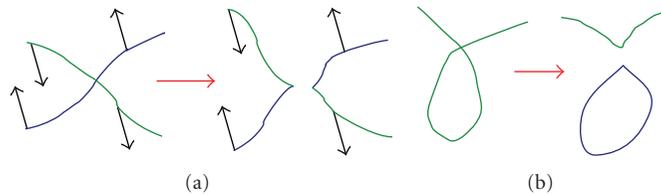


FIGURE 3: Intercommuting strings exchanging their partners (a). On (b), the same mechanism produces a loop from a self-intercommutation. The arrows represent velocity vectors.

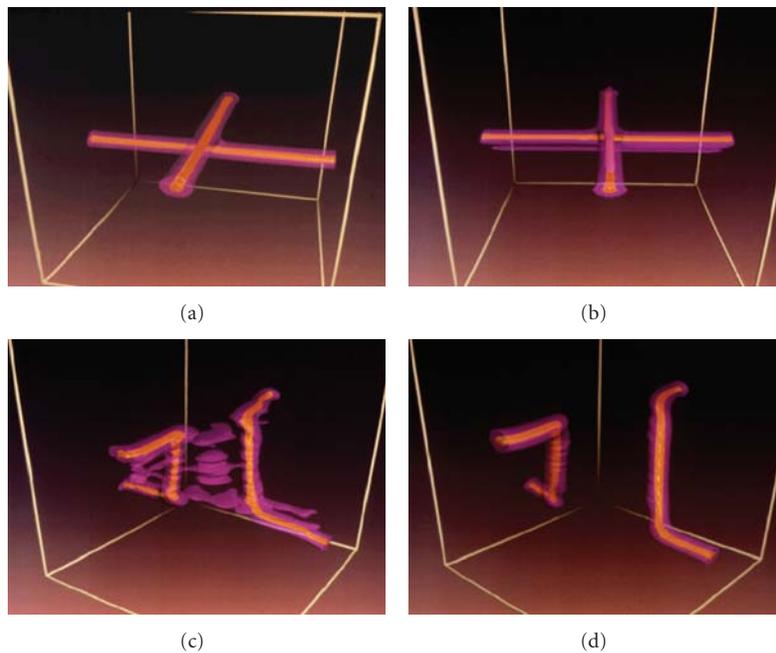


FIGURE 4: Numerical simulations of the intersection between two Abelian Higgs strings [154]. The Higgs and gauge field profiles are represented as the inner orange and outer pink tubes. Generically, Abelian Higgs strings with  $m_h > m_b$  intercommute.

which can be significantly smaller than unity [159]. In the case of  $(p, q)$ -string collisions, Y-junctions can be formed under some kinematic constraints [126, 157, 160–163]. Concerning NG simulations, string collisions are actually implemented through a phenomenological probability  $P_e$  of intercommutation at each intersection event.

**3.3. Initial Conditions.** Solving the cosmological evolution of a NG string network amounts to solving (34) and (35) along each string, finding all of their intersection points and implementing an intercommutation, or not, with the probability  $P_e$ . The network evolution is now uniquely determined once the initial conditions are specified. The simplest way to set initial conditions is through the Vachaspati-Vilenkin (VV) algorithm [164]. Motivated by the Kibble mechanism, one assumes a  $U(1)$  Higgs field to be uncorrelated above a given correlation length  $\ell_c$ . A cosmic string will cross a given plane if one can find a closed loop along which its phase runs from 0 to a multiple of  $2\pi$ . On a discrete three-dimensional lattice, of  $\ell_c$ -spacing, it is sufficient to approximate  $U(1)$  by  $Z_3$  and randomly choose

the phase at each corner from three values 0,  $2\pi/3$  and  $4\pi/3$  to decide if a string crosses the associated face. Other symmetry breaking schemes and lattice can be approximated in a similar way [165–169]. In Figure 5, we have shown the initial string network configuration obtained from the VV algorithm. The string paths have been smoothed by replacing the right angles by circles of radius  $\ell_c$ . The initial network configuration obviously depends on the physical parameter  $\ell_c$ , the network correlation length. In FLRW spacetime, there is, however, another physical parameter which has to be specified; the distance to the horizon  $d_h$ . From those, the initial string energy density is now uniquely determined. A random transverse velocity field can also be added along each string since one does not expect the strings to be initially at rest in any realistic setup. At this point, let us mention that the numerical implementation of the VV initial conditions introduce two additional purely numerical parameters: the size of the periodic box which contains the simulation, usually normalised to unity in comoving coordinates, and the discretisation step required to represent a string, usually given by  $N_{\text{ppcl}}$ , the number of points per correlation length.

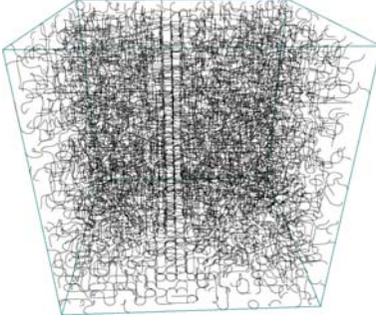


FIGURE 5: Initial string network configuration from the Vachaspati-Vilenkin algorithm. The box is a small lattice of  $25\ell_c$  for illustration purpose only.

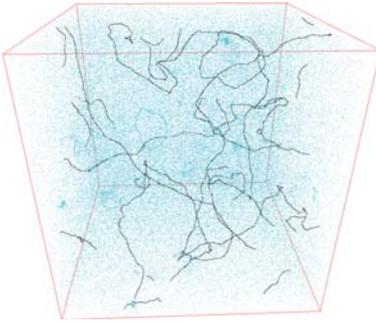


FIGURE 6: String network configuration in the matter era when the horizon fills the whole simulation box. String loops with a length smaller than the distance to the horizon appear in blue.

### 3.4. Cosmological Scaling

**3.4.1. Long Strings.** By switching on the evolution from the initial network, string motion and intersections drastically change the shape of the strings as well as the network aspect (see Figure 6). Naively, without any collisional process, one would expect the string network to dominate the energy density of the universe. In a volume  $V$ , denoting by  $\ell_\infty$  the typical correlation length of the network at a given time (initially  $\ell_\infty = \ell_c$ ), the number of strings should be roughly given by  $V/\ell_\infty^3$ . The resulting energy density should therefore be

$$\rho_\infty \simeq \frac{V}{\ell_\infty^3} \times (U\ell_\infty) \times \frac{1}{V} = \frac{U}{\ell_\infty^2}. \quad (37)$$

Due to cosmological expansion one has  $\ell_\infty \propto a$  and  $\rho_\infty \propto 1/a^2$ . As noted by Kibble, this domination does not occur due to intercommutation processes which allow the formation of loops. In the so-called ‘‘one scale model’’, Kibble [3] assumes that loops of typical size  $\ell_\infty$  are formed at a rate equals to  $\ell_\infty^{-4}$  (for relativistic speeds, one expects one intercommutation per string during the time  $\ell_\infty$ ). As a result, during a time interval  $\delta t$ , the energy density transferred to loops is

$$\delta\rho_{\text{loops}} \simeq \ell_\infty^{-4} \delta t U \ell_\infty. \quad (38)$$

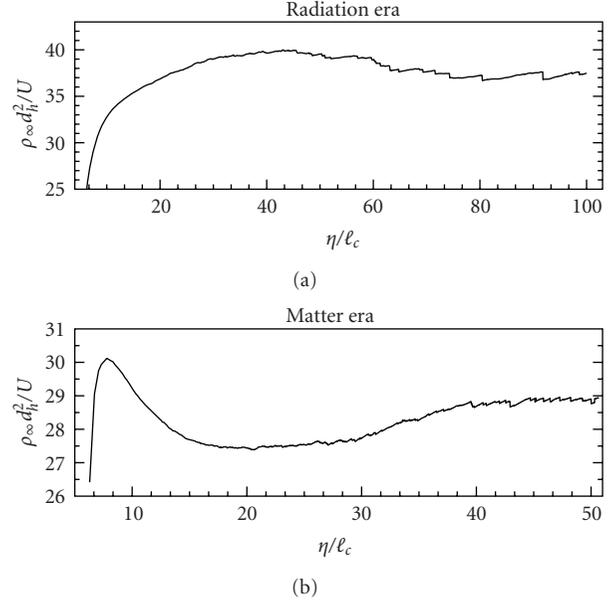


FIGURE 7: Energy density of super-horizon sized strings as a function of the conformal time (in unit of  $\ell_c$ ) in the radiation and matter era. After some relaxation time,  $\rho_\infty \propto 1/d_h^2$ , the proportionality factor is universal.

From (37), the energy density of strings which are not loops verifies

$$\frac{d\rho_\infty}{dt} \simeq -2H\rho_\infty - \frac{\rho_\infty}{\ell_\infty}, \quad (39)$$

where  $\ell_\infty$  is a function of the cosmic time. Defining  $C(t) = \ell_\infty(t)/t$ , this equation can be recast into

$$\frac{1}{C} \frac{dC}{dt} \simeq -\frac{1}{2t} \left( \frac{2+6w}{3+3w} - \frac{1}{C} \right), \quad (40)$$

where the background cosmological fluid sourcing the universe expansion has an equation of state  $P = w\rho$ . The constant solution  $C(t) = (3+3w)/(2+6w)$  is an attractor for which

$$\rho_\infty \propto \frac{U}{t^2} \propto \frac{U}{a^{3(1+w)}}. \quad (41)$$

The energy density associated with strings which are not loops ‘‘scales’’ as matter in the matter era and radiation in the radiation era. Therefore, it is prevented to dominate over the usual cosmological fluids and cannot overclose the universe. Notice however that the total energy density could still dominate the cosmological dynamics if the energy density under the form of loops is not evacuated by some extramechanism. For NG cosmic strings, loops are transformed into radiation due to the emission of gravitational waves [170–173]. Other types of loops may lose energy by different radiative processes, such as particle emission, or even energy leakage into the extra-dimensions in the case of cosmic superstrings [174, 175]. In Figure 6, we have represented an evolved string network at the end of a matter era run.

For FLRW simulations within a fixed comoving box with periodic boundary conditions, one cannot evolve the system indefinitely; at some point, periodic boundaries become causally connected. Usually, one stops the run when the distance to the horizon fills the simulation volume, or more rigorously half of it. In this figure, one sees that only a few super-horizon strings remain (black long strings) whereas the box is also filled with a lot of small loops (blue) and a few larger loops having a size typical of distance between two long strings. The latter are freshly formed Kibble loops whereas the existence of the small ones cannot be explained in the framework of the one scale model [144]. Concerning the energy density associated with the super-horizon strings (also called infinite strings), their evolution in the matter and radiation era have been plotted in Figure 7; they “scale” as expected. From [146], one has

$$\rho_\infty \frac{d_h^2}{U} \Big|_{\text{mat}} = 28.4 \pm 0.9, \quad \rho_\infty \frac{d_h^2}{U} \Big|_{\text{rad}} = 37.8 \pm 1.7. \quad (42)$$

As the behaviour of the energy density associated with long strings suggests, the time evolution drives the string network towards a stable cosmological configuration which does not seem to depend on its initial configuration, at least for the long strings. Figure 6, therefore, displays what a cosmological string network should look like inside a horizon volume, at any time during the matter era. As is clear from Figure 7, the relaxation time required for the energy density of long strings to reach the attractor is small. Concerning the cosmic string loops, their existence and behaviour have been the subject of various claims and analytical works [146, 148, 176–181]. In the following, we present recent results [146, 182] showing that the energy density of loops also reaches a scaling evolution similar to (42).

**3.4.2. Loops.** As previously mentioned, the small loops observed in NG simulation cannot be explained in the framework of the one scale model. These loops find their origin from the self-intercommutation of strings on length scales typical of their small size. The building of a small scale structure on strings is the outcome of the successive intersection events during which new kinks are produced and propagate along the intercommuted segments. Correlations between the kinks induce, from (33), auto- and cross-

correlations between  $\vec{X}(\sigma_1)$  and  $\vec{X}(\sigma_2)$  from which small loops can be produced [176, 179]. In Figure 8, we have plotted the energy density distribution under the form of loops with respect to the conformal time during the radiation and matter era. The simulation performed is one of the largest up to date; the box contains  $100\ell_c$  whereas the redshift simulation range reaches almost two orders of magnitude. The loop energy density distribution  $d\rho_\circ/d\alpha$  is defined such that  $d\rho_\circ(\alpha)$  is the energy density carried by all loops having a physical length  $\ell$  in the range  $\alpha d_h$  to  $(\alpha + d\alpha)d_h$ . In other words, we measure loop size in unit of the horizon length. (This is the relevant physical length scale of the problem.) A logarithmic binning in  $\alpha$  of resolution  $\Delta\alpha/\alpha \simeq 10^{-1}$  has been

used in the range  $[10^{-5}, 10^2]$  to compute these quantities. From this plot, it is clear that after an overproduction regime characterized by the bump of Figure 8, the energy density distribution of loops of given size  $\alpha$  relaxes towards a stationary regime in which it scales as  $1/d_h^2$ . Such an observation implies that, once relaxed, the loop number density distribution is of the form

$$\frac{dn}{d\alpha} = \frac{\mathcal{S}(\alpha)}{\alpha d_h^2}, \quad (43)$$

where the “scaling function”  $\mathcal{S}(\alpha)$  is found to be well fitted by the power laws [146]  $\mathcal{S}(\alpha) = C_\circ \alpha^{-p}$  with

$$\left. \begin{array}{l} p = 1.41 \quad {}^{+0.08}_{-0.07} \\ C_\circ = 0.09 \quad {}^{-0.03}_{+0.03} \end{array} \right|_{\text{mat}}, \quad \left. \begin{array}{l} p = 1.60 \quad {}^{+0.21}_{-0.15} \\ C_\circ = 0.21 \quad {}^{-0.12}_{+0.13} \end{array} \right|_{\text{rad}} \quad (44)$$

for the matter and radiation era, respectively. The loop number density distribution, for the matter era run, has been plotted in Figure 9. As Figure 8 already shows, the loop distribution takes more time to reach the scaling regime for the small loops. The relaxation bump is all the more so high and long than  $\alpha$  is small. In the loop number density distribution, this effect appears as a minimal time decreasing value  $\alpha_{\min}(\eta)$  such that the loop distribution is in scaling at  $\alpha > \alpha_{\min}$ . The redshift range probed by a FLRW string simulation is typically  $\Delta z \simeq 10^2$ , while for strings formed at the GUT energy scales, one expects a  $\Delta z \simeq 10^{18}$  at nucleosynthesis. It is clear that, in the cosmological context, the string network has quite a time to relax; on all of the relevant observable length scales the loop distribution should be in scaling, that is,  $\alpha_{\min} \ll 1$ . Since a power law distribution is scale-free, one concludes that a cosmologically stable string network *does not* exhibit loops of a particular size; this is not surprising since the only length scale involved is the distance to the horizon. These numerical results can be analytically recovered in the framework of the Polchinski-Rocha model [179]. The expected loop number density distribution have been explicitly derived by Rocha in [182] with a predicted power  $p = 1.5$  for the matter era and  $p = 1.8$  in the radiation era. If not due to statistical errors, these small differences may be explained by the existence of additional fractal microstructure along the strings not considered in the analytical approach (see also Section 4). Of course, the previous statements hold provided the other physical effects which are not included in the simulation do not enter the game, as gravitational radiation and gravitational backreaction. The typical length scales at which they should play a role is typically a multiple of  $GU$ , or some positive power of it [183, 184]. As shown in [182], gravitational radiation indeed cures the energy density divergence that one can extrapolate from Figure 9 when  $\alpha \rightarrow 0$ . Let us finally notice that although the long strings are defined by  $\alpha > 1$ , there is also a small population of Kibble loops. Their typical size being the horizon-sized correlation length of the long string network, they can be defined to be those having  $\alpha_\infty < \alpha < 1$ , where

$$\alpha_\infty = \frac{1}{d_h} \left( \frac{U}{\rho_\infty} \right)^{1/2}. \quad (45)$$

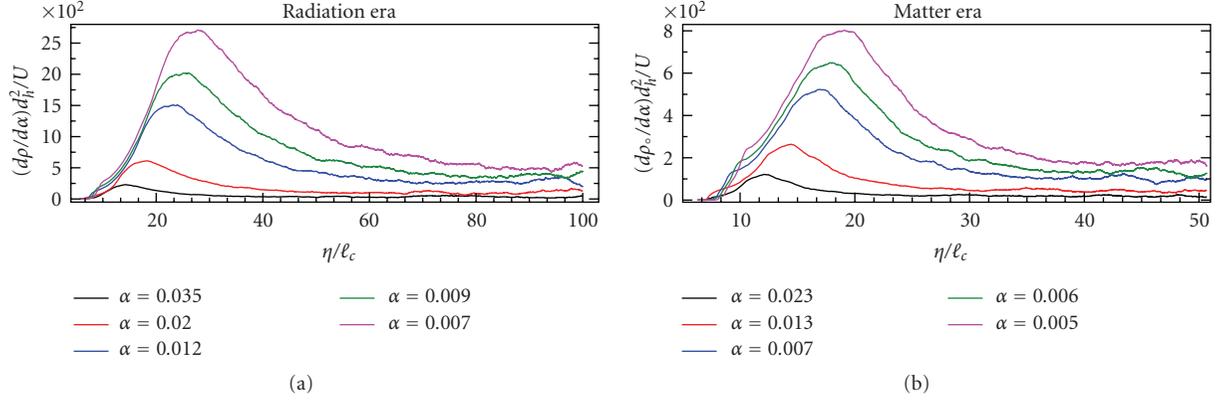


FIGURE 8: Evolution of the energy density distribution of  $\alpha d_h$ -sized loops as a function of the conformal time in the radiation and matter era. After an overproduction regime, the energy density distribution scales as  $1/d_h^2$ , as the long strings.

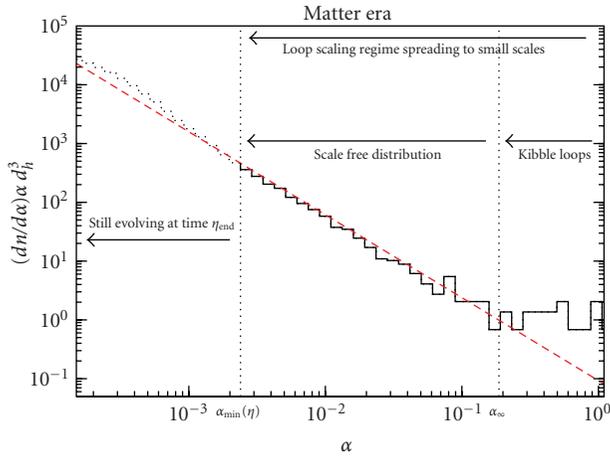


FIGURE 9: Loop number density distribution at the end of a  $(100\ell_c)^3$  matter era numerical simulation. Apart for a few Kibble loops ( $\alpha_\infty < \alpha < 1$ ), all loops of size  $\ell = \alpha d_h$  with  $\alpha > \alpha_{\min}$  follow a scaling regime in which their number density is a power law. Smaller loops are still in the relaxation regime and will enter their scaling regime later (see Figure 10). The red dashed line if the best power law fit of the scaling function  $\mathcal{S}(\alpha)$  is given by (44).

**3.5. Relaxation Towards Scaling: Memory of The Initial Conditions.** Although of less-cosmological interest, the relaxation of the loop energy density distribution towards its scaling regime shows interesting properties which could explain some of the differences observed between Abelian Higgs field simulation and NG simulations. In the left panel of Figure 10, we have plotted the loop energy density distribution for loops smaller than the ones in scaling. At the end of the numerical simulation, these length scales are those having  $\alpha < \alpha_{\min}$ . This plot shows that the formation of the smallest loops is a delayed mechanism which suggests that a cascading process takes place from the initial string network configuration. For  $\alpha \approx 10^{-5}$ , only the increasing part of the relaxation bump appears at the end of the simulation whereas the decaying towards scaling is still visible for the larger  $\alpha$ . On the right panel, we have plotted the loop number density distribution

at the end of the simulation, that is, at the time  $\eta/\ell_c = 50$  in the left panel. The loop number density distribution deviates from the asymptotic scaling distribution on two typical length scales.

Firstly, an overdensity compared to scaling is situated at  $\alpha_c = \ell_c/d_h \propto 1/t$ . In other words, there is an overproduction of loops with a typical size equal to the initial correlation length of the string network. Although one expects the system to retain some memory of the initial conditions during the relaxation, it may appear surprising that, in spite of the expansion of the universe, the physical length scale of these loops remains the same. A physical interpretation is that  $\ell_c \ll d_h$  which suggests that, at those small length scales, the system decouples from the Hubble flow. More quantitatively, this effect can be explained in the context of the three scale models [176]. Under some assumptions, the string small-scale correlations can indeed sustain a constant physical length.

Then, there is the overall peak of the loop number density distribution at  $\alpha_r = \ell_r/d_h \propto 1/t$ . Most of the loops which are not in scaling have this size at the end of the simulation. In fact, one can check that these loops start appearing soon after the beginning of the string evolution. This length scale is, again, at a constant physical length  $\ell_r$  and is associated with a purely numerical effect [146]. As discussed, a numerical string is discretised with  $N_{\text{ppcl}}$  points. The Bennett-Bouchet code at the basis of the simulations presented here uses an adaptive gridding algorithm meaning that loops of any physical size can be formed [144]. The only restriction is that, at a given time, a loop is an object of at least three points. Consequently, when the initial string network starts its evolution, loops smaller than  $\ell_r = \ell_c \times 3/N_{\text{ppcl}}$  cannot be formed. The existence of a finite numerical resolution therefore adds some unwanted initial correlations of length  $\ell_r$ . Notice that this is not a cutoff but indeed an extra-correlation.

As a basic consequence, one should not trust an NG simulation at those length scales. However, the fact that the initial string network violently relaxes towards scaling by emitting loops at the smallest available correlation lengths has still some physical significance [185]. What happens if we

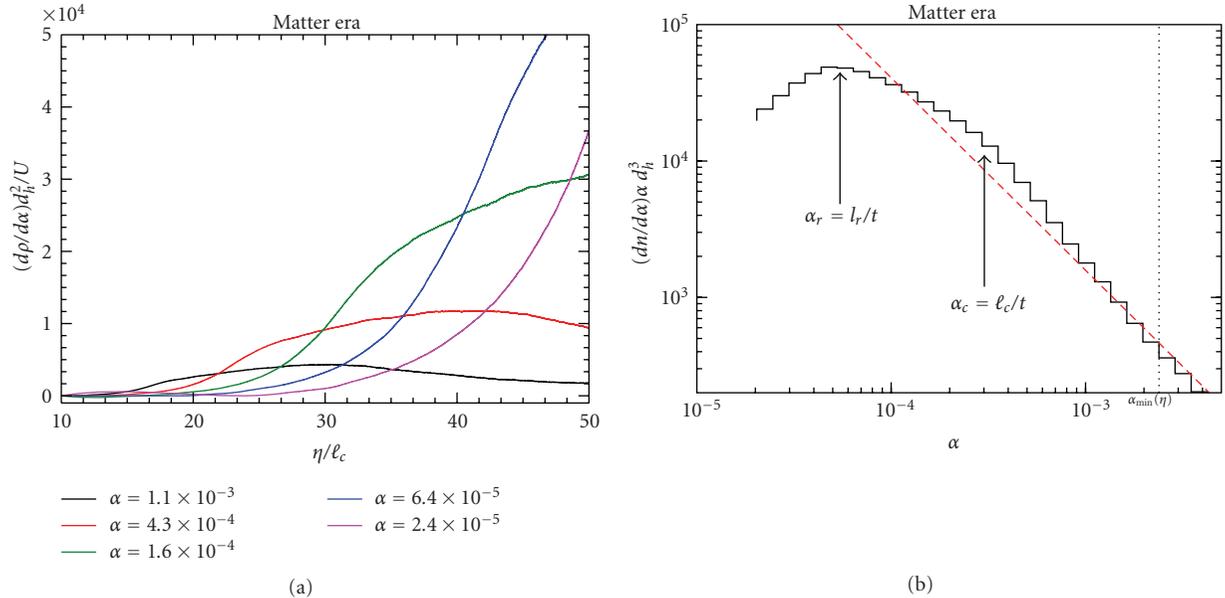


FIGURE 10: Loop energy density distribution with respect of the conformal time for the smallest loops. By the end of the numerical simulation, these loops have not yet reached their scaling behaviour (a). (b) is the corresponding loop number density distribution at the end of the simulation ( $\eta/\ell_c = 50$ ). Compared to the asymptotic scaling power law (red dashed), the loops in the relaxation regime are preferentially produced at fixed physical lengths given by the initial correlations present in the VV network.

increase  $N_{\text{ppcl}} \rightarrow \infty$ ? As discussed in [146], the larger length scales are not affected and only the overall peak is shifted around the new  $\ell_r \rightarrow 0$ . At these length scales, it is clear that using an NG string to describe a network of topological defects would break down and a reasonable assumption is to assume that the network will now relax by losing energy through the relevant physical mechanism available at those  $\ell_r \rightarrow 0$  distances. In Abelian Higgs simulation, most of the network energy is emitted through field radiation, up to the point that almost no loop are observed in the simulations [75, 77]. Abelian simulation suffers from low resolution compared to NG ones and this has been a subject of debate to decide whether or not this could explain the absence of loops [76, 77]. The above NG results clearly support that particle and/or gravitational waves; emission is an important mechanism which certainly dominates the relaxation regime. However, when the scaling regime progressively takes place, from large to small length scales, the loop formation mechanism should become dominant. One may speculate that it is not clearly observed in Abelian simulations due to its delayed appearance, as it is actually the case in NG simulations when we compare it to the formation of  $\ell_r$ -sized loops. (A hint for this is the compatibility of the Abelian Higgs string correlators with the Polchinski-Rocha model [77]; this one explaining the NG loop distribution [182].)

#### 4. Cosmic Microwave Background Anisotropies and Non-Gaussianities

At this point, numerical simulations of cosmic string evolution give us the means to derive observable predictions. As should be clear from the previous section, some structures

in the numerical simulations are not supposed to be present after a realistic expansion factor of  $10^{18}$ ; these are the structures still in the relaxation regime and such that  $\alpha < \alpha_{\text{min}}$ , at any simulation time. In the following, we will denote by “systematic errors”, the uncertainties inherent to the presence of such nonscaling structures when deriving observable predictions from NG numerical simulations.

*4.1. Unequal Time Correlators.* The first method used to derive CMB anisotropies has been introduced in [186] and applied in [79, 186–191] for global topological defects and recently in [192, 193] for the Abelian strings.

Cosmic strings are active sources of gravitational perturbations [41, 195] which means that the equations of motion of their induced linear perturbation is of the form

$$\mathcal{D}\mathcal{P} = \mathcal{S}, \quad (46)$$

where  $\mathcal{D}$  is a time differential operator,  $\mathcal{P}$  the perturbation in the energy density, or velocity, and so forth, which is directly related to the CMB temperature anisotropies. Here  $\mathcal{S}$  denotes the source terms, that is, the string stress tensor. From the Green’s function  $\mathcal{G}$  of this equation, one gets, today (at  $\eta_0$ ) and in Fourier space

$$\mathcal{P}(k, \eta_0) \propto \int \mathcal{G}_k(\eta) \mathcal{S}(\eta, k) d\eta. \quad (47)$$

The two-point correlator reads

$$\begin{aligned} & \langle \mathcal{P}^\dagger(\eta_0, k) \mathcal{P}(\eta_0, k) \rangle \\ & \propto \int \int^{\eta_0} \mathcal{G}_k^\dagger(\eta') \mathcal{G}_k(\eta) \langle \mathcal{S}^\dagger(\eta', k) \mathcal{S}(\eta, k) \rangle d\eta d\eta', \end{aligned} \quad (48)$$

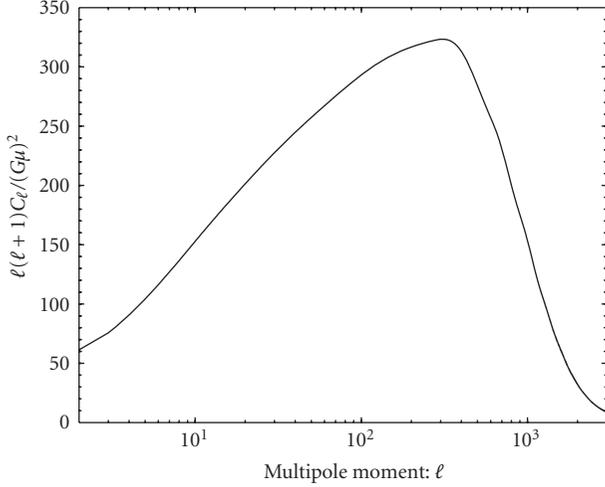


FIGURE 11: CMB temperature power spectrum induced by a network of Abelian cosmic strings and derived from the unequal time correlators method. Figure extracted from [194].

and its determination requires a full-time knowledge of the source term  $\mathcal{S}(\eta, k)$  for each mode. Since it is impossible to carry out a simulation over the whole cosmological history, the scaling properties of the cosmic string network can be used to analytically extrapolate the source terms over the required ranges. As shown in [189], as long as the cosmic string network is in a scaling regime, the source terms are the stress tensor components and assume the form

$$\langle T_{\mu\nu}(k, \eta) T_{\rho\sigma}^*(k, \eta') \rangle \propto \frac{1}{\sqrt{\eta\eta'}} f_{\mu\nu\rho\sigma} \left( k\sqrt{\eta\eta'}, \frac{\eta}{\eta'} \right). \quad (49)$$

Numerical simulations are actually used to determine the scaling functions  $f_{\mu\nu\rho\sigma}$ . In Figure 11, we have represented the temperature anisotropies derived in [194] using such a method from Abelian Higgs string simulations. The current CMB constraint on  $GU$  comes from this power spectrum: at two-sigma,  $GU < 7 \times 10^{-7}$  [196].

**4.2. Simulated Small Angle Maps.** The previous constraint typically corresponds to a string contribution which cannot exceed 10% at the multipole moment  $\ell = 10$ . On current observable angular scales, cosmic strings may only be a subdominant fraction of the overall CMB anisotropies. However, string-induced perturbations being non-Gaussian, as opposed to inflationary perturbations of quantum origin, one can go further than deriving the two-point function. Notice that, in principle, the unequal time correlator approach could be used to extrapolate the three- and higher  $n$ -point function by using the scaling properties of the string network. Another approach is to produce simulated maps of string-induced CMB anisotropies. Again, we face the problem of the small redshift range probed by the numerical simulations. By putting an observer inside the numerical simulation, such maps can only include stringy effects up to a finite redshift, typically  $z \simeq 10^2$ . The CMB anisotropies computed in this way are therefore only accurate on large

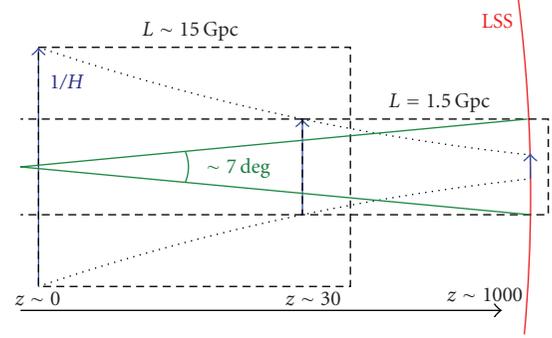


FIGURE 12: Stacking of two numerical simulations to compute the small angles CMB anisotropies induced by NG strings [197]. The dashed rectangles represent the redshift and angular extension of two numerical simulations used to evaluate (52). The first starts at last scattering and stops at  $z \simeq 30$  while the second starts at  $z \simeq 30$  and ends now, up to a small overlapping to ensure relaxation towards scaling.

angular scales but can produce full sky maps [198–201]. This limitation can be avoided by stacking maps from different redshifts, an approach outlined in [37] and applied in [197, 202].

Simulations with the observer outside of the numerical box are not well suited for a full-sky map reconstruction, but are perfectly designed for the small angular scales. The reason being that cosmic strings are incessantly sourcing the CMB fluctuations since the last scattering surface, and contrary to the perturbations of inflationary origin, this part cannot be affected by Silk damping. Therefore, at small angular scales, one expects the strings' signature in the CMB temperature fluctuations to be dominated by their integrated Sachs-Wolfe (ISW) effect from the last scattering surface [202]. In the temporal gauge ( $\tau = X^0 = \eta$ ), the NG stress tensor derived from (26) reads

$$T^{\mu\nu} = \frac{U}{\sqrt{-g}} \int d\sigma \left( \varepsilon \dot{X}^\mu \dot{X}^\nu - \frac{1}{\varepsilon} \dot{X}'^\mu \dot{X}'^\nu \right) \delta^3(\vec{x} - \vec{X}). \quad (50)$$

In the flat sky approximation, well suited for angles typically smaller than the Hubble angular size at the epoch of interest, Hindmarsh has shown that the ISW temperature anisotropies induced by NG strings can be simplified to [203, 204]

$$\hat{\Theta}_{\mathbf{k}} \simeq \frac{8\pi i GU}{k^2} \int_{\vec{X} \cap \vec{x}_y} (\mathbf{u} \cdot \mathbf{k}) e^{-i\mathbf{k} \cdot \mathbf{X}} \varepsilon d\sigma, \quad (51)$$

where  $\Theta(\mathbf{x}) \equiv \delta T(\mathbf{x})/T_{\text{CMB}}$ . The wave vector  $\mathbf{k}$  denotes the *transverse* component of the three-dimensional vector  $\vec{k}$  with respect to the line of sight  $\hat{n}$ , whereas, in the temporal gauge,  $\vec{u}$  encodes the string stress tensor distortions of the photon temperature and reads

$$\vec{u} = \frac{\dot{\vec{X}}}{1 + \hat{n} \cdot \dot{\vec{X}}}. \quad (52)$$

As can be seen in (51), only the strings that intercept the photon path  $\vec{x}_\gamma$  can imprint their signature in the CMB temperature fluctuations. The previous expression is nothing but the Gott-Kaiser-Stebbins effect in the temporal gauge [54, 203, 205]. As a result, the knowledge of  $\mathbf{u}$ , and therefore of the string trajectories  $\mathbf{X}$ , is only required on our past light cone. In the context of string numerical simulations, the trajectories of all strings are computed during all of the numerical simulation time. Therefore, to compute  $\mathbf{u}$ , one only needs to determine which parts of the string network intercept our past light cone and at what time.

In order to be able to generate a significant amount of maps using such a method, it is more convenient to stack two medium sized  $(50\ell_c)^3$  simulations along the lines sketched in Figure 12. The first one starts at the last scattering surface and ends at a redshift fixed by the maximum expansion factor achievable in the numerical box. For the simulations we performed, initially  $d_h \simeq 0.185$  (in unit of the fixed comoving simulation volume), which corresponds to 1.7 Gpc and a field of view of  $\theta_{\text{fov}} \simeq 7.2^\circ$  (for the current fiducial cosmological parameters [44]). Such a run ends after a 30-fold increase in expansion factor, corresponding to a redshift  $z \simeq 36$ . We then propagate the photons perturbed by the first run into a second numerical simulation of the same size but starting at  $z_1 \simeq 36$ . For another 30-fold increase in expansion factor, this run ends at  $z \simeq 0.3$ . As can be seen in Figure 12, the second simulation represents a much larger real volume than the first one and therefore subtends a greater angle in the sky. As a result, only the subpart of the second run that matches the angle subtended by the first simulation is actually used. As we will see later on, the CMB temperature maps are weakly sensitive to the string network at low redshifts, simply because there are almost no strings intercepting our past light cone in a recent past, which makes this technique perfectly acceptable. In practice, each of these numerical simulations is started before the redshifts mentioned, in order to give the cosmic string network enough time to relax toward its stable cosmological configuration. As discussed in the previous section, one has to make sure that the structures (strings and loops) we are interested in have indeed reached their scaling behavior during the numerical runs. We switch on the photon propagation inside the runs only after making sure all the large structures (infinite strings and loops) are in their scaling regime. This can be checked by monitoring the evolution of the energy density distributions, and we have chosen to start the photons' propagation when all loops larger than a third of the horizon size are in scaling. This cutoff is then dynamically pushed to smaller values to include all the loops entering the scaling regime at later times. The cutoff time dependence is simply the function  $\alpha_{\text{min}}(\eta)$  and can be deduced from the loop distribution relaxation times derived in Section 3.4.2. The resulting CMB temperature map is displayed in the left panel of Figure 13 whereas the right panel shows the string paths projected onto our past light cone. Again, these maps are only representative at small angular scales. On larger angles, they represent only the ISW contribution to the overall string anisotropies; for instance, Doppler contributions coming from photon decoupling at

last scattering are dominant around  $\ell = 300 - 400$  (see Figure 11).

The discussion on systematic effects coming from the presence of loops not yet in scaling can be found in [46]. In fact, they have only a small effect. The physical reason being that, due to scaling, the long strings are still the main source of CMB anisotropies even at (reasonably) small angles. Indeed, there are always roughly ten strings per Hubble volume at each time, which means that a patch of  $0.8^\circ$  is at least crossed by the ten long strings being there at last scattering, plus a few others from lower redshifts.

*4.3. Skewness and Kurtosis.* The most basic statistical test that can be performed from a set of small angle CMB maps is to plot the one-point function of the temperature anisotropies. As can be seen in Figure 14, the temperature anisotropies induced by cosmic strings are clearly non-Gaussian. From a set of 300 independent CMB maps, one finds the mean sample skewness to be negative

$$g_1 \equiv \left\langle \frac{\overline{(\Theta - \bar{\Theta})^3}}{\sigma^3} \right\rangle \simeq -0.22 \pm 0.12, \quad (53)$$

where the brackets stand for the mean over different realisations while the bar denotes averaging on each map. The variance itself averages to

$$\sigma^2 \equiv \left\langle \overline{(\Theta - \bar{\Theta})^2} \right\rangle \simeq (150.7 \pm 18)(GU)^2. \quad (54)$$

The quoted errors are statistical and refer to the square root of the variance between the different realisations. Similarly, the mean kurtosis averages to

$$g_2 \equiv \left\langle \frac{\overline{(\Theta - \bar{\Theta})^4}}{\sigma^4} \right\rangle - 3 \simeq 0.69 \pm 0.29. \quad (55)$$

An analytical approach extending these results to cosmic superstrings can be found in [206]. A simple way to look for strings is to search for large (but rare) temperature fluctuations. Deviations from Gaussianity start to be significant, let us say by a factor of two, only in the tails when the probability distribution becomes typically lower than  $10^{-6}$ .

*4.4. Real Space Methods.* Strings induce step-like discontinuities in the CMB anisotropies and various methods have been designed to probe the non-Gaussianities associated with them.

Multifractal analysis [207] has the advantage of being directly applicable to the time-ordered data retrieved when a CMB telescope scans the sky. As opposed to the fractal dimension of a set which measures how sparse it is, the multifractal spectrum of a measure defined over a set gives how many and which fractal dimensions there are. In the context of cosmic strings, this method has been applied in [208] on one dimensional scan of maps similar to the one in Figure 13, the measure being defined by

$$\mu(i) = [\Theta(i) - \Theta(i+1)]^2, \quad (56)$$

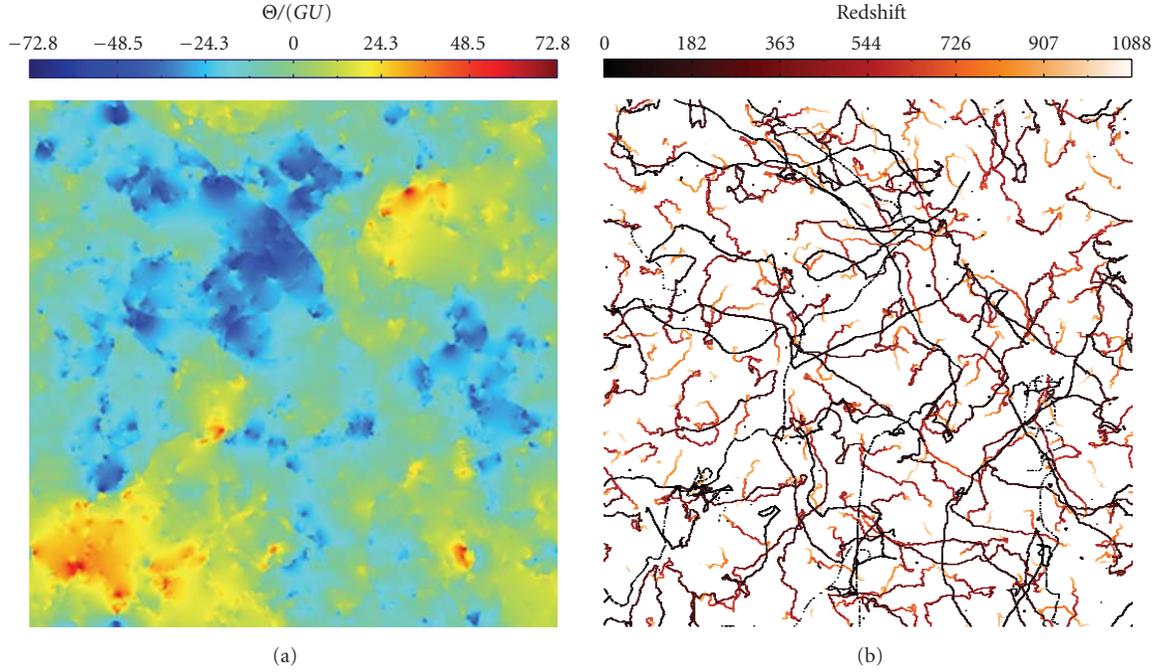


FIGURE 13: String-induced CMB temperature fluctuations on a  $7.2^\circ$  field of view [46]. Because of their cosmological scaling, most of the long strings intercept our past light cone close to the last scattering surface. As can be seen in the right image, the edges in the temperature patterns can be identified to strings intercepting our past light cone. Note that active regions corresponding to string intersection and loop formation events lead to the bright spots in these maps.

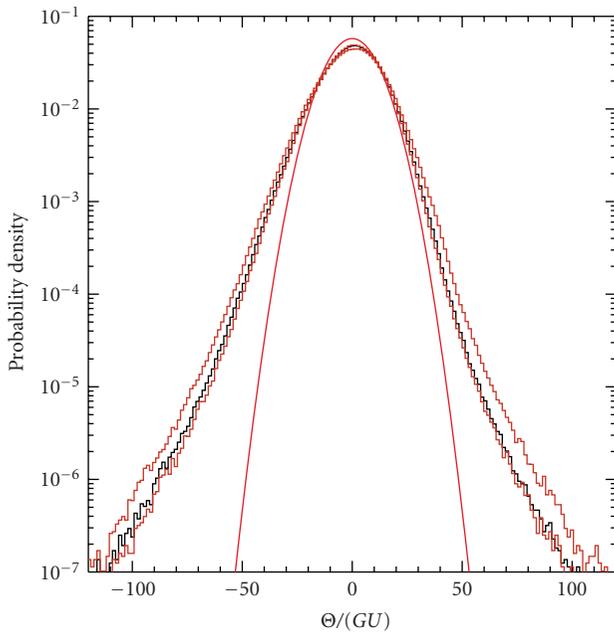


FIGURE 14: The probability distribution function of CMB temperature fluctuations induced by NG cosmic strings. The orange curves quantify systematic errors coming from the string simulations by including nonscaling loops, or by removing all loops. Deviations from Gaussianity are clearly apparent in the tails of the distribution, as well as from the negative skewness.

where the integer  $i$  labels a point along the scan. The multifractal properties of this measure have been shown to be distinctive enough to detect strings, compared to a Gaussian signal, but only when the detector resolution is sufficiently good. One may wonder, under multifractality, how a non-Gaussian string pattern could be distinguished from other non-Gaussian sources. In fact, (56) consists in taking the gradient of the induced CMB fluctuations along the scan. Step-like discontinuities, passed over a gradient filter, become one-dimensional delta functions, and this is a definite string feature that can only be altered by the beam experiment. Denoting by  $\alpha$  and  $\beta$  the horizontal and vertical angular coordinates, the gradient magnitude  $|\nabla\Theta|$  of the temperature anisotropies is defined by

$$|\nabla\Theta| \equiv \sqrt{\left(\frac{d\Theta}{d\alpha}\right)^2 + \left(\frac{d\Theta}{d\beta}\right)^2}. \quad (57)$$

This definition makes it clear that for a finite temperature step, let us say  $\Theta(\alpha, \beta) = \Theta_0 H(\alpha - \alpha_0)$ ,  $H$  being the Heaviside function, the resulting gradient magnitude is a Dirac distribution at the string location. In Figure 15, we have plotted the gradient magnitude of the temperature maps of Figure 13 as well as its convolved version with a Gaussian beam typical of the Planck satellite at 217 GHz. With a finite resolution beam, the discontinuities are now smoothed. Real space methods applied to string are therefore strongly sensitive to the angular resolution. Let us mention

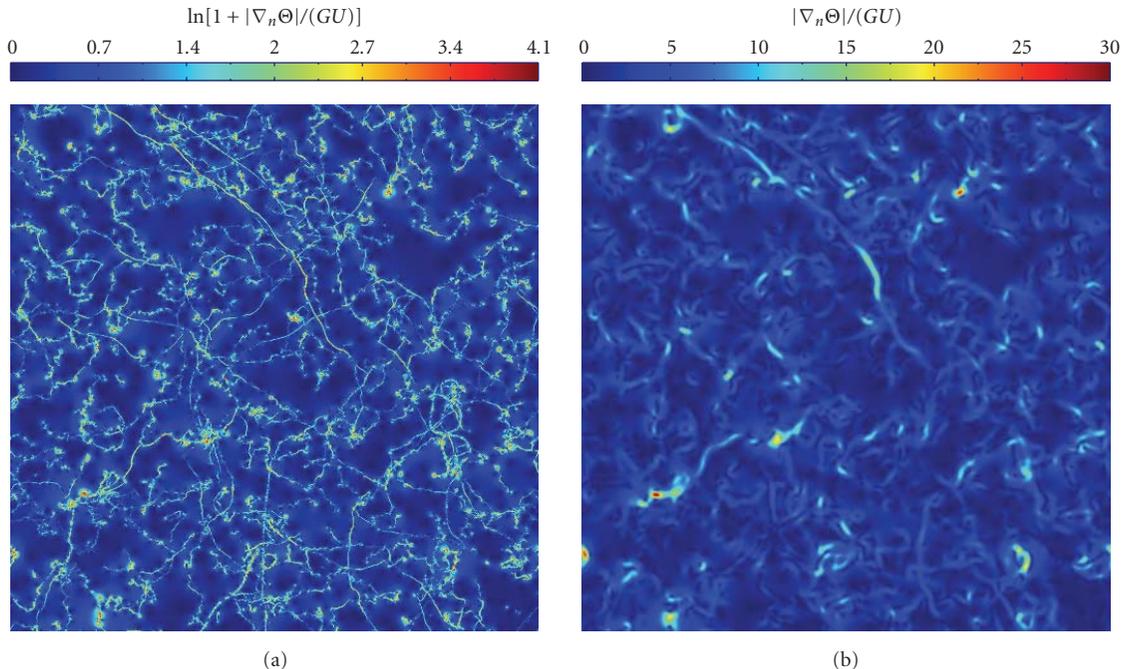


FIGURE 15: Normalized gradient magnitude of the string-induced temperature anisotropies shown in Figure 13(a). A logarithmic scale has been used to enhance the contrast by preventing the bright spots from saturating the color scale. The right panel is the gradient magnitude obtained after convolution by a Planck-like Gaussian beam of resolution  $5'$ . Notice that the color scale is back to linear, most of the bright spots being now smoothed by the beam.

that wavelet analysis methods have been also explored in this context [209, 210] or to produce cleaner maps [211].

Directional gradients, obtained by variations with respect to either  $\alpha$  or  $\beta$ , have been discussed in [205] in the context of Minkowski functionals. They are again found to provide a more distinctive non-Gaussian behaviour than the original temperature map for the contour length and genus.

*4.5. Temperature Power Spectrum.* Moving to Fourier space, the small angle CMB maps also permit a determination of the power spectrum at large multipoles. In Figure 16, we have plotted its mean value over the different maps as well as the one-sigma statistical error around the mean. The overall power at  $\ell = 1000$  is [46]

$$\frac{\ell(\ell+1)C_\ell}{2\pi} \Big|_{\ell=1000} \simeq 14 (GU)^2, \quad (58)$$

which is close to the value obtained in Abelian Higgs field simulation (see Figure 11). This is not so surprising since the long strings in both NG and Abelian Higgs simulation have a similar scaling evolution, and as explained above, long strings are the main sources of CMB anisotropies even at the small angles. The power law tail in Figure 16 is the direct consequence of the presence of strings at all times since the last scattering surface: one finds for  $\ell \gg 1$  [46]

$$\ell(\ell+1)C_\ell \propto \ell^{-p}, \quad \text{with } p = 0.889_{-0.090}^{+0.001}, \quad (59)$$

where only the systematic errors have been reported. Such a power law shows that cosmic strings have to become the

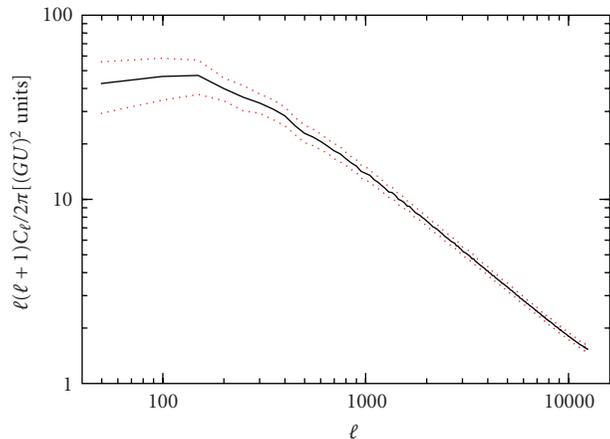


FIGURE 16: Mean angular power spectrum of the string-induced CMB anisotropies at small angular scales and its one sigma statistical errors (averaged over 300 maps). At “small” multipoles  $\ell < 500$ , Doppler contributions from the last scattering surface are expected to be significant and this plot gives only the ISW component [46].

dominant primary source of CMB anisotropies at the small angular scales, the fluctuations of inflationary origin being killed by Silk damping at those multipoles. In Figure 17, we have plotted the respective contributions of strings and adiabatic anisotropies of inflationary origin. The cosmological parameters have been set to their fiducial values in the Lambda-Cold-Dark-Matter (LCDM) model and the string

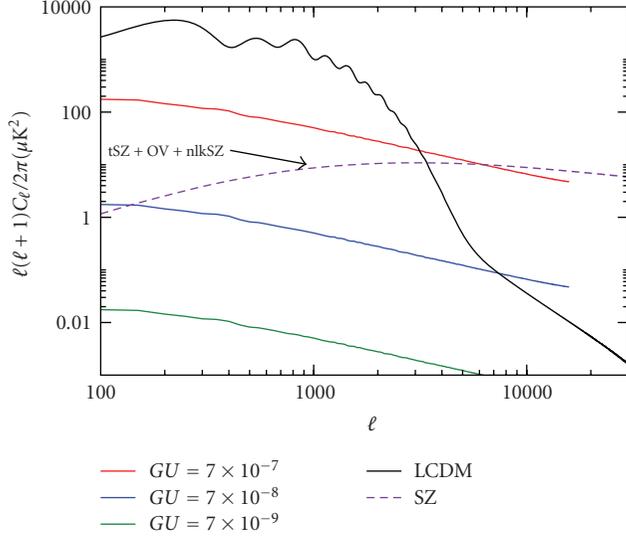


FIGURE 17: CMB temperature anisotropies from various sources, compared to the expected string contribution. Due to the Silk damping of primordial perturbations, string-induced anisotropies always become dominant for the large multipoles. However, an unresolved SZ effect may compromise such a clean signature [46].

energy density  $U$  is compatible with the current bounds. For the current upper limit on  $GU = 7 \times 10^{-7}$ , CMB anisotropies should become dominated by cosmic strings at  $\ell \gtrsim 3000$ . An unresolved Sunyaev-Zel'dovich (SZ) component may, however, compromise such a signal. Nevertheless, string induced anisotropies do not depend on the signal frequency whereas the SZ does, and one may hope to disentangle both [46].

**4.6. Hindmarsh Approximation.** As shown by Hindmarsh in [203], the power law behaviour of the NG string power spectrum at small angles can be analytically recovered. In Fourier space, the power spectrum of the string-induced CMB anisotropies  $\Theta$  is defined by

$$\langle \hat{\Theta}_{\mathbf{k}_1} \hat{\Theta}_{\mathbf{k}_2} \rangle = P(k_1) (2\pi)^2 \delta(\mathbf{k}_1 + \mathbf{k}_2), \quad (60)$$

the expression of  $\hat{\Theta}_{\mathbf{k}}$  being given by (51). Remembering that the string positions and velocity vectors have to be evaluated on the past light cone, it is more convenient to use the so-called light cone gauge. Instead of identifying the timelike worldsheet coordinate  $\tau$  with the background time at the string event, one chooses instead to identify  $\tau = X^+ \equiv X^0 + X^3$ . In this gauge, (51) simplifies to

$$-k^2 \hat{\Theta}_{\mathbf{k}} = i\epsilon k_A \int d\sigma \dot{X}^A(\sigma) e^{i\mathbf{k} \cdot \mathbf{X}(\sigma)}, \quad (61)$$

where we have defined

$$\epsilon = 8\pi GU, \quad (62)$$

and where the capital indices are two-dimensional. The time parameter  $\tau$  then labels the intersections of a set of

null hyperplanes with the worldsheet. For our problem, all quantities have to be evaluated at  $\tau = x^+ = \eta + z$ . In a field of view of formal area  $\mathcal{A} = (2\pi)^2 \delta(0)$ , one can express the power spectrum as

$$P(k) = \epsilon^2 \frac{k_A k_B}{\mathcal{A} k^4} \int d\sigma d\sigma' \langle \dot{X}^A(\sigma) \dot{X}^B(\sigma') e^{i\mathbf{k} \cdot [\mathbf{X}(\sigma) - \mathbf{X}(\sigma')]} \rangle. \quad (63)$$

Adding the assumptions that both  $\dot{X}^A$  and  $\dot{X}^B$  obey Gaussian statistics, all of the correlation functions of  $\hat{\Theta}_{\mathbf{k}}$  can now be written in terms of two-point functions only. Using the same notation as in [203], the nonvanishing two-point functions are

$$\langle \dot{X}^A(\sigma) \dot{X}^B(\sigma') \rangle = \frac{1}{2} \delta^{AB} V(\sigma - \sigma'), \quad (64)$$

$$\langle \dot{X}^A(\sigma) \dot{X}^B(\sigma') \rangle = \frac{1}{2} \delta^{AB} M(\sigma - \sigma'), \quad (65)$$

$$\langle \dot{X}^A(\sigma) \dot{X}^B(\sigma') \rangle = \frac{1}{2} \delta^{AB} T(\sigma - \sigma'), \quad (66)$$

as well as the quantities

$$\begin{aligned} \Gamma(\sigma - \sigma') &\equiv \langle [\mathbf{X}(\sigma) - \mathbf{X}(\sigma')]^2 \rangle \\ &= \int_{\sigma'}^{\sigma} d\sigma_1 \int_{\sigma'}^{\sigma} d\sigma_2 T(\sigma_1 - \sigma_2), \end{aligned} \quad (67)$$

$$\begin{aligned} \Pi(\sigma - \sigma') &\equiv \langle [\mathbf{X}(\sigma) - \mathbf{X}(\sigma')] \cdot \dot{\mathbf{X}}(\sigma') \rangle \\ &= \int_{\sigma'}^{\sigma} d\sigma_1 M(\sigma_1 - \sigma'). \end{aligned} \quad (68)$$

The leading terms are given by [203, 212]

$$V(\sigma) \rightarrow \begin{cases} \bar{v}^2, & \sigma \rightarrow 0 \\ 0, & \sigma \rightarrow \infty \end{cases}, \quad \Gamma(\sigma) \rightarrow \begin{cases} \bar{t}^2 \sigma^2, & \sigma \rightarrow 0, \\ \hat{\xi} \sigma, & \sigma \rightarrow \infty, \end{cases} \quad (69)$$

$$\Pi(\sigma) \rightarrow \begin{cases} \frac{1}{2} \frac{c_0}{\hat{\xi}} \sigma^2, & \sigma \rightarrow 0, \\ 0, & \sigma \rightarrow \infty, \end{cases} \quad (70)$$

where we have defined

$$\hat{\xi} = \Gamma'(\infty), \quad \bar{v}^2 = \langle \dot{\mathbf{X}}^2 \rangle, \quad \bar{t}^2 = \langle \dot{X}'^2 \rangle, \quad c_0 = \hat{\xi} \langle \dot{\mathbf{X}}'' \cdot \dot{\mathbf{X}} \rangle. \quad (71)$$

The correlation length  $\hat{\xi}$  is the projected correlation length on the past light cone,  $\bar{t}^2$  is the mean square projected tangent vector,  $\bar{v}^2$  is the mean square projected velocity and  $c_0$  the correlation between projected velocity and curvature. From these assumptions, (63) reduces to

$$\begin{aligned} P(k) &= \frac{\epsilon^2}{2\mathcal{A}k^2} \int d\sigma d\sigma' \left[ V(\sigma - \sigma') + \frac{1}{2} k^2 \Pi^2(\sigma - \sigma') \right] e^{-k^2 \Gamma(\sigma - \sigma')/4}. \end{aligned} \quad (72)$$

When  $k\hat{\xi}$  gets large, the terms involving the mixed correlator  $M$  can be shown to be subdominant and only the first term remains

$$P(k) = \frac{\epsilon^2}{4\mathcal{A}k^2} \int d\sigma_+ d\sigma_- V(\sigma_-) e^{-k^2\Gamma(\sigma_-)/4}, \quad (73)$$

where  $\sigma_{\pm} = \sigma \pm \sigma'$ . Denoting by  $L$  the total transverse light-cone gauge length of string in the box of area  $\mathcal{A}$ , one gets

$$k^2 P(k) \simeq \epsilon^2 \sqrt{\pi} \frac{L\hat{\xi} \bar{v}^2}{\mathcal{A} \bar{t}} \frac{1}{(k\hat{\xi})}. \quad (74)$$

At small angles, the wave number  $k^2 \simeq \ell(\ell + 1)$  and (74) predicts that  $\ell(\ell + 1)C_{\ell} \propto \ell^{-1}$ . The small difference with (59) is suggestive of a cloud of zero-dimensional objects along the string worldsheet which may be the signature of small loop production in the NG numerical simulations. Let us stress that (74) is not ‘‘primordial’’ but directly approximate the observed angular power spectrum of the CMB temperature anisotropies.

#### 4.7. Bispectrum

**4.7.1. Analytical Approach.** The success of Hindmarsh approximation to describe the small angular CMB anisotropies power spectrum suggests it can be applied to higher  $n$ -point functions. In [212], this method was used to derive the bispectrum defined from the three points function by

$$\langle \hat{\Theta}_{\mathbf{k}_1} \hat{\Theta}_{\mathbf{k}_2} \hat{\Theta}_{\mathbf{k}_3} \rangle = B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) (2\pi)^2 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3). \quad (75)$$

Plugging (61) into the previous expression gives

$$B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = i\epsilon^3 \frac{1}{\mathcal{A}} \frac{k_{1A} k_{2B} k_{3C}}{k_1^2 k_2^2 k_3^2} \int d\sigma_1 d\sigma_2 d\sigma_3 \langle \dot{X}_1^A \dot{X}_2^B \dot{X}_3^C e^{i\delta^{ab} \mathbf{k}_a \cdot \mathbf{X}_b} \rangle, \quad (76)$$

with  $\dot{X}_a^A = \dot{X}^A(\sigma_a)$ ,  $a, b \in \{1, 2, 3\}$ , and  $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$ . With the Gaussian assumption, the ensemble average of the string observables is lengthy but straightforward and the final result reads [212]

$$B(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) = -\epsilon^3 \pi c_0 \frac{\bar{v}^2}{\bar{t}^4} \frac{L\hat{\xi}}{\mathcal{A}} \frac{1}{\hat{\xi}^2} \frac{1}{k_1^2 k_2^2 k_3^2} \left[ \frac{k_1^4 \kappa_{23} + k_2^4 \kappa_{31} + k_3^4 \kappa_{12}}{(\kappa_{23} \kappa_{31} + \kappa_{12} \kappa_{31} + \kappa_{12} \kappa_{23})^{3/2}} \right]. \quad (77)$$

The quantities  $\kappa_{ab}$  are shorthand for the scalar products  $\kappa_{ab} \equiv -\mathbf{k}_a \cdot \mathbf{k}_b$ . In the same way as for the power spectrum, this expression directly gives the bispectrum of the CMB temperature anisotropies. Its overall dependence varies as  $1/k^6$ . Its sign depends on the sign of  $c_0$  defined in (71), and contrary to what one could naively expect  $c_0 \neq 0$ : the projected string velocity and curvature vectors are correlated. This can be shown by starting again from the equations of

motion (29), but this time, in the light cone gauge. The equation of motion for  $X^+$  gives

$$\frac{\dot{\xi}}{\epsilon} + 2\mathcal{H}(\dot{X}^0 + \dot{X}^2) = 0, \quad (78)$$

whereas the equation for the transverse components is

$$\ddot{\mathbf{X}} + 2\mathcal{H} \frac{1}{\epsilon^2} (\dot{\mathbf{X}}^2) \dot{\mathbf{X}} - \frac{1}{\epsilon} \frac{\partial}{\partial \sigma} \left( \frac{1}{\epsilon} \frac{\partial \mathbf{X}}{\partial \sigma} \right) - 2\mathcal{H} \frac{1}{\epsilon^2} (\dot{\mathbf{X}} \cdot \dot{\mathbf{X}}) \dot{\mathbf{X}} = 0. \quad (79)$$

In a FLRW background, assuming that  $\langle \dot{\mathbf{X}}^2 \rangle$  is constant, and neglecting higher-order correlations between  $\mathcal{H}$ ,  $\dot{\mathbf{X}}$ , and  $\dot{\mathbf{X}}$ , we find

$$\left\langle \frac{\partial^2 \mathbf{X}}{\partial s^2} \cdot \dot{\mathbf{X}} \right\rangle = 2\overline{\mathcal{H}} \left\langle \left( \frac{\partial \mathbf{X}}{\partial s} \right)^2 \dot{\mathbf{X}}^2 \right\rangle - 2\overline{\mathcal{H}} \left\langle \left( \dot{\mathbf{X}} \cdot \frac{\partial \mathbf{X}}{\partial s} \right)^2 \right\rangle, \quad (80)$$

where we have defined  $ds = \epsilon d\sigma$ , and where  $\overline{\mathcal{H}}$  is the averaged conformal Hubble parameter. Still assuming that the ensemble is approximately Gaussian in  $\dot{\mathbf{X}}$  and  $\dot{\mathbf{X}}/\epsilon$ , the right hand side reduces to

$$\left\langle \frac{\partial^2 \mathbf{X}}{\partial s^2} \cdot \dot{\mathbf{X}} \right\rangle = \overline{\mathcal{H}} \left( \langle \dot{\mathbf{X}}^2 \rangle \langle \dot{\mathbf{X}}^2 \rangle - \langle \dot{\mathbf{X}} \cdot \dot{\mathbf{X}} \rangle^2 \right). \quad (81)$$

The last term vanishes and the cross correlator  $\overline{\mathcal{H}} \bar{v}^2 \bar{t}^2$  is positive; from (71), we deduce that  $c_0 > 0$ . It is interesting to notice that  $c_0$  would vanish in Minkowski spacetime, which can be viewed as a consequence of time reversal invariance. The existence of a cosmic string bispectrum is the consequence of the breaking of the time reversal invariance in a FLRW background.

An illustrative example is to apply (77) to the isosceles triangle configurations in Fourier space such that

$$|\mathbf{k}_1| = |\mathbf{k}_2| = k, \quad |\mathbf{k}_3| = 2k \sin \frac{\theta}{2}, \quad (82)$$

where  $\theta$  denotes the angle between the wavevectors  $\mathbf{k}_1$  and  $\mathbf{k}_2$ . The isosceles bispectrum reads

$$B_{\ell\ell\theta}(k, \theta) = -\epsilon^3 \pi c_0 \frac{\bar{v}^2}{\bar{t}^4} \frac{L\hat{\xi}}{\mathcal{A}} \frac{1}{\hat{\xi}^2 k^6} \frac{1 + 4 \cos \theta \sin^2(\theta/2)}{\sin^3 \theta}. \quad (83)$$

Notice that for  $\theta = \pi/3$ , we obtain the peculiar case of an equilateral triangle. In Figure 18, we have plotted the angle dependency of the isosceles bispectrum. These configurations are amplified as  $1/\theta^3$  in the two flat triangle limits for which either  $\theta \rightarrow 0$  or  $\theta \rightarrow \pi$ . Both of these configurations are therefore better suited than the equilateral one to characterize the strings. As suggested by the real space searches, the strings can produce a strong bispectrum signal only if the detector resolution is sufficiently good. Assuming a beam resolution of  $5'$  means that the  $7.2^\circ$  field of view would contain at maximum roughly  $80^2$  Fourier modes. Consequently, the smallest values of  $\theta$  achievable would be around  $\theta > 0.03$  radians, with only a few modes saturating this bound.

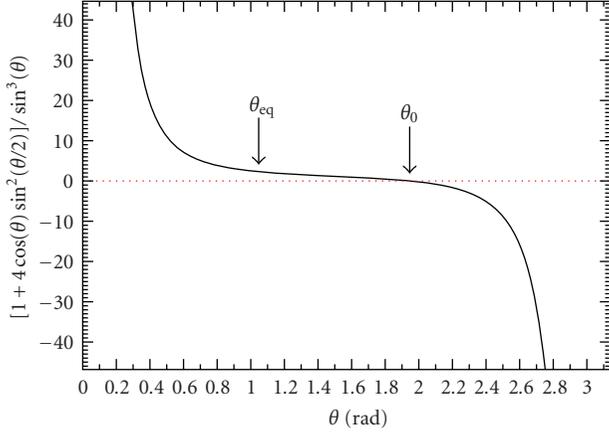


FIGURE 18: Angular dependency of the isosceles bispectrum as a function of the angle  $\theta$  in between the wavevectors  $\mathbf{k}_1$  and  $\mathbf{k}_2$ . The particular values  $\theta_{\text{eq}} = \pi/3$  corresponds to the equilateral configuration and  $\theta_0$  makes the bispectrum vanishing. Notice the amplification for flat triangle configurations at  $\theta \rightarrow 0$  (squeezed) and  $\theta \rightarrow \pi$  (collapsed).

**4.7.2. Numerical Results.** The previous analytical results can be compared to the CMB temperature bispectrum derived from the simulated maps of Section 4.2. Numerically, one can use the scale convolution method introduced in [213, 214] and applied to the string bispectrum in [212]. This method relies on the choice of unity window functions in Fourier space  $W_u(l)$  peaked around a particular wavenumber  $u$ . Defining

$$\Theta_u(\mathbf{x}) \equiv \int \frac{d\mathbf{l}}{(2\pi)^2} \hat{\Theta}_1 W_u(l) e^{-i\mathbf{l} \cdot \mathbf{x}}, \quad (84)$$

one can construct an estimator of the three point function in Fourier space by remarking that

$$\begin{aligned} & \int \Theta_{k_1}(\mathbf{x}) \Theta_{k_2}(\mathbf{x}) \Theta_{k_3}(\mathbf{x}) d\mathbf{x} \\ &= \int \frac{d\mathbf{p} d\mathbf{q} d\mathbf{k}}{(2\pi)^6} \hat{\Theta}_p \hat{\Theta}_q \hat{\Theta}_k W_{k_1}(p) W_{k_2}(q) W_{k_3}(k) (2\pi)^2 \\ & \quad \times \delta(\mathbf{p} + \mathbf{q} + \mathbf{k}). \end{aligned} \quad (85)$$

For thin enough window functions,  $\hat{\Theta}_k$  remains constant over the window function width and we construct our reduced bispectrum estimator as

$$b_{k_1 k_2 k_3} = \frac{1}{S_{k_1 k_2 k_3}^{(w)}} \left\langle \int \Theta_{k_1}(\mathbf{x}) \Theta_{k_2}(\mathbf{x}) \Theta_{k_3}(\mathbf{x}) d\mathbf{x} \right\rangle. \quad (86)$$

The function  $S^{(w)}$  is the flat sky equivalent of the inverse Wigner-3j symbols and reads

$$S_{k_1 k_2 k_3}^{(w)} = \int \frac{d\mathbf{p} d\mathbf{q}}{(2\pi)^4} W_{k_1}(p) W_{k_2}(q) W_{k_3}(|\mathbf{p} + \mathbf{q}|). \quad (87)$$

For the window functions such that  $W_u(k) = 1$  for  $u - w/2 < k < u + w/2$ , one can approximate

$$W_u(k) \simeq w \delta(k - u), \quad (88)$$

for small enough values of  $w$  compared to the wavenumber  $k$ . In this case, (87) can be worked out into

$$S^{(w)} \simeq \left( \frac{w}{2\pi} \right)^3 \frac{4k_1 k_2 k_3}{\sqrt{[(k_1 + k_2)^2 - k_3^2][k_3^2 - (k_1 - k_2)^2]}}. \quad (89)$$

In the left panel of Figure 19, we have plotted the mean string bispectrum and its standard deviation obtained by this method over the 300 string CMB maps. For illustration purpose, this plot is for the isosceles configuration having  $\theta = 0.2$  radians. The right frame of Figure 19 shows the same mean bispectrum but multiplied by  $\theta^3$ , for various small values of  $\theta$ . As expected from the analytical results, we recover the  $1/\theta^3$  behaviour. The wavenumber dependency also matches with the analytical calculations, up to similar slight power differences as we found for the power spectrum. A power law fit against the mean numerical estimator gives

$$[\ell(\ell + 1)]^{3/2} b_{\ell\ell\theta} \propto \ell^{-q}, \quad \text{with } q \simeq 2.8, \quad (90)$$

while the overall amplitude can be evaluated around the minimum variance multipole. At  $\ell = 5000$ , one gets

$$\left[ \frac{\ell(\ell + 1)}{2\pi} \right]^{3/2} b_{\ell\ell\theta} \Big|_{\ell=5000} \simeq (-2.7 \pm 1.4) \times 10^{-3} \left( \frac{GU}{\theta} \right)^3, \quad (91)$$

which also matches with (83) under some crude estimation of the string parameters [212]. Finally, as suggested by Figure 18, the string bispectrum is mostly negative. Integrated over all possible configurations, one recovers the mean negative sample skewness of (53), thereby explaining its origin as a direct consequence of the breaking of the time reversal symmetry in FLRW spacetimes.

**4.8. Trispectrum.** The trispectrum of the string induced CMB temperature anisotropies can be derived in a similar way. Starting from the definition of the four-point functions

$$\begin{aligned} & \langle \hat{\Theta}_{\mathbf{k}_1} \hat{\Theta}_{\mathbf{k}_2} \hat{\Theta}_{\mathbf{k}_3} \hat{\Theta}_{\mathbf{k}_4} \rangle \\ &= T(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) (2\pi)^2 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4), \end{aligned} \quad (92)$$

we define the trispectrum as

$$\begin{aligned} & T(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\ &= \frac{\epsilon^4}{\mathcal{A}} \frac{k_{1A} k_{2B} k_{3C} k_{4D}}{k_1^2 k_2^2 k_3^2 k_4^2} \int d\sigma_1 d\sigma_2 d\sigma_3 d\sigma_4 \langle \dot{X}_1^A \dot{X}_2^B \dot{X}_3^C \dot{X}_4^D e^{i\delta^{ab} \mathbf{k}_a \cdot \mathbf{x}_b} \rangle, \end{aligned} \quad (93)$$

with  $\dot{X}_a^A = \dot{X}^A(\sigma_a)$ ,  $(a, b) \in \{1, 2, 3, 4\}$  and  $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4 = 0$ . (Notice that our denomination ‘‘trispectrum’’ here contains the unconnected part. This one is however non-vanishing only for parallelogram configurations of the wavevectors.) As shown in [215], the trispectrum and the higher  $n$ -point functions exhibit unfactorable ‘‘flat directions’’ in the  $n$ -dimensional space of the integration variables  $\{\sigma_a\}$ . Physically, it means that the leading order part of the

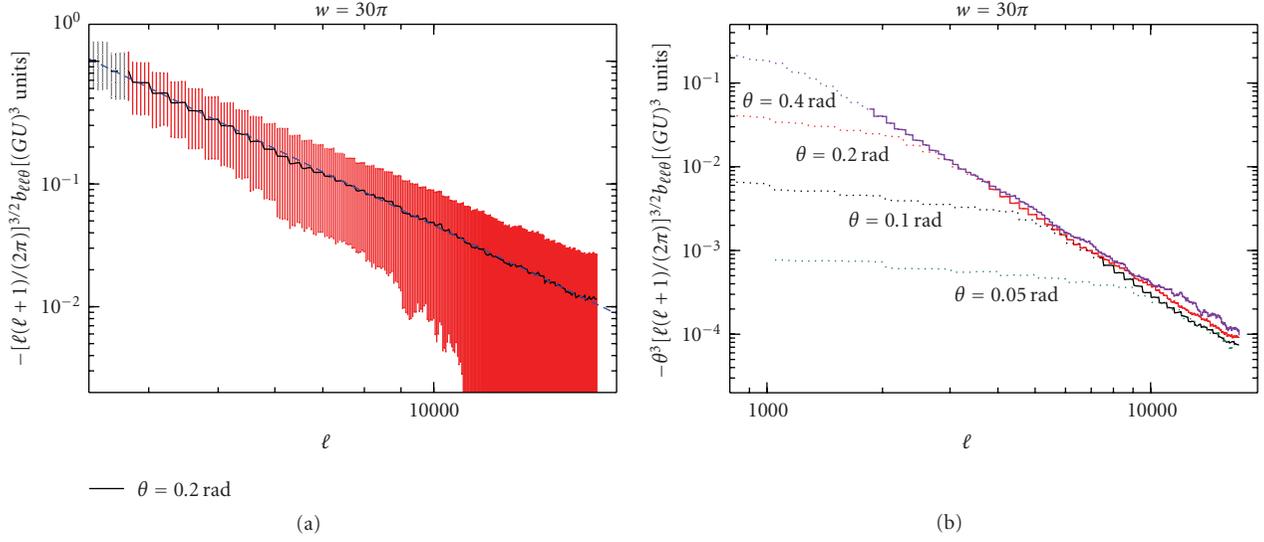


FIGURE 19: (a): mean value and standard deviation of the squeezed isosceles bispectrum  $[\ell(\ell + 1)/(2\pi)]^{3/2} b_{\ell\ell\theta}$  for  $\theta = 0.2$  radians. The dashed line is the best power law fit. (b) shows its rescaling by  $\theta^3 [\ell(\ell + 1)/(2\pi)]^{3/2} b_{\ell\ell\theta}$  showing the  $1/\theta^3$  dependency. The spurious plateau (dotted) for the lower multipoles comes from a numerical cutoff associated with the window functions and occurs at  $\ell_{\min}(\theta) \simeq 30\pi/(\theta_{\text{fov}}\theta)$ , the field of view being  $\theta_{\text{fov}} = 7.2^\circ$ .

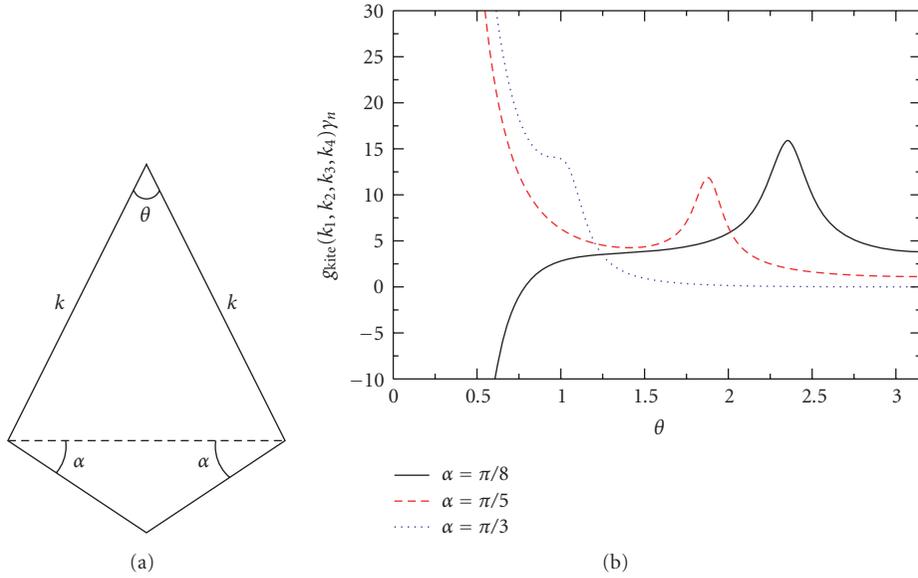


FIGURE 20: Trispectrum geometrical factor for the kite quadrilaterals (represented on the left panel) as a function of the opening angle  $\theta$ , and plotted for various values of  $\alpha$ . The trispectrum is enhanced in the squeezed limit  $\theta \rightarrow 0$ . The bump for  $\theta_p = \pi - 2\alpha$  corresponds to the parallelogram limit for which the unconnected part is no longer vanishing.

(connected) trispectrum is sensitive to the higher orders of the correlators in (64) to (66). For the trispectrum, the correlator  $T(\sigma)$  has to be expanded at next-to-leading order, and following the Polchinski-Rocha model [179], we assume a nonanalytical behaviour for  $T(\sigma)$  at small scales

$$T(\sigma) \simeq \bar{t}^2 - c_1 \left( \frac{\sigma}{\bar{\xi}} \right)^{2\chi}. \quad (94)$$

In the light-cone gauge, we leave  $c_1$  and  $\chi$  as undetermined parameters since they cannot be straightforwardly inferred from the numerics performed in the temporal gauge. Nevertheless, because the correlation should be smaller as  $\sigma$  becomes larger, one should have  $c_1 > 0$ . In the temporal gauge,  $\chi$  is directly related to the power law exponent of the scaling loop distribution functions through  $\chi = 1 - p/2$ . As we are bound to show, the mode dependence of the trispectrum will also be uniquely given by this parameter. Once the tangent vector correlator expressed as in (94), the

integrations in (93) can be performed explicitly, except for parallelogram configurations which have to be dealt with as a special case. After some tedious calculations, an interpolating expression for the trispectrum is [215]

$$T(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \simeq \epsilon^4 \frac{\bar{v}^4}{\bar{t}^2} \frac{L\hat{\xi}}{\mathcal{A}} (c_1 \hat{\xi}^2)^{-1/(2\chi+2)} f(\chi) \\ \times \gamma_n \left( \frac{1}{2\chi+2}, c_2 Y^2 \Lambda^{2\chi+2} \right) g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4). \quad (95)$$

In this equation,  $f(\chi)$  is a number depending only on the parameter  $\chi$

$$f(\chi) = \frac{\pi}{\chi+1} \Gamma \left( \frac{1}{2\chi+2} \right) \left[ 4(2\chi+1)(\chi+1) \right]^{1/(2\chi+2)}, \quad (96)$$

while  $g(\{\mathbf{k}_a\})$  is the trispectrum geometrical factor defined by

$$g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \frac{\kappa_{12}\kappa_{34} + \kappa_{13}\kappa_{24} + \kappa_{14}\kappa_{23}}{k_1^2 k_2^2 k_3^2 k_4^2} [Y^2]^{-1/(2\chi+2)}, \quad (97)$$

where

$$Y^2(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \equiv -\kappa_{12}(k_3^2 k_4^2 - \kappa_{34}^2)^{\chi+1} + \circlearrowleft, \quad (98)$$

and  $\circlearrowleft$  stands for cyclic permutations over the indices. The function  $\gamma_n(a, x)$  stands for the normalised incomplete lower gamma function defined by

$$\gamma_n(a, x) \equiv \frac{\gamma(a, x)}{\Gamma(a)} = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt, \quad (99)$$

and, finally,  $\Lambda$  has been defined by

$$\Lambda(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \equiv \frac{2L}{(k_1^2 k_2^2 - \kappa_{12}^2)^{1/2} + \circlearrowleft} \\ \times \frac{k_1 k_2 k_3 k_4}{\kappa_{12}\kappa_{34} + \kappa_{13}\kappa_{24} + \kappa_{14}\kappa_{23}}. \quad (100)$$

As an application, the trispectrum over parallelogram configurations is obtained when  $Y^2 = 0$  and the leading term of (95) simplifies to

$$T_0(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \simeq \frac{\pi \epsilon^4 \bar{v}^4}{\bar{t}^2} \frac{L^2}{\mathcal{A} k_1^3 k_2^3 |\sin \theta|}, \quad (101)$$

where  $\theta$  now refers to the parallelogram angle. Under the scaling transformation  $\mathbf{k}_a \rightarrow \lambda \mathbf{k}_a$ , the parallelogram trispectrum scales as

$$T_0(\lambda \mathbf{k}_1, \lambda \mathbf{k}_2, \lambda \mathbf{k}_3, \lambda \mathbf{k}_4) = \lambda^{-6} T_0(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4). \quad (102)$$

For parallelograms, as already mentioned, the trispectrum also gets a contribution from the unconnected part of the four-point function, which is Gaussian and reads

$$T_0^{\text{uc}}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) = \mathcal{A} P(k_1) P(k_2) + \circlearrowleft. \quad (103)$$

Using (63), one sees that the unconnected part also behaves as  $\lambda^{-6}$ . Therefore, the non-Gaussian contributions for parallelogram configurations remain of the same order of magnitude as the Gaussian ones, with however, and again, an exception in the squeezed limit  $\theta \rightarrow 0$ .

The most interesting situations come when  $Y^2 \neq 0$ . For these quadrilaterals, the Gaussian contribution vanishes and solely a non-Gaussian statistics can source the trispectrum. At large wavenumber (small angles), one has  $Y^2 \gg 1$  such that the normalised lower incomplete gamma function in (95) is close to unity:

$$T_\infty(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\ \simeq \epsilon^4 \frac{\bar{v}^4}{\bar{t}^2} \frac{L\hat{\xi}}{\mathcal{A}} (c_1 \hat{\xi}^2)^{-1/(2\chi+2)} f(\chi) g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4). \quad (104)$$

Under the scaling transformation  $\mathbf{k}_a \rightarrow \lambda \mathbf{k}_a$ , the geometric factor, and thus (104), scales as

$$g(\lambda \mathbf{k}_1, \lambda \mathbf{k}_2, \lambda \mathbf{k}_3, \lambda \mathbf{k}_4) = \lambda^{-\rho} g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4), \quad (105)$$

with

$$\rho = 6 + \frac{1}{\chi+1}. \quad (106)$$

As claimed, for NG strings,  $\rho$  is directly given by the power law of the loop distribution [146]. Since this exponent is different from the one associated with parallelogram configurations it may actually be used to distinguish a trispectrum sourced by cosmic strings with the one generated by other non-Gaussian effects.

In Figure 20, we have plotted the geometrical factor  $g(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4)$  for the kite quadrilaterals (represented in the same figure), as a function of their opening angles  $\theta$  and  $\alpha$ .

As for the bispectrum, the trispectrum is enhanced on the squeezed configuration obtained when  $\theta$  becomes small. In this limit (97) can be expanded as

$$g \sim \frac{8 \cos^2(\alpha)}{k^p \theta^{\rho-3}} (1 - 2 \cos 2\alpha) \\ \times \left[ 2(1 + \chi) \tan^2(\alpha) - 1 + 4^\chi (1 - \tan^2 \alpha) \right]^{-1/(2\chi+2)}, \quad (107)$$

and one recovers the mode dependency in  $k^{-\rho}$  while the amplitude is amplified as  $\theta^{\rho-3}$ . As discussed in the previous section, the singular limit  $\theta \rightarrow 0$  is never reached with a finite resolution beam.

**4.9. Comparison with Data.** The cosmic string bispectrum and trispectrum associated with the flat polygonal configurations are the best suited to look for string signatures. However, it is not easy to compare with existing constraints as much of the literature focuses on particular models of *primordial* non-Gaussianity. For instance, in the local type of primordial non-Gaussianities, the parameter  $f_{\text{NL}}$  characterises the primordial bispectrum and maximal amplitude occurs for squeezed triangle configurations, as

it is the case for the cosmic strings [216]. However, as a result of the CMB transfer functions, a given value of  $f_{\text{NL}}$  corresponds to oscillating damped patterns of the CMB temperature bispectrum, which are completely different of the power laws we have found for the string bispectrum at small scales. The current bounds on  $f_{\text{NL}}$  being precisely obtained from template matching procedures, they cannot be applied to the strings [44, 217, 218]. For this reason, the parameters used to quantify primordial non-Gaussianities are not well suited here, precisely because we expect the string non-Gaussianities to be nonprimordial. An efficient approach would be to use a template matching procedure with the formulae derived in the previous sections. Another approach might be to estimate what values the primordial parameters, such as  $f_{\text{NL}}$  and  $\tau_{\text{NL}}$ , would assume if the non-Gaussianities were actually due to strings. Notice that asking such a question would be close to find the best amplitude of a sine function to fit a power law. However, since primordial non-Gaussianities are and will be tested in CMB data anyway, one could answer this question by performing a Fisher matrix analysis along the lines of [55, 57].

## 5. Conclusion and Perspectives

The results presented in this paper were essentially concerned with Nambu-Goto type of cosmic strings, which is the simplest realisation of a one-dimensional spatially extended object. As a result, they should not be blindly extrapolated to other types of string, although, as argued in Section 2, some of them are expected to be generic. In particular, due to the scaling of the long strings, cosmic string loops do not influence significantly the CMB observables. Changing the intercommuting probability is expected to rescale some of the presented results [219, 220], but in a way which remains to be quantified.

In Section 3, we have briefly reviewed the current understanding of the cosmological evolution of a string network by means of FLRW numerical simulations, which is a nontrivial problem even for NG strings. Observable predictions crucially depend on this step. Numerical simulations can be avoided by making some assumptions on the string distribution but at the expense of introducing unnecessary extra parameters. When approximate analytical models are then used to derive observable predictions, one should keep in mind that the results are as uncertain as the values assumed for the additional parameters. Provided one is interested in length scales not affected by gravitational back-reaction effects, all of the statistical properties of an NG string network in scaling depends only on one unknown physical parameter: the string energy density per unit length  $U$ , not more. (The expansion rate is supposed to be known.)

In this context, Section 4 discusses the non-Gaussian effects induced by a cosmological string network in the CMB temperature anisotropies. We have shown that string induced CMB fluctuations have a negative skewness and a non-vanishing kurtosis. On a CMB temperature anisotropies map, these non-Gaussianities imprint characteristic signatures in a multifractal analysis as well as in the gradient

magnitude, both being more significant at small angles. This property is recovered in Fourier space: the CMB angular power spectrum decays at most as  $1/\ell$ , for the large multipoles  $\ell$ , and strings become the dominant sources of primary fluctuations. The skewness appears to be the direct consequence of the breaking of the time reversal symmetry in an expanding universe, and implies the existence of a non-vanishing bispectrum. Using analytical approximations, tested and confirmed by numerical simulations, we then derived the expected bispectrum and trispectrum of string induced CMB temperature anisotropies for the large multipoles. Although the bispectrum decays not faster than  $\ell^{-6}$ , the trispectrum multipole dependency is in  $\ell^{-\rho}$ , where  $\rho = 6 + 1/(\chi + 1)$  and  $\chi$  is a small number related to the tangent vector correlator and the NG loop distribution. Due to the line-like CMB patterns induced by the strings, both the bispectrum and trispectrum are enhanced on all elongated triangle and quadrilateral configurations of the wavevectors. These ones may constitute the best configurations to look for a non-Gaussian string signal while being experimentally limited by finite beam resolution. Let us note that our expressions have been derived in the flat sky approximation. String non-Gaussianities at small multipoles is still an open problem which could be dealt with full sky string maps [201, 221, 222]. However, if, as the current constraints suggest, cosmic strings marginally contribute to the large-scale CMB anisotropies, then they should show up at large multipoles in all of the above mentioned observables. This is precisely where the experimental efforts are directed. At very small angular scales, the difficulties will certainly be to separate the string signals from the astrophysical sources. Interestingly, the very soon accessible intermediate angles probed by the Planck satellite, and the other ground-based telescopes, may not suffer from this problem and could be an open window on cosmic strings.

## Acknowledgments

It is a pleasure to thank Patrick Peter, Mairi Sakellariadou, Danielle Steer and Teruaki Suyama for a careful reading of the paper and their enlightening comments. This paper is supported by the Belgian Federal Office for Scientific, Technical and Cultural Affairs, under the Inter-University Attraction Pole Grant P6/11.

## References

- [1] D. A. Kirzhnits and A. D. Linde, "Macroscopic consequences of the Weinberg model," *Physics Letters B*, vol. 42, no. 4, pp. 471–474, 1972.
- [2] I. Y. Kobsarev, L. B. Okun, and Y. B. Zeldovich, "Spontaneous CP-violation and cosmology," *Physics Letters B*, vol. 50, no. 3, pp. 340–342, 1974.
- [3] T. W. B. Kibble, "Topology of cosmic domains and strings," *Journal of Physics A*, vol. 9, no. 8, pp. 1387–1398, 1976.
- [4] M. B. Hindmarsh and T. W. B. Kibble, "Cosmic strings," *Reports on Progress in Physics*, vol. 58, no. 5, pp. 477–562, 1995.

- [5] A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects*, Cambridge University Press, Cambridge, UK, 2000.
- [6] M. Sakellariadou, "Cosmic strings," *Lecture Notes in Physics*, vol. 718, pp. 247–288, 2007.
- [7] P. Peter and J.-P. Uzan, *Primordial Cosmology*, Oxford Graduate Texts, Oxford University Press, Oxford, UK, 2009.
- [8] S. Sarkar, "Big bang nucleosynthesis and physics beyond the standard model," *Reports on Progress in Physics*, vol. 59, no. 12, pp. 1493–1609, 1996.
- [9] R. H. Cyburt, B. D. Fields, K. A. Olive, and E. Skillman, "New BBN limits on physics beyond the standard model from  ${}^4\text{He}$ ," *Astroparticle Physics*, vol. 23, no. 3, pp. 313–323, 2005.
- [10] T. W. B. Kibble, G. Lazarides, and Q. Shafi, "Walls bounded by strings," *Physical Review D*, vol. 26, no. 2, pp. 435–439, 1982.
- [11] P. Langacker and S.-Y. Pi, "Magnetic monopoles in grand unified theories," *Physical Review Letters*, vol. 45, no. 1, pp. 1–4, 1980.
- [12] A. H. Guth, "Inflationary universe: a possible solution to the horizon and flatness problems," *Physical Review D*, vol. 23, no. 2, pp. 347–356, 1981.
- [13] A. A. Starobinsky, "A new type of isotropic cosmological models without singularity," *Physics Letters B*, vol. 91, no. 1, pp. 99–102, 1980.
- [14] A. D. Linde, "A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems," *Physics Letters B*, vol. 108, no. 6, pp. 389–393, 1982.
- [15] A. A. Starobinsky, "Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations," *Physics Letters B*, vol. 117, no. 3–4, pp. 175–178, 1982.
- [16] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, "Theory of cosmological perturbations, part I: classical perturbations," *Physics Reports*, vol. 215, no. 5–6, pp. 203–256, 1992.
- [17] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, "Theory of cosmological perturbations, part II: quantum perturbations," *Physics Reports*, vol. 215, no. 5–6, pp. 257–295, 1992.
- [18] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, "Theory of cosmological perturbations, part III: extensions," *Physics Reports*, vol. 215, no. 5–6, pp. 296–333, 1992.
- [19] J. Martin and C. Ringeval, "Inflation after WMAP3: confronting the slow-roll and exact power spectra with CMB data," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 8, article 9, 2006.
- [20] P. P. Avelino, C. J. A. P. Martins, and E. P. S. Shellard, "Effects of inflation on a cosmic string loop population," *Physical Review D*, vol. 76, no. 8, Article ID 083510, 7 pages, 2007.
- [21] R. Jeannerot, J. Rocher, and M. Sakellariadou, "How generic is cosmic string formation in supersymmetric grand unified theories," *Physical Review D*, vol. 68, no. 10, Article ID 103514, 20 pages, 2003.
- [22] J. Rocher and M. Sakellariadou, " $D$ -term inflation, cosmic strings, and consistency with cosmic microwave background measurements," *Physical Review Letters*, vol. 94, no. 1, Article ID 011303, 2005.
- [23] J. Rocher and M. Sakellariadou, " $D$ -term inflation in non-minimal supergravity," *Journal of Cosmology and Astroparticle Physics*, no. 11, article 1, 2006.
- [24] R. A. Battye, B. Garbrecht, and A. Moss, "Constraints on supersymmetric hybrid inflation models," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 9, article 7, 2006.
- [25] K. Becker, M. Becker, and J. H. Schwarz, *String Theory and M-Theory: A Modern Introduction*, Cambridge University Press, Cambridge, UK, 2007.
- [26] G. Dvali and S.-H. H. Tye, "Brane inflation," *Physics Letters B*, vol. 450, no. 1–3, pp. 72–82, 1999.
- [27] S. H. S. Alexander, "Inflation from  $D-\bar{D}$  brane annihilation," *Physical Review D*, vol. 65, no. 2, Article ID 023507, 2002.
- [28] S. Kachru, R. Kallosh, A. Linde, J. Maldacena, L. McAllister, and S. P. Trivedi, "Towards inflation in string theory," *Journal of Cosmology and Astroparticle Physics*, vol. 2003, no. 10, article 13, 2003.
- [29] C. P. Burgess, M. Majumdar, D. Nolte, F. Quevedo, G. Rajesh, and R.-J. Zhang, "The inflationary brane-antibrane universe," *Journal of High Energy Physics*, vol. 5, no. 7, article 47, 2001.
- [30] S. Sarangi and S.-H. Tye, "Cosmic string production towards the end of brane inflation," *Physics Letters B*, vol. 536, no. 3–4, pp. 185–192, 2002.
- [31] G. Dvali and A. Vilenkin, "Formation and evolution of cosmic  $D$  strings," *Journal of Cosmology and Astroparticle Physics*, vol. 2004, no. 3, article 10, 2004.
- [32] N. T. Jones, H. Stoica, and S.-H. H. Tye, "The production, spectrum and evolution of cosmic strings in brane inflation," *Physics Letters B*, vol. 563, no. 1–2, pp. 6–14, 2003.
- [33] A.-C. Davis and T. W. B. Kibble, "Fundamental cosmic strings," *Contemporary Physics*, vol. 46, no. 5, pp. 313–322, 2005.
- [34] E. J. Copeland and T. W. B. Kibble, "Cosmic strings and superstrings," *Proceedings of the Royal Society A*, vol. 466, no. 2115, pp. 623–657, 2010.
- [35] M. Sakellariadou, "Cosmic superstrings," *Philosophical Transactions of the Royal Society A*, vol. 366, no. 1877, pp. 2881–2894, 2008.
- [36] M. Sakellariadou, "Cosmic strings and cosmic superstrings," *Nuclear Physics B—Proceedings Supplements*, vol. 192–193, pp. 68–90, 2009.
- [37] F. R. Bouchet, D. P. Bennett, and A. Stebbins, "Patterns of the cosmic microwave background from evolving string networks," *Nature*, vol. 335, no. 6189, pp. 410–414, 1988.
- [38] B. Carter, "Duality relation between charged elastic strings and superconducting cosmic strings," *Physics Letters B*, vol. 224, no. 1–2, pp. 61–66, 1989.
- [39] R. Durrer, M. Kunz, and A. Melchiorri, "Cosmic structure formation with topological defects," *Physics Report*, vol. 364, no. 1, pp. 1–81, 2002.
- [40] R. Durrer, A. Gangui, and M. Sakellariadou, "Doppler peaks in the angular power spectrum of the cosmic microwave background: a fingerprint of topological defects," *Physical Review Letters*, vol. 76, no. 4, pp. 579–582, 1996.
- [41] J. Magueijo, A. Albrecht, P. Ferreira, and D. Coulson, "The structure of Doppler peaks induced by active perturbations," *Physical Review D*, vol. 54, no. 6, pp. 3727–3744, 1996.
- [42] A. Albrecht, R. A. Battye, and J. Robinson, "The case against scaling defect models of cosmic structure formation," *Physical Review Letters*, vol. 79, no. 24, pp. 4736–4739, 1997.
- [43] D. N. Spergel, R. Bean, O. Doré et al., "Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: implications for cosmology," *Astrophysical Journal, Supplement Series*, vol. 170, no. 2, pp. 377–408, 2007.
- [44] E. Komatsu, J. Dunkley, M. R. Nolte et al., "Five-year wilkinson microwave anisotropy probe observations: cosmological

- interpretation,” *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [45] F. R. Bouchet, P. Peter, A. Riazuelo, and M. Sakellariadou, “Evidence against or for topological defects in the BOOMERanG data?” *Physical Review D*, vol. 65, no. 2, Article ID 021301, 4 pages, 2001.
- [46] A. A. Fraisse, “Limits on defects formation and hybrid inflationary models with three-year WMAP observations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 3, article 8, 2007.
- [47] N. Bevis, M. Hindmarsh, M. Kunz, and J. Urrestilla, “Fitting cosmic microwave background data with cosmic strings and inflation,” *Physical Review Letters*, vol. 100, no. 2, Article ID 021301, 2008.
- [48] A. M. Gilbert and L. Perivolaropoulos, “Spectra and statistics of cosmic string perturbations on the microwave background: a Monte Carlo approach,” *Astroparticle Physics*, vol. 3, no. 3, pp. 283–294, 1995.
- [49] E. Jeong and G. F. Smoot, “The validity of the cosmic string pattern search with the cosmic microwave background,” *Astrophysical Journal*, vol. 661, no. 1, pp. L1–L4, 2007.
- [50] S. Amsel, J. Berger, and R. H. Brandenberger, “Detecting cosmic strings in the CMB with the canny algorithm,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article 15, 2008.
- [51] A. Gangui, L. Pogosian, and S. Winitzki, “CMB bispectrum from active models of structure formation,” *Physical Review D*, vol. 64, no. 4, Article ID 043001, 7 pages, 2001.
- [52] L. Pogosian and M. Wyman, “B-modes from cosmic strings,” *Physical Review D*, vol. 77, no. 8, Article ID 083509, 11 pages, 2008.
- [53] J. R. Gott III, “Gravitational lensing effects of vacuum strings—exact solutions,” *Astrophysical Journal*, vol. 288, no. 8, pp. 422–427, 1985.
- [54] N. Kaiser and A. Stebbins, “Microwave anisotropy due to cosmic strings,” *Nature*, vol. 310, no. 5976, pp. 391–393, 1984.
- [55] E. Sefusatti and E. Komatsu, “The bispectrum of galaxies from high-redshift galaxy surveys: primordial non-Gaussianity and non-linear galaxy bias,” *Physical Review D*, vol. 76, no. 8, Article ID 083004, 17 pages, 2007.
- [56] C. Pitrou, J.-P. Uzan, and F. Bernardeau, “Cosmic microwave background bispectrum on small angular scales,” *Physical Review D*, vol. 78, no. 6, Article ID 063526, 2008.
- [57] D. Nitta, E. Komatsu, N. Bartolo, S. Matarrese, and A. Riotto, “CMB anisotropies at second order III: bispectrum from products of the first-order perturbations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 14, 2009.
- [58] L. Boubekeur, P. Creminelli, G. D’Amico, J. Norea, and F. Vernizzi, “Sachs-wolfe at second order: the CMB bispectrum on large angular scales,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 29, 2009.
- [59] S. Weinberg, “Conceptual foundations of the unified theory of weak and electromagnetic interactions,” *Reviews of Modern Physics*, vol. 52, no. 3, pp. 515–523, 1980.
- [60] A. J. Gill and R. J. Rivers, “Dynamics of vortex and monopole production by quench-induced phase separation,” *Physical Review D*, vol. 51, no. 12, pp. 6949–6958, 1995.
- [61] G. Karra and R. J. Rivers, “Initial vortex densities after a temperature quench,” *Physics Letters B*, vol. 414, no. 1-2, pp. 28–33, 1997.
- [62] L. M. A. Bettencourt, T. S. Evans, and R. J. Rivers, “Winding number correlation functions and cosmic string formation,” *Physical Review D*, vol. 53, no. 2, pp. 668–680, 1996.
- [63] E. Kavoussanaki, R. Monaco, and R. J. Rivers, “Testing the kibble-zurek scenario with annular Josephson tunnel junctions,” *Physical Review Letters*, vol. 85, no. 16, pp. 3452–3455, 2000.
- [64] N. K. Nielsen and P. Olesen, “Dynamical properties of superconducting cosmic strings,” *Nuclear Physics B*, vol. 291, pp. 829–846, 1987.
- [65] P. Peter, “Superconducting cosmic string: equation of state for spacelike and timelike current in the neutral limit,” *Physical Review D*, vol. 45, no. 4, pp. 1091–1102, 1992.
- [66] C. Ringeval, “Equation of state of cosmic strings with fermionic current carriers,” *Physical Review D*, vol. 63, no. 6, Article ID 063508, 24 pages, 2001.
- [67] A. Vilenkin, “Gravitational field of vacuum domain walls and strings,” *Physical Review D*, vol. 23, no. 4, pp. 852–857, 1981.
- [68] L. M. A. Bettencourt and R. J. Rivers, “Interactions between U(1) cosmic strings: an analytical study,” *Physical Review D*, vol. 51, no. 4, pp. 1842–1853, 1995.
- [69] N. D. Antunes, L. M. A. Bettencourt, and M. Hindmarsh, “Thermodynamics of cosmic string densities in U(1) scalar field theory,” *Physical Review Letters*, vol. 80, no. 5, pp. 908–911, 1998.
- [70] T. Vachaspati and A. Vilenkin, “Formation and evolution of cosmic strings,” *Physical Review D*, vol. 30, no. 10, pp. 2036–2045, 1984.
- [71] M. Hindmarsh and A. Rajantie, “Phase transition dynamics in the hot Abelian Higgs model,” *Physical Review D*, vol. 64, no. 6, Article ID 065016, 13 pages, 2001.
- [72] M. Donaire and A. Rajantie, “Heavy cosmic strings,” *Physical Review D*, vol. 73, no. 6, Article ID 063517, 4 pages, 2006.
- [73] A. Rajantie, “Super-horizon cosmic string correlations,” *Physical Review D*, vol. 79, no. 4, Article ID 043515, 2009.
- [74] M. Sakellariadou and H. Stoica, “Dynamics of F/D networks: the role of bound states,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 38, 2008.
- [75] G. Vincent, N. D. Antunes, and M. Hindmarsh, “Numerical simulations of string networks in the Abelian-Higgs model,” *Physical Review Letters*, vol. 80, no. 11, pp. 2277–2280, 1998.
- [76] J. N. Moore, E. P. S. Shellard, and C. J. A. P. Martins, “Evolution of Abelian-Higgs string networks,” *Physical Review D*, vol. 65, no. 2, Article ID 023503, 19 pages, 2001.
- [77] M. Hindmarsh, S. Stuckey, and N. Bevis, “Abelian Higgs cosmic strings: small-scale structure and loops,” *Physical Review D*, vol. 79, no. 12, Article ID 123504, 2009.
- [78] R. L. Davis, “Goldstone bosons in string models of galaxy formation,” *Physical Review D*, vol. 32, no. 12, pp. 3172–3177, 1985.
- [79] R. Durrer, M. Kunz, and A. Melchiorri, “Cosmic microwave background anisotropies from scaling seeds: global defect models,” *Physical Review D*, vol. 59, no. 12, Article ID 123005, 26 pages, 1999.
- [80] M. Yamaguchi, “Scaling property of the global string in the radiation dominated universe,” *Physical Review D*, vol. 60, no. 10, Article ID 103511, 10 pages, 1999.
- [81] M. Yamaguchi, J. Yokoyama, and M. Kawasaki, “Evolution of a global string network in a matter-dominated universe,” *Physical Review D*, vol. 61, no. 6, Article ID 061301, 5 pages, 2000.
- [82] P. McGraw, “Evolution of a non-abelian cosmic string network,” *Physical Review D*, vol. 57, no. 6, pp. 3317–3339, 1998.

- [83] A. Vilenkin, "String-dominated universe," *Physical Review Letters*, vol. 53, no. 10, pp. 1016–1018, 1984.
- [84] G. Dvali and G. Senjanović, "Flavor changing strings and domain walls," *Physical Review Letters*, vol. 72, no. 1, pp. 9–12, 1994.
- [85] D. Spergel and U.-L. Pen, "Cosmology in a string-dominated universe," *Astrophysical Journal*, vol. 491, no. 2, pp. L67–L71, 1997.
- [86] M. Bucher and D. Spergel, "Is the dark matter a solid?" *Physical Review D*, vol. 60, no. 4, Article ID 043505, 11 pages, 1999.
- [87] K. Hashimoto and D. Tong, "Reconnection of non-abelian cosmic strings," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 9, article 4, pp. 53–72, 2005.
- [88] M. Eto, et al., "Universal reconnection of non-Abelian cosmic strings," *Physics Review Letters*, vol. 98, no. 9, Article ID 091602, 4 pages, 2007.
- [89] M. Eto, Y. Isozumi, M. Nitta, K. Ohashi, and N. Sakai, "Solitons in the Higgs phase: The moduli matrix approach," *Journal of Physics A*, vol. 39, no. 26, pp. R315–R392, 2006.
- [90] M. Shifman and A. Yung, "Supersymmetric solitons and how they help us understand non-abelian gauge theories," *Reviews of Modern Physics*, vol. 79, no. 4, pp. 1139–1196, 2007.
- [91] M. Kobayashi, Y. Kawaguchi, M. Nitta, and M. Ueda, "Collision dynamics and rung formation of non-Abelian vortices," *Physics Review Letters*, vol. 103, no. 11, Article ID 115301, 4 pages, 2009.
- [92] T. Vachaspati and A. Achúcarro, "Semilocal cosmic strings," *Physical Review D*, vol. 44, no. 10, pp. 3067–3071, 1991.
- [93] A. Achúcarro, J. Borrill, and A. R. Liddle, "Formation rate of semilocal strings," *Physical Review Letters*, vol. 82, no. 19, pp. 3742–3745, 1999.
- [94] A. Achúcarro and T. Vachaspati, "Semilocal and electroweak strings," *Physics Report*, vol. 327, no. 6, pp. 347–426, 2000.
- [95] M. Hindmarsh, "Existence and stability of semilocal strings," *Physical Review Letters*, vol. 68, no. 9, pp. 1263–1266, 1992.
- [96] X. Zhang, T. Huang, and R. H. Brandenberger, "Pion and  $\eta'$  strings," *Physical Review D*, vol. 58, no. 2, Article ID 027702, 3 pages, 1998.
- [97] A. P. Balachandran and S. Dital, "Topological string defect formation during the chiral phase transition," *International Journal of Modern Physics A*, vol. 17, no. 8, pp. 1149–1158, 2002.
- [98] K. B. W. Buckley and A. R. Zhitnitsky, "Superconducting K strings in high density QCD," *Journal of High Energy Physics*, vol. 6, no. 8, pp. 235–249, 2002.
- [99] R. H. Brandenberger, B. Carter, and A.-C. Davis, "Microwave background constraints on decaying defects," *Physics Letters B*, vol. 534, no. 1–4, pp. 1–7, 2002.
- [100] M. Nitta and N. Shiiki, "Non-Abelian global strings at chiral phase transition," *Physics Letters B*, vol. 658, no. 4, pp. 143–147, 2008.
- [101] M. Eto, E. Nakano, and M. Nitta, "Non-Abelian global vortices," *Nuclear Physics B*, vol. 821, no. 1-2, pp. 129–150, 2009.
- [102] E. Nakano, M. Nitta, and T. Matsuura, "Non-Abelian strings in hot or dense QCD," *Progress of Theoretical Physics Supplement*, vol. 174, pp. 254–257, 2008.
- [103] M. Eto, M. Nitta, and N. Yamamoto, "Instabilities of non-abelian vortices in dense QCD," *Physical Review Letters*, vol. 104, no. 16, Article ID 161601, 4 pages, 2010.
- [104] E. Babichev, "Global topological  $k$ -defects," *Physical Review D*, vol. 74, no. 8, Article ID 085004, 7 pages, 2006.
- [105] E. Babichev, "Gauge  $k$ -vortices," *Physical Review D*, vol. 77, no. 6, Article ID 065021, 2008.
- [106] S. Sarangi, "DBI global strings," *Journal of High Energy Physics*, vol. 2008, no. 7, article 18, 2008.
- [107] E. Babichev, P. Brax, C. Caprini, J. Martin, and D. A. Steer, "Dirac Born Infeld (DBI) cosmic strings," *Journal of High Energy Physics*, vol. 2009, no. 3, article 91, 2009.
- [108] E. Witten, "Superconducting strings," *Nuclear Physics B*, vol. 249, no. 4, pp. 557–592, 1985.
- [109] B. Carter, "Essentials of classical brane dynamics," *International Journal of Theoretical Physics*, vol. 40, no. 12, pp. 2099–2130, 2001.
- [110] R. L. Davis, "Semitopological solitons," *Physical Review D*, vol. 38, no. 12, pp. 3722–3730, 1988.
- [111] R. Brandenberger, B. Carter, A.-C. Davis, and M. Trodden, "Cosmic vortons and particle physics constraints," *Physical Review D*, vol. 54, no. 10, pp. 6059–6071, 1996.
- [112] C. Ringeval, "Fermionic massive modes along cosmic strings," *Physical Review D*, vol. 64, no. 12, Article ID 123505, 2001.
- [113] B. Carter and X. Martin, "Dynamical instability criterion for circular (vorton) string loops," *Annals of Physics*, vol. 227, no. 1, pp. 151–171, 1993.
- [114] X. Martin, "Zones of dynamical instability for rotating string loops," *Physical Review D*, vol. 50, no. 12, pp. 7479–7492, 1994.
- [115] J. Polchinski, "Introduction to cosmic F- and D-strings," in *Lectures Given at NATO Advanced Study Institute and EC Summer School on String Theory: From Gauge Interactions to Cosmology*, pp. 229–253, Cargèse, France, June 2004.
- [116] A.-C. Davis, P. Brax, and C. Van De Bruck, "Brane inflation and defect formation," *Philosophical Transactions of the Royal Society A*, vol. 366, no. 1877, pp. 2833–2842, 2008.
- [117] E. Witten, "Cosmic superstrings," *Physics Letters B*, vol. 153, no. 4-5, pp. 243–246, 1985.
- [118] L. Lorenz, J. Martin, and C. Ringeval, "Constraints on kinetically modified inflation from WMAP5," *Physical Review D*, vol. 78, no. 6, Article ID 063543, 2008.
- [119] S. B. Giddings, S. Kachru, and J. Polchinski, "Hierarchies from fluxes in string compactifications," *Physical Review D*, vol. 66, no. 10, Article ID 106006, 2002.
- [120] I. R. Klebanov and M. J. Strassler, "Supergravity and a confining gauge theory: duality cascades and  $\chi$ SB-resolution of naked singularities," *Journal of High Energy Physics*, vol. 4, no. 8, pp. 21–35, 2000.
- [121] L. Lorenz, J. Martin, and C. Ringeval, "Brane inflation and the WMAP data: a Bayesian analysis," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article 1, 2008.
- [122] H. Firouzjahi and S.-H. H. Tye, "Brane inflation and cosmic string tension in superstring theory," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 3, pp. 115–131, 2005.
- [123] S. S. Gubser, C. P. Herzog, and I. R. Klebanov, "Symmetry breaking and axionic strings in the warped deformed conifold," *Journal of High Energy Physics*, vol. 8, no. 9, pp. 795–820, 2004.
- [124] H. Firouzjahi, L. Leblond, and S.-H. H. Tye, "The  $(p, q)$  string tension in a warped deformed conifold," *Journal of High Energy Physics*, vol. 2006, no. 5, article 47, 2006.
- [125] M. G. Jackson, "A note on Cosmic  $(p, q, r)$  strings," *Physical Review D*, vol. 75, no. 8, Article ID 087301, 2007.
- [126] E. J. Copeland, T. W. B. Kibble, and D. A. Steer, "Constraints on string networks with junctions," *Physical Review D*, vol. 75, no. 6, Article ID 065024, 12 pages, 2007.

- [127] R. J. Rivers and D. A. Steer, "Statistical mechanics of strings with Y-junctions," *Physical Review D*, vol. 78, no. 2, Article ID 023521, 13 pages, 2008.
- [128] N. Bevis, E. J. Copeland, P.-Y. Martin, et al., "Evolution and stability of cosmic string loops with Y-junctions," *Physical Review D*, vol. 80, no. 12, Article ID 125030, 14 pages, 2009.
- [129] P. M. Saffin, "A practical model for cosmic  $(p, q)$  superstrings," *Journal of High Energy Physics*, vol. 2005, no. 9, pp. 283–295, 2005.
- [130] I. Cho, Y. Kim, and B. Kyae, "DF-strings from  $D3\bar{D}3$  as cosmic strings," *Journal of High Energy Physics*, vol. 2006, no. 4, article 12, 2006.
- [131] M. Hindmarsh and P. M. Saffin, "Scaling in a  $SU(2)/\mathbb{Z}_3$  model of cosmic superstring networks," *Journal of High Energy Physics*, vol. 2006, no. 8, article 66, 2006.
- [132] A. Rajantie, M. Sakellariadou, and H. Stoica, "Numerical experiments with  $p$  F- and  $p$  D-strings: the formation of  $(p, q)$  bound states," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, article 21, 2007.
- [133] B. Carter, "Stability and characteristic propagation speeds in superconducting cosmic and other string models," *Physics Letters B*, vol. 228, no. 4, pp. 466–470, 1989.
- [134] B. Carter, "Outer curvature and conformal geometry of an imbedding," *Journal of Geometry and Physics*, vol. 8, pp. 53–88, 1992.
- [135] R. M. Wald, *General Relativity*, University of Chicago Press, Chicago, Ill, USA, 1984.
- [136] B. Carter, M. Sakellariadou, and X. Martin, "Cosmological expansion and thermodynamic mechanisms in cosmic string dynamics," *Physical Review D*, vol. 50, no. 2, pp. 682–699, 1994.
- [137] B. Carter, "Integrable equation of state for noisy cosmic string," *Physical Review D*, vol. 41, no. 12, pp. 3869–3872, 1990.
- [138] B. Carter, "Transonic elastic model for wiggly Goto-Nambu string," *Physical Review Letters*, vol. 74, no. 16, pp. 3098–3101, 1995.
- [139] B. Carter and D. A. Steer, "Symplectic structure for elastic and chiral conducting cosmic string models," *Physical Review D*, vol. 69, no. 12, Article ID 125002, 9 pages, 2004.
- [140] B. Carter and P. Peter, "Dynamics and integrability property of the chiral string model," *Physics Letters B*, vol. 466, no. 1, pp. 41–49, 1999.
- [141] T. Goto, "Relativistic quantum mechanics of one-dimensional mechanical continuum and subsidiary condition of dual resonance model," *Progress of Theoretical Physics*, vol. 46, no. 5, pp. 1560–1569, 1971.
- [142] D. P. Bennett and F. R. Bouchet, "Cosmic-string evolution," *Physical Review Letters*, vol. 63, no. 26, pp. 2776–2779, 1989.
- [143] A. Albrecht and N. Turok, "Evolution of cosmic string networks," *Physical Review D*, vol. 40, no. 4, pp. 973–1001, 1989.
- [144] D. P. Bennett and F. R. Bouchet, "High-resolution simulations of cosmic-string evolution. I. Network evolution," *Physical Review D*, vol. 41, no. 8, pp. 2408–2433, 1990.
- [145] B. Allen and E. P. S. Shellard, "Cosmic-string evolution: a numerical simulation," *Physical Review Letters*, vol. 64, no. 2, pp. 119–122, 1990.
- [146] C. Ringeval, M. Sakellariadou, and F. R. Bouchet, "Cosmological evolution of cosmic string loops," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 2, article 23, 2007.
- [147] C. J. A. P. Martins and E. P. S. Shellard, "Fractal properties and small-scale structure of cosmic string networks," *Physical Review D*, vol. 73, no. 4, Article ID 043515, 6 pages, 2006.
- [148] K. D. Olum and V. Vanchurin, "Cosmic string loops in the expanding universe," *Physical Review D*, vol. 75, no. 6, Article ID 063521, 2007.
- [149] J. Urrestilla, N. Bevis, M. Hindmarsh, M. Kunz, and A. R. Liddle, "Cosmic microwave anisotropies from BPS semilocal strings," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 7, article 10, 2008.
- [150] J. Urrestilla and A. Vilenkin, "Evolution of cosmic superstring networks: a numerical simulation," *Journal of High Energy Physics*, vol. 2008, no. 2, article 37, 2008.
- [151] A. Hanany and K. Hashimoto, "Reconnection of colliding cosmic strings," *Journal of High Energy Physics*, vol. 2005, no. 6, article 21, 2005.
- [152] A. Achúcarro and R. De Putter, "Effective nonintercommutation of local cosmic strings at high collision speeds," *Physical Review D*, vol. 74, no. 12, Article ID 121701, 2006.
- [153] E. P. S. Shellard, "Cosmic string interactions," *Nuclear Physics B*, vol. 283, pp. 624–656, 1987.
- [154] R. A. Matzner, "Interaction of  $U(1)$  cosmic strings: numerical intercommutation," *Computers in Physics*, vol. 2, pp. 51–64, 1988.
- [155] L. M. A. Bettencourt and T. W. B. Kibble, "Non-intercommuting configurations in the collisions of type I  $U(1)$  cosmic strings," *Physics Letters B*, vol. 332, no. 3–4, pp. 297–304, 1994.
- [156] L. M. A. Bettencourt, P. Laguna, and R. A. Matzner, "Non-intercommuting cosmic strings," *Physical Review Letters*, vol. 78, no. 11, pp. 2066–2069, 1997.
- [157] P. Salmi, A. Achúcarro, E. J. Copeland, T. W. B. Kibble, R. De Putter, and D. A. Steer, "Kinematic constraints on formation of bound states of cosmic strings: field theoretical approach," *Physical Review D*, vol. 77, no. 4, Article ID 041701, 2008.
- [158] P. Laguna and R. A. Matzner, "Numerical simulation of bosonic-superconducting-string interactions," *Physical Review D*, vol. 41, no. 6, pp. 1751–1763, 1990.
- [159] M. G. Jackson, N. T. Jones, and J. Polchinski, "Collisions of cosmic F- and D-strings," *Journal of High Energy Physics*, vol. 2005, no. 10, article 13, 2005.
- [160] E. J. Copeland, T. W. B. Kibble, and D. A. Steer, "Collisions of strings with y junctions," *Physical Review Letters*, vol. 97, no. 2, Article ID 021602, 2006.
- [161] E. J. Copeland, H. Firouzjahi, T. W. B. Kibble, and D. A. Steer, "Collision of cosmic superstrings," *Physical Review D*, vol. 77, Article ID 063521, 2008.
- [162] N. Bevis and P. M. Saffin, "Cosmic string Y-junctions: a comparison between field theoretic and Nambu-Goto dynamics," *Physical Review D*, vol. 78, no. 2, Article ID 023503, 2008.
- [163] P. Binetruy, A. Bohe, T. Hertog, and D. A. Steer, "Proliferation of sharp kinks on cosmic (super-)string loops with junctions," *Journal of High Energy Physics*, <http://arxiv.org/abs/1005.2426>.
- [164] T. Vachaspati and A. Vilenkin, "Formation and evolution of cosmic strings," *Physical Review D*, vol. 30, no. 10, pp. 2036–2045, 1984.
- [165] T. W. B. Kibble, "Configuration of  $Z_2$  strings," *Physics Letters B*, vol. 166, no. 3, pp. 311–313, 1986.
- [166] A. Yates and T. W. B. Kibble, "An extension to models for cosmic string formation," *Physics Letters B*, vol. 364, no. 3, pp. 149–156, 1995.
- [167] J. Borrill, T. W. B. Kibble, T. Vachaspati, and A. Vilenkin, "Defect production in slow first-order phase transitions," *Physical Review D*, vol. 52, no. 4, pp. 1934–1943, 1995.

- [168] R. J. Scherrer and A. Vilenkin, ““Lattice-free” simulations of topological defect formation,” *Physical Review D*, vol. 58, no. 10, Article ID 103501, 8 pages, 1998.
- [169] R. J. Scherrer and A. Vilenkin, “Cosmic string formation from correlated fields,” *Physical Review D*, vol. 56, no. 2, pp. 647–652, 1997.
- [170] A. Vilenkin, “Gravitational radiation from cosmic strings,” *Physics Letters B*, vol. 107, no. 1-2, pp. 47–50, 1981.
- [171] T. Damour and A. Vilenkin, “Gravitational wave bursts from cusps and kinks on cosmic strings,” *Physical Review D*, vol. 64, no. 6, Article ID 064008, 2001.
- [172] X. Siemens, J. Creighton, I. Maor, S. R. Majumder, K. Cannon, and J. Read, “Gravitational wave bursts from cosmic (super)strings: quantitative analysis and constraints,” *Physical Review D*, vol. 73, no. 10, Article ID 105001, 2006.
- [173] S. Olmez, V. Mandic, and X. Siemens, “Gravitational-wave stochastic background from kinks and cusps on cosmic strings,” *Physical Review D*, vol. 81, Article ID 104028, 10 pages, 2010.
- [174] T. Vachaspati, “Cosmic sparks from superconducting strings,” *Physical Review Letters*, vol. 101, no. 14, Article ID 141301, 2008.
- [175] T. Vachaspati, “Cosmic rays from cosmic strings with condensates,” *Physical Review D*, vol. 81, no. 4, Article ID 043531, 2010.
- [176] D. Austin, E. J. Copeland, and T. W. B. Kibble, “Evolution of cosmic string configurations,” *Physical Review D*, vol. 48, no. 12, pp. 5594–5627, 1993.
- [177] E. J. Copeland, T. W. B. Kibble, and D. A. Steer, “Evolution of a network of cosmic string loops,” *Physical Review D*, vol. 58, no. 4, Article ID 043508, 14 pages, 1998.
- [178] C. J. A. P. Martins and E. P. S. Shellard, “Fractal properties and small-scale structure of cosmic string networks,” *Physical Review D*, vol. 73, no. 4, Article ID 043515, 6 pages, 2006.
- [179] J. Polchinski and J. V. Rocha, “Analytic study of small scale structure on cosmic strings,” *Physical Review D*, vol. 74, no. 8, Article ID 083504, 2006.
- [180] F. Dubath, J. Polchinski, and J. V. Rocha, “Cosmic string loops, large and small,” *Physical Review D*, vol. 77, no. 12, Article ID 123528, 2008.
- [181] E. J. Copeland and T. W. B. Kibble, “Kinks and small-scale structure on cosmic strings,” *Physical Review D*, vol. 80, no. 12, Article ID 123523, 7 pages, 2009.
- [182] J. V. Rocha, “Scaling solution for small cosmic string loops,” *Physical Review Letters*, vol. 100, no. 7, Article ID 071601, 2008.
- [183] B. Allen and E. P. S. Shellard, “Gravitational radiation from cosmic strings,” *Physical Review D*, vol. 45, no. 6, pp. 1898–1912, 1992.
- [184] X. Siemens, K. D. Olum, and A. Vilenkin, “Size of the smallest scales in cosmic string networks,” *Physical Review D*, vol. 66, no. 4, Article ID 043501, 4 pages, 2002.
- [185] G. R. Vincent, M. Hindmarsh, and M. Sakellariadou, “Scaling and small-scale structure in cosmic string networks,” *Physical Review D*, vol. 56, no. 2, pp. 637–646, 1997.
- [186] R. Durrer and Z.-H. Zhou, “Large-scale structure formation with global topological defects,” *Physical Review D*, vol. 53, no. 10, pp. 5394–5410, 1996.
- [187] N. Turok, “Causality and the Doppler peaks,” *Physical Review D*, vol. 54, no. 6, pp. R3686–R3689, 1996.
- [188] U.-L. Pen, U. Seljak, and N. Turok, “Power spectra in global defect theories of cosmic structure formation,” *Physical Review Letters*, vol. 79, no. 9, pp. 1611–1614, 1997.
- [189] R. Durrer and M. Kunz, “Cosmic microwave background anisotropies from scaling seeds: generic properties of the correlation functions,” *Physical Review D*, vol. 57, no. 6, pp. R3199–R3203, 1998.
- [190] C. Contaldi, M. Hindmarsh, and J. Magueijo, “Power spectra of the cosmic microwave background and density fluctuations seeded by local cosmic strings,” *Physical Review Letters*, vol. 82, no. 4, pp. 679–682, 1999.
- [191] J.-H. P. Wu, P. P. Avelino, E. P. S. Shellard, and B. Allen, “Cosmic strings, loops, and linear growth of matter perturbations,” *International Journal of Modern Physics D*, vol. 11, no. 1, pp. 61–102, 2002.
- [192] N. Bevis, M. Hindmarsh, M. Kunz, and J. Urrestilla, “CMB power spectrum contribution from cosmic strings using field-evolution simulations of the Abelian Higgs model,” *Physical Review D*, vol. 75, no. 6, Article ID 065015, 2007.
- [193] N. Bevis, M. Hindmarsh, M. Kunz, and J. Urrestilla, “CMB power spectra from cosmic strings: predictions for the Planck satellite and beyond,” *Journal of High Energy Physics*, <http://arxiv.org/abs/1005.2663>.
- [194] N. Bevis, M. Hindmarsh, M. Kunz, and J. Urrestilla, “CMB power spectrum contribution from cosmic strings using field-evolution simulations of the Abelian Higgs model,” *Physical Review D*, vol. 75, no. 6, Article ID 065015, 22 pages, 2007.
- [195] J. Magueijo, A. Albrecht, D. Coulson, and P. Ferreira, “Doppler peaks from active perturbations,” *Physical Review Letters*, vol. 76, no. 15, pp. 2617–2620, 1996.
- [196] N. Bevis, M. Hindmarsh, M. Kunz, and J. Urrestilla, “Fitting cosmic microwave background data with cosmic strings and inflation,” *Physical Review Letters*, vol. 100, no. 2, Article ID 021301, 4 pages, 2008.
- [197] D. P. Bennett, A. Stebbins, and F. R. Bouchet, “The implications of the COBE diffuse microwave radiation results for cosmic strings,” *Astrophysical Journal*, vol. 399, no. 1, pp. L5–L8, 1992.
- [198] U.-L. Pen, D. N. Spergel, and N. Turok, “Cosmic structure formation and microwave anisotropies from global field ordering,” *Physical Review D*, vol. 49, no. 2, pp. 692–729, 1994.
- [199] B. Allen, R. R. Caldwell, S. Dodelson, L. Knox, E. P. S. Shellard, and A. Stebbins, “Cosmic microwave background anisotropy induced by cosmic strings on angular scales  $\gtrsim 15^\circ$ ,” *Physical Review Letters*, vol. 79, no. 14, pp. 2624–2627, 1997.
- [200] M. Landriau and E. P. S. Shellard, “Fluctuations in the cosmic microwave background induced by cosmic strings: methods and formalism,” *Physical Review D*, vol. 67, no. 10, Article ID 103512, 2003.
- [201] M. Landriau and E. P. S. Shellard, “Large angle cosmic microwave background fluctuations from cosmic strings with a cosmological constant,” *Physical Review D*, vol. 69, no. 2, Article ID 023003, 2004.
- [202] A. A. Fraisse, C. Ringeval, D. N. Spergel, and F. R. Bouchet, “Small-angle CMB temperature anisotropies induced by cosmic strings,” *Physical Review D*, vol. 78, no. 4, Article ID 043535, 2008.
- [203] M. Hindmarsh, “Small-scale microwave background fluctuations from cosmic strings,” *Astrophysical Journal*, vol. 431, no. 2, pp. 534–542, 1994.
- [204] A. Stebbins and S. Veeraraghavan, “Beyond the small-angle approximation for MBR anisotropy from seeds,” *Physical Review D*, vol. 51, no. 4, pp. 1465–1478, 1995.

- [205] J. R. Gott III, C. Park, R. Juskiewicz et al., “Topology of microwave background fluctuations: theory,” *Astrophysical Journal*, vol. 352, no. 1, pp. 1–14, 1990.
- [206] K. Takahashi, A. Naruko, Y. Sendouda, D. Yamauchi, C.-M. Yoo, and M. Sasaki, “Non-Gaussianity in the cosmic microwave background temperature fluctuations from cosmic (super-)strings,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 10, article 3, 2009.
- [207] K. Falconer, *Fractal Geometry*, John Wiley & Sons, Chichester, UK, 2006.
- [208] M. P. Pompilio, F. R. Bouchet, G. Murante, and A. Provenzale, “Multifractal analysis of string-induced cosmic microwave background radiation anisotropies,” *Astrophysical Journal*, vol. 449, no. 1, pp. 1–8, 1995.
- [209] M. P. Hobson, A. W. Jones, and A. N. Lasenby, “Wavelet analysis and the detection of non-Gaussianity in the cosmic microwave background,” *Monthly Notices of the Royal Astronomical Society*, vol. 309, no. 1, pp. 125–140, 1999.
- [210] R. B. Barreiro and M. P. Hobson, “The discriminating power of wavelets to detect non-Gaussianity in the cosmic microwave background,” *Monthly Notices of the Royal Astronomical Society*, vol. 327, no. 3, pp. 813–828, 2001.
- [211] D. K. Hammond, Y. Wiaux, and P. Vanderghaynst, “Wavelet domain Bayesian denoising of string signal in the cosmic microwave background,” *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 3, pp. 1317–1332, 2009.
- [212] M. Hindmarsh, C. Ringeval, and T. Suyama, “CMB temperature bispectrum induced by cosmic strings,” *Physical Review D*, vol. 80, no. 8, Article ID 083501, 2009.
- [213] D. N. Spergel and D. M. Goldberg, “Microwave background bispectrum. I. Basic formalism,” *Physical Review D*, vol. 59, no. 10, Article ID 103001, 8 pages, 1999.
- [214] N. Aghanim, M. Kunz, P. G. Castro, and O. Forni, “Non-gaussianity: comparing wavelet and Fourier based methods,” *Astronomy and Astrophysics*, vol. 406, no. 3, pp. 797–816, 2003.
- [215] M. Hindmarsh, C. Ringeval, and T. Suyama, “CMB temperature trispectrum of cosmic strings,” *Physical Review D*, vol. 81, no. 6, Article ID 063505, 2010.
- [216] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, “Non-Gaussianity from inflation: theory and observations,” *Physics Reports*, vol. 402, no. 3-4, pp. 103–266, 2004.
- [217] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $f_{NL}^{local}$  from WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, 2009.
- [218] D. M. Regan and E. P. S. Shellard, “General CMB and primordial trispectrum estimation,” *Journal of Cosmology and Astroparticle Physics*, <http://arxiv.org/abs/1004.2915>.
- [219] M. Sakellariadou, “A note on the evolution of cosmic string/superstring networks,” *Journal of Cosmology and Astroparticle Physics*, no. 4, article 3, pp. 103–116, 2005.
- [220] A. Avgoustidis and E. P. S. Shellard, “Velocity-dependent models for non-Abelian/entangled string networks,” *Physical Review D*, vol. 78, no. 10, Article ID 103510, 21 pages, 2008.
- [221] D. M. Regan and E. P. S. Shellard, “Cosmic string power spectrum, bispectrum and trispectrum,” *Journal of Cosmology and Astroparticle Physics*, <http://arxiv.org/abs/0911.2491>.
- [222] M. Landriau and E. P. S. Shellard, “Cosmic string induced CMB maps,” *Journal of Cosmology and Astroparticle Physics*, <http://arxiv.org/abs/1004.2885>.

## Research Article

# A Comprehensive Overview of the Cold Spot

**Patricio Vielva**

*Instituto de Física de Cantabria (CSIC—Universidad de Cantabria), Avenida Los Castros s/n, 39005 Santander, Spain*

Correspondence should be addressed to Patricio Vielva, vielva@ifca.unican.es

Received 27 May 2010; Accepted 17 August 2010

Academic Editor: Dragan Huterer

Copyright © 2010 Patricio Vielva. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The report of a significant deviation of the CMB temperature anisotropies distribution from Gaussianity (soon after the public release of the WMAP data in 2003) has become one of the most solid WMAP *anomalies*. This detection grounds on an excess of the kurtosis of the Spherical Mexican HatWavelet coefficients at scales of around 10 degrees. At these scales, a prominent feature—located in the southern Galactic hemisphere—was highlighted from the rest of the SMHW coefficients: the Cold Spot. This paper presents a comprehensive overview related to the study of the Cold Spot, paying attention to the non-Gaussianity detection methods, the morphological characteristics of the Cold Spot, and the possible sources studied in the literature to explain its nature. Special emphasis is made on the Cold Spot compatibility with a *cosmic texture*, commenting on future tests that would help to give support or discard this hypothesis.

## 1. Introduction

Besides the great success of the NASA WMAP satellite on providing a detailed knowledge of the cosmological parameters that define the physical properties of the Universe (e.g., [1, 2]), some unexpected results have attracted the attention of the cosmological community soon after the first release of the WMAP data: the so-called *WMAP anomalies*. Some of these anomalies are related to hemispherical asymmetries (e.g., [3–11]), an anomalous alignment of the quadrupole and octopole components (e.g., [12–21]), significantly low variance of the CMB temperature fluctuations (e.g., [22–25]), or anomalous alignment of the CMB features toward the Ecliptic poles (e.g., [26, 27]). Some of these aspects are addressed in this special issue. In addition to the previous findings, the prominent cold spot (hereinafter, the Cold Spot) detected in the southern hemisphere by [28] became one of the most studied anomalies of the WMAP data. The Cold Spot was identified after testing that the Spherical Mexican Hat Wavelet (SMHW) coefficients of the WMAP data presented an excess of kurtosis (at scales of around  $10^\circ$  in the sky), as compared to the distribution derived from isotropic and Gaussian CMB simulations. This paper presents a complete review on the detection and characterization of this nonstandard signature, and a description of the different attempts made so far in understanding what could be the

cause behind such departure from the standard inflationary paradigm. It is organized as follows: in Section 2, the author justifies the use of wavelets as a natural tool for probing the Gaussianity of the CMB temperature fluctuations. The author also presents the different statistics applied in the wavelet space that had led to point out the WMAP data incompatibility with the standard model. In Section 3, the author briefly describes the morphological characteristics of the Cold Spot. The important question of the actual significance of the detection of the Cold Spot and the aspects associated with a *posteriori* interpretations are addressed in Section 4. Some of the different sources that have been considered in the literature to explain the Cold Spot feature are discussed in Section 5. In Section 6, the author explains in detail a plausible hypothesis to accommodate the existence of the Cold Spot together with the standard cosmological model: a cosmic texture. In addition, the author also describe possible followup tests that could help to confirm or discard such hypothesis. Finally, the author's conclusions are given in Section 7.

## 2. The Non-Gaussianity Detection

The Cold Spot was firstly identified through a *blind* Gaussianity test of the WMAP first-year data [28]. This test

was designed to probe the isotropic and Gaussian nature of the CMB, as predicted by the standard inflationary model (see, e.g., [29]), and it was based in a multiresolution analysis performed with the Spherical Mexican Hat Wavelet (SMHW). In this section, the author summarizes the non-Gaussianity detection that led to the identification of the Cold Spot. The author starts by justifying why an analysis based on wavelets was proposed and presenting the main characteristics of the wavelet used in the analysis: the SMHW. Afterwards, the author explains which statistics (all of them based on the SMHW coefficients) reported the original deviations from Gaussianity of the WMAP data.

**2.1. Why a Wavelet?** Nowadays, the CMB scientific community is already very familiar with the application of wavelets (and other members of the *-lets* zoo, like curvelets, ridglets or needlets) to data analysis. However, it is worth recalling that this is a relatively new custom. Although wavelet applications in cosmology shyly started already in mid-80s, it was not until 1997 that the first application to CMB was presented [30], precisely in an exercise devoted to probe the Gaussianity of the CMB, and it was just a year afterwards that the first CMB data analyses with wavelets came to light [31, 32], in particular, with two applications to COBE [33] data. During the last decade, the application of wavelets to extract cosmological information from CMB data has growth considerably, in many different branches: compact source detection (e.g., [34–39]), Gaussianity (e.g., [28, 40–51]), cross-correlation with large scale structure (e.g., [52–55]), decomposition of the coupled E/B signals [56], probing isotropy (e.g., [9, 26, 27]), cosmic string detection [57], microwave sky recovery (e.g., [58–61]), CMB denoising (e.g., [34, 62]), CMB power-spectrum determination [63], and primordial power-spectrum recovery (e.g., [64]) are some of the application fields. The author refers the reader to [65, 66] for some reviews on the wavelets applications to CMB data, with a particular emphasis on data analysis on the sphere.

The wavelet transform (e.g., [67, 68]) has become very popular for a major reason: they offer a unique opportunity to probe scale-dependent phenomena, but keeping, at the same time, information about spatial localization. This is a clear advantage—for many purposes—over classical Fourier or harmonic transforms: physical processes typically exhibit a clear scale-dependent behaviour, and, often, such behaviour differs enough from one phenomenon to another (e.g., in a microwave image, the localized emission due to cluster of galaxies has very different properties as compared to the large-scale signal produced by the Galactic components).

The capability of emphasising or amplifying some features (at a particular scale) makes wavelets unique to probe the Gaussianity of the CMB: there are many different physical processes that might introduce non-Gaussian signatures into a CMB signal, at a very particular scale range (e.g., primordial non-Gaussianity due to nonstandard inflationary scenarios, cosmic defects like strings or textures, secondary anisotropies, foreground emissions, etc.).

In this sense, when a Gaussianity analysis of the CMB is performed in wavelet space, we are in a very adequate

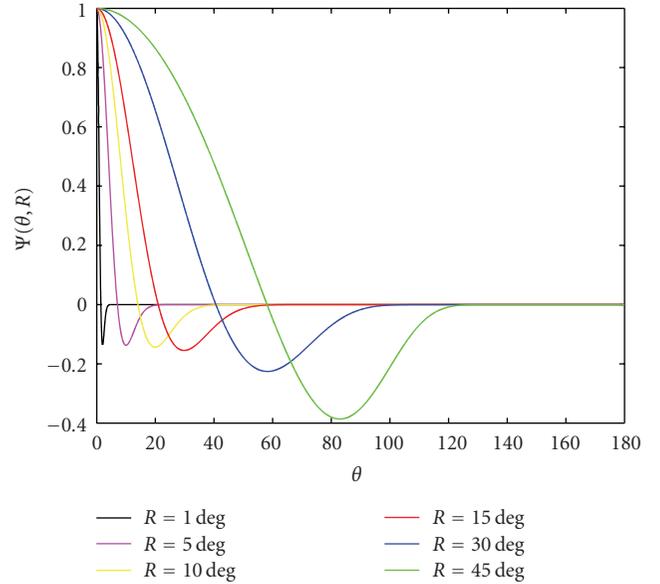


FIGURE 1: Figure shows the profile on the SMHW as a function of the angular distance  $\theta$  for different wavelet scales  $R$ , ranging from  $1^\circ$  to  $45^\circ$ . Notice that the amplitude of the SMHW at  $\theta \equiv 0$  has been fixed to 1, to allow for an easier comparison.

framework to probe, almost separately, any potential non-Gaussian signatures present in the data. The ability of wavelets to amplify a given signature is explained because they can be seen as compensated filters (i.e.,  $\int_{\mathcal{R}^n} \Psi = 0$ , where  $\mathcal{R}^n$  is the space where the wavelet  $\Psi$  is defined). This property is not satisfied by other standard filters, as the top-hat or the Gaussian functions. As mentioned above, wavelets do not only provide us with the capability of selecting a given scale range, but they also allow one to keep the information about spatial localization: we do not just study, for instance, the compatibility with Gaussianity, but we are also able to identify where in the data a deviation might be spatially localized. This intrinsic property of the wavelet transform (spatial localization) is also unique to explore not only Gaussianity, but also isotropy, since the statistical properties can be studied (almost independently) from one region in the data to another, in a self-consistent way. Hereinafter, the author will focus on the use of the two-dimensional (2D) continuous wavelet transform (CWT, see, e.g., Chapter 2 in [68]) and, in particular, on the 2D CWT defined on the sphere (e.g., [69–72]); most of the works related to the Cold Spot have been performed using the definition of the isotropic Spherical Mexican Hat Wavelet (SMHW) proposed by [45] which is a stereographic projection of the Mexican Hat Wavelet, as proposed by [69]

$$\Psi(\theta; R) = \frac{1}{\sqrt{2\pi R N_R}} \left[ 1 + \left( \frac{y}{2} \right)^2 \right]^2 \left[ 2 - \left( \frac{y}{R} \right)^2 \right] e^{-y^2/2R^2}, \quad (1)$$

where  $y \equiv 2 \tan(\theta/2)$  is the stereographic projection variable,  $\theta \in [0, \pi)$  is the colatitude, and the constant  $N_R \equiv \sqrt{1 + (R^2/2) + (R^4/4)}$  is chosen such as the square of the

wavelet function  $\Psi(\theta; R)$  is normalized to unity. In Figure 1, the radial profile of the SMHW, for different wavelet scales  $R$ , is shown. For a signal  $T(\theta, \phi)$  defined on the sphere, its spherical harmonic coefficients  $t_{\ell m}$  are defined as

$$t_{\ell m} = \int d\Omega Y_{\ell m}^*(\theta, \phi) T(\theta, \phi), \quad (2)$$

where  $d\Omega = d\theta \sin\theta d\phi$ , the spherical coordinates are, as mentioned above, the colatitude  $\theta$  (related to the latitude  $b$  as  $b = \pi/2 - \theta$ ), and  $\phi \in [0, 2\pi)$  is the longitude. The function  $Y_{\ell m}(\theta, \phi)$  is the spherical harmonic of order  $\ell$  and  $m$ , and  $*$  denotes complex conjugation. For an isotropic 2D CWT on the sphere, the wavelet coefficients  $w(\theta, \phi; R)$  are obtained as

$$w(\theta, \phi; R) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{\ell} t_{\ell m} \Psi_{\ell}(R) Y_{\ell m}(\theta, \phi) T(\theta, \phi), \quad (3)$$

where  $\Psi_{\ell}(R)$  is the window function associated with the wavelet function (e.g.,  $\Psi(\theta; R)$  in (1)), and  $\ell_{\max}$  represents the maximum multipole associated with a given resolution of the signal  $T(\theta, \phi)$ , typically limited by the size of the pixel adopted to represent such signal on the sphere.

**2.2. The Statistics.** The Cold Spot was firstly detected via a positive deviation of the kurtosis of the SMHW coefficients at scales of around  $R \approx 300^\circ$ . The inspection of the map of the SMHW coefficients at these scales revealed the presence of a very large and cold spot in the southern hemisphere. The comparison of the amplitude and the area of this cold spot as compared with Gaussian simulations showed that it was particularly anomalous. Finally, a higher criticism test of the SMHW coefficients also indicated a deviation, at around the same wavelet scales, also showing that the major source for such deviation was located in the position of the Cold Spot. In the following subsections, the author describes briefly these statistics: the kurtosis, the amplitude, the area and the higher criticism.

**2.2.1. The Kurtosis.** The two most obvious indicators for a possible deviation from Gaussianity of a given data set  $z = \{z_1, z_2, \dots, z_N\}$  are the third ( $\mu_3$ ) and fourth ( $\mu_4$ ) central moments. The  $n$ -central moment of a random set of  $N$  numbers is defined as

$$\mu_n = \frac{1}{N} \sum_{i=1}^N (z_i - \bar{z})^n, \quad (4)$$

where  $\bar{z} = (1/N) \sum_{i=1}^N z_i$  is the mean value of the data set. For a Gaussian random data set of zero mean (i.e.,  $\mu_1 = \bar{z} \equiv 0$ ) and dispersion  $\sigma$  (i.e.,  $\mu_2 \equiv \sigma^2$ ), it is trivial to prove that higher order central moments are either zero (in the case of  $\mu_{2n+1}$ , for  $n \geq 1$ ) or a given function of the dispersion (in the case of  $\mu_{2n}$ , for  $n \geq 1$ ). Usually, it is much more convenient to work with normalized central moments  $\nu_n$ , such as

$$\nu_n = \frac{\mu_n}{\sigma^n}. \quad (5)$$

The normalized central moments are more convenient than central moments, since they are referred to the intrinsic

fluctuations of the random data set (represented by the dispersion). This helps to absorb into  $\nu_n$  possible uncertainties on the knowledge on the amplitude of the fluctuations of the random sample. The normalized central moment  $\nu_3$  is normally referred to as the skewness ( $S$ ) whereas  $K \equiv \nu_4 - 3$  is called kurtosis. The reason for the subtraction of the number 3 in the previous expression comes from the fact that for a Gaussian random variable,  $\mu_4 = 3\sigma^4$  and, therefore,  $\nu_4 = 3$ . The previous definition assures that as it happens for the skewness ( $S$ ), the kurtosis ( $K$ ) of a Gaussian field is zero. Applying these concepts to the SMHW coefficients  $w(\theta, \phi; R)$ , the skewness  $S_R$  and the kurtosis  $K_R$  of the wavelet coefficients, as a function of the wavelet scale  $R$ , can be defined as follows:

$$S_R = \frac{1}{\sigma_R^3} \frac{1}{N_{\text{pix}}(R)} \sum_{i=1}^{N_{\text{pix}}(R)} w(\theta_i, \phi_i; R)^3, \quad (6)$$

$$K_R = \frac{1}{\sigma_R^4} \frac{1}{N_{\text{pix}}(R)} \sum_{i=1}^{N_{\text{pix}}(R)} w(\theta_i, \phi_i; R)^4 - 3,$$

where at each scale, it is assumed that the coefficients  $w(\theta, \phi; R)$  have zero mean.  $\sigma_R$  is the dispersion of the wavelet coefficients at the scale  $R$

$$\sigma_R = \left[ \frac{1}{N_{\text{pix}}(R)} \sum_{i=1}^{N_{\text{pix}}(R)} w(\theta_i, \phi_i; R)^2 \right]^{1/2}. \quad (7)$$

In the previous expressions,  $N_{\text{pix}}(R)$  represents the number of wavelet coefficients at a given scale  $R$ . Notice that very often, CMB data is not available in the full celestial sphere (e.g., because strong contamination from astrophysical foregrounds has to be masked). In most of the works in the CMB field, the HEALPix tessellation [73] is adopted. In this scheme, the resolution of a given image represented on the sphere is given by the  $N_{\text{SIDE}}$  parameter, which indicates how many divisions of the 12 basic pixels are required to achieve such resolution. The  $N_{\text{SIDE}}$  parameter is related with the number of the pixels ( $N_{\text{pix}}$ ) required to fill the sphere at that resolution as:  $N_{\text{pix}} = 12N_{\text{SIDE}}^2$ . In Figure 2, the major result of the seminal work [28] on the Cold Spot is shown. It represents the kurtosis of the wavelet coefficients, as a function of the scale, for the first release of the WMAP data (blue stars). The solid yellow line represents the mean value obtained from 10,000 CMB Gaussian random simulations, taking into account the instrumental properties of the analyzed data, and generated from an angular power spectrum derived from the best-fit cosmological model. The coloured regions (red, green, and magenta) represent the acceptance intervals at 32%, 5%, and 1%, respectively. Notice that at  $R$  scales of 250 and 300 arcmin, the kurtosis of the WMAP data was above the 1% acceptance interval. In detail, the excess of kurtosis is given by a  $p$ -value of  $\approx 4 \times 10^{-3}$  for the two scales. No significant deviations were found related to the skewness of the wavelet coefficients. The analysis was repeated in the two Galactic hemispheres separately. This was motivated by previous findings of asymmetries related to the genus [74] and the N-point correlation function [3].

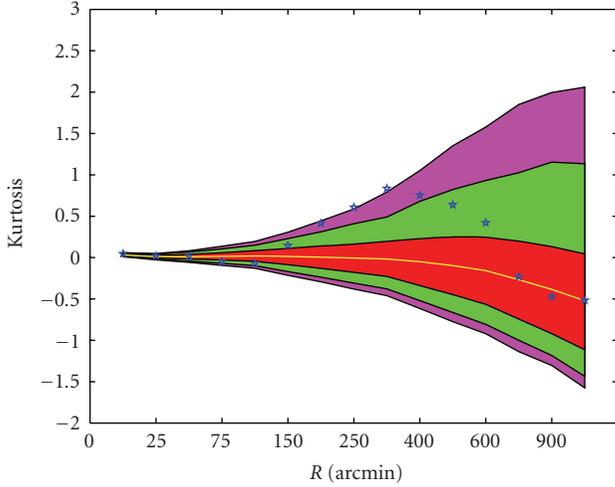


FIGURE 2: Figure showing the positive deviation of the kurtosis of the wavelet coefficients at scales  $R$  of 250 and 300 arcmin found by [28]. Red, green and magenta regions represent the acceptance intervals at 32%, 5%, and 1%, respectively.

It was found that whereas there was not deviation on the northern hemisphere, the excess of kurtosis was even more remarkable in the southern region. In particular, at the scale of 250 arcmin, the kurtosis of the SMHW coefficients was associated with a  $p$ -value of  $\approx 2 \times 10^{-3}$ . Again, no deviation on the skewness was noticed. Hence, these analyses indicated that the source for the deviation on the kurtosis of the wavelet coefficient was related to feature/s with a typical length of around  $10^\circ$  in the sky and located in the southern hemisphere. Thanks to the frequency/spatial properties of the SMHW decomposition, it was possible to study in detail such features. In the left panel of Figure 3, the cleaned CMB map obtained from the fourth release of the WMAP data is given. In particular, the author plot the optimal combination for cosmological analysis (hereinafter, the QVW map), obtained as a noise-weighted linear combination of the Q-, V-, and W-bands, previously cleaned via a template fitting (e.g., [75]). Regions highly contaminated by Galactic foregrounds, nearby clusters, and extragalactic point sources have been masked. In the right panel, the SMHW coefficients of the previous map are represented, at the scale  $R = 250$  arcmin. It is evident the presence of the Cold Spot in the southwest side of the image. In particular, the centre of the Cold Spot is estimated to be  $\theta = 147^\circ$  and  $\phi = 209^\circ$ . The study of the Cold Spot, through the application of follow-up tests, provided further evidences for its anomalous nature. The author reviews these tests in the following subsections.

**2.2.2. The Amplitude.** One of the most trivial statistics to study extreme values (as the Cold Spot) in a random sample is the largest/smallest observation. In [28], it was established that the temperature of the Cold Spot was  $-4.57$  times the dispersion of the SMHW coefficients at  $R = 250$  arcmin. This cold value represented a  $p$ -value of .01 (relative to Monte Carlo simulations). A more robust statistic related to the

extreme values is the MAX statistic, understood as the largest observation (in absolute value). For the particular case of the SMHW coefficients, MAX is defined, at scale  $R$ , as

$$\text{MAX}_R = \max\{|w(\theta_i, \phi_i; R)|\}. \quad (8)$$

The MAX statistic is more robust than selecting the coldest of the extrema, since the selection of the lowest values could be seen as an *a posteriori* selection. This statistic was studied in [76], showing that the Cold Spot was always the maximum absolute observation of the WMAP data at scales around 300 arcmin, representing an upper tail probability of 0.38% (relative to Monte Carlo simulations). This value was less significant than the one mentioned in the previous subsection. The reason for this change is, as commented, that the MAX statistic is more robust than simply selecting the smallest values of the observations.

**2.2.3. The Area.** The area above or below a given threshold is one of the most common statistics used to characterize the properties of a random field. In particular, the area is the most commonly Minkowski functional used in the literature (see, e.g., [77–79]). (For 2D images, there are three Minkowski functionals, namely, the contour or length, the area, and the genus. These three quantities are defined above/below a given threshold.) Generalizing this concept to the case of the wavelet coefficients, we can define cold ( $A_R^{-\nu}$ ) and hot ( $A_R^{+\nu}$ ) areas, at a given threshold  $\nu$  and a given scale  $R$ , as

$$\begin{aligned} A_R^{-\nu} &= \#\{w(\theta_i, \phi_i; R) < -\nu\}, \\ A_R^{+\nu} &= \#\{w(\theta_i, \phi_i; R) \geq +\nu\}, \end{aligned} \quad (9)$$

where the *number operator*  $\#\{\text{condition}_i\}$  indicates how many times  $\text{condition}_i$  is satisfied, for  $i$  ranging from 1 to  $N_{\text{pix}}(R)$ . The cold and hot areas of the WMAP data were analyzed by [46]. It was reported that whereas the hot area was consistent with the expected behaviour for the standard Gaussian model (at all the scales  $R$  and thresholds  $\nu$ ), the cold area was not compatible. In particular, deviations from Gaussianity were found, again, at SMHW scales of  $R \approx 300^\circ$ . The deviation took place for thresholds equal or smaller than  $-3\sigma_R$  (see Figure 4). The analysis per different regions of the sky confirmed that the anomaly on the cold area was localized in the southern-west Galactic quadrant of the sky, and that the Cold Spot was responsible for this anomaly. In particular, the cold area of the WMAP data (at the mentioned scales, and below a threshold of  $-3\sigma_R$ ) was found anomalous with a probability of  $\approx 99.7\%$  whereas it became fully compatible with Gaussian simulations, once the Cold Spot was not considered in the analysis. As for the case of the MAX statistic, a more conservative estimator (i.e., less dependent from the fact that the Cold Spot is negative) can be considered, just by selecting the maximum value of the previous cold and hot areas

$$A_R^\nu = \max\{A_R^{-\nu}, A_R^{+\nu}\}. \quad (10)$$

This new statistic was used by [76], finding, again, that the WMAP data was anomalous about thresholds larger than  $|3\sigma_R|$ , for scales of the SMHW of around 300 arcmin.

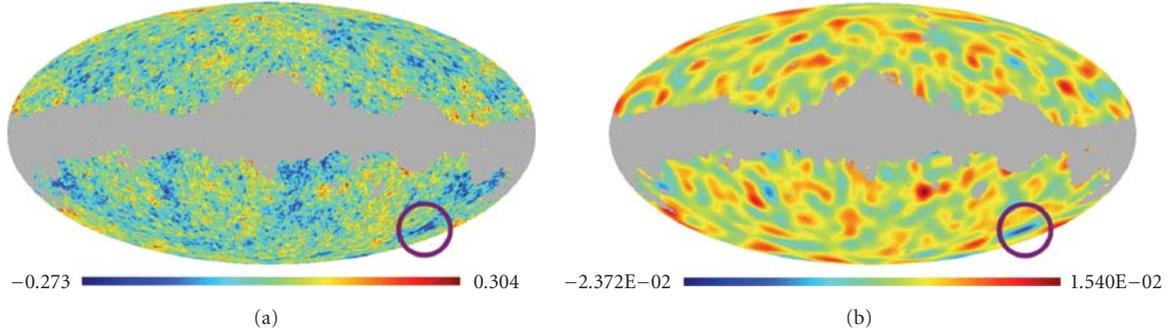


FIGURE 3: (a) CMB cleaned map derived from the 4th release of the WMAP data, obtained via the template fitting technique described in [75]. (b) wavelet coefficients of the previous map, obtained after the SMHW convolution at a scale of  $R = 250$  arcmin. The location of the Cold Spot is indicated in both panels by the circle. The centre of the Cold Spot is  $\theta = 147^\circ$  and  $\phi = 210^\circ$ .

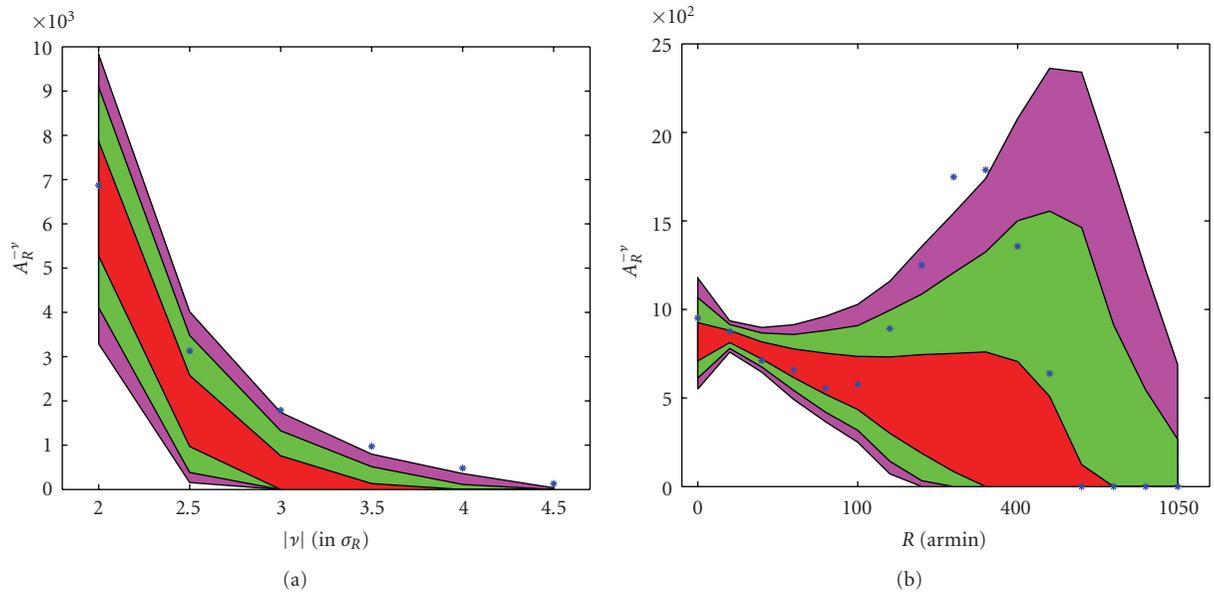


FIGURE 4: (a) cold area of the SMHW coefficients (at  $R = 300$  arcmin), as a function of the threshold ( $\nu$ ). (b) cold area ( $A_R^{-\nu}$ ) of the SMHW coefficients (at  $\nu = 3\sigma_R$ ), as a function of the scale ( $R$ ). As in Figure 2, the red, green, and magenta regions represent the 32%, 5%, and 1% acceptance intervals, respectively. These plots correspond to the analysis done by [46].

**2.2.4. The Higher Criticism.** Higher criticism (HC) is a relatively new statistic introduced in 2004 by [80], and firstly applied to the context of probing the Gaussianity of the CMB only a year after by [81]. Although there is not a unique definition for the HC, all the forms proposed in the literature satisfy the same key concept: HC is a measurement of the distance between a given sample of  $n$  elements to a Gaussian probability density distribution, established by means of the difference between the  $p$ -value  $p_i$  of a given observation  $X_i$ —assuming it comes from a  $N(0, 1)$ —, and its cardinal position on the sorted list (in increasing order) of  $p$ -values  $p_i$  (i.e.,  $p_{i-1} < p_i < p_{i+1}$ , for all  $i = 1, \dots, n$ ). The HC associated with the sample is just defined as the largest value of such differences.

This concept can be applied to the SMHW coefficients of a given signal (e.g., the QVW map) at a given scale  $R$ . This was the analysis proposed by [81]. Let us adopt

the following definition for the HC associated with  $N_{\text{pix}}(R)$  wavelet coefficients  $w(\theta_i, \phi_i; R)$ , at scale  $R$ :

$$\text{HC}_{N_{\text{pix}}}(R) = \max\{\text{HC}_{N_{\text{pix}}}^i(R)\}, \quad (11)$$

where the maximization is made over the quantity  $\text{HC}_{N_{\text{pix}}}^i(R)$ , that provides the difference between the *experimental* probability of the wavelet coefficients  $w(\theta_i, \phi_i; R)$  at scale  $R$  and the corresponding *theoretical*  $p$ -value. Such quantity reads as

$$\text{HC}_{N_{\text{pix}}}^i(R) = \sqrt{N_{\text{pix}}} \frac{\left| \left( \frac{i}{N_{\text{pix}}(R)} \right) - p_i(R) \right|}{\sqrt{p_i(R)(1 - p_i(R))}}, \quad (12)$$

where the  $p$ -value is given by  $p_i(R) = P\{|N(0, 1)| > |\hat{w}(\theta_i, \phi_i; R)|\}$ . The  $\hat{\cdot}$  operator indicates that the  $N_{\text{pix}}(R)$  SMHW coefficients at the scale  $R$  have been transformed into

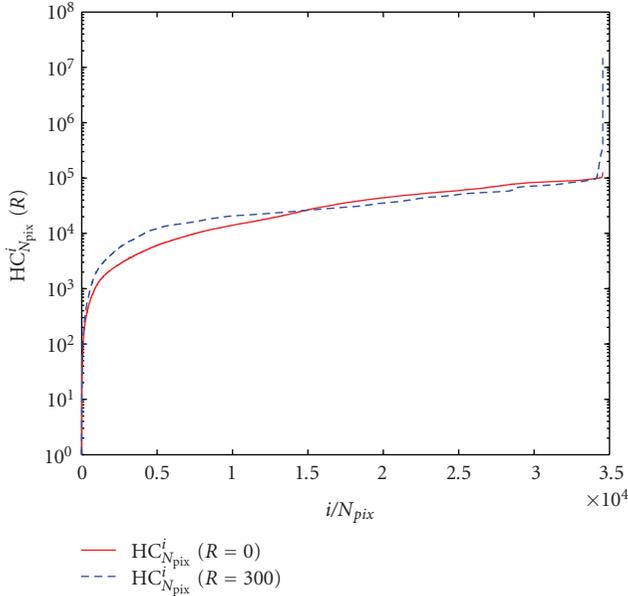


FIGURE 5: The  $HC_{N_{pix}}^i$  values obtained from the analysis of the QVW map. The solid red line corresponds to the application of (12) to the CMB map in the real space (i.e.,  $R \equiv 0$ ) whereas the dot-dash blue line corresponds to the analysis of the SMHW coefficients at scale  $R = 300$  arcmin.  $HC_{N_{pix}}^i$  curves are normalized by their respective minimum values.

a zero mean and unit variance sample. The author remarks that the  $p$ -values have been sorted in increasing order.

As explained in [81], the HC represents, under certain conditions, some advantages with respect to more traditional statistics designed to study the Gaussianity of a given sample. In particular, HC seems to be a better estimator than the MAX statistic: whereas the latter is designed to capture Gaussian deviations caused by very large values of the distribution, the HC is also sensitive to anomalies produced by moderate values. In addition, the HC can identify which values (in a given sample) are the ones that differ from the theoretical Gaussian distribution.

In [81], it was reported that the  $HC_{N_{pix}}(R)$  was above the 1% acceptance interval, again, at the SMHW scale  $R = 300$  arcmin. They found that, actually, all the SMHW coefficients associated with  $HC_{N_{pix}}^i(R)$  values above the 1% acceptance interval set by CMB Gaussian simulations, were localized in the position of the Cold Spot. This extra test was an additional support to the anomalous nature of the WMAP data, and of the Cold Spot in particular. Results were confirmed by [76] for the analysis of the second WMAP data release, reporting an upper tail probability even lower than for the 1-year data.

As an illustration of the HC statistic, in Figure 5, the author represents (in solid red) the  $HC_{N_{pix}}^i$  values obtained from the QVW map in the real space, and (in dot-dash blue) the corresponding curve for  $R = 300$  arcmin. These quantities are normalized to their minimum values, for a better comparison. They are represented in an increasing order. It is remarkable that, for the case of the analyses

performed on the SMHW coefficients, there is a tail of very large values of  $HC_{N_{pix}}^i$ , that are not present for the real space case. This is due to the ability of the SMHW transform to enhance features of a given scale and shape. In Figure 6, these values are represented on the celestial sphere (left panel for the real space case, and right panel for the SMHW coefficients). The figure indicates that there are not particular signatures in the real space whereas the SMHW coefficients at  $R = 300$  arcmin allow us for a clear identification of the features causing the anomalous values of the  $HC_{N_{pix}}(R)$ . In particular, the key role played by the Cold Spot is highlighted.

### 3. The Characteristics of the Cold Spot

In this section, the author summarizes briefly some of the most important properties of the Cold Spot. The author will focus in two major aspects: its morphology and its frequency dependence.

The morphological properties of the Cold Spot are different depending whether we do the analysis in the real or wavelet space. As it was pointed out in [46], the region associated with the Cold Spot, in the real space, appears as formed by several small cold spots. The amplitude of the most prominent of these spots is  $\lesssim -350 \mu\text{K}$  with a size of  $\approx 1^\circ$ . None of these structures is particularly anomalous. The image of the Cold Spot in the real space is shown in the left panel of Figure 7. It is, however, in the wavelet space where the Cold Spot appears more interesting. In the right panel of Figure 7, the author presents a close view of the Cold Spot after convolution with the SMHW at a scale  $R = 250$  arcmin. Besides all the anomalous characteristics previously discussed (i.e., area, HC, and MAX), the Cold Spot appears as a very symmetric feature. However, this effect could be biased since, after all, the SMHW is an isotropic filter and, therefore, the symmetric features of the Cold Spot could be amplified, erasing any possible intrinsic anisotropy. This issue was studied in detail by [82]. Instead of applying an isotropic wavelet, the anisotropic Elliptical Spherical Mexican Hat Wavelet (ESMHW) was adopted. The cleaned CMB map derived from the WMAP was transformed into ESMHW coefficients (at the scales for which the WMAP data appeared as anomalous), for different ratios  $\zeta$  between the smallest and the largest axes of the ESMHW, and for different orientations. This work proved that the maximum matching between the Cold Spot and the ESMHW took place when  $\zeta \in [0.875, 1]$  and, hence, indicated that the Cold Spot structure was quite close to be isotropic (assuming that the ratio of the ESMHW axes mimics, somehow, the ratio between the Cold Spot axes).

The frequency dependence of the Cold Spot has been a matter of study soon after its discovery. Already in [28], the SMHW coefficients of the cleaned WMAP frequency channels (namely, Q, V, and W bands) were computed, and the mean value of the coefficients associated with the Cold Spot at the scale  $R = 250$  arcmin was estimated. No obvious frequency dependence of this mean value of the wavelet coefficients was found, hence, being fully consistent with the expected behaviour for the CMB emission and,

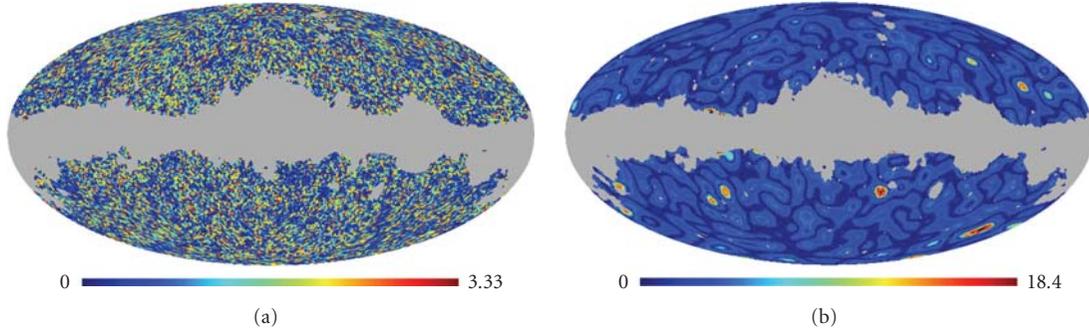


FIGURE 6: Maps of  $HC_{N_{\text{pix}}}^i$  obtained from the analysis of the QVW map derived from the WMAP data. Left panel corresponds to the study of the real space case, while right panel shows the outcome of the analysis of the SMHW coefficients at  $R = 300$  arcmin. Whereas for the former there are not particular signatures in the map, the wavelet analysis shows some prominent features, being the Cold Spot the most pronounced one.

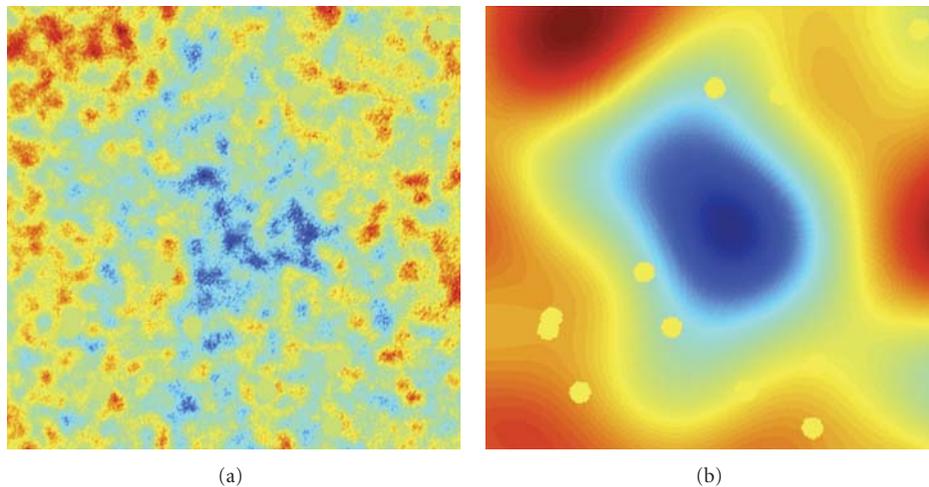


FIGURE 7: Image of the Cold Spot ( $\theta = -147^\circ$ ,  $\phi = 209^\circ$ ) in the real (a) and wavelet (b) spaces. The homogeneously filled circles correspond to positions where known extragalactic point sources have been masked.

therefore, quite different to the typical frequency dependence of the Galactic foregrounds. (The analyzed WMAP frequency channels were in thermodynamical temperature and, therefore, the CMB appears as a frequency independent emitter. Notice that, since the SMHW transform is a linear operation, the same behaviour is expected for the wavelet coefficients.)

#### 4. The Significance of the Detection: The A Posteriori Issue

One of the most questioned aspects of the *WMAP anomalies* in general, and of the Cold Spot in particular, is the issue of the actual significance of the detection. This is a very important point that is intimately linked to the *blind* nature of all the Gaussianity/isotropy tests that led to the report of such anomalies.

The author reviews where this problem comes from: if many tests are performed in a given data set, it is not strange that some of them report some deviation from the null hypothesis. It is quite usual to face the following

situation: a set of *blind* tests (i.e., tests that just challenge the compatibility of the data with a given null hypothesis,  $H_0$ , and not confronting such hypothesis with an alternative one,  $H_1$ ) claim a given incompatibility of the WMAP data. A subsequent test is performed, taking into account the previous finding and, usually, in such a way that the initial reported deviation is now found at higher significance. In this procedure, there are two weak points: the first one, already mentioned, is to assess the probability of finding a deviation as the one claimed during the first step, taking into account all the possible tests that were performed. The second one is the *credibility* of the probability for the follow-up test, where a particularity was studied in greater detail.

As mentioned above, this is a common situation for the *WMAP anomalies* works and, therefore, the Cold Spot is not an exception. Several tests were made in the first work by [28], namely, the estimation of the skewness and the kurtosis at several scales of the SMHW. A particular deviation was highlighted: the excess of kurtosis at several scales around  $R = 250$  arcmin. After that, the Cold Spot was identified as a prominent feature, and further tests (the MAX, the cold

area, the HC) were applied. The author believes that there is not a unique and clear way to solve these ambiguities and, to the author's view, this point is not usually addressed in the literature. However, whereas for the latter aspect (i.e., the significance for the follow-up tests) the solution is hard, for the former there could be some possible getaway, at least depending on the complexity of the preliminary analysis. Actually, the Cold Spot is one of the few *WMAP anomalies* where this particular aspect has been considered with deeper interest. In fact, it was the matter of several papers [47, 82] and, in particular, of [76].

In this last work, the significance of the first detection was addressed, focusing in the *a posteriori* selection of the statistic (the kurtosis) and the scale range ( $\approx 250$  arcmin). A conservative procedure to establish the  $p$ -value of the non-Gaussian detection, based on the characteristics of the analysis, was proposed. More specifically, since 30 statistics were applied to the QVW map (i.e., the skewness and the kurtosis of the SMHW coefficients at 15 scales), and only 3 out of these 30 statistics were found as anomalous (i.e., the kurtosis of the SMHW at scales  $R = 200, 250$ , and  $300$  arcmin were outside the 1% acceptance interval—see [76] for details), then, it was decided to estimate the significance of the non-Gaussianity detection by exploring in how many out of 10,000 CMB Gaussian simulations it was observed that the skewness or the kurtosis of the SMHW coefficients were outside the 1% acceptance interval, at least, at three scales. The  $p$ -value obtained in this manner was 0.0185. This  $p$ -value can serve, as explained in [76], as a conservative probability related to the non-Gaussianity associated with the SMHW analysis.

In this spirit, the following up tests (e.g., the amplitude, the area, or the HC) can be just seen as additional probes to explore/understand the previous deviation, rather than as independent sources for establishing a proper significance level for the detection.

Recently, [83] has questioned the non-Gaussianity found by [28], since the excess of the kurtosis was clearly found with the SMHW, but it was not the case with other analyzing kernels (proposed in [83]), as the top-hat and the Gaussian filters. The authors argued that these tools are more *natural* than a wavelet like the SMHW and that, therefore, the selection of the SMHW is somehow *a posteriori*. Contrary to this reasoning, the author found that the results obtained by [83] imply a different conclusion: the lack of detection when analyzing with the top-hat and the Gaussian filter is a proof of the issue discussed in Section 2.1, namely, that any filtering kernel is not necessary suitable for the detection of any non-Gaussian feature. It is clear that some features (like point sources, cosmic strings, or textures) are much better detected after applying optimal or targeted filters, rather than general ones (like the top-hat or the Gaussian functions).

The reason why a compensate filter as the SMHW gets a much larger amplification as compared to uncompensated kernels as the previous ones is that it is much more efficient to remove the background fluctuations above and below a given scale interval. Even more, it can be shown (e.g., [84]) that the SMHW is close to the optimal or *Matched* filter

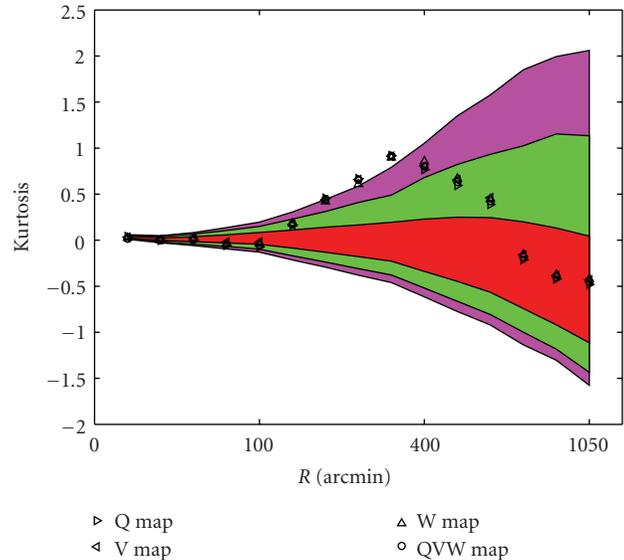


FIGURE 8: Frequency dependence of the excess of the kurtosis of the SMHW coefficients. The plot shows the variation of the kurtosis as a function of the scale  $K_R$  for cleaned CMB maps obtained from the WMAP data at three cosmological frequencies (Q-, V-, and W-bands), and for the optimal CMB cleaned map provided by WMAP, as noise-weighted combinations of the previous maps (the QVW map). The coloured regions represent, as for other figures, the 32%, 5% and 1% acceptance intervals provided by simulations.

to detect objects with a Gaussian-like profile embedded in CMB-dominated background (as it is the case in the region of the Cold Spot). The form of a *matched filter* designed for a given situation is defined, not only by the shape of the feature to be detected, but also by the statistical properties of the background. In particular, the *matched filter* (defined in the Harmonic space) is proportional to the shape of the feature and inversely proportional to the angular power spectrum of the background. Therefore, bearing in mind that (at degree scales) the CMB is well described by an angular power spectrum close to  $C_\ell \propto \ell^{-2}$ , it is trivial to show that the SMHW is near to be an optimal tool for detecting a feature described by a Gaussian-like profile and embedded in such background.

The author remarks that as it was said in Section 2.1, the selection of wavelets as a suitable tool for non-Gaussian analysis, as it was shown by several authors in the past (e.g., [40, 45, 85]), cannot be considered as an *a posteriori* choice, but, rather, as a natural option for studying scale-dependent phenomena. In particular, the compensation property satisfied by wavelets makes them extremely good analyzing kernels to amplify certain features, precisely because it assures a strong suppression of the large-scale fluctuations of the background.

Finally, it is worth recalling that the Cold Spot has been identified as an anomalous feature by other tools different from the SMHW: by directional wavelets [47], scalar indices [86, 87], steerable wavelets [27], needlets [48], and the Kolmogorov stochastic parameter [88].

## 5. Some Possible Sources to Explain the Cold Spot

To find an explanation for the non-Gaussianity deviation associated with the Cold Spot is the next step, once its anomalous nature (i.e., noncompatible with the standard inflationary scenario) is accepted. With this aim, many efforts have been done in the last years, considering different sources for the observed anomaly. The possible causes addressed so far account for *systematics* effects, mostly due to instrumental aspects that are not well understood/modelled, such as spurious emissions due to *foregrounds* or contaminants of the cosmological signal, nonaccounted *secondary anisotropies* induced on the CMB photons, as the interaction with the ionized medium (e.g., the Sunyaev-Zeldovich effect) or the nonlinear evolution of the gravitational potential (e.g., the Rees-Sciama effect), and, of course, alternative (or complementary) models to the standard inflationary scenario (as cosmic defects). In the following subsection, and also in Section 6, the author addresses these possibilities, starting from those hypotheses that are less dramatic from the point of view of strong implications for the standard cosmological model.

**5.1. Systematics.** To test the influence of unknown/unmodelled systematics on the non-Gaussianity deviation is, as one could imagine, a very hard task. The analyses that can be carried out to probe such sources are, basically of two types.

One of these analyses are *consistency tests*. As in most of the CMB experiments, WMAP satellite can provide us with cleaned CMB maps (e.g., following the template fitting approach described by [75]) for several detectors. Therefore, focusing in the non-Gaussian deviation associated with the Cold Spot, an obvious procedure would be to check whether the application of the different statistical tools reveals that such a feature is associated with only one detector, or a smaller set of detectors. If this were the case that would be a clear indication for a lack of consistency and, therefore, that the non-Gaussianity detection is associated with a given instrumental feature. This was done by [28, 46] for the kurtosis ( $K_R$ ) and the area ( $A_R^v$ ) of the SMHW wavelet coefficients. No inconsistency was found: the excess of kurtosis and of area was found to be the same for every difference assembly. As an illustration, in 8, the kurtosis of the SMHW coefficients (as a function of the scale  $K_R$ ) for 4 different CMB maps is presented. In particular, results for the Q-, V-, and W-band cleaned CMB maps are shown, together with the curve obtained from the analysis of the optimal QVW-map. These curves are quite similar, which indicates that the non-Gaussian signal is presented in all the WMAP detectors, at a similar level.

The second type of analyses are *null tests*. The cosmological frequencies of the WMAP satellite (i.e., Q-band at 41 GHz, V-band at 61 GHz, and W-band at 94 GHz) are made from more than one difference assembly. Hence, just subtracting difference assemblies at the corresponding band can produce noise maps per frequency. Neglecting small differences from the optical beams and the band-pass widths,

the CMB and foreground emissions have been cancelled out in this new map. Therefore, the application of the statistical tools to these difference maps helps to check whether the non-Gaussian signal is a *noisy artifact* (if such signal is still present) or not (if consistency with Monte Carlo simulations is found). These were done by [28, 46], again, for the kurtosis and the area of the SMHW coefficients, respectively. As for the previous type of systematics probe, there was a clear indication that the non-Gaussian signal was not related to any instrumental signature. As an example, in Figure 9, the result obtained by [28] for  $K_R$  is presented. On the left panel, the author represents the variation of the SMHW wavelet coefficients for the difference map constructed at 41 GHz as  $Q1 - Q2$ . Similarly, the results for the  $V1 - V2$  map at 61 GHz are provided in the middle panel. Finally, in the right panel, the author gives the output for the difference map obtained at the 94 GHz band as the combination  $W1 - W2 + W3 - W4$ . Results for the WMAP data is given as blue asterisks whereas, as for previous figures, the red, green, and magenta regions provide the 32%, 5%, and 1% acceptance intervals, respectively.

Summarizing, *consistency* and *null tests* do not reveal the presence of systematics behind the non-Gaussianity associated with the Cold Spot. Besides these tests, it is important to remark that the angular size associated with this feature is  $\approx 10^\circ$  in the sky. It is not trivial to think in a systematic effect affecting at this scale, and providing such localized feature in the sky as the Cold Spot.

Finally, the subsequent releases of the WMAP data (where the modelling of the instrumental properties have been improving with time) have shown that there are not changes in the non-Gaussianity deviation, except for a slight increasing on its significance, which reflects the higher signal-to-noise ratio that WMAP data is getting as observational time increases.

**5.2. Foregrounds.** Astrophysical contaminants or *foregrounds* are the next possible origin for the non-Gaussianity associated with the Cold Spot. It is well known that foregrounds are highly non-Gaussian signals. It is worth commenting that although the Cold Spot is negative, it is still possible to think in an *additive* source (as foregrounds are) as a feasible explanation. To understand this point, it is important to recall that the QVW map (that, as the author said before, is commonly adopted in the literature for cosmological analyses) is obtained as a noise-weighted linear combination of cleaned CMB maps at different frequencies. These maps (at 41, 61, and 94 GHz) are produced, as mentioned previously, via a template fitting (see, e.g., [75]). Therefore, any oversubtraction of foregrounds templates could cause a *foreground residual* in the form of a *cold* emission. (In some works, a VW map is adopted, that is, a map built as a noise-weighted linear combination of the cleaned V- and W-bands.)

Since, as mentioned before, the size of the Cold Spot is of several degrees, it is really hard to believe that this feature could be associated with residuals from point sources (i.e., from radio and infrared galaxies). Therefore, only galactic emissions (as synchrotron, free-free, and thermal

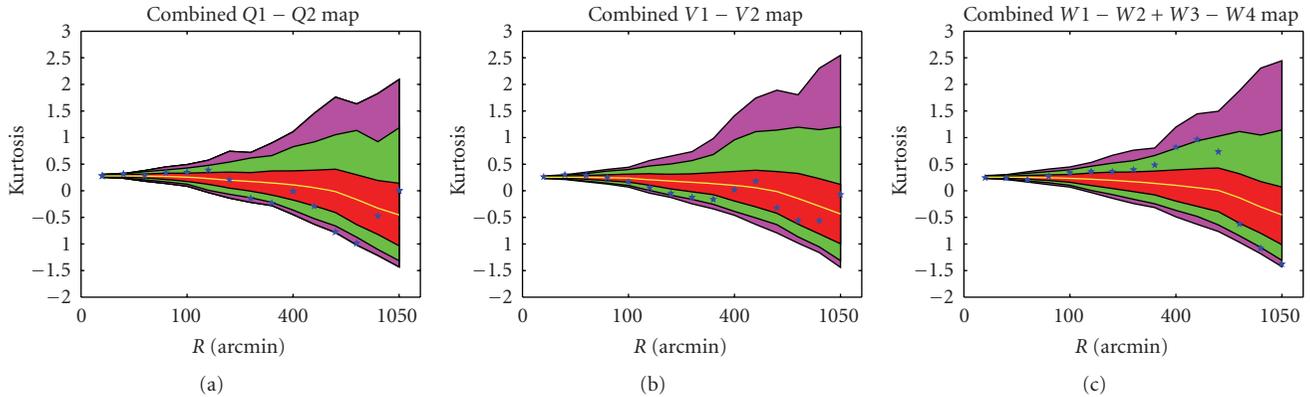


FIGURE 9: *Null* tests performed on difference maps, free of any foreground and cosmological signal. (a) represents the kurtosis of the SMHW coefficients, as a function of the scale  $R$ , for the Q1 – Q2 map, that is, for a map built as the difference of the difference assemblies observations at 41 GHz. (b) is as the previous one, but for the V1 – V2 difference map, that is, at 61 GHz. Finally, (c) provides the variation of the SMHW coefficients as a function of the wavelet scale, but for the difference map obtained with the 4 difference assemblies of WMAP at 90 GHz, that is,  $W1 - W2 + W3 - W4$ . The blue asterisks represent values for the data whereas red, green and magenta regions represented the acceptance intervals at 32%, 5%, and 1%, respectively, determined by simulations.

and spinning dust) could be responsible for a large feature as the Cold Spot. However, notice that the Cold Spot is located at  $57^\circ$  from the Galactic plane, and at a longitude of  $209^\circ$ . In other words, the Cold Spot is placed in a region of low Galactic contamination. According to the previous reasoning, it is already hard to make compatible the presence of the Cold Spot with a given Galactic emission. Even though, of course, the issue has been a matter of discussion. The author reviews here some of these analyses.

The most obvious test is, of course, to check whether there is or not any frequency dependence of the statistical estimators that indicated the Gaussian deviation. For instance, in [28, 82], the kurtosis of the SMHW coefficients ( $K_R$ ) was studied for the different CMB-cleaned maps obtained at the Q-, V-, and W-bands.

The results obtained for this study are given in Figure 8. The kurtosis  $K_R$  is presented for the 41, 61, and 94 GHz channels, and for the noise-weighted lineal combination (the QVW map). It is remarkable the high similarity of the curves. The pattern of the kurtosis, as a function of the SMHW scale, is the same for all the maps. The same is observed for its normalization. An equivalent test can be done for the area of the SMHW coefficients ( $A_R^\nu$ ), as proposed by [46, 82]. Results, at  $R = 300$  arcmin, are presented in Figure 10. Notice that the agreement of the area of the SMHW coefficients (above threshold  $\nu = 3, 3.5, 4\sigma_R$ ) among different frequency bands (Q, V, and W) and the combined QVW map (represented by the solid lines in the figure) is very high. This kind of tests shows that there is not any evident frequency dependence of the statistics associated with the non-Gaussian deviation and, therefore, that such anomaly is fully consistent with the expected behaviour for a CMB feature.

Additional tests supporting this idea have been proposed in the literature. First [28], the kurtosis of the SMHW coefficients was analyzed for a CMB-free map, constructed as the combination of the 4 difference assemblies at the W-band

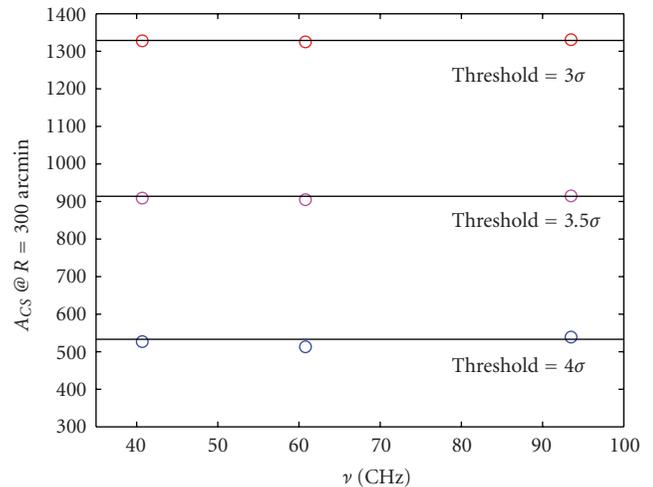


FIGURE 10: Frequency dependence of area of the SMHW coefficients. The plot shows the area of the wavelet coefficients at  $R = 300$  arcmin for different thresholds (from top to bottom:  $3, 3.5,$  and  $4\sigma_R$ ), as a function of the frequency. The circles represent the values obtained for cleaned CMB maps obtained from the WMAP data at three cosmological frequencies (Q-, V-, and W-bands). The solid lines represent the areas, at different thresholds, for the noise-weighted QVW map.

minus the sum of the 4 ones at Q- and V-band (i.e.,  $W1 + W2 + W3 + W4 - Q1 - Q2 - V1 - V2$ ). This kind of map could have a contribution of the Galactic contaminants (outside a given observing mask), since foreground emissions are not frequency independent. This analysis did not show any significant deviation from the expected behaviour from Gaussian simulations, and, therefore, discarding the presence of significant foreground residuals.

Second, as suggested by [82], different CMB recoveries from the WMAP data (where independent and alternative

component separation approaches were followed) could be analyzed. In particular, the cleaned CMB maps obtained by [89] were probed. The kurtosis, the area, and the MAX of the SMHW coefficients do not change significantly from the different CMB maps.

Reference [82] explored a more complicate scenario: a situation in which combinations of different foreground emissions could mimic, in the region of the Cold Spot, the behaviour associated with the CMB, that is, a frequency independent global emission. To check this possibility, several templates were used as tracers of the Galactic foregrounds, namely, the Rodhes/HartRa0 2326 MHz [90] radio survey for synchrotron, the  $H_\alpha$  by [91] for the free-free, and the thermal dust model by [92]. Authors studied the expected contribution of foregrounds in the region of the Cold Spot, and they found that taking into account the uncertainties in the extrapolation of these templates from their original observations to the WMAP frequency range, it was possible to find a global Galactic emission that was nearly frequency-independent from 41 to 94 GHz (i.e., from Q- to W-bands). However, it was found that the emission was at a level of one order of magnitude below the Cold Spot temperature. It was checked that even accounting twice for that hypothetical foreground emission, it was not possible to reconcile observations with the Gaussian model.

All these tests on the impact of the foregrounds indicated that the non-Gaussian signal associated with the Cold Spot was fully consistent with a CMB like frequency dependence, and that the role played by astrophysical contaminants was negligible.

**5.3. The Sunyaev-Zeldovich Effect.** After checking that the possible impact of systematics and foregrounds on the non-Gaussian detection is very low, the next step is to study whether secondary anisotropies of the CMB could explain the anomalous nature of the Cold Spot.

The Sunyaev-Zeldovich effect (SZ) could be a potential candidate to explain the anomaly (it is produced by the inverse Compton interaction of the CMB photons, as they cross the hot electron gas that is found in clusters of galaxies). Two major reasons support this possibility: first, the size of the Cold Spot is nearly compatible with the fluctuations caused by the nearest clusters of galaxies, and second, this fluctuations (in the frequency range covered by WMAP) produce cold spots [93].

There is no evidence for the existence of any large cluster in the direction of the Cold Spot. However, [94] reported the presence of a large concentration of galaxies in that direction—referred to as the *Eridanus super-group*—that could account for a total mass of  $\approx 10^{14} M_\odot$  (see [95] for details).

This scenario implies, therefore, that the SZ could explain the nature of the Cold Spot, at least, partially. In other words, it could be possible that a combination of a SZ contribution plus a large (but not anomalous) CMB fluctuation could account for the Cold Spot emission. This was studied by [82], and the results are summarized in Figure 11. In this figure, the values of the SMHW coefficients  $w(\theta, \phi; R)$ , at

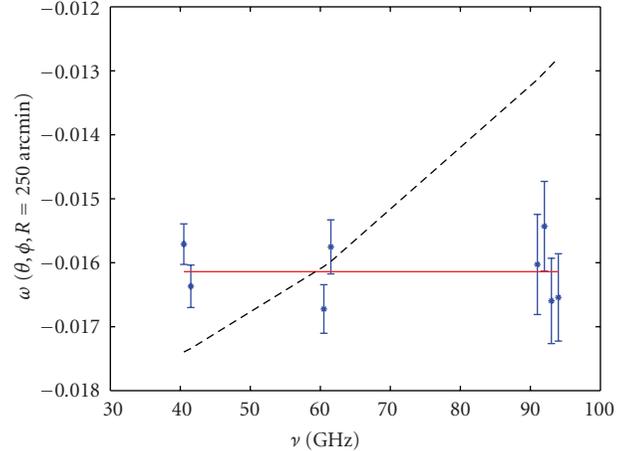


FIGURE 11: Best-fit of the WMAP data—across the 8 difference assemblies for the Q-, V-, and W-bands—to a CMB (solid line) and a SZ spectra (dashed line). The fit is performed to the minimum amplitude of the Cold Spot in wavelet space, at a scale of  $R = 300$  arcmin. Error bars are obtained from Gaussian simulations.

the position of the Cold Spot and at  $R = 300$  arcmin, are given for the eight CMB cleaned maps at the Q-, V-, and W-bands. The corresponding error bars were computed from CMB plus noise simulations, corresponding to the instrumental properties of these detectors, and convolved with the appropriate SMHW kernel.

Three different fits to the data were explored. First, a pure CMB spectrum was used, giving a very good fit with a reduced  $\chi^2 = 1.00$ . Such a fit is represented by the solid line in Figure 11. Second, a pure SZ spectrum was tested, obtaining a very poor fit (dashed line) with a reduced  $\chi^2 = 9.12$ . Finally, a joint fit to a CMB plus SZ spectra was explored, obtaining an amplitude for the SZ spectrum consistent with zero and a reduced  $\chi^2$  quite similar to the first case. These results confirm that the frequency dependence of the Cold Spot is consistent with a CMB-like spectrum. They also rule out the possibility that the SZ is playing any significant role.

**5.4. The Late Evolution of the Large-Scale Structure.** Another secondary anisotropy that could explain the anomalous nature of the Cold Spot is the one due to the nonlinear evolution of the gravitational field: the so-called Rees-Sciama effect (RS). In particular, it is known [96, 97] that voids in the large-scale structure could induce a negative nonlinear anisotropy in the CMB photons. The size of such secondary anisotropies depends on the proper size of the void and its redshift. Therefore, as for the SZ, the RS is another potential candidate to explain the anomalous nature of the Cold Spot.

Extra support for this hypothesis came from two different paths. On the one hand, theoretical works as [98, 99] proposed that a very large void (of  $\approx 300 h^{-1}$  Mpc) and located at low redshift ( $z \ll 1$ ) could produce large negative CMB fluctuations such as the Cold Spot, even with modest density contrast values (i.e., in a quasilinear regime). On the other hand, it was suggested [100] that the NVSS catalogue

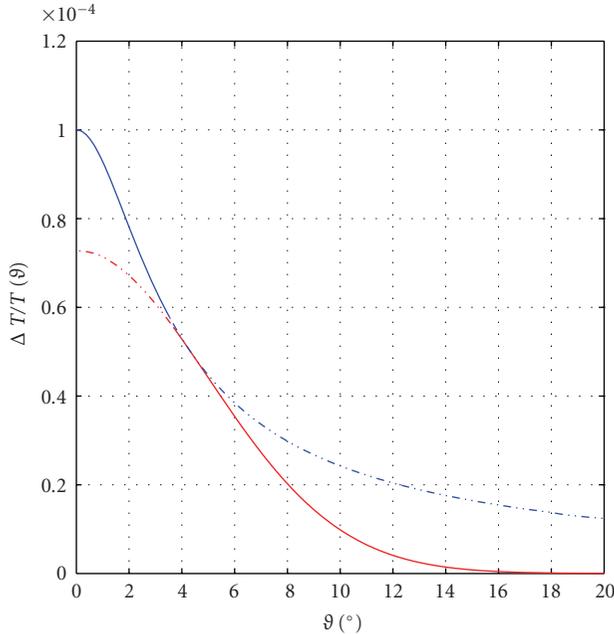


FIGURE 12: Radial profile (solid blue/red lines) of a cosmic texture, as given in (14). The parameters defining this profile are  $\varepsilon = 10^{-4}$  and  $\vartheta_c = 5^\circ$ . The dashed lines represent the truncated values of the profiles given by (13) (blue) and the Gaussian function (red). See text for details.

[101] seems to show, at the position of the Cold Spot, a lack in the number count of radio galaxies.

Against these ideas, some criticisms can be made. First, this kind of voids are not observed and, even more, according to current  $N$ -body simulations [102] they are extremely rare events ( $\approx 13\sigma$ , i.e., much more rare than the Cold Spot deviation itself, that was a 98.15% event!) in the standard cosmological framework. Second, the claim made on the NVSS data has been recently questioned by [103], suggesting that such finding was an artefact caused by possible systematics related with the NVSS data processing, the statistical procedure and the *a posteriori* selection of the Cold Spot position.

In addition, observational campaigns on the region of the Cold Spot were made by [104] with the MegaCam on the Canada-France-Hawaii Telescope and by [105] with the VIMOS spectrograph on the very large telescope. Both works have reported that the large-scale structure in that direction, and up to  $z \approx 1$ , is fully consistent with the standard model and that there is no evidence of such large voids as those required by the nonlinear evolution of the gravitational potential hypothesis.

## 6. A Plausible Explanation: Cosmic Textures

In the previous section, the author has presented an overview of the several works carried out to establish whether the non-Gaussianity detection associated with the Cold Spot could be explained in terms of systematics, foregrounds, and secondary anisotropies as the Sunyaev-Zeldovich and

the Rees-Sciama effects. None of these possibilities seem to provide a satisfactory explanation, and, therefore, other sources should be investigated.

In this section, the author pays attention to the suggestion made by [106]: the Cold Spot could be caused by a *cosmic texture*. Cosmic textures [107] are a type of cosmic defects. They are supposed to be generated at some stage of the early Universe, associated with the symmetry-breaking phase transitions that are predicted by certain theoretical models of high energy physics (see, e.g., [108] and references therein for a much more detailed explanation). In short, defects can be understood as space regions of a given phase state, surrounded by a space already in a new phase. In some cases, as for textures, these regions could collapse and, therefore, left an imprint on the CMB photons.

Among the different types of cosmic defects, textures are the most plausible candidate to explain the anomalous nature of the Cold Spot, since the interaction of the CMB photons with the time variation of the gravitational potential, associated with an eventual collapse of the texture, produces spots in the CMB fluctuations [109]. Even more, cosmic textures are expected to be a possible source of kurtosis deviation whereas the expected level of skewness is almost negligible (at least for values of the symmetry-breaking energy scale compatible with current observations). This is caused by the even probability of textures producing cold and hot spots and, therefore, providing a nearly symmetric distribution of temperature fluctuations. In fact, the equilateral  $f_{\text{NL}}$  expected from textures goes as  $f_{\text{NL}} \approx 1.5 \times 10^{-100} \psi_0^6$  [110], where  $\psi_0$  is the symmetry-breaking energy scale, measured in GeV. For typical limits in  $\psi_0$  imposed by the CMB angular power spectrum analysis (e.g., [111, 112]), the expected equilateral  $f_{\text{NL}}$  is  $\approx 10^{-9}$ , that is, a tiny value well below the current constraints [113].

The isotropic shape of the temperature fluctuations related to these spots can be approximated, at least at small angular distances  $\vartheta$ , as [114]

$$\frac{\Delta T}{T}(\vartheta) = \pm \varepsilon \frac{1}{\sqrt{1 + 4(\vartheta/\vartheta_c)^2}}, \quad (13)$$

where  $\vartheta$  represents the angular distance from the centre of the spot, and  $\vartheta_c$  is a characteristic scale parameter of the spot—that is related to the redshift of the spot and the dynamics of the Universe [106]. The amplitude  $\varepsilon$  is proportional to the symmetry-breaking energy scale  $\psi_0$ :  $\varepsilon = 8\pi^2 G \psi_0^2$ . It is worth remarking that, according to cosmic texture models, the amplitude  $\varepsilon$  is the same for every single spot generated by the collapsing defects.

As mentioned above, the profile proposed in (13) is only valid up to small relative distances (e.g.,  $\vartheta \approx \vartheta_c$ ) (the author knowledge, there is not any analytical or numerical solution for the full profile of a given cosmic texture yet). In order to have a profile valid at larger angular distances, [106] proposed to extend the profile given by (13) from its half-maximum, following a Gaussian function. The extension is done by imposing continuity, both, of the profile itself and of its first derivative.

Taking into account these conditions, the assumed profile for the CMB temperature fluctuation caused by a collapsing cosmic texture would be given by

$$\frac{\Delta T}{T} = \pm \begin{cases} \frac{\varepsilon}{\sqrt{1 + 4(\vartheta/\vartheta_c)^2}}, & \text{if } \vartheta \leq \vartheta_* \\ \frac{\varepsilon}{2} e^{-(1/2\vartheta_c^2)(\vartheta^2 + \vartheta_*^2)}, & \text{if } \vartheta > \vartheta_*, \end{cases} \quad (14)$$

where  $\vartheta_* = \sqrt{3}/2\vartheta_c$ .

In Figure 12, the author shows the radial section of the above profile (solid line). The red and blue parts of the solid line correspond to the Gaussian function and to the profile of (13), respectively. The dashed lines represent the truncated regions for both curves. The cosmic texture parameters used in this profile are  $\vartheta = 5^\circ$  (i.e., similar to the SMHW scale at which the Cold Spot appears as anomalous) and  $\varepsilon = 10^{-4}$ . This value corresponds to a symmetry-breaking energy scale of  $\psi_0 = 1.13 \times 10^{16}$  GeV that corresponds to a conservative upper limit imposed by CMB measurements (e.g., [111, 115]).

The author remarks that the results obtained with a profile as the one given in (14), and that are reviewed in the next section, do not depend very much with the specific function adopted for the extrapolation. Similar results are obtained, for instance, when an exponential function or a SMHW-like kernel are used.

**6.1. The Bayesian Framework.** Attending to the issues discussed in the previous subsection, a cosmic texture could be a strong candidate to explain the Cold Spot: textures produce spots on the CMB temperature fluctuations, they are non-Gaussian signals, and, depending on their amplitude (or the symmetry-breaking scale), they could be compatible with current constraints on the role played by cosmic defects on the structure formation and evolution of the Universe.

Reference [106] proposed to make use of the texture profile of (14) to perform a hypothesis test to decide whether the WMAP data (in the position of the Cold Spot) is more likely to be described by a large (but not anomalous) CMB spot (i.e., the *null*  $H_0$  hypothesis) or by a cosmic texture of amplitude  $\varepsilon$  and size  $\vartheta$  added to a random Gaussian and isotropic CMB field (i.e., the *alternative*  $H_1$  hypothesis).

The optimal way of performing such hypotheses test is within the Bayesian framework. Bayes' theorem states that, given a data set  $D$  and some unknown parameters  $\Theta$  (defining a given model in the context of a given hypothesis  $H_i$ ), the posterior probability of the parameters/model given the data  $P(\Theta | D, H_i)$ , is related to the likelihood  $P(D | \Theta, H_i)$  (i.e., the probability of the data given the parameters/model) as

$$P(\Theta | D, H_i) = \frac{P(D | \Theta, H_i)P(\Theta | H_i)}{P(D | H_i)}, \quad (15)$$

where  $P(\Theta | H_i)$  is a measurement of our *a priori* knowledge about the parameters/model (i.e., the prior), and  $P(D | H_i)$  is a constant (i.e., it does not depend on the parameters/model) called *Bayesian evidence* (BE). The BE is nothing but the average likelihood with respect to the prior

$$P(D | H_i) = \int P(D | \Theta, H_i)P(\Theta | H_i)d\Theta, \quad (16)$$

and it is a largely used mechanism to perform hypotheses test. In particular, its role on different cosmology fields has been quite remarkable during the last years (e.g., [116] for dark energy studies, [117] for anisotropic models of the Universe expansion, [118] for studying different re-ionization models, [119] for point source detection, and [120] for exploring nonstandard inflationary models).

The importance of BE for hypotheses test is clear. First, it is obvious that the quantity that we would like to obtain is a measurement of the probability of a given hypothesis  $H_0$ , given the data, that is,  $P(H_0 | D)$ . This probability can be written, attending to the *probability multiplication rule*, as

$$P(H_0 | D) = P(D | H_0) \frac{P(H_0)}{P(D)}, \quad (17)$$

that is, it is proportional to the BE and to the probability of the hypothesis and inversely proportional to the probability of the data. Under certain circumstances, the probability of the hypothesis could be known, but, however, it is not the case for the probability of the data. In other words, we only can learn about the probability of the hypothesis  $H_0$  up to a factor. Therefore, what we can extract is a *relative* measurement of the probability of two hypotheses ( $H_0$  and  $H_1$ ), given the same data set  $D$ . This relative measurement is called the *posterior probability ratio*,  $\rho$ , and reads

$$\rho \equiv \frac{P(H_1 | D)}{P(H_0 | D)} = \frac{E_1 P(H_1)}{E_0 P(H_0)}, \quad (18)$$

where, for simplicity, the author rewrite the BE,  $P(D | H_i)$ , as  $E_i$ . Hence, if  $\rho > 1$ , we can conclude that the hypothesis  $H_1$  is favoured by the data with respect to  $H_0$ . In some cases, there is not a clear choice for the probability of the hypotheses. In this case, empirical rules for the ratio of evidences—as the Jeffreys' scale [121]—are usually adopted.

Therefore, the procedure required to explore whether the Cold Spot is more likely to be explained in terms of a cosmic texture ( $H_1$ ) rather than by a Gaussian CMB fluctuation ( $H_0$ ) is clear: first, the likelihood is computed for both hypotheses; second, the BE is estimated, taking into account the adequate priors (16); finally, the posterior probability ratio (18) is evaluated, making use of suitable a priori probabilities for  $H_0$  and  $H_1$ .

This was the procedure followed in [106]. It is straightforward to show that since the *noise term* is caused by standard CMB Gaussian fluctuations and instrumental Gaussian

noise, the likelihood function reads as

$$P(D | \Theta, H_i) \propto \exp\left(-\frac{\chi^2}{2}\right), \quad (19)$$

where  $\chi^2 = (D - T(\Theta))C^{-1}(D - T(\Theta))^T$ . The correlation matrix  $C$  accounts for the full Gaussian CMB and noise correlations—that is,  $C = S + N$ , where  $s_{ij} \propto \sum C_\ell P_\ell(\cos \theta_{ij})$  and  $n_{ij} = \sigma_i^2 \delta_{ij}$ , being  $P_\ell$  the Legendre polynomials,  $\theta_{ij}$  the angular distance between the pixels  $i$  and  $j$ ,  $\sigma_i^2$  the instrumental noise contribution to pixel  $i$ , and  $\delta_{ij}$  represents the Kronecker delta.  $D$  represents the data (i.e., the QVW map), and the function  $T(\Theta)$  represents the model behind the hypotheses—that is, (14) for  $H_1$ , and  $\equiv 0$  for  $H_0$ . Finally,  $T$  denotes standard matrix transpose.

The priors adopted by [106] for the parameters  $\Theta \equiv (\varepsilon, \vartheta_c)$  were chosen attending to observational constraints and cosmic texture simulations. In particular, the prior on the amplitude was  $|\varepsilon| \leq 10^{-4}$  whereas the prior of the size  $1^\circ \leq \vartheta_c \leq 15^\circ$  was assumed. The amplitude prior was uniformly distributed, and, as mentioned above, it is a conservative constraint imposed from the contribution of cosmic defects to the CMB angular power spectrum. The size  $\vartheta_c$  follows a scale-invariant law, and the limits come from texture simulations. Textures below  $1^\circ$  should be smeared out by photon diffusion, and, in addition, they would be related to collapsing events above redshift  $\approx 1000$ , which would not affect the CMB image. The upper limit is due to the unlikely probability of generating such large textures in the finite celestial sphere. Even so, [106] tested that results were not specially sensitivity to the prior selection, since the likelihood was clearly peaked, within a region of the parameter space clearly allowed by observations and texture models.

The marginalization of the posterior probability in (15), led to the determination of the texture parameters, obtaining  $\varepsilon = 7.3_{-3.6}^{+2.5} \times 10^{-5}$  and  $\vartheta_c = 4.9_{-2.4}^{+2.8^\circ}$  at 95% confidence. The BE ratio was 150, which, in terms of the empirical rules [121], is a *strong* indication that the texture hypothesis for the Cold Spot is favoured over the isotropic and Gaussian CMB fluctuation option. Adopting a ratio for the probability of the hypotheses given by the fraction of the sky that is covered by a cosmic texture as large as the one required for the Cold Spot ( $\approx 0.017$ ), the posterior probability ratio was  $\rho = 2.5$ , which also favours the texture hypothesis.

It is worth mentioning that the estimated value for the texture amplitude could be affected by selection bias. In [106], it is established that such bias could provide an overestimate of the texture amplitude by a factor of 2. This bias is caused because the texture amplitude is estimated in a low signal-to-noise regime, where the features placed in large background fluctuations are more easily detected. Even so, the estimated value for the texture amplitude,  $\varepsilon = 7.3 \times 10^{-5}$ , would imply a symmetry-breaking energy scale of  $\psi_0 = 8.7 \times 10^{15}$  GeV, which, on the one hand, is fully compatible with more recent constraints imposed from the analysis of the CMB temperature and polarization power spectra (e.g., [112]), and, on the other hand, is in agreement with the predictions of most of the models for particle physics. In addition, by relating the angular size of the CMB texture

profile to the cosmological parameters defining the geometry and evolution of the Universe, it was possible to establish that the texture collapse (that generated the CMB profile in (14)) occurred at redshift  $z \approx 6$ .

Figure 13 shows the effect of correcting the QVW map from the texture emission. In the left panel, the region of the sky where the Cold Spot was identified is shown. The middle panel presents the best fit of the texture profile, according to the parameters previously mentioned. On the right panel, the author presents the resulting map after the subtraction of the estimated cosmic texture contribution. The Cold Spot is noticeably reduced. More quantitative measurements were made by [106]. It was proved that if cosmic textures (adequate to the parameters fixed by the Cold Spot analysis) were added to isotropic and Gaussian simulations of CMB signals, as seen in the QVW WMAP map, then the kurtosis of the SMHW became compatible. Even more, [122] showed that the WMAP map (corrected from the cosmic texture contribution in the location of the Cold Spot) was not compatible with anisotropic patterns for nonstandard expansions of the Universe (in particular, for Bianchi VII<sub>h</sub> models), as it was previously the case for the uncorrected data (e.g., [117, 123]).

Finally, the author remarks that [95] proposed a similar approach as the one described in this section to study (attending to spatial templates) whether SZ and RS could provided more suitable hypotheses than the standard isotropic and Gaussian model. This Bayesian analysis indicated that neither of these hypotheses is favoured.

**6.2. Followup Tests.** The studies described in Section 5 and in the previous subsection, clearly indicates that among all the realistic sources that could explain the anomalous nature of the Cold Spot, only the cosmic texture hypothesis remains as a feasible option. The results obtained by [106], and reviewed in the previous subsection, have to be understood as a (clear) indication that the cosmic texture is plausible. However, before accepting it *as the final explanation*, it should be confirmed by additional tests. In particular, if the cosmic texture hypothesis would be the right one, then there are some clear predictions that could be tested (at least in the near future, once ongoing/upcoming experiments as SPT, ACT, QUIJOTE, and ALMA are fully operative).

In this subsection, the author comments on the three most obvious followup tests that could help to discard, or accept, the cosmic texture. These follow-up tests are: the searching of more textures, the local polarization of the CMB, and the local CMB lensing. These foreseen effects were firstly pointed out by [106].

**6.2.1. Looking for More Textures.** If a texture were found in the location of the Cold Spot, then, attending to cosmic texture models, there should be more cold and hot spots randomly distributed across the sky. In fact, the distribution of CMB spots caused by cosmic textures follows a scale-invariant law

$$N_{\text{sp}}(> \vartheta_c) = \frac{4\pi\nu\kappa^3}{3\vartheta_c^2}, \quad (20)$$

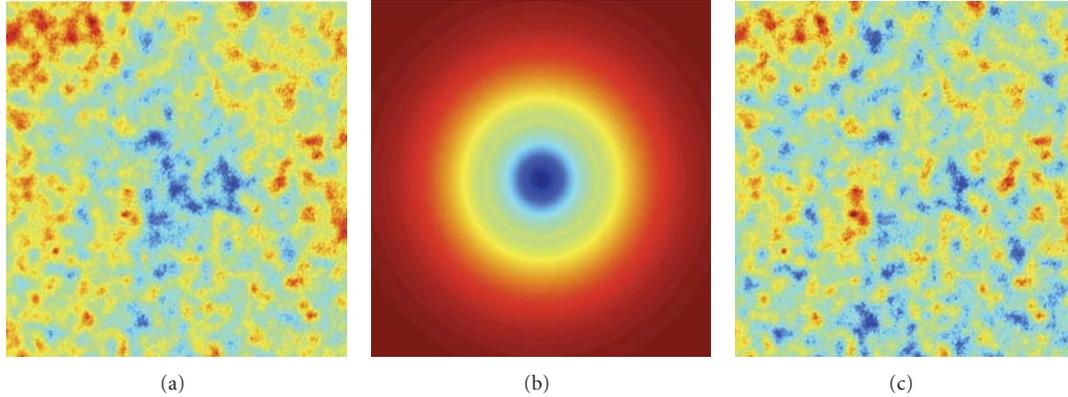


FIGURE 13: (a) CMB temperature fluctuations obtained from WMAP data, centred in the position of the Cold Spot. (b) best fit of a texture profile to the data, according to (14). (c) CMB map, at the position of the Cold Spot, after the subtraction of the best-fit texture model.

that is, the number of cold/hot spots with a scale equal or larger than  $\vartheta_c$ , is inversely proportional to  $\vartheta_c^2$  (see [124] for the grounds and [106] for a derivation). The  $\nu$  and  $\kappa$  parameters are associated with the physics of the cosmic texture models, and their specific meaning is out of the scope of this paper. The author remarks here that according to state-of-the-art simulations, these parameters are well consistent with  $\nu \approx 2$  and  $\kappa \approx 0.1$  values, when  $\vartheta_c$  is expressed in radians. Before continuing the discussion, the author stresses that the number of expected cold/hot spots with a scale  $\vartheta_c$  equal or larger than  $5^\circ$  (i.e., as the Cold Spot scale) is  $\approx 1$ . In other words, the fact that we only found a Cold Spot of a scale similar or larger than the Cold Spot size is fully consistent with the cosmic texture scenario. This fact, of course, is an extra support for the texture hypothesis causing the Cold Spot emission.

Let us come back to (20). A straightforward calculation tells as that if the cosmic texture hypothesis is correct, then the CMB temperature fluctuations should contain  $\approx 28$  spots with a scale  $\vartheta_c \geq 1^\circ$ , and  $\approx 7$  spots with a scale  $\vartheta_c \geq 2^\circ$ . Poisson errors can be safely assumed for these numbers. Therefore, the number of cold/hot spots expected in the WMAP data are  $1 \lesssim N_{\text{sp}}(>2^\circ) \lesssim 13$  at the 95% confidence level. Current work is in progress to check this prediction, by using a fast cluster nesting sampling algorithm—MULTINEST [125]—to explore the posterior probability ratio. There are well-founded hopes to find new textures. In particular, some non-Gaussian analyses as [27, 48, 88] reported some hot/cold spots (in addition to the Cold Spot) as potentially anomalous.

**6.2.2. The Polarization of the CMB.** The effect of a collapsing texture on the passing by CMB photons is nothing but a secondary anisotropy of the CMB fluctuations, whose origin is merely gravitational. Hence, the effect of such gravitational phenomenon on the E-mode polarization is almost negligible (only vector modes would be affected, which are well below the scalar mode contribution).

Strictly speaking, this lack of polarization is not a unique signature produced by cosmic textures. As the author said,

any secondary anisotropy of gravitational origin would cause it. However, these other effects (as huge voids) are quite implausible explanations (see Section 5.4). For that reason, this effect is a valid follow-up test to probe the texture hypothesis.

The procedure is simple: to compare the E-mode polarization in the position of a temperature spot (as large and extreme as the Cold Spot is), under two different hypotheses, the *null* or  $H_0$  one (i.e., the temperature spot is caused by a Gaussian fluctuation) and the *alternative* or  $H_1$  option (i.e., the temperature spot is a secondary anisotropy caused by the collapse of an evolving texture). In fact, as proposed by [126], the best discriminating measurement is the T-E correlation, rather than simply the E signal. This cross-correlation is expected to be close to zero for the  $H_1$  hypothesis. The approach suggested in this work was to estimate the correlation of the T and E profiles around the position of the spot temperature signal. This statistic was computed for many simulations according to the  $H_0$  and  $H_1$  hypotheses, and a hypothesis test was performed, via the definition of an optimal Fisher discriminant statistic (e.g., [127]).

The method was applied to probe the capabilities of current and upcoming CMB experiments for discriminating between the two hypotheses. In particular, the cases of WMAP, Planck [128], and QUIJOTE [129] were considered.

As the major conclusion, it can be established that the discrimination power of the T-E correlation is not very high. In fact, for an ideal noise-free experiment, at a power of the test of 0.5, the significance level is (up most) 0.8%. The reason for this limitation is that the characteristic size of the Cold Spot is  $\approx 10^\circ$ , which, roughly, corresponds to multipoles of  $\ell \approx 40$ . It happens that at this multipole scales, the T-E angular cross-power spectrum is very close to zero, already for  $H_0$  and, therefore, it is hard to discriminate it from  $H_1$ .

The results are graphically summarized in Figure 14. The author represent the significance level or  $p$ -value (for a power of the test of 0.5) as a function of the instrumental noise sensitivity for the E-mode polarization. The three experiments previously mentioned are indicated as vertical lines. Notice that the WMAP instrumental characteristics

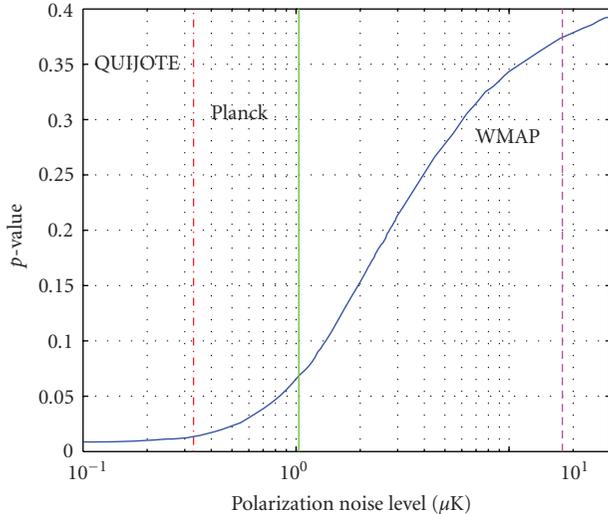


FIGURE 14: The curve indicates the probability (the  $p$ -value) for rejecting the texture hypothesis, attending to the T-E correlation, as a function of the instrumental noise sensitivity in polarization. From left to right, the vertical lines indicate the noise level associated with QUIJOTE, Planck, and WMAP experiments.

do not allow for any significant discrimination between hypotheses. With Planck, it would be possible to reach a modest 7% significance level. Finally, the upcoming QUIJOTE experiment would allow for a 1.4% detection. However, if the significance of the Cold Spot obtained via the first non-Gaussianity wavelet analysis (i.e., 1.85%, as mentioned in Section 4) is also considered, [126] claims that a joint T and T-E significance detections of 0.025% and 0.12% could be imposed on the texture hypothesis, for the QUIJOTE and Planck experiments, respectively.

**6.2.3. Gravitational Lensing.** Besides the lack of polarization discussed in the previous subsection, the lensing of the CMB photons would be another foreseen effect caused by the gravitational field generated by the collapsing texture. This point was recently addressed by [130]. This work studied the capabilities of small-scale CMB experiments as ACT to detect a possible lensing effect occurring at the position of the Cold Spot, caused by the gravitational field of a texture placed at  $z = 6$ , and with a typical scale of  $5^\circ$ , associated with a symmetry-breaking energy scale of  $\psi_0 = 4.5 \times 10^{15}$  GeV (i.e., comparable to the parameters determined by [106]).

As for the polarization test, the power of the lensing analysis to probe the existence of the texture is relatively modest. For instance, detection is made at the  $3\sigma$  level after 1000 minutes of integration time. In other words, this test would require of a dedicated observational campaign.

## 7. Conclusions

In this paper the author has presented a comprehensive overview of the Cold Spot. Since its detection in 2003 by

[28], this feature proved to be one of the most intriguing anomalies found in the WMAP data.

The Cold Spot was detected after performing a non-Gaussianity test on a cleaned CMB (obtained via a template fitting of the WMAP difference assemblies at Q-, V-, and W-bands). The non-Gaussian analysis was performed by comparing the values obtained for the skewness and the kurtosis of the SMHW coefficients (at several scales) for such cleaned CMB map to the distribution expected from isotropic and Gaussian CMB realizations. This analysis indicated an excess of the kurtosis at SMHW scales  $R$  of around 300 arcmin. Subsequent analysis of the SMHW coefficients, based on the area above/below a given threshold, the *Higher Criticism*, and the maximum value, agreed in detecting the non-Gaussian deviation occurring at the same scales, and confirmed the peculiar role played by the Cold Spot. Avoiding any possible *a posteriori* choice of statistics, a conservative significance detection level of 1.85% was placed by [76].

The Cold Spot was found to be highly isotropic [82], and the impact of possible systematics and residual foreground contamination was discarded [28, 46, 76, 82]. Equivalently, some secondary anisotropies (potentially responsible for the Cold Spot emission) as the Sunyaev-Zeldovich and the Rees-Sciama effects, that, in principle, could be accepted as valid solutions, were shown to be unlikely explanations [82, 95].

A plausible explanation in terms of a cosmic defect was addressed by [106]. It would imply the presence of a collapsing cosmic texture at redshift  $z \approx 6$ , with a typical scale of  $\approx 5^\circ$ , corresponding to a symmetry-breaking energy scale of  $\psi_0 \approx 9 \times 10^{15}$ . In addition, some follow-up tests were proposed to confirm or discard the cosmic texture hypothesis. In particular, one should expect to have more spots generated by evolving textures in CMB observations (e.g.,  $\approx 7$  with scales  $\vartheta \gtrsim 2^\circ$ ), a lack of E-mode polarization is predicted, as compared to the values associated with spots derived from a Gaussian CMB field; and, finally, a lensing of the CMB photons is potentially detectable with future small scale CMB experiments as ACT or SPT.

Summarizing, the study of the Cold Spot has provided a wealth of information, and the upcoming high-quality CMB data, as the one expected from Planck, guarantee that still more knowledge will come from the analysis of this very interesting feature.

## Acknowledgments

The author thanks R. Belén Barreiro, Marcos Cruz, and Enrique Martínez-González for a careful reading of this paper. The issue reviewed in this paper is the result of a very fruitful and exciting collaboration with them (and with other colleagues) during the last 7 years, and the author presumes it is not over yet, since new data is expected to come in the near future. The author acknowledges partial financial support from the Spanish Ministerio de Ciencia e Innovación Project no. AYA2007-68058-C03-02. The author also thanks financial support from the *Ramón y Cajal* programme.

## References

- [1] E. Komatsu, J. Dunkley, M. R.olta et al., “Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [2] D. Larson, J. Dunkley, G. Hinshaw et al., “Seven-year Wilkinson microwave anisotropy probe (WMAP) observations: power spectra and WMAP-derived parameters,” <http://arxiv.org/abs/1001.4635>.
- [3] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Górski, and P. B. Lilje, “Asymmetries in the cosmic microwave background anisotropy field,” *The Astrophysical Journal*, vol. 605, no. 1, pp. 14–20, 2004.
- [4] F. K. Hansen, P. Cabella, D. Marinucci, and N. Vittorio, “Asymmetries in the local curvature of the wilkinson microwave anisotropy probe data,” *The Astrophysical Journal*, vol. 607, no. 2, pp. L67–L70, 2004.
- [5] H. K. Eriksen, D. I. Nivikov, P. B. Lilje, A. J. Banday, and K. M. Górski, “Testing for non-Gaussianity in the Wilkinson microwave anisotropy probe data: Minkowski functionals and the length of the skeleton,” *The Astrophysical Journal*, vol. 612, no. 1, pp. 64–80, 2004.
- [6] F. K. Hansen, A. J. Banday, and K. M. Górski, “Testing the cosmological principle of isotropy: local power-spectrum estimates of the WMAP data,” *Monthly Notices of the Royal Astronomical Society*, vol. 354, no. 3, pp. 641–665, 2004.
- [7] E. P. Donoghue and J. F. Donoghue, “Isotropy of the early universe from CMB anisotropies,” *Physical Review D*, vol. 71, no. 4, Article ID 043002, 2005.
- [8] J. Hoftuft, H. K. Eriksen, A. J. Banday, K. M. Górski, F. K. Hansen, and P. B. Lilje, “Increasing evidence for hemispherical power asymmetry in the five-year wmap data,” *The Astrophysical Journal*, vol. 699, no. 2, pp. 985–989, 2009.
- [9] D. Pietrobon, P. Cabella, A. Balbi, R. Crittenden, G. de Gasperis, and N. Vittorio, “Needlet bispectrum asymmetries in the WMAP 5-year data,” *Monthly Notices of the Royal Astronomical Society*, vol. 402, no. 1, pp. L34–L38, 2010.
- [10] P. Vielva and J. L. Sanz, “Constraints on fNL and gNL from the analysis of the N-pdf of the CMB large-scale anisotropies,” *Monthly Notices of the Royal Astronomical Society*, vol. 404, no. 2, pp. 895–907, 2010.
- [11] F. Paci, A. Gruppuso, F. Finelli et al., “Power asymmetries in the cosmic microwave background temperature and polarization patterns,” *Monthly Notices of the Royal Astronomical Society*, vol. 407, no. 1, pp. 399–404, 2010.
- [12] P. Bielewicz, K. M. Górski, and A. J. Banday, “Low-order multipole maps of cosmic microwave background anisotropy derived from WMAP,” *Monthly Notices of the Royal Astronomical Society*, vol. 355, no. 4, pp. 1283–1302, 2004.
- [13] D. J. Schwarz, G. D. Starkman, D. Huterer, and C. J. Copi, “Is the low- $\ell$  microwave background cosmic?” *Physical Review Letters*, vol. 93, no. 22, Article ID 221301, 2004.
- [14] C. J. Copi, D. Huterer, and G. D. Starkman, “Multipole vectors: a new representation of the CMB sky and evidence for statistical anisotropy or non-Gaussianity at  $2 \leq \ell \leq 8$ ,” *Physical Review D*, vol. 70, no. 4, Article ID 043515, 13 pages, 2004.
- [15] A. de Oliveira-Costa, M. Tegmark, M. Zaldarriaga, and A. Hamilton, “Significance of the largest scale CMB fluctuations in WMAP,” *Physical Review D*, vol. 69, no. 6, Article ID 063516, 2004.
- [16] P. Bielewicz, H. K. Eriksen, A. J. Banday, K. M. Górski, and P. B. Lilje, “Multipole vector anomalies in the first-year WMAP data: a cut-sky analysis,” *The Astrophysical Journal*, vol. 635, no. 2, pp. 750–760, 2005.
- [17] K. Land and J. Magueijo, “Cubic anomalies in the Wilkinson microwave anisotropy probe,” *Monthly Notices of the Royal Astronomical Society*, vol. 357, no. 3, pp. 994–1002, 2005.
- [18] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, “On the large-angle anomalies of the microwave sky,” *Monthly Notices of the Royal Astronomical Society*, vol. 367, no. 1, pp. 79–102, 2006.
- [19] L. R. Abramo, A. Bernui, I. S. Ferreira, T. Villela, and C. A. Wuensche, “Alignment tests for low CMB multipoles,” *Physical Review D*, vol. 74, no. 6, Article ID 063506, 2006.
- [20] A. Gruppuso and C. Burigana, “Large scale alignment anomalies of CMB anisotropies: a new test for residuals applied to WMAP 5yr maps,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 004, 2009.
- [21] M. Frommert and T. A. Enßlin, “The axis of evil—a polarization perspective,” *Monthly Notices of the Royal Astronomical Society*, vol. 403, no. 4, pp. 1739–1748, 2010.
- [22] C. Monteserín, R. B. Barreiro, P. Vielva, E. Martínez-González, M. P. Hobson, and A. N. Lasenby, “A low cosmic microwave background variance in the Wilkinson microwave anisotropy probe data,” *Monthly Notices of the Royal Astronomical Society*, vol. 387, no. 1, pp. 209–219, 2008.
- [23] Y. Ayaita, M. Weber, and C. Wetterich, “Too few spots in the cosmic microwave background,” *Physical Review D*, vol. 81, no. 2, Article ID 023507, 2010.
- [24] L. Cayón, “Variograms of the cosmic microwave background temperature fluctuations: confirmation of deviations from statistical isotropy,” *Monthly Notices of the Royal Astronomical Society*, vol. 405, no. 2, pp. 1084–1088, 2010.
- [25] M. Cruz, P. Vielva, E. Martínez-González, and R. B. Barreiro, “Anomalous variance in the WMAP data caused by Galactic Foreground residuals,” *Monthly Notices of the Royal Astronomical Society*, <http://arxiv.org/abs/1005.1264>. Submitted.
- [26] Y. Wiaux, P. Vielva, E. Martínez-González, and P. Vanderghaynst, “Global universe anisotropy probed by the alignment of structures in the cosmic microwave background,” *Physical Review Letters*, vol. 96, no. 15, Article ID 151303, 2006.
- [27] P. Vielva, Y. Wiaux, E. Martínez-González, and P. Vanderghaynst, “Alignment and signed-intensity anomalies in Wilkinson microwave anisotropy probe data,” *Monthly Notices of the Royal Astronomical Society*, vol. 381, no. 3, pp. 932–942, 2007.
- [28] P. Vielva, E. Martínez-González, R. B. Barreiro, J. L. Sanz, and L. Cayón, “Detection of non-gaussianity in the Wilkinson microwave anisotropy probe first-year data using spherical wavelets,” *The Astrophysical Journal*, vol. 609, no. 1, pp. 22–34, 2004.
- [29] A. Liddle and D. H. Lyth, *Cosmological Inflation and Large-Scale Structure*, Cambridge University Press, Cambridge, UK, 2000.
- [30] P. G. Ferreira, J. Magueijo, and J. Silk, “Cumulants as non-Gaussian qualifiers,” *Physical Review D*, vol. 56, no. 8, pp. 4592–4603, 1997.
- [31] L. Popa, “Power fluctuations in the wavelet space: large-scale non-gaussian statistics,” *New Astronomy*, vol. 3, no. 7, pp. 563–570, 1998.
- [32] J. Pando, D. Valls-Gabaud, and L.-Z. Fang, “Evidence for scale-scale correlations in the cosmic microwave background

- radiation,” *Physical Review Letters*, vol. 81, no. 21, pp. 4568–4571, 1998.
- [33] C. L. Bennett, A. Banday, K. M. Gorski et al., “Four-year COBE DMR cosmic microwave background observations: maps and basic results,” *Astrophysical Journal Letters*, vol. 464, pp. L1–L4, 1996.
- [34] L. Tenorio, A. H. Jaffe, S. Hanany, and C. H. Lineweaver, “Applications of wavelets to the analysis of cosmic microwave background maps,” *Monthly Notices of the Royal Astronomical Society*, vol. 310, no. 3, pp. 823–834, 1999.
- [35] L. Cayón, J. L. Sanz, R. B. Barreiro et al., “Isotropic wavelets: a powerful tool to extract point sources from cosmic microwave background maps,” *Monthly Notices of the Royal Astronomical Society*, vol. 315, no. 4, pp. 757–761, 2000.
- [36] P. Vielva, E. Martínez-González, J. E. Gallegos, L. Toffolatti, and J. L. Sanz, “Point source detection using the Spherical Mexican Hat Wavelet on simulated all-sky Planck maps,” *Monthly Notices of the Royal Astronomical Society*, vol. 344, no. 1, pp. 89–104, 2003.
- [37] J. González-Nuevo, F. Argüeso, M. López-Cañiego et al., “The Mexican hat wavelet family: application to point-source detection in cosmic microwave background maps,” *Monthly Notices of the Royal Astronomical Society*, vol. 369, no. 4, pp. 1603–1610, 2006.
- [38] M. López-Cañiego, D. Herranz, J. González-Nuevo et al., “Comparison of filters for the detection of point sources in Planck simulations,” *Monthly Notices of the Royal Astronomical Society*, vol. 370, no. 4, pp. 2047–2063, 2006.
- [39] S. Pires, J. B. Juin, D. Yvon, Y. Moudren, S. Anthoine, and E. Pierpaoli, “Sunyaev-Zel’dovich cluster reconstruction in multiband bolometer camera surveys,” *Astronomy & Astrophysics*, vol. 455, no. 2, pp. 741–755, 2006.
- [40] N. Aghanim and O. Forni, “Searching for the non-Gaussian signature of the CMB secondary anisotropies,” *Astronomy & Astrophysics*, vol. 347, no. 2, pp. 409–418, 1999.
- [41] P. Mukherjee, M. P. Hobson, and A. N. Lasenby, “Do wavelets really detect non-Gaussianity in the 4-year COBE data?” *Monthly Notices of the Royal Astronomical Society*, vol. 318, no. 4, pp. 1157–1163, 2000.
- [42] R. B. Barreiro, M. P. Hobson, A. N. Lasenby, A. J. Banday, K. M. Górski, and G. Hinshaw, “Testing the Gaussianity of the COBE DMR data with spherical wavelets,” *Monthly Notices of the Royal Astronomical Society*, vol. 318, no. 2, pp. 475–480, 2000.
- [43] R. B. Barreiro and M. P. Hobson, “The discriminating power of wavelets to detect non-Gaussianity in the cosmic microwave background,” *Monthly Notices of the Royal Astronomical Society*, vol. 327, no. 3, pp. 813–828, 2001.
- [44] L. Cayón, J. L. Sanz, E. Martínez-González et al., “Spherical Mexican hat wavelet: an application to detect non-Gaussianity in the COBE-DMR maps,” *Monthly Notices of the Royal Astronomical Society*, vol. 326, no. 4, pp. 1243–1248, 2001.
- [45] E. Martínez-González, J. E. Gallegos, F. Argüeso, L. Cayón, and J. L. Sanz, “The performance of spherical wavelets to detect non-Gaussianity in the cosmic microwave background sky,” *Monthly Notices of the Royal Astronomical Society*, vol. 336, no. 1, pp. 22–32, 2002.
- [46] M. Cruz, E. Martínez-González, P. Vielva, and L. Cayón, “Detection of a non-Gaussian spot in WMAP,” *Monthly Notices of the Royal Astronomical Society*, vol. 356, no. 1, pp. 29–40, 2005.
- [47] J. D. McEwen, M. P. Hobson, A. N. Lasenby, and D. J. Mortlock, “A high-significance detection of non-Gaussianity in the Wilkinson microwave anisotropy probe 1-yr data using directional spherical wavelets,” *Monthly Notices of the Royal Astronomical Society*, vol. 359, no. 4, pp. 1583–1596, 2005.
- [48] D. Pietrobon, A. Amblard, A. Balbi, P. Cabella, A. Cooray, and D. Marinucci, “Needlet detection of features in the WMAP CMB sky and the impact on anisotropies and hemispherical asymmetries,” *Physical Review D*, vol. 78, no. 10, Article ID 103504, 2008.
- [49] Y. Wiaux, P. Vielva, R. B. Barreiro, E. Martínez-González, and P. Vanderghenst, “Non-Gaussianity analysis on local morphological measures of WMAP data,” *Monthly Notices of the Royal Astronomical Society*, vol. 385, no. 2, pp. 939–947, 2008.
- [50] A. Curto, E. Martínez-González, and R. B. Barreiro, “Improved constraints on primordial non-gaussianity for the wilkinson microwave anisotropy probe 5-year data,” *The Astrophysical Journal*, vol. 706, no. 1, pp. 399–403, 2009.
- [51] Ø. Rudjord, F. K. Hansen, X. Lan, M. Liguori, D. Marinucci, and S. Matarrese, “An estimate of the primordial non-gaussianity parameter  $f_{NL}$  using the needlet bispectrum from wmap,” *The Astrophysical Journal*, vol. 701, no. 1, pp. 369–376, 2009.
- [52] P. Vielva, E. Martínez-González, and M. Tucci, “Cross-correlation of the cosmic microwave background and radio galaxies in real, harmonic and wavelet spaces: detection of the integrated Sachs-Wolfe effect and dark energy constraints,” *Monthly Notices of the Royal Astronomical Society*, vol. 365, no. 3, pp. 891–901, 2006.
- [53] D. Pietrobon, A. Balbi, and D. Marinucci, “Integrated Sachs-Wolfe effect from the cross correlation of WMAP 3 year and the NRAO VLA sky survey data: new results and constraints on dark energy,” *Physical Review D*, vol. 74, no. 4, Article ID 043524, 2006.
- [54] X. Liu and S. -N. Zhang, “A cross-correlation analysis of WMAP and egret data in wavelet space,” *The Astrophysical Journal*, vol. 636, no. 1, pp. L1–L4, 2006.
- [55] J. D. McEwen, Y. Wiaux, M. P. Hobson, P. Vanderghenst, and A. N. Lasenby, “Probing dark energy with steerable wavelets through correlation of WMAP and NVSS local morphological measures,” *Monthly Notices of the Royal Astronomical Society*, vol. 384, no. 4, pp. 1289–1300, 2008.
- [56] L. Cao and L.-Z. Fang, “A wavelet-galerkin algorithm of the E/B decomposition of cosmic microwave background polarization maps,” *The Astrophysical Journal*, vol. 706, no. 2, p. 1545, 2009.
- [57] D. K. Hammond, Y. Wiaux, and P. Vanderghenst, “Wavelet domain Bayesian denoising of string signal in the cosmic microwave background,” *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 3, pp. 1317–1332, 2009.
- [58] K. Maisinger, M. P. Hobson, and A. N. Lasenby, “Maximum-entropy image reconstruction using wavelets,” *Monthly Notices of the Royal Astronomical Society*, vol. 347, no. 1, pp. 339–354, 2004.
- [59] Y. Moudren, P. Abrial, P. Vielva et al., “Independent component separation from incomplete spherical data using wavelets. Application to CMB data analysis,” in *Proceedings of the International Conference on Physics in Signal and Image Processing (PSIP ’05)*, Toulouse, France, January 2005.
- [60] F. K. Hansen, A. J. Banday, H. K. Eriksen, K. M. Górski, and P. B. Lilje, “Foreground subtraction of cosmic microwave background maps using WI-fit (wavelet-based high-resolution fitting of internal templates),” *The Astrophysical Journal*, vol. 648, no. 2, pp. 784–796, 2006.

- [61] J. Delabrouille, J.-F. Cardoso, M. Le Jeune, M. Betoule, G. Fay, and F. Guillaoux, "A full sky, low foreground, high resolution CMB map from WMAP," *Astronomy & Astrophysics*, vol. 493, no. 3, pp. 835–837, 2009.
- [62] J. L. Sanz, F. Argüeso, L. Cayón, E. Martínez-González, R. B. Barreiro, and L. Toffolatti, "Wavelets applied to cosmic microwave background maps: a multiresolution analysis for denoising," *Monthly Notices of the Royal Astronomical Society*, vol. 309, no. 3, pp. 672–680, 1999.
- [63] G. Faÿ, F. Guillaoux, M. Betoule, J.-F. Cardoso, J. Delabrouille, and M. Le Jeune, "CMB power spectrum estimation using wavelets," *Physical Review D*, vol. 78, no. 8, Article ID 083013, 2008.
- [64] P. Mukherjee and Y. Wang, "Model-independent reconstruction of the primordial power spectrum from Wilkinson microwave anisotropy probe data," *The Astrophysical Journal*, vol. 599, no. 1, pp. 1–6, 2003.
- [65] J. D. McEwen, P. Vielva, Y. Wiaux et al., "Cosmological applications of a wavelet analysis on the sphere," *Journal of Fourier Analysis and Applications*, vol. 13, no. 4, pp. 495–510, 2007.
- [66] P. Vielva, "Probing the Gaussianity and the statistical isotropy of the CMB with spherical wavelets," in *Wavelets XII*, D. van de Ville, V. K. Goyal, and M. Papadakis, Eds., vol. 6701 of *Proceedings of the SPIE*, p. 19, San Diego, Calif, USA, 2007.
- [67] R. T. Odgen, *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhauser, Boston, Mass, USA, 1997.
- [68] J.-P. Antoine, R. Murenzi, P. Vandergheynst, and S. T. Ali, *Two-Dimensional Wavelets and Their Relatives*, Cambridge University Press, Cambridge, UK, 2004.
- [69] J.-P. Antoine and P. Vandergheynst, "Wavelets on the  $n$ -sphere and related manifolds," *Journal of Mathematical Physics*, vol. 39, no. 8, pp. 3987–4008, 1998.
- [70] Y. Wiaux, L. Jacques, and P. Vandergheynst, "Correspondence principle between spherical and euclidean wavelets," *The Astrophysical Journal*, vol. 632, no. 1, pp. 15–28, 2005.
- [71] J. L. Sanz, D. Herranz, M. López-Caniego, and F. Argüeso, "Wavelets on the sphere. Application to the detection problem," in *Proceedings of the 14th European Signal Processing Conference (EUSIPCO '06)*, F. Gini and E. E. Kuruoglu, Eds., Florence, Italy, September 2006.
- [72] Y. Wiaux, J. D. McEwen, and P. Vielva, "Complex data processing: fast wavelet analysis on the sphere," *Journal of Fourier Analysis and Applications*, vol. 13, no. 4, pp. 477–493, 2007.
- [73] M. Górski, E. Hivon, A. J. Banday et al., "HEALPix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere," *The Astrophysical Journal*, vol. 622, no. 2, pp. 759–771, 2005.
- [74] C.-G. Park, "Non-Gaussian signatures in the temperature fluctuation observed by the Wilkinson microwave anisotropy probe," *Monthly Notices of the Royal Astronomical Society*, vol. 349, no. 1, pp. 313–320, 2004.
- [75] E. L. Wright, X. Chen, N. Odegard et al., "Five-year wilkinson microwave anisotropy probe observations: source catalog," *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 283–295, 2009.
- [76] M. Cruz, L. Cayón, E. Martínez-González, P. Vielva, and J. Jin, "The non-gaussian cold spot in the 3 year Wilkinson microwave anisotropy probe data," *The Astrophysical Journal*, vol. 655, no. 1, pp. 11–20, 2007.
- [77] P. Coles, "Statistical geometry and the microwave background," *Monthly Notices of the Royal Astronomical Society*, vol. 234, pp. 509–531, 1988.
- [78] J. R. Gott III, C. Park, R. Juszkiewicz et al., "Topology of microwave background fluctuations: theory," *The Astrophysical Journal*, vol. 352, no. 1, pp. 1–14, 1990.
- [79] J. Schmalzing and K. M. Górski, "Minkowski functionals used in the morphological analysis of cosmic microwave background anisotropy maps," *Monthly Notices of the Royal Astronomical Society*, vol. 297, no. 2, pp. 355–365, 1998.
- [80] D. Donoho and J. Jin, "Higher criticism for detecting sparse heterogeneous mixtures," *Annals of Statistics*, vol. 32, no. 3, pp. 962–994, 2004.
- [81] L. Cayón, J. Jin, and A. Treaster, "Higher criticism statistic: detecting and identifying non-Gaussianity in the WMAP first-year data," *Monthly Notices of the Royal Astronomical Society*, vol. 362, no. 3, pp. 826–832, 2005.
- [82] M. Cruz, M. Tucci, E. Martínez-González, and P. Vielva, "The non-Gaussian cold spot in Wilkinson microwave anisotropy probe: significance, morphology and foreground contribution," *Monthly Notices of the Royal Astronomical Society*, vol. 369, no. 1, pp. 57–67, 2006.
- [83] R. Zhang and D. Huterer, "Disks in the sky: a reassessment of the WMAP "cold spot"," *Astroparticle Physics*, vol. 33, no. 2, pp. 69–74, 2010.
- [84] D. Herranz, J. L. Sanz, R. B. Barreiro, and E. Martínez-González, "Scale-adaptive filters for the detection/separation of compact sources," *The Astrophysical Journal*, vol. 580, no. 1, pp. 610–625, 2002.
- [85] M. P. Hobson, A. W. Jones, and A. N. Lasenby, "Wavelet analysis and the detection of non-Gaussianity in the cosmic microwave background," *Monthly Notices of the Royal Astronomical Society*, vol. 309, no. 1, pp. 125–140, 1999.
- [86] C. Räth, P. Schuecker, and A. J. Banday, "A scaling index analysis of the Wilkinson microwave anisotropy probe three-year data: signatures of non-Gaussianities and asymmetries in the cosmic microwave background," *Monthly Notices of the Royal Astronomical Society*, vol. 380, no. 2, pp. 466–478, 2007.
- [87] G. Rossmannith, C. Räth, A. J. Banday, and G. Morfill, "Non-gaussian signatures in the five-year WMAP data as identified with isotropic scaling indices," *Monthly Notices of the Royal Astronomical Society*, vol. 399, no. 4, pp. 1921–1933, 2009.
- [88] V. G. Gurzadyan, A. E. Allahverdyan, T. Ghahramanyan et al., "Kolmogorov cosmic microwave background sky," *Astronomy & Astrophysics*, vol. 497, no. 2, pp. 343–346, 2009.
- [89] M. Tegmark, A. de Oliveira-Costa, and A. J. S. Hamilton, "High resolution foreground cleaned CMB map from WMAP," *Physical Review D*, vol. 68, no. 12, Article ID 123523, 2003.
- [90] J. L. Jonas, E. E. Baart, and G. D. Nicolson, "The Rhodes/HartRAO 2326-MHz radio continuum survey," *Monthly Notices of the Royal Astronomical Society*, vol. 297, no. 4, pp. 977–989, 1998.
- [91] D. P. Finkbeiner, "A full-sky  $H\alpha$  template for microwave foreground prediction," *Astrophysical Journal, Supplement Series*, vol. 146, no. 2, pp. 407–415, 2003.
- [92] D. P. Finkbeiner, M. Davis, and D. J. Schlegel, "Extrapolation of galactic dust emission at 100 microns to cosmic microwave background radiation frequencies using FIRAS," *The Astrophysical Journal*, vol. 524, no. 2, pp. 867–886, 1999.
- [93] R. A. Sunyaev and Ya. B. Zeldovich, "Small-scale fluctuations of relic radiation," *Astrophysics and Space Science*, vol. 7, no. 1, pp. 3–19, 1970.
- [94] S. Brough, D. A. Forbes, V. A. Kilborn, W. Couch, and M. Colless, "Eridanus—a supergroup in the local Universe?" *Monthly Notices of the Royal Astronomical Society*, vol. 369, no. 3, pp. 1351–1374, 2006.

- [95] M. Cruz, E. Martínez-González, P. Vielva, J. M. Diego, M. Hobson, and N. Turok, “The CMB cold spot: texture, cluster or void?” *Monthly Notices of the Royal Astronomical Society*, vol. 390, no. 3, pp. 913–919, 2008.
- [96] E. Martínez-González and J. L. Sanz, “CMB anisotropies generated by cosmic voids and great attractors,” *Monthly Notices of the Royal Astronomical Society*, vol. 247, no. 3, pp. 473–478, 1990.
- [97] E. Martínez-González, J. L. Sanz, and J. Silk, “Anisotropies in the microwave sky due to nonlinear structures,” *The Astrophysical Journal*, vol. 355, no. 1, pp. L5–L9, 1990.
- [98] K. T. Indue and J. Silk, “Local voids as the origin of large-angle cosmic microwave background anomalies. I,” *The Astrophysical Journal*, vol. 648, no. 1, pp. 23–30, 2006.
- [99] K. T. Inoue and J. Silk, “Local voids as the origin of large-angle cosmic microwave background anomalies: the effect of a cosmological constant,” *The Astrophysical Journal*, vol. 664, no. 2, pp. 650–659, 2007.
- [100] L. Rudnick, S. Brown, and L. R. Williams, “Extragalactic radio sources and the WMAP cold spot,” *The Astrophysical Journal*, vol. 671, no. 1, pp. 40–44, 2007.
- [101] J. J. Condon, W. D. Cotton, E. W. Greisen et al., “The NRAO VLA sky survey,” *Astronomical Journal*, vol. 115, no. 5, pp. 1693–1716, 1998.
- [102] J. M. Colberg, R. K. Sheth, A. Diaferio, L. Gao, and N. Yoshida, “Voids in a  $\Lambda$ CDM universe,” *Monthly Notices of the Royal Astronomical Society*, vol. 360, no. 1, pp. 216–226, 2005.
- [103] K. M. Smith and D. Huterer, “No evidence for the cold spot in the NVSS radio survey,” *Monthly Notices of the Royal Astronomical Society*, vol. 403, no. 1, pp. 2–8, 2010.
- [104] B. R. Granett, I. Szapudi, and M. C. Neyrinck, “Galaxy counts on the cosmic microwave background cold spot,” *The Astrophysical Journal*, vol. 714, no. 1, p. 825, 2010.
- [105] M. N. Bremer, J. Silk, L. J. M. Davies, and M. D. Lehnert, “A redshift survey towards the cosmic microwave background cold spot,” *Monthly Notices of the Royal Astronomical Society*, vol. 404, no. 1, pp. L69–L73, 2010.
- [106] M. Cruz, N. Turok, P. Vielva, E. Martínez-González, and M. Hobson, “A cosmic microwave background feature consistent with a cosmic texture,” *Science*, vol. 318, no. 5856, pp. 1612–1614, 2007.
- [107] N. Turok, “Global texture as the origin of cosmic structure,” *Physical Review Letters*, vol. 63, no. 24, pp. 2625–2628, 1989.
- [108] A. Vilenkin and E. P. S. Shellard, *Cosmic Strings and Other Topological Defects*, Cambridge University Press, Cambridge, UK, 2000.
- [109] N. Turok and D. Spergel, “Global texture and the microwave background,” *Physical Review Letters*, vol. 64, no. 23, pp. 2736–2739, 1990.
- [110] A. Silvestri and M. Trodden, “Non-Gaussian signatures from the postinflationary early Universe,” *Physical Review Letters*, vol. 103, no. 25, Article ID 251301, 2009.
- [111] N. Bevis, M. Hindmarsh, and M. Kunz, “WMAP constraints on inflationary models with global defects,” *Physical Review D*, vol. 70, no. 4, Article ID 043508, 9 pages, 2004.
- [112] J. Urrestilla, N. Bevis, M. Hindmarsh, M. Kunz, and A. R. Liddle, “Cosmic microwave anisotropies from BPS semilocal strings,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 7, article 010, 2008.
- [113] E. Komatsu, K. M. Smith, J. Dunkley et al., “Seven-year Wilkinson microwave anisotropy probe (WMAP) observations: cosmological interpretation,” <http://arxiv.org/abs/1001.4538>.
- [114] U.-L. Pen, D. N. Spergel, and N. Turok, “Cosmic structure formation and microwave anisotropies from global field ordering,” *Physical Review D*, vol. 49, no. 2, pp. 692–729, 1994.
- [115] R. Durrer, M. Kunz, and A. Melchiorri, “Cosmic microwave background anisotropies from scaling seeds: global defect models,” *Physical Review D*, vol. 59, no. 12, Article ID 123005, 26 pages, 1999.
- [116] P. Mukherjee, D. Parkinson, and A. R. Liddle, “A nested sampling algorithm for cosmological model selection,” *The Astrophysical Journal*, vol. 638, no. 2, pp. L51–L54, 2006.
- [117] M. Bridges, J. D. McEwen, A. N. Lasenby, and M. P. Hobson, “Markov chain Monte Carlo analysis of Bianchi VIII models,” *Monthly Notices of the Royal Astronomical Society*, vol. 377, no. 4, pp. 1473–1480, 2007.
- [118] P. Mukherjee and A. R. Liddle, “Planck and re-ionization history: a model selection view,” *Monthly Notices of the Royal Astronomical Society*, vol. 389, no. 1, pp. 231–236, 2008.
- [119] P. Carvalho, G. Rocha, and M. P. Hobson, “A fast Bayesian approach to discrete object detection in astronomical data sets—PowellSnakes I,” *Monthly Notices of the Royal Astronomical Society*, vol. 393, no. 3, pp. 681–702, 2009.
- [120] P. Vielva and J. L. Sanz, “Analysis of non-Gaussian cosmic microwave background maps based on the N-pdf. Application to Wilkinson microwave anisotropy probe data,” *Monthly Notices of the Royal Astronomical Society*, vol. 397, no. 2, pp. 837–848, 2009.
- [121] H. Jeffreys, *Theory of Probability*, Oxford University Press, Oxford, UK, 3rd edition, 1961.
- [122] M. Bridges, J. D. McEwen, M. Cruz et al., “Bianchi VIII models and the cold spot texture,” *Monthly Notices of the Royal Astronomical Society*, vol. 390, no. 4, pp. 1372–1376, 2008.
- [123] T. R. Jaffe, A. J. Banday, H. K. Eriksen, K. M. Górski, and F. K. Hansen, “Evidence of vorticity and shear at large angular scales in the WMAP data: a violation of cosmological isotropy?” *The Astrophysical Journal*, vol. 629, no. 1, pp. L1–L4, 2005.
- [124] D. N. Spergel, N. Turok, W. H. Press, and B. S. Ryden, “Global texture as the origin of large-scale structure: numerical simulations of evolution,” *Physical Review D*, vol. 43, no. 4, pp. 1038–1046, 1991.
- [125] F. Feroz, M. P. Hobson, and M. Bridges, “MultiNest: an efficient and robust Bayesian inference tool for cosmology and particle physics,” *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 4, pp. 1601–1614, 2009.
- [126] P. Vielva, E. Martínez-González, M. Cruz, R. B. Barreiro, and M. Tucci, “CMB polarization as a probe of the anomalous nature of the Cold Spot,” *Monthly Notices of the Royal Astronomical Society*, <http://fr.arxiv.org/abs/1002.4029>. In press.
- [127] G. Cowan, *Statistical Data Analysis*, Oxford University Press, Oxford, UK, 1998.
- [128] J. Tauber et al., “Planck pre-launch status: The Planck mission,” *Astronomy & Astrophysics*, Vol. 520, id.A1.
- [129] J. A. Rubiño-Martín, R. Rebolo, M. Tucci et al., “The Quijote CMB experiment,” in *Highlights of Spanish Astrophysics V, Proceedings of the VIII Scientific Meeting of the Spanish Astronomical Society (SEA)*, J. Gorgas, L. J. Goicoechea, J. I. Gonzalez-Serrano, and J. M. Diego, Eds., pp. 127–135, Santander, Spain, July 2010.
- [130] S. Das and D. N. Spergel, “CMB lensing and the WMAP cold spot,” *Physical Review D*, vol. 79, no. 4, Article ID 043007, 2009.

## Review Article

# Primordial Non-Gaussianities from Inflation Models

**Xingang Chen**

*Center for Theoretical Cosmology, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK*

Correspondence should be addressed to Xingang Chen, x.chen@damtp.cam.ac.uk

Received 15 February 2010; Accepted 9 June 2010

Academic Editor: Sarah Shandera

Copyright © 2010 Xingang Chen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This is a pedagogical review on primordial non-Gaussianities from inflation models. We introduce formalisms and techniques that are used to compute such quantities. We review different mechanisms which can generate observable large non-Gaussianities during inflation, and distinctive signatures they leave on the non-Gaussian profiles. They are potentially powerful probes to the dynamics of inflation. We also provide a nontechnical and qualitative summary of the main results and underlying physics.

## 1. Introduction

An ambitious goal of modern cosmology is to understand the origin of our Universe all the way to its very beginning. To what extent this can be achieved largely depends on what type of observational data we are able to get. Thanks to many modern experiments, we are really making progress in this direction.

One of the representative experiments is the Wilkinson Microwave Anisotropy Probe (WMAP) satellite [1–6]. It measures the light coming from the last scattering surface about 13.7 billions years ago. This cosmic microwave background (CMB) is emitted at about 379,000 years after the Big Bang, when electrons and protons combine to form neutral hydrogen atoms and photons start to travel freely through the space. Our Universe was very young at that moment and the large scale fluctuations were still developing at linear level. So the CMB actually carries valuable information much earlier than itself, which can potentially tell us about the origin of the Big Bang.

There are two amazing facts about the CMB temperature map. On the one hand, it is extremely isotropic, despite the fact that the causally connected region at the time when CMB formed spans an angle of only about 0.8 degree on the sky today. On the other hand, we do observe small fluctuations, with  $\Delta T/T \sim 10^{-5}$ .

The inflationary scenario [7–9] naturally solves the above two puzzles. It was proposed nearly thirty years ago

to address some of the basic problems of the Big Bang cosmology, namely, why the universe is so homogeneous and isotropic. In this scenario, our universe was once dominated by dark energy and had gone through an accelerated expansion phase, during which a Hubble size patch was stretched by more than 60 e-folds or so. Inhomogeneities and large curvature were stretched away by this inflationary epoch, making our current observable universe very homogenous and flat. In the mean while, the fields that were responsible for and participated in this inflationary phase did have quantum fluctuations. These fluctuations also got stretched and imprinted at superhorizon scales. Later they reentered the horizon and seeded the large scale structures today [10–14].

The inflationary scenario has several generic predictions on the properties of the density perturbations that seed the large scale structures.

- (i) They are primordial. Namely, they were laid down at superhorizon scales and entering the horizon after the Big Bang.
- (ii) They are approximately scale-invariant. This is because, during the 60 e-folds, each mode experiences the similar expansion when they are stretched across the horizon.
- (iii) They are approximately Gaussian. In simplest slow-roll inflation models, the inflaton is freely propagating in the inflationary background at the leading

order. This is also found to be true in most of the other models and for different inflationary mechanisms. So the tiny primordial fluctuations can be treated as nearly Gaussian.

The CMB temperature anisotropy is the ideal data that we can use to test these predictions. The obvious first step is to analyze their two-point correlation functions, that is, the power spectrum. All the above predictions are verified to some extent [1]. The presence of the baryon acoustic oscillations proves that the density perturbations are indeed present at the superhorizon scales and reentering the horizon as the horizon expands after the Big Bang. The spectral index,  $n_s = 0.963 \pm 0.012$ , is very close to one, therefore, the density perturbations are nearly scale invariant. Several generic types of non-Gaussianities are constrained to be less than one thousandth of the leading Gaussian component.

But is this enough?

Experimentally, the amplitude and the scale-dependence of the power spectrum consist of about 1000 numbers for WMAP. For the Planck satellite, this will be increased up to about 3000. However, we have about one million pixels in the WMAP temperature map alone, and 60 millions for Planck. So the information that we got so far is highly compressed comparing to what the data could offer in principle. This high compression is only justified if the density perturbations are Gaussian within the ultimate sensitivities of our experiments, so all the properties of the map is determined by the two-point function. Otherwise, we are expecting a lot more information in the non-Gaussian components.

Theoretically, inflation still remains as a paradigm. We do not know what kind of fields are responsible for the inflation. We do not know their Lagrangian. We also would like to distinguish inflation from other alternatives. Being our very first data on quantum gravity, we would like to extract the maximum number of information from the CMB map to understand aspects of the quantum gravity. All these motivate us to go beyond the power spectrum.

To give an analogy, in particle physics, two-point correlation functions of fields describe freely propagating particles in Minkowski spacetime. More interesting objects are their higher-order correlations. Measuring these are the goals of particle colliders. Similarly, the power spectrum here describes the freely propagating particles in the inflationary background. To find out more about their interaction details and break the degeneracies among models, we need higher-order correlation functions, namely, non-Gaussianities. So the role non-Gaussianities play for the very early universe is similar to the role colliders play for particle physics.

With these motivations in mind, in this paper, we explore various mechanisms that can generate potentially observable primordial non-Gaussianities, and are consistent with the current results of power spectrum. We will not take the approach of reviewing models one by one. Rather, we divide them into different categories, such that models in each category share the same physical aspect which leaves a unique fingerprint on primordial non-Gaussianities. On the one hand, if any such non-Gaussianity is observed, we know what

we have learned concretely in terms of fundamental physics. On the other hand, explicit forms of non-Gaussianities resulted from this exploration provide important clues on how they should be searched in data. Even if the primordial density perturbations were perfectly Gaussian, to test it, we would still go through these analyses until various well-motivated non-Gaussian forms are properly constrained.

*1.1. Road Map.* The following is the outline of the paper. For readers who would like to get a quick and qualitative understanding of the main results instead of technical details, we also provide a shortcut after the outline.

In Section 2, we review the essential features of the inflation model and density perturbations.

In Section 3, we review the first-principle in-in formalism and related techniques that will be used to calculate the correlation functions in time-dependent background.

In Section 4, we compute the scalar three-point function in the simplest slow-roll model. We list the essential assumptions that lead to the conclusion that the non-Gaussianities in this model is too small to be observed.

In Section 5, we review aspects of inflation model building, emphasizing various generic problems which suggest that the realistic model may not be the algebraically simplest. We also introduce some terminologies used to describe properties of non-Gaussianities.

Sections 6, 7, and 8 contain the main results of this paper. We study various mechanisms that can lead to large non-Gaussianities, and their distinctive predictions in terms of the non-Gaussian profile.

In Section 9, we give a qualitative summary of the main results in this paper. Before conclusion, we discuss several useful relations among different non-Gaussianities.

Here is a road map for readers who wish to have a non-technical explanation and understanding of our main results. After reading the short review on the inflation model and density perturbations in Section 2, one may read the first and the last paragraph of Section 4 to get an idea of the no-go statement, and then directly proceed to read Section 5. After that, one may jump to Section 9 where the main results are summarized in non-technical terms.

The subject of the primordial non-Gaussianities is a fast-growing one. There exists many nice reviews and books in this and closely related subjects. The introductions to inflation and density perturbations can be found in many textbooks [15–20] and reviews [21–25]. Inflationary model buildings in particle physics, supergravity, and string theory are reviewed in [26–32]. Comprehensive reviews on the developments of theories and observations of primordial non-Gaussianities up to mid 2004 can be found in [33, 34]. Recent comprehensive reviews on theoretical and observational developments on the bispectrum detection in CMB and large-scale structure has appeared in [35, 36]. A recent comprehensive review on non-Gaussianities from the second-order postinflationary evolution of CMB, which acts as contaminations of the primordial non-Gaussianities, has

appeared in [37]. A recent review on how primordial non-Gaussianities can be generated in alternatives to inflation can be found in [38].

## 2. Inflation and Density Perturbations

In this section, we give a quick review on basic elements of inflation and density perturbations. We consider the simplest slow-roll inflation. The action is

$$S = \frac{M_{\text{P}}}{2} \int d^4x \sqrt{-g} R + \int d^4x \sqrt{-g} \left[ -\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right]. \quad (1)$$

The first term is the Einstein-Hilbert action. The second term describes a canonical scalar field coupled to gravity through the metric  $g_{\mu\nu}$ . This is the inflaton, which stays on the potential  $V(\phi)$  and creates the vacuum energy that drives the inflation. We first study the zero-mode background evolution of the spacetime and the inflaton. The background metric is

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a^2(t) d\mathbf{x}^2, \quad (2)$$

where  $a(t)$  is the scale factor and  $\mathbf{x}$  is the comoving spatial coordinates. The equations of motion are

$$H^2 = \frac{1}{3M_{\text{P}}^2} \left( \frac{1}{2} \dot{\phi}_0^2 + V \right), \quad (3)$$

$$\dot{H} = -\frac{\dot{\phi}_0^2}{2M_{\text{P}}^2}, \quad (4)$$

$$\ddot{\phi}_0 + 3H\dot{\phi}_0 + V' = 0. \quad (5)$$

The first equation determines the Hubble parameter  $H$ , which is the expansion rate of the universe. The second equation is the continuity condition. The third equation describes the evolution of the inflaton. Only two of them are independent.

The requirement of having at least  $\mathcal{O}(60)$  e-fold of inflation imposes some important conditions. By definition, to have this amount of inflation, the Hubble parameter cannot change much within a Hubble time  $H^{-1}$ . This gives the first condition

$$\epsilon \equiv -\frac{\dot{H}}{H^2} \ll \mathcal{O}(1). \quad (6)$$

We also require that the parameter  $\epsilon$  does not change much within a Hubble time,

$$\eta \equiv \frac{\dot{\epsilon}}{\epsilon H} \ll \mathcal{O}(1). \quad (7)$$

In principle,  $\eta$  can be close to  $\mathcal{O}(1)$  but  $\epsilon$  kept small. In such a case,  $\epsilon$  grows exponentially with e-folds and the inflation period tends to be shorter. More importantly, such a case will not generate a scale-invariant spectrum, as we will see shortly, thus cannot be responsible for the CMB. The two

conditions (6) and (7) are called the *slow-roll conditions*. Using the background equations of motion, we can see that the slow-roll conditions impose restrictions on the rolling velocity of the inflaton. The first condition (6) implies that

$$\frac{\dot{\phi}_0^2}{2H^2 M_{\text{P}}^2} = \epsilon \ll \mathcal{O}(1). \quad (8)$$

So the energy driving the inflation on the right-hand side of (3) is dominated by the potential. Adding the second condition (7) further implies that

$$\frac{\ddot{\phi}_0}{\dot{\phi}_0 H} = -\epsilon + \frac{\eta}{2} \ll \mathcal{O}(1). \quad (9)$$

So the first term  $\ddot{\phi}_0$  in (5) is negligible and the evolution of the zero-mode inflaton is determined by the attractor solution

$$3H\dot{\phi}_0 + V' = 0. \quad (10)$$

Using (10), the slow-roll conditions can also be written in a form that restricts the shape of the potential,

$$\epsilon_V \equiv \frac{M_{\text{P}}^2}{2} \left( \frac{V'}{V} \right)^2 \ll \mathcal{O}(1), \quad \eta_V \equiv M_{\text{P}}^2 \frac{V''}{V} \ll \mathcal{O}(1). \quad (11)$$

They are related to  $\epsilon$  and  $\eta$  by

$$\epsilon = \epsilon_V, \quad \eta = -2\eta_V + 4\epsilon_V. \quad (12)$$

So the shape of the potential has to be rather flat relative to its height. We emphasize that, although in this example several definitions of the slow-roll conditions are all equivalent, the definitions (6) and (7) are more general. In other cases that we will encounter later in this paper, these two conditions are still necessary to ensure a prolonged inflation and generate a scale-invariant spectrum, but the others no longer have to be satisfied. For example, the shape of potential can be steeper, or the inflationary energy can be dominated by the kinetic energy.

Now let us study the perturbations. To keep things simple but main points illustrated, in this section, we will ignore the perturbations in the gravity sector and only perturb the inflaton,

$$\phi(\mathbf{x}, t) = \phi_0(t) + \delta\phi(\mathbf{x}, t). \quad (13)$$

We also ignore terms suppressed by the slow-roll parameters, which we often denote collectively as  $\mathcal{O}(\epsilon)$ . For example, the mass of the inflaton is  $V'' \sim \mathcal{O}(\epsilon)H^2$ , and will be ignored. The quadratic Lagrangian for the perturbation theory is simply

$$L = \int d^3x \left[ \frac{a^3}{2} \dot{\delta\phi}^2 - \frac{a}{2} (\partial_i \delta\phi)^2 \right], \quad (14)$$

and the equation of motion follows:

$$\delta\ddot{\phi}(\mathbf{k}, t) + 3H\delta\dot{\phi}(\mathbf{k}, t) + \frac{k^2}{a^2} \delta\phi(\mathbf{k}, t) = 0, \quad (15)$$

where we have written it in the comoving momentum space,

$$\delta\phi(\mathbf{k}, t) = \int d^3x \delta\phi(\mathbf{x}, t) e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (16)$$

The solution to the differential equation (15),  $u(\mathbf{k}, t)$ , is called the mode function. It is not difficult to check that

$$a^3 u(\mathbf{k}, t) \dot{u}^*(\mathbf{k}, t) - \text{c.c.} = t\text{-independent const.} \quad (17)$$

To quantize the perturbations according to the canonical commutation relations between  $\delta\phi$  and its conjugate momentum  $\delta\pi \equiv \partial L / \partial \dot{\delta\phi}$ ,

$$\begin{aligned} [\delta\phi(\mathbf{x}, t), \delta\pi(\mathbf{y}, t)] &= i\delta(\mathbf{x} - \mathbf{y}), \\ [\delta\phi(\mathbf{x}, t), \delta\phi(\mathbf{y}, t)] &= 0, \quad [\delta\pi(\mathbf{x}, t), \delta\pi(\mathbf{y}, t)] = 0, \end{aligned} \quad (18)$$

we decompose

$$\delta\phi = u(\mathbf{k}, t) a_{\mathbf{k}} + u^*(-\mathbf{k}, t) a_{-\mathbf{k}}^\dagger, \quad (19)$$

$$\delta\pi = a^3 \dot{u}(\mathbf{k}, t) a_{\mathbf{k}} + a^3 \dot{u}^*(-\mathbf{k}, t) a_{-\mathbf{k}}^\dagger, \quad (20)$$

with the commutation relations

$$\begin{aligned} [a_{\mathbf{p}}, a_{-\mathbf{q}}^\dagger] &= (2\pi)^3 \delta^3(\mathbf{p} + \mathbf{q}), \\ [a_{\mathbf{p}}, a_{-\mathbf{q}}] &= 0, \quad [a_{\mathbf{p}}^\dagger, a_{-\mathbf{q}}^\dagger] = 0. \end{aligned} \quad (21)$$

One can check that the commutation relations (18) and (21) are equivalent because of (17), given that the constant on the right-hand side of (17) is specified to be  $i$ . This gives the normalization condition for the mode function.

We now write down the mode function explicitly by solving (15):

$$u(\mathbf{k}, \tau) = C_+ \frac{H}{\sqrt{2k^3}} (1 + ik\tau) e^{-ik\tau} + C_- \frac{H}{\sqrt{2k^3}} (1 - ik\tau) e^{ik\tau}, \quad (22)$$

where we have used the conformal time  $\tau$  defined as  $dt \equiv ad\tau$ . The infinite past corresponds to  $\tau \rightarrow -\infty$  and the infinite future  $\tau \rightarrow 0$ . We also used the relation  $\tau = -1/Ha + \mathcal{O}(\epsilon)$ . This mode function is a superposition of two linearly independent solutions with the normalization condition

$$|C_+|^2 - |C_-|^2 = 1 \quad (23)$$

followed from (17). Consider the limit in which the mode is well within the horizon, that is, its wavelength  $a/k$  much shorter than the Hubble length  $1/H$ , and consider a time period much shorter than a Hubble time. In these limits, the mode effectively feels the Minkowski spacetime, and the first component in (22) with the positive frequency asymptotes to the vacuum mode of the Minkowski spacetime as we can see from (23). We choose this component as our vacuum choice, and it is usually called the Bunch-Davies state. The annihilation operator  $a_{\mathbf{p}}$  annihilates the corresponding Bunch-Davies vacuum,  $a_{\mathbf{p}}|0\rangle = 0$ .

The mode function

$$u(\mathbf{k}, \tau) = \frac{H}{\sqrt{2k^3}} (1 + ik\tau) e^{-ik\tau} \quad (24)$$

has the following important properties. It is oscillatory within the horizon  $k|\tau| \gg 1$ . As it gets stretched out of the horizon  $k|\tau| \ll 1$ , the amplitude becomes a constant and frozen. Physically this means that, if we look at different comoving patches of the universe that have the superhorizon size, and ignore the shorter wavelength fluctuations, they all evolve classically but with different  $\delta\phi$ . This difference makes them arrive at  $\phi_f$ , the location of the end of inflation, at different times. This space-dependent time difference  $\delta t \approx \delta\phi/\dot{\phi}_0$  leads to the space-dependent inflationary  $e$ -fold difference

$$\zeta \approx H\delta t \approx H \frac{\delta\phi}{\dot{\phi}_0}. \quad (25)$$

Again we ignore terms that are suppressed by the slow-roll parameters. This  $e$ -fold difference is the conserved quantity after the mode exits the horizon, and remains so until the mode reenters the horizon sometime after the Big Bang. It is the physical quantity that we can measure, for example, by measuring the temperature anisotropy in the CMB,  $\zeta \approx -5\Delta T/T$  [39]. The information about the primordial inflation is then encoded in the statistical properties of this variable. So we would like to calculate the correlation functions of this quantity. Using (25), (19), (24), and (8), we get the following two-point function:

$$\langle \zeta^2 \rangle \equiv \langle 0 | \zeta(\mathbf{k}_1, 0) \zeta(\mathbf{k}_2, 0) | 0 \rangle = \frac{P_\zeta}{2k_1^3} (2\pi)^5 \delta(\mathbf{k}_1 + \mathbf{k}_2), \quad (26)$$

where  $P_\zeta$  is defined as the power spectrum and in this case it is

$$P_\zeta = \frac{H^2}{8\pi^2 M_{\text{pl}}^2 \epsilon}. \quad (27)$$

The spectrum index is defined to be

$$n_s - 1 \equiv \frac{d \ln P_\zeta}{d \ln k} = -2\epsilon - \eta, \quad (28)$$

where the relation  $d \ln k = H dt$  is used. If  $n_s = 1$ , the spectrum is scale invariant. The current data from CMB tells us that  $n_s = 0.963 \pm 0.012$  [1]. So as we have mentioned, this requires a small  $\eta$ , which is also a value that tends to give more  $e$ -folds of inflation.

If this were the end of story, all the primordial density perturbations would be determined by this two-point function and they are Gaussian. The rest of the paper will be devoted to making the above procedure rigorous and to the calculations of higher-order non-Gaussian correlation functions in this and various other models.

### 3. In-In Formalism and Correlation Functions

In this section, we review the in-in formalism and the related techniques that are used to calculate the correlation functions in time-dependent background. The main procedure is summarized in the last subsection.

3.1. *In-In Formalism.* We start with the in-in formalism [40–45], following Weinberg’s presentation [45].

We are interested in the correlation functions of super-horizon primordial perturbations generated during inflation. So our goal is to calculate the expectation value of an operator  $Q$ , which is a product in terms of field perturbations  $\delta\phi_a$  and  $\delta\pi_a$ , at the end of inflation. The subscript  $a$  labels different fields. In inflation models, these fields are, for example, the fluctuations of the scalars and metric and their conjugate momenta. In the Heisenberg picture,

$$\langle Q \rangle \equiv \langle \Omega | Q(t) | \Omega \rangle, \quad (29)$$

where  $t$  is the end of inflation and  $|\Omega\rangle$  is the vacuum state for this interacting theory at the far past  $t_0$ .

We start by looking at how the time-dependence in  $Q(t)$  is generated.

The Hamiltonian of the system

$$H[\phi(t), \pi(t)] \equiv \int d^3x \mathcal{H}[\phi_a(\mathbf{x}, t), \pi_a(\mathbf{x}, t)] \quad (30)$$

is a functional of the fields  $\phi_a(\mathbf{x}, t)$  and their conjugate momenta  $\pi_a(\mathbf{x}, t)$  at a fixed time  $t$ . On the left-hand side of (30), we have suppressed the variable  $\mathbf{x}$  and index  $a$  which are integrated or summed over. The  $\phi_a(\mathbf{x}, t)$  and  $\pi_a(\mathbf{x}, t)$  satisfy the canonical commutation relations

$$\begin{aligned} [\phi_a(\mathbf{x}, t), \pi_b(\mathbf{y}, t)] &= i\delta_{ab}\delta^3(\mathbf{x} - \mathbf{y}), \\ [\phi_a(\mathbf{x}, t), \phi_b(\mathbf{y}, t)] &= [\pi_a(\mathbf{x}, t), \pi_b(\mathbf{y}, t)] = 0, \end{aligned} \quad (31)$$

and their evolution is generated by  $H$  following the equations of motion:

$$\begin{aligned} \dot{\phi}_a(\mathbf{x}, t) &= i[H[\phi(t), \pi(t)], \phi_a(\mathbf{x}, t)], \\ \dot{\pi}_a(\mathbf{x}, t) &= i[H[\phi(t), \pi(t)], \pi_a(\mathbf{x}, t)]. \end{aligned} \quad (32)$$

We consider a time-dependent background,  $\bar{\phi}_a(\mathbf{x}, t)$  and  $\bar{\pi}_a(\mathbf{x}, t)$  which are c-numbers and commute with everything, and the perturbations,  $\delta\phi_a(\mathbf{x}, t)$  and  $\delta\pi_a(\mathbf{x}, t)$ ,

$$\begin{aligned} \phi_a(\mathbf{x}, t) &\equiv \bar{\phi}_a(\mathbf{x}, t) + \delta\phi_a(\mathbf{x}, t), \\ \pi_a(\mathbf{x}, t) &\equiv \bar{\pi}_a(\mathbf{x}, t) + \delta\pi_a(\mathbf{x}, t). \end{aligned} \quad (33)$$

The background evolution is determined by the classical equations of motion,

$$\dot{\bar{\phi}}_a(\mathbf{x}, t) = \frac{\partial \mathcal{H}}{\partial \bar{\pi}_a}, \quad \dot{\bar{\pi}}_a(\mathbf{x}, t) = -\frac{\partial \mathcal{H}}{\partial \bar{\phi}_a}. \quad (34)$$

The commutation relations (31) become those for the perturbations,

$$\begin{aligned} [\delta\phi_a(\mathbf{x}, t), \delta\pi_b(\mathbf{y}, t)] &= i\delta_{ab}\delta^3(\mathbf{x} - \mathbf{y}), \\ [\delta\phi_a(\mathbf{x}, t), \delta\phi_b(\mathbf{y}, t)] &= [\delta\pi_a(\mathbf{x}, t), \delta\pi_b(\mathbf{y}, t)] = 0. \end{aligned} \quad (35)$$

We expand the Hamiltonian as

$$\begin{aligned} H[\phi(t), \pi(t)] &= H[\bar{\phi}(t), \bar{\pi}(t)] + \sum_a \int d^3x \frac{\partial \mathcal{H}}{\partial \phi_a(\mathbf{x}, t)} \delta\phi_a(\mathbf{x}, t) \\ &\quad + \sum_a \int d^3x \frac{\partial \mathcal{H}}{\partial \pi_a(\mathbf{x}, t)} \delta\pi_a(\mathbf{x}, t) \\ &\quad + \tilde{H}[\delta\phi(t), \delta\pi(t); t], \end{aligned} \quad (36)$$

where we use  $\tilde{H}$  to denote terms of quadratic and higher-orders in perturbations.

Using (34), (35), and (36), the equations of motion (32) become

$$\begin{aligned} \delta\dot{\phi}_a(\mathbf{x}, t) &= i[\tilde{H}[\delta\phi(t), \delta\pi(t); t], \delta\phi_a(\mathbf{x}, t)], \\ \delta\dot{\pi}_a(\mathbf{x}, t) &= i[\tilde{H}[\delta\phi(t), \delta\pi(t); t], \delta\pi_a(\mathbf{x}, t)]. \end{aligned} \quad (37)$$

So the evolution of the perturbations,  $\delta\phi_a$  and  $\delta\pi_a$ , is generated by  $\tilde{H}$ . It is straightforward to verify that the solutions for (37) are

$$\begin{aligned} \delta\phi_a(\mathbf{x}, t) &= U^{-1}(t, t_0)\delta\phi_a(\mathbf{x}, t_0)U(t, t_0), \\ \delta\pi_a(\mathbf{x}, t) &= U^{-1}(t, t_0)\delta\pi_a(\mathbf{x}, t_0)U(t, t_0), \end{aligned} \quad (38)$$

where  $U$  satisfies

$$\frac{d}{dt}U(t, t_0) = -i\tilde{H}[\delta\phi(t_0), \delta\pi(t_0); t]U(t, t_0) \quad (39)$$

with the condition at an initial time  $t_0$  being

$$U(t_0, t_0) = 1. \quad (40)$$

To have a systematic scheme to do the perturbation theory, we split  $\tilde{H}$  into two parts,

$$\tilde{H}[\delta\phi(t), \delta\pi(t); t] = H_0[\delta\phi(t), \delta\pi(t); t] + H_I[\delta\phi(t), \delta\pi(t); t]. \quad (41)$$

The  $H_0$  is the quadratic kinematic part, which in the perturbation theory will describe the leading evolution of fields. Fields whose evolution are generated by  $H_0$  are called the *interaction picture fields*. We add a superscript “ $I$ ” to label such fields. They satisfy

$$\begin{aligned} \delta\dot{\phi}_a^I(\mathbf{x}, t) &= i[H_0[\delta\phi^I(t), \delta\pi^I(t); t], \delta\phi_a^I(\mathbf{x}, t)], \\ \delta\dot{\pi}_a^I(\mathbf{x}, t) &= i[H_0[\delta\phi^I(t), \delta\pi^I(t); t], \delta\pi_a^I(\mathbf{x}, t)]. \end{aligned} \quad (42)$$

The solutions are

$$\begin{aligned} \delta\phi_a^I(\mathbf{x}, t) &= U_0^{-1}(t, t_0)\delta\phi_a(\mathbf{x}, t_0)U_0(t, t_0), \\ \delta\pi_a^I(\mathbf{x}, t) &= U_0^{-1}(t, t_0)\delta\pi_a(\mathbf{x}, t_0)U_0(t, t_0), \end{aligned} \quad (43)$$

where  $U_0$  satisfies

$$\frac{d}{dt}U_0(t, t_0) = -iH_0[\delta\phi(t_0), \delta\pi(t_0); t]U_0(t, t_0) \quad (44)$$

with

$$U_0(t_0, t_0) = 1. \quad (45)$$

So the idea is to encode the leading kinematic evolution in terms of the interaction picture fields, and calculate the effects of the interaction through the series expansion in terms of powers of  $H_I$ . To do this, we rewrite (29) as

$$\begin{aligned} & \langle \Omega | Q[\delta\phi_a(\mathbf{x}, t), \delta\pi_a(\mathbf{x}, t)] | \Omega \rangle \\ &= \langle \Omega | U^{-1}(t, t_0) Q[\delta\phi_a(\mathbf{x}, t_0), \delta\pi_a(\mathbf{x}, t_0)] U(t, t_0) | \Omega \rangle \\ &= \langle \Omega | F^{-1}(t, t_0) U_0^{-1}(t, t_0) Q[\delta\phi_a(\mathbf{x}, t_0), \delta\pi_a(\mathbf{x}, t_0)] \\ & \quad \times U_0(t, t_0) F(t, t_0) | \Omega \rangle \\ &= \langle \Omega | F^{-1}(t, t_0) Q[\delta\phi_a^I(\mathbf{x}, t), \delta\pi_a^I(\mathbf{x}, t)] F(t, t_0) | \Omega \rangle, \end{aligned} \quad (46)$$

where

$$F(t, t_0) \equiv U_0^{-1}(t, t_0) U(t, t_0). \quad (47)$$

Using (39), (44), and (41), we get

$$\begin{aligned} \frac{d}{dt} F(t, t_0) &= -iU_0^{-1}(t, t_0) H_I[\delta\phi(t_0), \delta\pi(t_0); t] U_0(t, t_0) F(t, t_0) \\ &= -iH_I[\delta\phi^I(t), \delta\pi^I(t); t] F(t, t_0) \\ &\equiv -iH_I(t) F(t, t_0), \end{aligned} \quad (48)$$

with

$$F(t_0, t_0) = 1. \quad (49)$$

The solution to (48) and (49) can be written in the following way,

$$F(t, t_0) = T \exp\left(-i \int_{t_0}^t H_I(t) dt\right), \quad (50)$$

where the operator  $T$  means that, in each term in the Taylor series expansion of the exponential, the time variables have to be time-ordered. The operator  $\bar{T}$  will be used to mean the reversed time-ordering.

In summary, the expectation value (29) is

$$\begin{aligned} \langle Q \rangle &= \langle \Omega | F^{-1}(t, t_0) Q^I(t) F(t, t_0) | \Omega \rangle, \\ &= \langle \Omega | \left[ \bar{T} \exp\left(i \int_{t_0}^t H_I(t) dt\right) \right] Q^I(t) \\ & \quad \times \left[ T \exp\left(-i \int_{t_0}^t H_I(t) dt\right) \right] | \Omega \rangle. \end{aligned} \quad (51)$$

Notice that in

$$\begin{aligned} H_I(t) &\equiv H_I[\delta\phi^I(t), \delta\pi^I(t); t], \\ Q^I(t) &\equiv Q[\delta\phi_a^I(\mathbf{x}, t), \delta\pi_a^I(\mathbf{x}, t)], \end{aligned} \quad (52)$$

all the field perturbations are in the interaction picture.

The perturbation theory is also often done in terms of the Lagrangian formalism. In the following, we show that they are equivalent. In the above, we perform perturbations on the Hamiltonian, and define  $\delta\pi_a$  by perturbing  $\pi_a \equiv \partial L / \partial \dot{\phi}_a$ , (here we use  $\partial$  to denote the functional derivative)

$$\delta\pi_a = \frac{\partial L}{\partial \dot{\phi}_a}(\phi_a, \dot{\phi}_a) - \frac{\partial L}{\partial \dot{\phi}_a}(\bar{\phi}_a, \dot{\bar{\phi}}_a). \quad (53)$$

The Hamiltonian  $\tilde{H}$  is defined by (36). So using the definition

$$H \equiv \int d^3x \frac{\partial L}{\partial \dot{\phi}_a} \dot{\phi}_a - L, \quad (54)$$

together with the classical equations of motions (34) and  $\dot{\pi}_a = \partial L / \partial \bar{\phi}_a$ , the definition (36) for  $\tilde{H}$  becomes

$$\begin{aligned} \tilde{H} &= \int d^3x \frac{\partial L}{\partial \dot{\phi}_a}(\phi_a, \dot{\phi}_a) \delta\dot{\phi}_a \\ & \quad + \int d^3x \frac{\partial L}{\partial \dot{\phi}_a}(\bar{\phi}_a, \dot{\bar{\phi}}_a) \delta\dot{\phi}_a - L(\phi_a, \dot{\phi}_a) + L(\bar{\phi}_a, \dot{\bar{\phi}}_a). \end{aligned} \quad (55)$$

If we perturb the Lagrangian directly, we keep the part of the Lagrangian that is quadratic and higher in perturbations  $\delta\phi_a$  and  $\delta\dot{\phi}_a$ ,

$$\begin{aligned} \tilde{L}(\delta\phi_a, \delta\dot{\phi}_a, t) &\equiv L(\phi_a, \dot{\phi}_a) - L(\bar{\phi}_a, \dot{\bar{\phi}}_a) \\ & \quad - \int d^3x \frac{\partial L}{\partial \dot{\phi}_a}(\bar{\phi}_a, \dot{\bar{\phi}}_a) \delta\dot{\phi}_a \\ & \quad - \int d^3x \frac{\partial L}{\partial \dot{\phi}_a}(\bar{\phi}_a, \dot{\bar{\phi}}_a) \delta\dot{\phi}_a. \end{aligned} \quad (56)$$

The  $\delta\pi_a$  is defined directly as

$$\delta\pi_a \equiv \frac{\partial \tilde{L}}{\partial(\delta\dot{\phi}_a)} = \frac{\partial L}{\partial \dot{\phi}_a}(\phi_a, \dot{\phi}_a) - \frac{\partial L}{\partial \dot{\phi}_a}(\bar{\phi}_a, \dot{\bar{\phi}}_a), \quad (57)$$

where in the second step (56) has been used. So these two definitions of  $\delta\pi_a$  are equivalent. The Hamiltonian  $\tilde{H}$  is defined through  $\tilde{L}$ ,

$$\tilde{H} \equiv \int d^3x \frac{\partial \tilde{L}}{\partial \delta\dot{\phi}_a} \delta\dot{\phi}_a - \tilde{L}. \quad (58)$$

Again, using (56), we can see that the two definitions of  $\tilde{H}$  are equivalent.

**3.2. Mode Functions and Vacuum.** The Hamiltonian  $H_0$  in the above formalism is typically chosen to be the quadratic kinematic terms for field perturbations  $\delta\phi_a$  without mixing,

$$H_0 = \int d^3x \sum_a \left[ \frac{1}{2A} \delta\pi_a^2 + \frac{B}{2} (\partial_i \delta\phi_a)^2 + \frac{C}{2} \delta\phi_a^2 \right]. \quad (59)$$

So they describe free fields propagating in the time-dependent background. The  $A$ ,  $B$ , and  $C$  are some time-dependent background fields, and they are all positive. The solutions to the equations of motion (42) in momentum space,  $u_a(\mathbf{k}, t)$ , are called the *mode functions*, where  $\mathbf{k}$  denotes the comoving momentum. They satisfy the Wronskian condition

$$Au_a(\mathbf{k}, t)\dot{u}_a^*(\mathbf{k}, t) - \text{c.c.} = i, \quad (\text{no sum over } a). \quad (60)$$

Note that we have specified the time-independent constant on the right-hand side of (60) to be  $i$  for the same reason that we see in Section 2. Namely, we decompose  $\delta\phi_a^I$  as

$$\delta\phi_a^I(\mathbf{k}, t) = u_a(\mathbf{k}, t)a_a(\mathbf{k}) + u_a^*(-\mathbf{k}, t)a_a^\dagger(-\mathbf{k}), \quad (61)$$

where the annihilation and creation operators satisfy the following relations,

$$\begin{aligned} [a_a(\mathbf{k}), a_b^\dagger(-\mathbf{p})] &= (2\pi)^3 \delta_{ab} \delta^3(\mathbf{k} + \mathbf{p}), \\ [a_a(\mathbf{k}), a_b(-\mathbf{p})] &= 0, \quad [a_a^\dagger(\mathbf{k}), a_b^\dagger(-\mathbf{p})] = 0. \end{aligned} \quad (62)$$

These commutation relations are equivalent to (35) because of (60), but the constant needs to be  $i$ . This gives the normalization condition for the mode functions.

Being the solutions of the second-order differential equation, generally the mode function is a linear superposition of two independent solutions. So we need to specify the initial condition. For inflation models, as long as the field theory applies, one can always find an early time at which the physical momentum of the mode is much larger than the Hubble parameter and study a time interval much less than a Hubble time. Under these conditions, the equations of motion approach to those in the Minkowski limit, in which the mode function is a linear superposition of two independent plane waves, one with positive frequency and another negative. The ground state in the Minkowski spacetime is the positive one. The mode function which approaches this positive frequency state in the Minkowski limit is called the Bunch-Davies state. In physical coordinates, this limit is proportional to  $e^{-ik_{\text{ph}}t}$ , (for  $k_{\text{ph}} \gg m$ ), where  $k_{\text{ph}}$  is the physical momentum. In terms of the conformal time  $\tau \equiv \int dt/a(t)$  and the comoving momentum coordinate  $k \equiv k_{\text{ph}}/a(t)$  which we often use, this limit is proportional to  $e^{-ik\tau}$ . We have seen an example in Section 2 and will see more similar examples later with different  $A$ ,  $B$  and  $C$ . The corresponding vacuum  $|0\rangle$  is the Bunch-Davies vacuum and annihilated by  $a_a(\mathbf{k})$  defined in (61),  $a_a(\mathbf{k})|0\rangle = 0$ .

We also would like to write the vacuum of the interacting theory (51) in terms of the vacuum of the free theory  $|0\rangle$  defined above. Unlike the scattering theory where the vacuum of the free theory is generally different from the vacuum of the interaction theory, the process that we are studying here do not generate any nontrivial vacuum fluctuations through interactions. This is a direct consequence of the identity

$$F^{-1}F = 1. \quad (63)$$

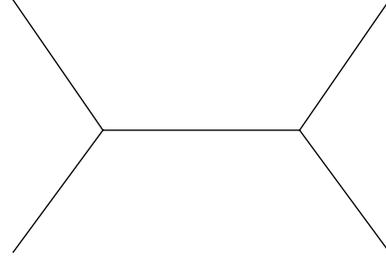


FIGURE 1: An example of Feynman diagram.

So we can replace  $|\Omega\rangle$  in (51) with the Bunch-Davies vacuum  $|0\rangle$  that we have specified above.

The integrand  $H_I(t)$  in (50) is highly oscillatory in the infinite past due to the behavior of the mode function  $\propto e^{-ik\tau}$ . Their contribution to the integral is averaged out. For the Bunch-Davies vacuum, this regulation can be achieved by introducing a small tilt to the integration contour  $\tau_0 \rightarrow -\infty(1 + i\epsilon)$  or performing a Wick rotation  $\tau \rightarrow i\tau$ . The imaginary component turns the oscillatory behavior into exponential decay, making the integral well defined.

**3.3. Contractions.** When evaluating (51), one encounters (anti)time-ordered integrals, of which the integrands are products of fields, such as  $\delta\phi_a^I$  and  $\delta\pi_a^I$ , or  $\delta\phi_a^I$  and  $\delta\phi_a^I$ , sandwiched between the vacua. In contrast to the Minkowski space, in the inflationary background, we do not have a simple analogous Feynman propagator which takes care of the time ordering. Therefore, we will just evaluate the integrands, but leave the complication of the time ordering to the final integration.

To evaluate the integrand, one can shift around the orders of fields in that product, following the rules of the commutation relations. A *contraction* is defined to be a nonzero commutator between the following components of two fields,  $[\delta\phi_a^+, \delta\phi_b^-]$ , where  $\delta\phi_a^+$  and  $\delta\phi_b^-$  denote the first and second term on the right-hand side of (61), respectively. After normal ordering, namely moving annihilation operators to the right-most and creation operators to the left-most so that they give zeros hitting the vacuum, it is not difficult to convince oneself that the only terms left are those with all fields contracted. Feynman diagrams can be used to keep track of what kind of contractions are necessary.

In the following, we demonstrate this using a simple example. We consider a field  $\delta\phi^I$  and quantize it as usual,

$$\begin{aligned} \delta\phi^I(\mathbf{k}, t) &\equiv \delta\phi^+ + \delta\phi^- \\ &= u(\mathbf{k}, t)a_{\mathbf{k}} + u^*(-\mathbf{k}, t)a_{-\mathbf{k}}^\dagger. \end{aligned} \quad (64)$$

So a contraction between the two terms,  $\delta\phi(\mathbf{k}, t')$  on the left and  $\delta\phi(\mathbf{p}, t'')$  on the right, is defined to be

$$[\delta\phi^+(\mathbf{k}, t'), \delta\phi^-(\mathbf{p}, t'')] = u(\mathbf{k}, t')u^*(-\mathbf{p}, t'')(2\pi)^3\delta^3(\mathbf{k} + \mathbf{p}). \quad (65)$$

For example, we want to compute a contribution to the four-point function  $\langle\delta\phi^4\rangle$  from a tree-diagram containing two three-point interactions of the following form:

$$H^I \propto \int \prod_{i=1}^3 d\mathbf{p}_i \delta\phi^I(\mathbf{p}_1, t) \delta\phi^I(\mathbf{p}_2, t) \delta\phi^I(\mathbf{p}_3, t). \quad (66)$$

$$\begin{aligned} & \delta\phi^I(\mathbf{p}_1, t') \delta\phi^I(\mathbf{p}_2, t') \delta\phi^I(\mathbf{p}_3, t') \delta\phi^I(\mathbf{k}_1, t) \delta\phi^I(\mathbf{k}_2, t) \delta\phi^I(\mathbf{k}_3, t) \delta\phi^I(\mathbf{k}_4, t) \delta\phi^I(\mathbf{q}_1, t'') \delta\phi^I(\mathbf{q}_2, t'') \delta\phi^I(\mathbf{q}_3, t'') \\ &= [\delta\phi^+(\mathbf{p}_1, t'), \delta\phi^-(\mathbf{k}_1, t)] [\delta\phi^+(\mathbf{p}_2, t'), \delta\phi^-(\mathbf{k}_2, t)] [\delta\phi^+(\mathbf{k}_3, t), \delta\phi^-(\mathbf{q}_1, t'')] [\delta\phi^+(\mathbf{k}_4, t), \delta\phi^-(\mathbf{q}_2, t'')] \\ & \quad \times [\delta\phi^+(\mathbf{p}_3, t'), \delta\phi^-(\mathbf{q}_3, t'')]. \end{aligned} \quad (67)$$

Note that all terms are contracted. The result can be further evaluated using (65). After integration over momenta indicated in (66), the final momentum conservation will always manifest itself as  $(2\pi)^3 \sum_i(\mathbf{k}_i)$ . There are other sets of contractions represented by the same diagram for the same term. Namely, there are three ways of picking two of the three  $\mathbf{p}_i$ s ( $\mathbf{q}_i$ s), so we have a symmetry factor 9; also, there are 24 permutations of the four  $\mathbf{k}_i$ s. We need to sum over all these possibilities. We also need to sum over all possible terms containing two  $H^I$ s in the perturbative series, which are not listed here, with their corresponding time ordered integral structure.

**3.4. Three Forms.** Now we deal with the time ordered integrals in the series expansion. There are two ways to expand (51).

In the first form, we simply expand the exponential in (50). For example, for an even  $n$ , the  $n$ th order term is

$$\begin{aligned} & i^n (-1)^{n/2} \int_{t_0}^t d\bar{t}_1 \int_{t_0}^{\bar{t}_1} d\bar{t}_2 \cdots \int_{t_0}^{\bar{t}_{n/2-1}} d\bar{t}_{n/2} \\ & \times \int_{t_0}^t dt_1 \int_{t_0}^{t_1} dt_2 \cdots \int_{t_0}^{t_{n/2-1}} dt_{n/2} \\ & \times \langle H_I(\bar{t}_{n/2}) \cdots H_I(\bar{t}_1) Q_I(t) H_I(t_1) \cdots H_I(t_{n/2}) \rangle \\ & + 2 \operatorname{Re} \sum_{m=1}^{n/2} i^n (-1)^{m+n/2} \int_{t_0}^t d\bar{t}_1 \int_{t_0}^{\bar{t}_1} d\bar{t}_2 \cdots \int_{t_0}^{\bar{t}_{n/2-1-m}} d\bar{t}_{n/2-m} \\ & \times \int_{t_0}^t dt_1 \int_{t_0}^{t_1} dt_2 \cdots \int_{t_0}^{t_{n/2-1+m}} dt_{n/2+m} \\ & \times \langle H_I(\bar{t}_{n/2-m}) \cdots H_I(\bar{t}_1) Q_I(t) H_I(t_1) \cdots H_I(t_{n/2+m}) \rangle. \end{aligned} \quad (68)$$

Each term in the above summation contains two factors of multiple integrals, one from  $F^{-1}$  and another from  $F$ . Each multiple integral is time ordered or antitime-ordered, but there is no time ordering between the two. We call this representation the *factorized form*.

These two  $H^I$ s come from expanding  $F^{-1}$  or  $F$  in (51). The corresponding Feynman diagram is Figure 1.

In Figure 1, the two cubic vertices each represent the three-point interaction (66). Each line represents a contraction. The four outgoing legs connect to the four  $\delta\phi(\mathbf{p}_i, t)$  ( $i = 1, 2, 3, 4$ ) in  $\langle\delta\phi^4\rangle$ . The following is a term from the perturbative series expansion of (51). We demonstrate in the following one set of contractions represented by the diagram in Figure 1,

In the second form, we rearrange the factorized form so that all the time variables are time-ordered, and all the integrands are under a unique integral. The  $n$ th order term in this form is [45]

$$\begin{aligned} & i^n \int_{t_0}^t dt_1 \int_{t_0}^{t_1} dt_2 \cdots \int_{t_0}^{t_{n-1}} dt_n \\ & \times \langle [H_I(t_n), [H_I(t_{n-1}), \dots, [H_I(t_1), Q_I(t)] \cdots]] \rangle. \end{aligned} \quad (69)$$

We call this representation the *commutator form*.

Each representation has its computational advantages and disadvantages.

The factorized form is most convenient to achieve the UV ( $t_i \rightarrow t_0$ ) convergence. As mentioned, after we tilt or rotate the integration contour into the positive imaginary plane for the left integral, and negative imaginary plane for the right integral, all the oscillatory behavior in the UV becomes well-behaved exponential decay. However, this form is not always convenient to deal with the IR ( $t_i \rightarrow t$ ) behavior. For cases where the correlation functions have some nontrivial evolution after modes exit the horizon, as typically happens for inflation models with multiple fields, the convergence in the IR is slow. Cancellation of spurious leading contributions from different terms in the sum (68) can be very implicit in this representation, and could easily lead to wrong leading order results in analytical estimation or numerical evaluation.

The commutator form is most convenient to get the correct leading order behavior in the IR. The mutual cancellation between different terms are made explicit in terms of the nested commutators, before the multiple integral is performed. However, such a regrouping of integrands makes the UV convergence very implicit. Recall that the contour deformation is made to damp the oscillatory behavior in the infinite past. In the commutator form, for any individual term in the integrand, we can still generically choose a unique convergence direction in terms of contour deformation. Although the directions are different for different terms, they can be separately chosen for each of them. But now the problem is, if these different terms have to be grouped as in the

nested commutator so that the IR cancellation is explicit, the two directions get mixed. Hence, the explicit IR cancellation is incompatible with the explicit UV convergence in this case.

To take advantage of both forms, we introduce a cutoff  $t_c$  and write the IR part of the in-in formalism in terms of the commutator form and the UV part in terms of the factorized form [46],

$$\begin{aligned} & \sum_{i=1}^n \int_{t_c}^t dt_1 \cdots \int_{t_c}^{t_{i-1}} dt_i \{ \text{commutator form} \} \\ & \times \int_{-\infty}^{t_c} dt_{i+1} \cdots \int_{-\infty}^{t_{n-1}} dt_n \{ \text{factorized form} \}. \end{aligned} \quad (70)$$

This representation is called the *mixed form*. This form is particularly efficient in numerical computations when combined with the Wick-rotations in the UV.

We will not always encounter all these subtleties in every model, but there does exist such interesting examples, as we will see in Section 7.1.

**3.5. Summary.** To end this section, we summarize the procedure that we need to go through to calculate the correlation functions in the in-in formalism.

Starting with the Lagrangian  $L[\phi(t), \dot{\phi}(t)]$ , we perturb it around the homogenous solutions  $\bar{\phi}_a$  and  $\dot{\bar{\phi}}_a$  of the classical equations of motion,

$$\begin{aligned} \phi_a(\mathbf{x}, t) &= \bar{\phi}_a(t) + \delta\phi_a(\mathbf{x}, t), \\ \dot{\phi}_a(\mathbf{x}, t) &= \dot{\bar{\phi}}_a(t) + \delta\dot{\phi}_a(\mathbf{x}, t). \end{aligned} \quad (71)$$

Keep the part of the Lagrangian that is quadratic and higher in perturbations and denote it as  $\tilde{L}$ . Define the conjugate momentum densities as  $\delta\pi_a = \partial\tilde{L}/\partial(\delta\dot{\phi}_a)$ . We can also equivalently expand the Hamiltonian  $H[\phi(t), \pi(t)]$  by perturbing  $\phi_a(\mathbf{x}, t)$  and  $\pi_a(\mathbf{x}, t)$ .

Work out the Hamiltonian in terms of  $\delta\phi_a$  and  $\delta\pi_a$ , and separate them into the quadratic kinematic part  $H_0$ , which describes the free fields in the time-dependent background, and the interaction part  $H_I$ . Relabel  $\delta\phi_a$ s and  $\delta\pi_a$ s in the Hamiltonian density as the interaction picture fields,  $\delta\phi_a^I$ s and  $\delta\pi_a^I$ s. These latter variables satisfy the equations of motion followed from the  $H_0$ . We quantize  $\delta\phi_a^I$  and  $\delta\pi_a^I$  in terms of the annihilation and creation operators as in (61) and (62). The mode functions  $u_a(\mathbf{k}, t)$  are solutions of the equations of motion from  $H_0$ , normalized according to the Wronskian conditions (60) and specified by an initial condition such as the Bunch-Davies vacuum. The correlation function for  $Q(t)$  is given by

$$\begin{aligned} \langle Q(t) \rangle &\equiv \langle 0 | \left[ \bar{T} \exp \left( i \int_{t_0}^t H_I(t) dt \right) \right] Q^I(t) \\ &\times \left[ T \exp \left( -i \int_{t_0}^t H_I(t) dt \right) \right] | 0 \rangle, \end{aligned} \quad (72)$$

where  $Q(t)$  is a product in terms of  $\delta\phi_a^I(\mathbf{x}, t)$  and  $\delta\pi_a^I(\mathbf{x}, t)$ . If we want to work with  $\delta\phi_a$  and  $\delta\dot{\phi}_a$  instead of  $\delta\phi_a^I$  and  $\delta\pi_a^I$ , we replace  $\delta\pi_a^I$  with  $\delta\dot{\phi}_a^I$  using the relation  $\delta\dot{\phi}_a^I = \partial H_0 / \partial(\delta\pi_a^I)$ .

Choose appropriate forms in Section 3.4 and series-expand the integrand in powers of  $H_I$  to the desired orders. Perform contractions defined in Section 3.3 for each term in this expansion. Each term gives a nonzero contribution only when all fields are contracted. Draw Feynman diagrams that represent the correlation functions, and use them as a guidance to do contractions. Finally sum over all possible contractions and perform the time-ordered integrations.

## 4. A No-Go Theorem

Simplest inflation models generate negligible amount of non-Gaussianities that are well below our current experimental abilities [47, 48]. By simplest, we mean

- (i) single scalar field inflation
- (ii) with canonical kinetic term
- (iii) always slow-rolls
- (iv) in Bunch-Davies vacuum
- (v) in Einstein gravity.

This list is extracted based on Maldacena's computation of three-point functions in an explicit slow-roll model [47]. We now review this proof. The notations here follow [49, 50] and will be consistently used later in this paper.

The Lagrangian for the *single scalar field inflation* with *canonical kinetic term* is the following:

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_{\text{P}}}{2} R + X - V(\phi) \right], \quad (73)$$

where  $\phi$  is the inflaton field,  $X = -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$  is the canonical kinetic term and  $V$  is the slow-roll potential. The first term is the *Einstein gravity* and  $M_{\text{P}} = (8\pi G)^{-1/2}$  is the reduced Planck mass. For convenience we will set the reduced Planck mass  $M_{\text{P}} = 1$ . The signature of the metric is  $(-1, 1, 1, 1)$ .

The inflaton starts near the top of the potential and slowly rolls down. As we have reviewed in Section 2, to ensure that the inflation lasts for at least  $\mathcal{O}(60)$  e-folds, the potential is required to be flat so that the slow-roll parameters (11) are both much less than one most of the time. The energy of the universe is dominated by the potential energy, and the inflaton follows the slow-roll attractor solution (10). Also as discussed in Section 2, we will use the following more general slow-roll parameters:

$$\epsilon = -\frac{\dot{H}}{H^2}, \quad \eta = \frac{\dot{\epsilon}}{\epsilon H}. \quad (74)$$

To study the perturbation theory, it is convenient to use the ADM formalism, in which the metric takes the form

$$ds^2 = -N^2 dt^2 + h_{ij} (dx^i + N^i dt) (dx^j + N^j dt). \quad (75)$$

The action becomes

$$\begin{aligned} S &= \frac{1}{2} \int dt dx^3 \sqrt{h} N (R^{(3)} + 2X - 2V) \\ &+ \frac{1}{2} \int dt dx^3 \sqrt{h} N^{-1} (E_{ij} E^{ij} - E^2), \end{aligned} \quad (76)$$

where the index of  $N^i$  can be lowered by the 3D metric  $h_{ij}$  and  $R^{(3)}$  is the 3D Ricci scalar constructed from  $h_{ij}$ . The definitions of  $E_{ij}$  and  $E$  are

$$\begin{aligned} E_{ij} &= \frac{1}{2} (\dot{h}_{ij} - \nabla_i N_j - \nabla_j N_i), \\ E &= E_{ij} h^{ij}. \end{aligned} \quad (77)$$

In the ADM formalism, the variables  $N$  and  $N^i$  are Lagrangian multipliers whose equations of motion are easy to solve. In single field inflation, we have only one physical scalar perturbation [51]. We choose the uniform inflaton gauge (also called the comoving gauge) in which the scalar perturbation  $\zeta$  appears in the three dimensional metric  $h_{ij}$  in the following form:

$$h_{ij} = a^2 e^{2\zeta} \delta_{ij}, \quad (78)$$

and the inflaton fluctuation  $\delta\phi$  vanishes. The  $a(t)$  is the homogeneous scale factor of the universe, so  $\zeta$  is a space-dependent rescaling factor. In this paper we do not consider the tensor perturbations.

We plug (75) and (78) into the action (76) and solve the constraint equations for the Lagrangian multipliers  $N$  and  $N^i$ . We then plug them back to the action and expand up to the cubic order in  $\zeta$  in order to calculate the three-point functions. To do this, in the ADM formalism, it is enough to solve  $N$  and  $N^i$  to the first-order in  $\zeta$ . This is because their third-order perturbations will multiply the zeroth order constraint equation which vanishes, and their second-order perturbations will multiply the first-order constraint equation which again vanishes. After some lengthy algebra, we obtain the following expansions:

$$S_2 = \int dt d^3x [a^3 \epsilon \dot{\zeta}^2 - a\epsilon (\partial\zeta)^2], \quad (79)$$

$$\begin{aligned} S_3 = \int dt d^3x & \left[ a^3 \epsilon^2 \zeta \dot{\zeta}^2 + a\epsilon^2 \zeta (\partial\zeta)^2 - 2a\epsilon \dot{\zeta} (\partial\zeta) (\partial\chi) \right. \\ & + \frac{a^3 \epsilon}{2} \dot{\eta} \zeta^2 \dot{\zeta} + \frac{\epsilon}{2a} (\partial\zeta) (\partial\chi) \partial^2 \chi \\ & \left. + \frac{\epsilon}{4a} (\partial^2 \zeta) (\partial\chi)^2 + f(\zeta) \frac{\delta L}{\delta \zeta} \Big|_1 \right], \end{aligned} \quad (80)$$

where

$$\chi = a^2 \epsilon \partial^{-2} \dot{\zeta}, \quad (81)$$

$$\frac{\delta L}{\delta \zeta} \Big|_1 = 2a \left( \frac{d\partial^2 \chi}{dt} + H \partial^2 \chi - \epsilon \partial^2 \dot{\zeta} \right), \quad (82)$$

$$f(\zeta) = \frac{\eta}{4} \zeta^2 + \text{terms with derivatives on } \zeta. \quad (83)$$

Here  $\partial^{-2}$  is the inverse Laplacian and  $\delta L/\delta \zeta|_1$  is the variation of the quadratic action with respect to the perturbation  $\zeta$ . We now can follow Section 3 and proceed to calculate the correlation functions. For simplicity, we will always neglect the superscript “1” on various interaction picture fields.

We restrict to the case where the slow-roll parameters are *always small and featureless*. We first look at the quadratic action. In this case, we can analytically solve the equation of motion followed from (79) in terms of the Fourier mode of  $\zeta$ ,

$$u_k = \int d^3x \zeta(t, \mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}}, \quad (84)$$

and get the mode function

$$u_k = u(\mathbf{k}, \tau) = \frac{iH}{\sqrt{4\epsilon k^3}} (1 + ik\tau) e^{-ik\tau}, \quad (85)$$

where  $\tau \equiv \int dt/a \approx -(aH)^{-1}$  is the conformal time. The normalization is determined by the Wronskian condition (60). We have chosen the *Bunch-Davies vacuum* by imposing the condition that the mode function approaches the vacuum state of the Minkowski spacetime in the short wavelength limit  $k/a \gg 1/H$ ,

$$u_k \longrightarrow -\frac{H\tau}{\sqrt{4\epsilon k}} e^{-ik\tau}. \quad (86)$$

The dynamical behavior of  $\zeta$  that has been explained around (24) and (25) is made precise here. In particular,  $\zeta$  is exactly massless without dropping any  $\mathcal{O}(\epsilon)$  suppressed terms. In addition, from (78), we can see that, for superhorizon modes, the only effect of  $\zeta$  is to provide a homogeneous spatial rescaling. And  $\zeta$  is the only scalar perturbation. So the fact that  $\zeta$  is frozen after horizon exit will not be changed by higher-order terms.

If we choose the spatially flat gauge, we make  $\zeta$  disappear and the scalar in this perturbation theory becomes the perturbation of  $\phi$ . The relation between  $\zeta$  and  $\delta\phi$  in (25) (with  $\mathcal{O}(\epsilon)$  corrections) is thus a gauge transformation through a space-dependent time shift.

We quantize the field as

$$\zeta(\mathbf{k}, \tau) = u_k a_{\mathbf{k}} + u_k^* a_{-\mathbf{k}}^\dagger, \quad (87)$$

with the canonical commutation relation  $[a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = (2\pi)^3 \delta^3(\mathbf{k} - \mathbf{k}')$ . We can easily compute the two-point function at the tree level,

$$\langle \zeta(\mathbf{k}_1) \zeta(\mathbf{k}_2) \rangle = \frac{P_\zeta}{2k_1^3} (2\pi)^5 \delta^3(\mathbf{k}_1 + \mathbf{k}_2), \quad (88)$$

where

$$P_\zeta = \frac{H^2}{8\pi^2 \epsilon}. \quad (89)$$

Since  $\zeta$  remains constant after it exits the horizon, the  $H$  and  $\epsilon$  are both evaluated near the horizon exit.

We next look at the cubic action. For single field models,  $H_{I,3} = -L_3$ . Keeping in mind that  $\chi$  is proportional to  $\epsilon$ , one can see that the first line of (80) is proportional to  $\epsilon^2$ . For the featureless potential,  $\dot{\eta} = \mathcal{O}(\epsilon^2)$ , where  $\epsilon$  collectively denotes either  $\epsilon$  or  $\eta$ . So the second line of (80) is proportional to  $\epsilon^3$ , and negligible. The third line can be absorbed by a field redefinition  $\zeta \rightarrow \zeta_n + f(\zeta_n)$ . The only term in  $f(\zeta_n)$

that will contribute to the correlation function is written out explicitly in (83). All the others involve derivatives of  $\zeta$  so vanish outside the horizon. Thus this redefinition eventually introduces an extra term,

$$\begin{aligned} & \langle \zeta(\mathbf{k}_1)\zeta(\mathbf{k}_2)\zeta(\mathbf{k}_3) \rangle \\ &= \langle \zeta_n(\mathbf{k}_1)\zeta_n(\mathbf{k}_2)\zeta_n(\mathbf{k}_3) \rangle \\ &+ \frac{\eta}{4} (\langle \zeta_n^2(\mathbf{k}_1)\zeta_n(\mathbf{k}_2)\zeta_n(\mathbf{k}_3) \rangle + 2 \text{ perm.}) + \mathcal{O}\left(\eta^2 (P_k^\zeta)^3\right). \end{aligned} \quad (90)$$

According to (72), we expand the exponential to the first-order in  $H_{I,3}$  to get the leading result

$$\langle \zeta_n^3 \rangle = -i \langle 0 | \int_{t_0}^t dt [\zeta_n(\mathbf{k}_1)\zeta_n(\mathbf{k}_2)\zeta_n(\mathbf{k}_3), H_{I,3}] | 0 \rangle. \quad (91)$$

To estimate the order of magnitude of the bispectrum, we only need to keep track of the factors of  $H$  and  $\epsilon$ . For example, from the first term in (80), we have  $\int dt H_3(t) \supset -\int dx^3 d\tau a^2 \epsilon^2 \zeta' \zeta'^2$ , where we used the conformal time  $\tau$  and the prime denotes the derivative to  $\tau$ . Using  $a \propto H^{-1}$ ,  $\zeta \propto \zeta' \propto H/\sqrt{\epsilon}$ , we see that this three-point vertex contributes  $\propto H\sqrt{\epsilon}$ . Together with the three external legs  $\zeta^3$  and the definition  $P_\zeta \propto H^2/\epsilon$ , we get

$$\langle \zeta^3 \rangle = -i \int dt \langle [\zeta^3, H_{I,3}(t)] \rangle \propto \frac{H^4}{\epsilon} \propto \mathcal{O}(\epsilon) P_\zeta^2. \quad (92)$$

Similar results can be obtained for the other two terms in the first line of (80). As we will define more carefully later, the size of the three-point function is conventionally characterized by the number  $f_{\text{NL}}$ , which is defined as  $\langle \zeta^3 \rangle \sim f_{\text{NL}} P_\zeta^2$ . So the contribution from the first line of (80) is  $f_{\text{NL}} = \mathcal{O}(\epsilon)$ . The extra term due to the redefinition (90) contributes  $f_{\text{NL}} = \mathcal{O}(\eta)$ . This completes the order-of-magnitude estimate. To get the full non-Gaussian profile, we need to compute the integrals explicitly and get

$$\langle \zeta(\mathbf{k}_1)\zeta(\mathbf{k}_2)\zeta(\mathbf{k}_3) \rangle = (2\pi)^7 \delta^3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) P_\zeta^2 \frac{1}{\prod_i k_i^2} S, \quad (93)$$

where

$$\begin{aligned} S &= \frac{\epsilon}{8} \left[ -\left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right) + \left( \frac{k_1}{k_2} + 5 \text{ perm.} \right) \right. \\ &\quad \left. + \frac{8}{K} \left( \frac{k_1 k_2}{k_3} + 2 \text{ perm.} \right) \right] + \frac{\eta}{8} \left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right), \end{aligned} \quad (94)$$

where  $K = k_1 + k_2 + k_3$  and the permutations stand for those among  $k_1, k_2$ , and  $k_3$ .

The slow-roll parameters are of order  $\mathcal{O}(0.01)$ , so  $f_{\text{NL}} \sim \mathcal{O}(0.01)$  for these models. Even if we start with Gaussian primordial perturbations, nonlinear effects in CMB evolution will generate  $f_{\text{NL}} \sim \mathcal{O}(1)$  [37], and a similar number for large scale structures due to the nonlinear gravitational evolution or the galaxy bias [35]. It seems unlikely that we can disentangle all these contaminations and detect such small primordial non-Gaussianities in the near future.

## 5. Beyond the No-Go

*5.1. Inflation Model Building.* The following are two examples of slow-roll potentials in the simplest inflation models that we studied in Section 4:

$$V_{\text{small}} = V_0 - \frac{1}{2} m^2 \phi^2, \quad (95)$$

$$V_{\text{large}} = \frac{1}{2} m^2 \phi^2. \quad (96)$$

The first type (95) belongs to the small field inflation models. The slow-roll conditions (11) require the potential to be flat enough relative to its height, that is, the mass of the inflaton should satisfy  $m \ll H$ . The second type (96) belongs to the large field inflation models. The potential also needs to be flat relative to its height, but here one achieves this by making the field range  $\phi$  very large, typically  $\phi \gg M_{\text{p}}$ . The other conditions that we listed in Section 4 should also be satisfied by these models. These are the classic examples, which exhibit algebraic simplicities and illustrate many essential features of inflation.

However, when it comes to the more realistic model building in a UV complete setup, such as in supergravity and string theory, situations get much more complicated. For example, it is natural that we encounter multiple light and heavy fields, and the potentials for them form a complex landscape. These multiple fields live in an internal space, whose structure can be very sophisticated. In string theory, some of them manifest themselves as extra dimensions and can have intricate geometry and warping. All these elements have to coexist with the inflationary background that introduces profound back-reactions.

Even with varieties of model building ingredients, it has been proven to be very subtle to construct an explicit and self-consistent inflation model. Indeed various problems have been noticed over the years in the course of the inflation model building. For example, consider the following problems.

(i) *The  $\eta$ -Problem for Slow-Roll Inflation* [52]. As we have seen, in order to have slow-roll inflation [8, 9], the mass of the inflaton field has to be light enough,  $m \ll H$ , to maintain a flat potential. However, in the inflationary background, the natural mass of a light particle is of order  $H$ . This can be seen in many ways, and in some ideal situations they are equivalent to each other. For example, one way to see this is to consider the coupling between the Ricci scalar and the inflaton,  $\sim R\phi^2$ . In the inflationary background  $R \sim H^2$ . Unless we have good reasons to set the coefficient of this term to be much less than one, it will give inflaton a mass of order  $H$ , spoiling the inflation. Another way to see this is to note that the effective potential in supergravity takes the form  $V = V_0 \exp(K/M_{\text{p}}^2) \times \text{other terms}$ . Here schematically  $K \sim \phi^2 + \dots$  is the Kahler potential and its dependence on  $\phi$  is normalized as such to give the canonical kinetic term for  $\phi$ . So the first term in the expansion of  $V$  is of order  $V_0 \phi^2/M_{\text{p}}^2 \sim H^2 \phi^2$  and model independent. Therefore, either symmetry needs to be imposed or other tuning contributions introduced to solve this  $\eta$ -problem.

(ii) *The h-Problem for DBI Inflation* [53]. DBI inflation [54] is invented to generate inflation by a different mechanism. It makes use of the warped space in the internal field space [55, 56]. These warped space impose speed limits for the scalar field, so even if the potential is steep, the inflaton is not allowed to roll down the potential very quickly. A canonical example of warped space is

$$ds^2 = h(r)^2(-dt^2 + a(t)^2 dx^2) + h(r)^{-2} dr^2, \quad (97)$$

where  $r$  is the extra dimension (or internal space),  $h(r) = r/R$  is the warp factor, and  $R$  is the length scale of the warped space. The position of a 3 + 1 dimensional brane in the  $r$ -coordinate is the inflaton. So the inflaton velocity is limited by the speed limit in the  $r$ -direction,  $h^2$ . In order to provide a speed limit that is small enough for inflation, the warp factor has to be small enough,  $h \ll HR$ . However one of the Einstein equations with the metric (97) takes the following form

$$\left(\frac{dh}{dr}\right)^2 - H^2 h^{-2} = \frac{1}{R^2} + \text{other source terms}, \quad (98)$$

where the second term on the left-hand side is due to the back-reaction of the inflationary spacetime. It is easy to see that the naive  $h = r/R$  should be modified for  $h < HR$ , precisely where the inflation is supposed to happen. Without contributions from other source terms, such a deformed geometry closes up too quickly and leads to an unacceptable probe-brane back-reaction if we demand the inflaton still follow the speed limit. Therefore, either symmetry, or tuning using other source terms from the right-hand side of (98), is necessary to solve this  $h$ -problem. The  $\eta$ -problem and  $h$ -problem are closely related in an AdS/CFT setup.

(iii) *The Field Range Bound* [57, 58]. Large field inflation models require the field range to be much larger than  $M_{\text{P}}$ . In supergravity and string theory, starting from a ten-dimensional theory with 10-dim Planck mass  $M_{10}$ , the 4-dim Planck mass  $M_{\text{P}}$  is obtained by integrating out the six extra dimensions,

$$\begin{aligned} M_{(10)}^8 \int d^6 y d^4 x \sqrt{-G_{(10)}} R_{(10)} &\supset M_{(10)}^8 V_{(6)} \\ &\times \int d^4 x \sqrt{-g_{(4)}} R_{(4)} \equiv M_{\text{P}}^2 \int d^4 x \sqrt{-g_{(4)}} R_{(4)}, \end{aligned} \quad (99)$$

where we use  $L$  and  $V_{(6)} \sim L^6$  to denote the size and volume of the extra dimensions, respectively. The field range  $\Delta\phi$  often appears as the distance in the extra dimensions,  $\Delta\phi \sim \Delta L \cdot M_{(10)}^2$ , with the factor  $M_{(10)}^2$  being the proportional coefficient. Clearly,  $\Delta L \lesssim L$ . If the field range manifests itself within a warped throat with a length scale  $R$ , we still require  $R < L$ , and so  $\Delta L \lesssim L$ . Together with  $M_{\text{P}} = M_{(10)}^4 L^3$ , we get

$$\Delta\phi \lesssim \frac{M_{\text{P}}}{(M_{(10)} L)^2}. \quad (100)$$

We further note that the microscopic length scale  $L$  has to be much larger than the 10-dim Planck length  $M_{(10)}^{-1}$  for the field

theory to make sense. So  $M_{(10)} L \gg 1$ , and the field range  $\Delta\phi$  in these models is generically sub-Planckian. For example, for a warped throat with charge  $N$ ,  $(M_{(10)} L)^2 \gtrsim (M_{(10)} R)^2 \sim N^{1/2}$ , we have

$$\Delta\phi \lesssim \frac{M_{\text{P}}}{\sqrt{N}}. \quad (101)$$

We have ignored a detailed numerical coefficient appearing on the right-hand side of (101), which is model dependent. For example, considering the volume  $V_{(6)}$  to be the sum of the throat and a generic bulk volume, it is  $\mathcal{O}(0.01)$  [57]; considering an extreme case where the throat does not attach to a bulk, it is  $\mathcal{O}(1)$  [58]. Notice that, due to the dependence of  $M_{\text{P}}$  on the volume  $V_{(6)}$ , increasing the volume only makes the bound tighter.

(iv) *The Variation of Potential* [59]. Even in cases where there is no fundamental restriction on the excursion of fields, one encounters problems constructing the large field inflationary potential. Large field potentials that arise from a fundamental theory take the following general form:

$$V(\phi) = \sum_{n=0}^{\infty} \lambda_n m_{\text{fund}}^{4-n} \phi^n, \quad (102)$$

where  $m_{\text{fund}}$  represents typical scales in the theory. For field theory descriptions to hold, such scales are much less than  $M_{\text{P}}$ . For example,  $m_{\text{fund}}$  can be the higher dimensional Planck mass, string mass, or their warped scales. The  $\lambda_n$ s are dimensionless couplings of order  $\mathcal{O}(1)$ . Unless some symmetries are present to forbid an infinite number of terms in (102), or a high degree of fine-tuning is assumed, the shape of potential (102) varies over a scale of order  $m_{\text{fund}} \ll M_{\text{P}}$ . This variation is too dramatic for the potential to be a successful large field slow-roll potential.

None of the arguments in the above list is meant to show that the specific type of inflation is impossible. In fact, these have been the driving forces for the ingenuity and creativity in the field of inflation model building. This list is used to demonstrate some typical examples of complexities in reality. Often times, solving one problem will be accompanied by other structures that make the model step beyond the simplest one. So we may want to keep an open mind that the algebraic simplicity may not mean the simplicity in Nature.

Following is a partial list of possibilities that allow us to go beyond the no-go theorem in Section 4.

- (i) Instead of single field inflation, we can consider quasi-single field or multifield inflation models (Sections 7 and 8).
- (ii) Instead of canonical kinetic terms, there are models where the higher derivative kinetic terms dominate the dynamics (Section 6.1).
- (iii) Instead of following the attractor solution such as the slow-roll precisely, features can be present in the potentials or internal space, that temporarily break the attractor solution, or cause small but persistent perturbations on the background evolution (Sections 6.2 and 6.3).

- (iv) Instead of staying in the Bunch-Davies vacuum, other excitations can exist due to, for example, boundary conditions or low scales of new physics (Section 6.4).
- (v) Although strong constraints, from experimental results and theoretical consistencies, exist on non-Einstein gravities, early universe may provide an opportunity for their appearance. We use this category to include a variety of possibilities, such as modified gravities, noncommutativity, nonlocality and models beyond field theories.

There are also strong motivations from data analyses for us to search and study different large non-Gaussianities. The signal-to-noise ratio in the CMB data is not large enough for us to detect primordial non-Gaussianities model-independently. A well-established method is to start with a theoretical non-Gaussian ansatz, and construct optimal estimators that compare theory and data by taking into accounts all momenta configurations. This then gives constraints on the parameters characterizing the theoretical ansatz. Therefore, the following two important possibilities exist. First, the primordial non-Gaussianities exist in data could be missed if we did not start with a right theoretical ansatz. Second, even if a non-Gaussian signal was detected with one ansatz, it does not mean that we have found the right one. So different well-motivated non-Gaussian templates are needed for clues on how corresponding data analyses should be formed. From a different perspective, even if the primordial density perturbations were Gaussian, we would still do the similar amount of work and reach the conclusion after various well-motivated non-Gaussian forms are properly constrained.

**5.2. Shape and Running of Bispectra.** In this paper, we will be mainly interested in the three-point correlation functions of the scalar primordial perturbation  $\zeta$ . They are also called the *bispectra*. In this subsection, we introduce some simple terminologies that we often encounter in studies of bispectra.

The three-point function is a function of three momenta,  $\mathbf{k}_1$ ,  $\mathbf{k}_2$ , and  $\mathbf{k}_3$ , which form a triangle due to the translational invariance. Assuming also the rotational invariance, we are left with three variables, which are their amplitudes,  $k_1$ ,  $k_2$ , and  $k_3$ , satisfying the triangle inequalities. The information is encoded in a function  $S(k_1, k_2, k_3)$  that we define as

$$\langle \zeta^3 \rangle \equiv S(k_1, k_2, k_3) \frac{1}{(k_1 k_2 k_3)^2} \tilde{P}_\zeta^2 (2\pi)^7 \delta^3 \left( \sum_{i=1}^3 \mathbf{k}_i \right), \quad (103)$$

where  $\tilde{P}_\zeta$  is the fiducial power spectrum, and we fix it to be a constant  $\tilde{P}_\zeta \equiv P_\zeta(k_{\text{wmap}}) = 6.1 \times 10^{-9}$ , where  $k_{\text{wmap}} = 0.027 \text{ Mpc}^{-1}$ . We have chosen the above definition so that it can be uniformly applied to different types of bispectra that we will encounter in this paper. In the literature, different notations have been used. The differences are simple and harmless. For example, different functions such as  $\mathcal{A} = k_1 k_2 k_3 S$  or  $F = S/(k_1 k_2 k_3)^2$  are sometimes defined. We choose  $S$  since it is dimensionless and, for scale-invariant bispectra, it is invariant under a rescaling of all momenta.

This quantity is the combination that is used to plot the profiles of bispectra in the literature any way, despite of different conventions. Also, the precise power spectrum  $P_\zeta$  instead of  $\tilde{P}_\zeta$  is often used in the definition (103). Here, we absorb the momentum dependence of  $P_\zeta$  in  $S$ . This is because the three-point function is an independent statistic relative to the two-point. In cases where both the power spectrum and bispectrum have strong scale dependence, it is not convenient if they are defined in an entangled way.

Under different circumstances, different properties of  $S$  are emphasized. The conventions involved may not always be precisely consistent with each other, since they are chosen to best describe the case at hand. Following are some typical examples.

The dependence of  $S$  on  $k_1$ ,  $k_2$ , and  $k_3$  is usually split into two kinds.

One is called the *shape* of the bispectrum. This refers to the dependence of  $S$  on the momenta ratio  $k_2/k_1$  and  $k_3/k_1$ , while fixing the overall momentum scale  $K = k_1 + k_2 + k_3$ . Several special momentum configurations are shown in Figure 2.

Another is called the *running* of the bispectrum. This refers to the dependence of  $S$  on the overall momentum scale  $K = k_1 + k_2 + k_3$ , while fixing the ratio  $k_2/k_1$  and  $k_3/k_1$ .

For bispectra that are approximately scale invariant, the shape is a more important property [50, 60]. We will encounter such cases in Sections 6.1, 7.1, and 8.1. The amplitude, also called the *size*, of the bispectra is often denoted as  $f_{\text{NL}}$  by matching

$$S(k_1, k_2, k_3) \xrightarrow[\text{limit}]{k_1=k_2=k_3} \frac{9}{10} f_{\text{NL}}. \quad (104)$$

In this case,  $f_{\text{NL}}$  is approximately a constant but can also have a mild running, that is, a weak dependence on the overall momentum  $K$  [61, 62]. An index  $n_{\text{NG}} - 1 \equiv d \ln f_{\text{NL}} / d \ln k$  is introduced to describe this scale dependence. The power spectrum also has a mild running,  $P_\zeta = (k/k_0)^{n_s-1} \tilde{P}_\zeta$ . In this paper, when we give explicit forms of  $S$  in the approximately scale-invariant cases, for simplicity, we mostly ignore these mild scale dependence and concentrate on shapes. Shapes of bispectra have been given names according to the overall dependence of  $S$  on momenta. For example, for the *equilateral bispectrum*,  $S$  peaks at the equilateral triangle limit and vanishes as  $\sim k_3/k_1$  in the squeezed triangle limit ( $k_3 \ll k_1 = k_2$ ). The *local bispectrum* peaks at the squeezed triangle limit in the form  $\sim (k_3/k_1)^{-1}$ , such as the two shape components in (94). To visualize the shapes, we often draw 3D plots  $S(1, x_2, x_3)$ , where  $x_2$  and  $x_3$  vary from 0 to 1 and satisfy the triangle inequality  $x_2 + x_3 \geq 1$ .

There are also cases where the running becomes the most important property, while the shape is relatively less important [63, 64]. In such cases, the bispectra are mostly functions of  $K$ . So  $f_{\text{NL}}$  defined in (104) has strong scale dependence. Instead, one can choose a constant  $f_{\text{NL}}$  to describe the overall running amplitude. We will encounter such cases in Sections 6.2 and 6.3. In these cases, the shape plot  $S(1, x_2, x_3)$  may look nontrivial but this is because it does not fix  $K$ .

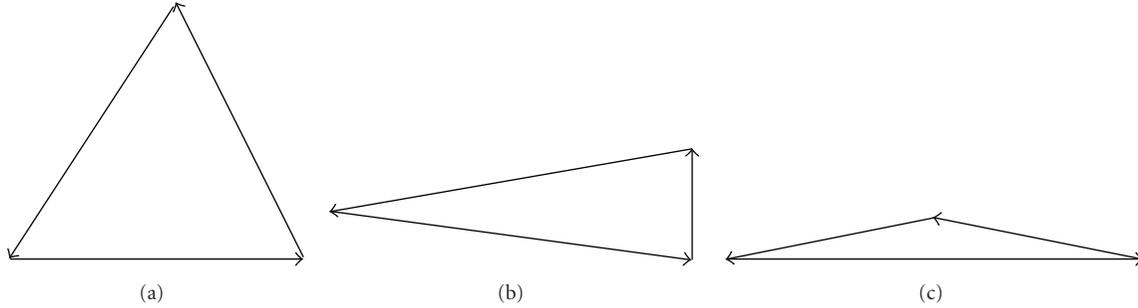


FIGURE 2: Momentum configurations: (a) equilateral, (b) squeezed, and (c) folded.

The above dissection will become less clean for cases where both properties become important.

One thing is clear. The  $f_{\text{NL}}$ , that is always used to quantify the level of non-Gaussianities, is only sensible with an extra label that specifies, at least qualitatively, the profile of the momentum dependence, such as shapes and runnings.

It is useful to quantify the correlations between different non-Gaussian profiles, because as we mentioned in data analyses an ansatz can pick up signals that are not completely orthogonal to it. In real data analyses this is performed in the CMB  $l$ -space. To have a simple but qualitative analogue in the  $k$ -space, we define the inner product of the two shapes as

$$S \cdot S' \equiv \int_{V_k} S(k_1, k_2, k_3) S'(k_1, k_2, k_3) w(k_1, k_2, k_3) dk_1 dk_2 dk_3, \quad (105)$$

and normalize it to get the shape correlator [60, 65]

$$C(S, S') \equiv \frac{S \cdot S'}{(S \cdot S)^{1/2} (S' \cdot S')^{1/2}}. \quad (106)$$

Following [65], we choose the weight function to be

$$w(k_1, k_2, k_3) = \frac{1}{k_1 + k_2 + k_3}, \quad (107)$$

so that the  $k$ -scaling is close to the  $l$ -scaling in the data analyses estimator. Later in this review, when we use this correlator to estimate the correlations between shapes, we take the ratio between the smallest and largest  $k$  to be  $2/800$ , close to that in WMAP. A more precise correlator should be computed in the  $l$ -space in the same way that the estimator is constructed. We refer to [35] for more details.

In typical data analyses [66–70], the estimator involves a triple integral of the bispectrum over the three momenta  $k_i$ . To have practical computational costs, it is necessary that this integral can be factorized into a multiplication of three integrals, each involves only an individual  $k_i$ . This requires the bispectrum to be of the form  $f_1(k_1) f_2(k_2) f_3(k_3)$ , or a sum of such forms. Such a form is called the *factorizable form* or *separable form*. The factor  $K^{-n}$  may be tolerable since it can be written as  $(1/\Gamma(n)) \int_0^\infty t^{n-1} e^{-Kt}$ . If the analytical result is too complicated, to make contact with experiments we will try to construct simple factorizable ansatz or template to capture the main features of the original one. New methods

that are applicable to nonfactorizable bispectrum forms and are more model-independent are under active development [71].

## 6. Single Field Inflation

In this section, we relax several restrictions of the no-go theorem on single field inflation models and study how large non-Gaussianities can arise. We present the formalisms and compute the three-point functions. We emphasize how different physical processes during inflation are imprinted as distinctive signatures in non-Gaussianities. Obviously, any mechanism that works for single field inflation can be generalized to multifield inflation models.

*6.1. Equilateral Shape: Higher Derivative Kinetic Terms.* In this subsection, we study large non-Gaussianities generated by noncanonical kinetic terms in general single field inflation models, following [50].

Consider the following action for the general single field inflation [72]:

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_{\text{P}}}{2} R + P(X, \phi) \right]. \quad (108)$$

Comparing to (73), we have replaced the canonical form  $X - V$  with an arbitrary function of  $X \equiv -(1/2)g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$  and  $\phi$ . This is the most general Lorentz-invariant Lagrangian as a function of  $\phi$  and its first derivative. It is useful to define several quantities that characterize the differential properties of  $P$  with respect to  $X$  [49, 72]:

$$\begin{aligned} c_s^2 &= \frac{P_{,X}}{P_{,X} + 2XP_{,XX}}, \\ \Sigma &= XP_{,X} + 2X^2P_{,XX} = \frac{H^2\epsilon}{c_s^2}, \\ \lambda &= X^2P_{,XX} + \frac{2}{3}X^3P_{,XXX}, \end{aligned} \quad (109)$$

where  $c_s$  is called the sound speed and the subscript “ $X$ ” denotes the derivative with respect to  $X$ . The third derivative is enough since we will only study the three-point function here.

It is a nontrivial question which forms of  $P$  will give rise to inflation. The model-independent approach we take here

is to list the conditions that an inflation model has to satisfy, no matter which mechanism is responsible for it. Namely, we generalize the slow-roll parameters in (74) to the following slow-variation parameters:

$$\epsilon = -\frac{\dot{H}}{H^2}, \quad \eta = \frac{\dot{\epsilon}}{\epsilon H}, \quad s = \frac{\dot{c}_s}{c_s H}, \quad (110)$$

and require them to be small most of the time during the inflation. The smallness of these parameters guarantees the Hubble constant  $H$ , the parameter  $\epsilon$ , and the sound speed  $c_s$  to vary slowly in terms of the Hubble time. Similar to the arguments given in the case of slow-roll inflation in Section 2, these are necessary to ensure a prolonged inflation as well as an approximately scale-invariant power spectrum that we observed in the CMB.

Following the same procedure that is outlined in Section 4, we get the quadratic and cubic action for the scalar perturbation  $\zeta$  [47, 49, 50]. The quadratic part is

$$S_2 = \int dt d^3x \left[ a^3 \frac{\epsilon}{c_s^2} \dot{\zeta}^2 - a\epsilon (\partial\zeta)^2 \right]. \quad (111)$$

If the slow-variation parameters are always small and featureless, we can analytically solve the equation of motion followed from (111) and get the following mode function:

$$u_k(\tau) = \frac{iH}{\sqrt{4\epsilon c_s k^3}} (1 + ikc_s \tau) e^{-ikc_s \tau}. \quad (112)$$

Notice the appearance of  $c_s$  comparing to (85). The two-point function is

$$\langle \zeta(\mathbf{k}_1) \zeta(\mathbf{k}_2) \rangle = \frac{P_\zeta}{2k_1^3} (2\pi)^5 \delta^3(\mathbf{k}_1 + \mathbf{k}_2), \quad (113)$$

with the power spectrum

$$P_\zeta = \frac{H^2}{8\pi^2 \epsilon c_s}, \quad (114)$$

where the variables are evaluated at the horizon crossing of the corresponding  $k$ -mode.

To calculate the bispectrum, we look at the cubic action. In the following, we list three terms that are most interesting for this subsection,

$$\begin{aligned} S_3 = \int dt d^3x \left\{ -\frac{a^3 \epsilon}{H c_s^2} \left[ \left( 1 - \frac{1}{c_s^2} \right) + \frac{2\lambda}{\Sigma} \right] \dot{\zeta}^3 \right. \\ \left. + \frac{3a^3 \epsilon}{c_s^2} \left( 1 - \frac{1}{c_s^2} \right) \zeta \dot{\zeta}^2 \right. \\ \left. - a\epsilon \left( 1 - \frac{1}{c_s^2} \right) \zeta (\partial\zeta)^2 + \dots \right\}. \end{aligned} \quad (115)$$

The full terms can be found in in [50, equations (4.26)–(4.28)].

The order of magnitude contribution from these three terms can be estimated similarly as we did in (92), but now we not only keep factors of  $H$  and  $\epsilon$ , but also factors of  $c_s$ .

Take the first term as an example, we write it in terms of the conformal time,

$$\int d\tau H_3(\tau) \supset \int d\tau d^3x \frac{a\epsilon}{H c_s^2} \left[ \left( 1 - \frac{1}{c_s^2} \right) + \frac{2\lambda}{\Sigma} \right] \dot{\zeta}^3. \quad (116)$$

Comparing (112) with (85), we see that there is an extra factor of  $c_s$  accompanying  $\tau$ . So we estimate  $d\tau \propto c_s^{-1}$  and  $a \approx -(H\tau)^{-1} \propto c_s H^{-1}$ . Also,  $\zeta \propto H/\sqrt{\epsilon c_s}$ , but  $\dot{\zeta} \propto c_s \zeta$ . Overall, the vertex (116) contributes

$$\propto \frac{H}{\sqrt{\epsilon c_s}} \left[ \left( 1 - \frac{1}{c_s^2} \right) + \frac{2\lambda}{\Sigma} \right]. \quad (117)$$

Multiplying the three external legs  $\zeta^3$ , and using the definition

$$\langle \zeta^3 \rangle \sim f_{\text{NL}} P_\zeta^2 \quad (118)$$

and  $P_\zeta \propto (H/\sqrt{\epsilon c_s})^2$ , we get

$$f_{\text{NL}} \sim \mathcal{O}\left(\frac{1}{c_s^2}\right) + \mathcal{O}\left(\frac{\lambda}{\Sigma}\right). \quad (119)$$

The other two terms are similar. A detailed calculation reveals that

$$\begin{aligned} \langle \zeta(\mathbf{k}_1) \zeta(\mathbf{k}_2) \zeta(\mathbf{k}_3) \rangle &= (2\pi)^7 \delta^3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) (P_\zeta)^2 \frac{1}{\prod_i k_i^2} \\ &\times (S_\lambda + S_c + S_o + S_\epsilon + S_\eta + S_s), \end{aligned} \quad (120)$$

where we have decomposed the shape of the three-point function into six parts. The first two come from the leading order terms that we listed in (115),

$$S_\lambda = \left( \frac{1}{c_s^2} - 1 - \frac{2\lambda}{\Sigma} \right) \frac{3k_1 k_2 k_3}{2K^3}, \quad (121)$$

$$\begin{aligned} S_c = \left( \frac{1}{c_s^2} - 1 \right) \left( -\frac{1}{K} \sum_{i>j} k_i^2 k_j^2 + \frac{1}{2K^2} \sum_{i \neq j} k_i^2 k_j^3 + \frac{1}{8} \sum_i k_i^3 \right) \\ \times \frac{1}{k_1 k_2 k_3}. \end{aligned} \quad (122)$$

In terms of  $f_{\text{NL}}$  their sizes are

$$f_{\text{NL}}^\lambda = \frac{5}{81} \left( \frac{1}{c_s^2} - 1 - \frac{2\lambda}{\Sigma} \right), \quad (123)$$

$$f_{\text{NL}}^c = -\frac{35}{108} \left( \frac{1}{c_s^2} - 1 \right). \quad (124)$$

The next four terms come from the subleading terms that we did not list explicitly in (115) as well as the subleading contributions from the first two terms. Their orders of magnitude are

$$f_{\text{NL}}^o = \mathcal{O}\left(\frac{\epsilon}{c_s^2}, \frac{\epsilon\lambda}{\Sigma}\right), \quad (125)$$

$$f_{\text{NL}}^{\epsilon, \eta, s} = \mathcal{O}(\epsilon, \eta, s).$$

The detailed profiles can be found in [50].

The full results we obtained can be used in different regimes.

- (i) If we look at the limit,  $c_s \ll 1$  or  $\lambda/\Sigma \gg 1$ , the leading order results give two shape components,  $S_\lambda$  and  $S_c$ . This result can also be obtained using a simple method of considering only the fluctuations in scalar field while neglecting those in gravity [73, 74]. Intuitively, this is because the higher derivative terms are responsible for the generation of large non-Gaussianities, and the gravity contribution is expected to be small as we saw in Section 4. Therefore, one expands  $P(X, \phi)$  using  $\phi(\mathbf{x}, t) = \phi_0(t) + \delta\phi(\mathbf{x}, t)$ . The derivatives of  $P$  with respect to  $\phi$  are ignored because the inflation and scale invariance imposes an approximate shift symmetry on  $P$  in terms of inflaton  $\phi$ . We then get two terms in the cubic Lagrangian density,

$$\begin{aligned} \mathcal{L}_3 = a^3 & \left( \frac{1}{2} P_{,XX} \dot{\phi}_0 + \frac{1}{6} P_{,XXX} \dot{\phi}_0^3 \right) \delta\phi^3 \\ & - \frac{a}{2} P_{,XX} \dot{\phi}_0 \delta\phi (\nabla \delta\phi)^2. \end{aligned} \quad (126)$$

This gives two leading bispectra the same as (121) and (122). The approach that we present here gives a rigorous justification to such a method. The subleading order component  $S_o$  may be observable as well. At this limit where the higher derivative terms of the inflaton field are dominant, the Lagrangian of the above effective field theory are generalized [75] to include, for example, the ghost inflation [76] whose Lagrangian cannot be written in a form of  $P(X, \phi)$ . Another two slightly different equilateral shapes arise. However, it is worth to mention that, generally in single field models and Einstein gravity, going beyond  $P(X, \phi)$  requires adding either terms that explicitly break the Lorentz symmetry, or terms with higher time derivatives on  $\phi$  which cannot be eliminated by partial integration, such as  $(\square\phi)^2$ . Different treatment of such terms and discussions on their effects can be found in [77–79].

- (ii) If we take the opposite, slow-roll limit,  $c_s \rightarrow 1$  and  $\lambda/\Sigma \rightarrow 0$ , we recover the two shape components  $S_\epsilon$  and  $S_\eta$  that we got in Section 4, with unobservable size  $f_{\text{NL}} \sim \mathcal{O}(\epsilon)$ .
- (iii) We can also look at the intermediate parameter space. In slow-roll inflation models, one can also add higher derivative terms [49, 80]. But in order not to spoil the slow-roll mechanism, the effect of these terms can only be subdominant. This corresponds to  $c_s \approx 1$  and  $\lambda/\Sigma < \mathcal{O}(1)$ . Using the full results, we can see that the size of the non-Gaussianity is  $f_{\text{NL}} < \mathcal{O}(1)$ . Therefore, it is important to emphasize that, for the class of models we consider here, *nonslow-roll* inflationary mechanisms, such as the example that will be given below, are necessary to generate observable large non-Gaussianities.

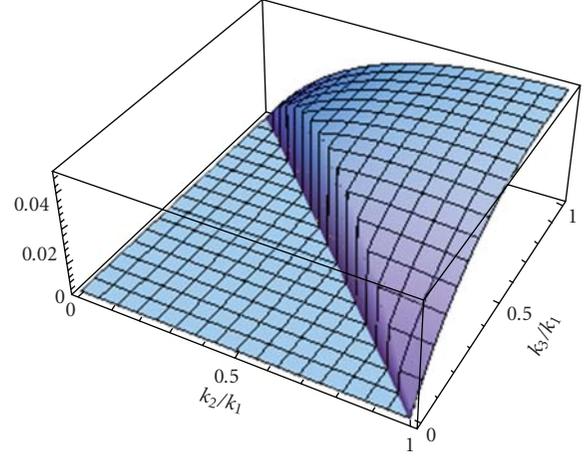


FIGURE 3: Shape of  $S_\lambda$  in (121).

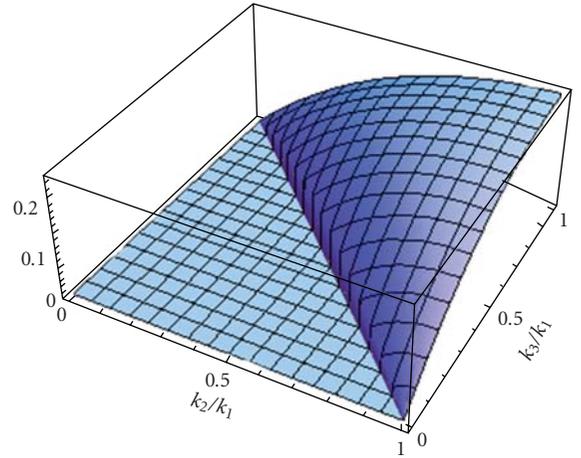


FIGURE 4: Shape of  $S_c$  in (122).

- (iv) The other terms that we did not list in (115) (see [50]) and their canonical limit (80) are also useful. These terms are exact for arbitrary values of  $\epsilon$ ,  $\eta$ , and  $s$ , so the usage of the action is beyond the category of models that we focus on in this subsection. As we will see in Sections 6.2 and 6.3, it can be applied to the cases of sharp or periodic features where these parameters do not always remain small.

In the rest of this subsection, we focus on the first case.

In Figures 3 and 4, we draw the shapes of  $S_\lambda$  and  $S_c$ . The two shapes are similar. They both peak at the equilateral limit, and behave as  $S \sim k_3/k_1$  in the squeezed limit  $k_3 \ll k_1 = k_2$ . We call these shapes the *equilateral shapes*. There are some small differences between  $S_\lambda$  and  $S_c$ , for example, around the folded triangle limit  $k_2 + k_3 = k_1$ . A factorizable shape ansatz for the equilateral shape that is often used in data analyses is the following [81]:

$$S_{\text{ansatz}}^{\text{eq}} = -6 \left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right) + 6 \left( \frac{k_1}{k_2} + 5 \text{ perm.} \right) - 12, \quad (127)$$

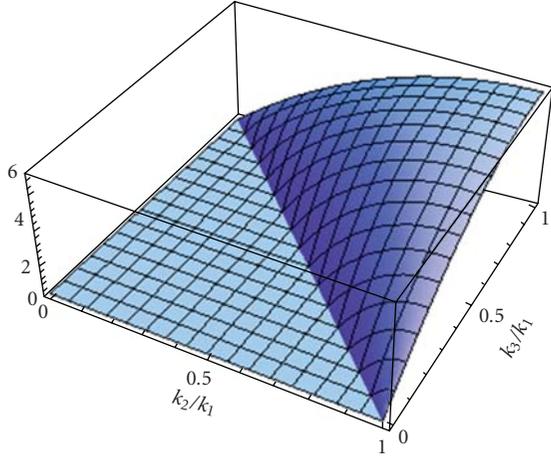


FIGURE 5: An ansatz (127) for the equilateral shape.

and is shown in Figure 5. As we can see, it represents the most important features of Figures 3 and 4.

The shape of  $S_0$  is more complicated, but we expect they have the similar shapes as the equilateral one because their squeezed limits behave the same [50]. The three other shapes  $S_{\epsilon}$ ,  $S_{\eta}$  and  $S_s$  are all close to the local shapes as their squeezed limit scale as  $k_1/k_3$  for  $k_3 \ll k_1 = k_2$ .

The scale dependence in  $P_\zeta$ ,  $c_s$  and  $\lambda/\Sigma$  will introduce mild running for the three-point function. We usually regard only the contributions from  $c_s$  and  $\lambda/\Sigma$  as the running of the non-Gaussianity.

The underlying physics of the equilateral shape can be readily understood in terms of their generation mechanism. In single field inflation, the long wavelength mode that exits the horizon are frozen and can have little interaction with modes within the horizon. The large interaction only occurs among modes that are crossing the horizon at about the same time. These modes then have similar wavelengths. This is why the shape of the non-Gaussianity peaks at the equilateral limit in momentum space.

This physical origin also suggests the caveat that, as long as there are large interactions involving modes with similar wavelengths, an equilateral-like shape may arise. For example, such cases can happen in multifield models where there are particle creation [82, 83] (see however [84]).

(i) *An Example: Dirac-Born-Infeld (DBI) Inflation.* An explicit example of the above general results is the DBI inflation [54, 73, 85–92]. These inflation models describe a 3 + 1 dimensional brane moving in warped extra dimensions. The location of the brane is a scalar field in 4D effective field theory, and it is the inflaton. The warped extra dimensions provide a nontrivial internal field space for the inflaton. In terms of the 4D effective field theory, the action is

$$\begin{aligned}
 & - \int d^4x \sqrt{-g} \\
 & \times \left[ f(\phi)^{-1} \sqrt{1 + f(\phi) g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi} - f(\phi)^{-1} + V(\phi) \right].
 \end{aligned} \tag{128}$$

The nontrivial part is the kinetic term involving the square-root. It can be understood as a generalization of the following two familiar situations. It is a higher dimensional generalization of the action of a relativistic point particle

$$\int dt \sqrt{1 - f \dot{\mathbf{x}}^2}, \tag{129}$$

where the speed of light  $f^{-1/2}$  varies with  $\mathbf{x}$ . It is also a relativistic generalization of the usual canonical kinetic term in the nonrelativistic limit  $|f(\phi) g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi| \ll 1$ ,

$$- \int d^4x \sqrt{-g} \left[ \frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi + V(\phi) \right]. \tag{130}$$

Because the speed limit of the inflaton  $f^{-1/2}$  can vary in the internal space, if it can be made small enough near the top of potential where the inflaton is about to roll down, the warped space restricts the rolling velocity even if the potential is too steep for slow-roll inflation to happen. So the inflaton rolls ultra relativistically, but with very small velocity, and this generates the DBI inflation.

The physical consequence is now easy to obtain using the general results in this subsection. In our notation the Lagrangian is

$$P = -f^{-1} \sqrt{1 - 2fX} + f^{-1} - V. \tag{131}$$

The sound speed is

$$c_s = \sqrt{1 - 2fX}, \tag{132}$$

which is the inverse of the Lorentz boost factor  $\gamma$ , so  $c_s \ll 1$ . The component (123) vanishes identically, and we have a large bispectrum of shape  $\mathcal{A}_c$  with size (124).

DBI inflation is still driven by the potential energy. The general single field inflation models also include the k-inflation [93], where the inflation is driven by the inflaton kinetic energy. Model construction of single field k-inflation can be found in [93–95]. The bispectra for such models are computed in [50, 94].

Multifield generalization have been studied in [96–102], where this type of kinetic terms are generalized to multiple fields. The three-point functions involving these different fields have the same or similar shapes.

The current CMB constraint on the equilateral ansatz (127) is  $-214 < f_{\text{NL}}^{\text{eq}} < 266$  [1].

*6.2. Sinusoidal Running: Sharp Feature.* Although various slow-variation parameters in (110) have to be small most of the time during inflation, they can become temporarily large. Such cases can happen if there are sharp features in inflaton potentials or internal field space, so the behavior of inflatons temporarily deviates from the attractor solution, and then relaxes back within several Hubble time, or stay longer but with small deviation amplitudes. Motivations for such models include the following. It may be possible explanations for features in power spectrum [103–106], and if so the associated non-Gaussianity is a cross-check. And there are brane inflation models that are very sensitive to

sharp features present in the potential or in the internal space [107].

As an example, we study a sharp feature in the slow-roll potential. The fact that a sharp feature in potential can enhance non-Gaussianities has long been anticipated and qualitative estimates have been made by different methods [108–110]. The precise method of analyzing the size, running and shape of such non-Gaussianities [63, 64] is made possible with the developments of the formalisms that we reviewed in Sections 3.1, 4, and 6.1. This will be the subject of this subsection.

We start by studying the behavior of the slow-roll parameters. We use a small step in potential as an example and will ignore numerical coefficients. We use  $c \sim \Delta V/V$  to denote the relative height of the step, and  $d$  the width of the step. In the attractor solution, the inflaton velocity is given by  $\dot{\phi} \sim V'/H \sim \sqrt{\epsilon V}$ . As it falls down the step, the potential energy  $cV$  gets converted to the kinetic energy, so we have

$$\dot{\phi} \lesssim \sqrt{V(c + \epsilon)}. \quad (133)$$

The amplitude of density perturbations is given by  $P_\zeta \sim H^4/\dot{\phi}^2$ , so such a sharp feature causes glitches in the power spectrum. It will leave a dip with relative size  $\Delta P_\zeta/P_\zeta \sim \sqrt{1 + c/\epsilon} - 1$  since  $\dot{\phi}$  increases first, followed by oscillations caused by a nonattractor component of the mode function before it settles down again in the attractor solution. To fit the CMB data,  $\dot{\phi}$  cannot change much. As we can see, the sensitivity of the power spectrum to the step size  $c$  is proportional to  $\epsilon$ , and we need  $c/\epsilon \lesssim 1$ . Reducing the width  $d$  of sharp feature increases the amplitude of the glitches, but this is only for a large  $d$  over which the inflaton spends more than one  $e$ -fold to cross. Further reducing  $d$  will not change the amplitude of the glitches since (133) is saturated, but the sharpness will determine how deep within the horizon the modes are affected.

So  $\epsilon$  does not change much,  $\Delta\epsilon \sim \Delta(\dot{\phi}^2)/H^2 \sim c$ . But it changes within a very short period,  $\Delta t \sim \Delta\phi/\dot{\phi} \sim d/\sqrt{V(c + \epsilon)}$ . So  $\eta$  can be very large,

$$\Delta\eta \sim \frac{\Delta\epsilon}{H\epsilon\Delta t} \sim \frac{c\sqrt{c + \epsilon}}{d\epsilon}. \quad (134)$$

It is also clear that the feature is associated with a characteristic physical scale and generates a scale-dependent power spectrum and higher-order correlation functions.

With these qualitative behavior in mind, we now study the three-point function. An important fact of the formalisms in Sections 4 and 6 is that the expansion is exact in terms of the slow-variation parameters. So it is valid even if these parameters are not always small, as long as the expansion in  $\zeta \sim \mathcal{O}(10^{-5})$  is perturbative.

In all terms in the cubic expansion (80),  $\zeta$  appears at most with one time derivative; the field redefinition gives a term that is proportional to  $\eta$  at the end of the inflation; and the other terms are all suppressed by powers of  $\epsilon$ , which remains small even in the presence of a sharp feature. So the most important term is

$$\int dt d^3x \frac{1}{2} a^3 \epsilon \dot{\eta} \zeta^2 \zeta, \quad (135)$$

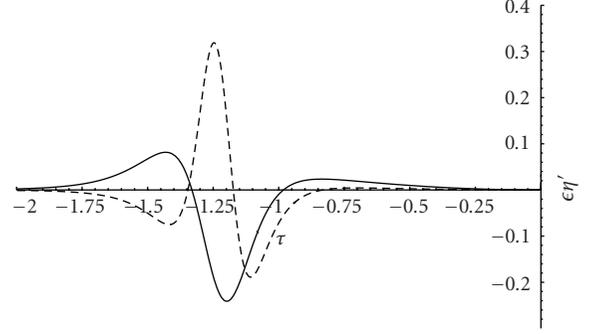


FIGURE 6: Behavior of the slow-roll parameters for a step (solid line,  $c = 0.0018$ ,  $d = 0.022M_{\text{p}}$ ) and bump (dashed line,  $c = 0.0005$ ,  $d = 0.01M_{\text{p}}$ ) sharp feature on  $(1/2)m^2\phi^2$  potential. Note that in the absence of the sharp feature,  $\epsilon\eta'$  is of order  $\mathcal{O}(10^{-4})$ .

in which the coupling is proportional to  $\dot{\eta}$ . The correlation function  $\langle \zeta(\mathbf{k}_1)\zeta(\mathbf{k}_2)\zeta(\mathbf{k}_3) \rangle$  is dominated by

$$i \left( \prod_i u_i(\tau_{\text{end}}) \right) \int_{-\infty}^{\tau_{\text{end}}} d\tau a^2 \epsilon \eta' \times \left( u_1^*(\tau) u_2^*(\tau) \frac{d}{d\tau} u_3^*(\tau) + \text{sym} \right) (2\pi)^3 \delta^3 \left( \sum_i \mathbf{k}_i \right) + \text{c.c.} \quad (136)$$

Precise evaluation of this expression has to be done numerically. But it is not difficult to see the generic properties of bispectra associated with a sharp feature.

For long wavelength modes that already crossed the horizon at the time of the sharp feature,  $k_i\tau \ll 1$ , the mode function is already frozen and the integration (136) gives vanishing contribution. For short wavelength modes that are still well within the horizon, the modes are not affected if their momenta are larger than the inverse of the time scale characterizing the sharpness of changes in slow-roll parameters. The modes most affected are those which are near the horizon crossing. These modes are all oscillatory,  $\sim e^{-ik_i\tau}$ . As we have studied,  $\eta'$  is temporarily boosted, so it can be roughly approximated as several hat-functions that satisfy  $\int d\tau \eta' = 0$ . Examples of such behavior are shown in Figure 6. If we simply approximate the hat-functions by several delta-functions,  $\eta' \propto \delta(\tau - \tau_*)$ , the integration (136) will give something like

$$S \sim f_{\text{NL}}^{\text{feat}} \sin \left( \frac{K}{k_*} + \phi_0 \right), \quad (137)$$

where  $k_* \equiv 1/\tau_*$  is the scale corresponding to the location of feature,  $\phi_0$  is a phase and

$$f_{\text{NL}}^{\text{feat}} \sim \Delta\eta \sim \mathcal{O} \left( \frac{c\sqrt{c + \epsilon}}{d\epsilon} \right). \quad (138)$$

Comparing with the effect on the power spectrum, one can keep the size of glitches in the power spectrum small while make  $f_{\text{NL}}$  large, for example, by fixing  $c/\epsilon$  and decreasing  $d$ .

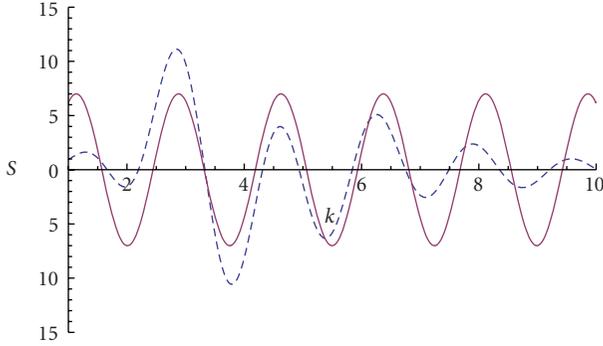


FIGURE 7: Numerical result (dashed line) for the bispectrum running for a sharp step ( $c = 0.0018, d = 0.022$ ) along the  $k_1 = k_2 = k_3 \equiv k$  direction, compared with the simple ansatz (137) (solid line).

This ansatz describes the most important running behavior of this bispectrum. Notice that the oscillatory frequency in the  $k$ -space is of order  $1/k_*$ , which is the scale of the feature. A rescale in  $k_*$  can be compensated by a rescale in all  $k_i$ . Also notice that the oscillatory frequency,  $3/k_*$ , along the  $k_1 = k_2 = k_3 \equiv k$  direction is  $3/2$  of that in the power spectrum,  $2/k_*$ . (For power spectrum, the sharp feature introduces a small non-Bunch-Davies component for the mode function. The oscillatory frequency in the power spectrum is determined by the phase of the coefficient for this component. This is obtained through matching conditions across the feature, and the phase is  $\sim 2k/k_*$ ; see Section 5.3 of [107] for an example. From this result we can see that, observationally, while sharp features located at large scales (such as  $\ell \sim 30$ ) introduce glitches that need to be distinguished from statistical fluctuations, those located at much shorter scales (such as  $\ell \sim 1000$ ) introduce oscillatory modulation that coherently shifts all points over several acoustic peaks in the same direction, which is completely different from statistical fluctuations of data points.)

In practice, (137) is a crude ansatz that needs to be refined. First of all, we have only considered the modes that have not exited the horizon. For those that did, as we mentioned, their correlation function is as small as usual. The ansatz needs to be cut off for the long wavelength modes  $K/k_* \ll 1$ . A more detailed analysis [64] reveals, using the hat functions as an approximation of the slow-roll parameter behavior, that the bispectrum falls off as  $K^2$  for these long wavelength modes. Secondly, the fact that in (137) all short wavelength modes are equally affected is due to the sharp-change approximation. Smoother functions will only affect a finite range of modes within the horizon. So the amplitude of the ansatz should decay and how fast depends on the sharpness of feature. To take into account both effects, empirically, we can multiply (137) with an envelope function

$$\propto \left(\frac{K}{k_*}\right)^n e^{-K/mk_*}, \quad (139)$$

where  $n$  and  $m$  are parameters chosen to fit the numerical results. For example,  $n = 2$  and  $m = 5$  for Figure 7. Lastly,

in the very squeezed limit,  $k_3 \ll k_1 |K/k_*|$ ,  $S$  can no longer be approximated as a function of  $K$  only and starts to have a nontrivial shape [64]. Here we concentrate on the signature running behavior.

A numerical result with

$$V(\phi) = \frac{1}{2} m^2 \phi^2 \left[ 1 + c \tanh\left(\frac{\phi - \phi_s}{d}\right) \right] \quad (140)$$

is shown in Figure 7. A subtlety encountered in the numerical integration is how to handle the oscillatory behavior at  $\tau \rightarrow -\infty$ . One can do a tilt into the imaginary plane,  $-\tau \rightarrow -\infty(1 + i\epsilon)$ , as prescribed in the analytical procedure in Section 3.2; or more efficiently, perform integration by part to increase the convergence of the integrand at the  $\tau \rightarrow -\infty$  end. One may also try the method of Wick rotation, but this will first require solving the background equations of motion in the Wick-rotated space, since here we do not have the analytical expression for the mode function.

Sharp features can also appear elsewhere instead of potentials, for example, in the internal warped space for DBI inflation [107]. The qualitative running behavior in bispectrum is similar, and overall large non-Gaussianities become a superposition of the approximate scale-invariant equilateral shape and the sinusoidal running.

Non-attractor initial conditions can be included as a case of sharp features, except that we only observe the relaxation part.

**6.3. Resonant Running: Periodic Features.** In this subsection, we consider a different type of features. These features may or may not be sharp, but the most important property is their periodicity. Such features will induce an oscillatory component to the background evolution, in particular, to the couplings in the interaction terms. We denote this oscillatory frequency as  $\omega$ . We know that each mode oscillates when it starts the life well within the horizon. This frequency keeps on decreasing as the mode gets stretched by the inflation, until it reaches  $H$  when the mode becomes frozen. So the mode scans through all frequencies that is larger than  $H$ , up to some very high cutoff scale such as  $M_p$ . Therefore, as long as

$$\omega > H, \quad (141)$$

the oscillatory frequency of the modes in the integral will hit  $\omega$  at some point during the inflation. This causes a resonance between the couplings and modes, hence a constructive contribution to the correlation function [64]. Without the resonance, as we encountered previously, the highly oscillatory modes simply average out within the horizon. In contrast to the previous mechanisms, here the non-Gaussianities are generated when modes are subhorizon.

We now study the properties of such a non-Gaussianity, following reference [64].

To estimate the integral, we use the unperturbed mode function. Similar to the sharp feature case, we get

$$\begin{aligned} & \langle \zeta(\mathbf{k}_1)\zeta(\mathbf{k}_2)\zeta(\mathbf{k}_3) \rangle \\ & \approx i \frac{H^4}{64\epsilon^3 \prod_i k_i^3} (2\pi)^3 \delta^3 \left( \sum_i \mathbf{k}_i \right) \\ & \quad \times \int_{-\infty}^0 \frac{d\tau}{\tau} \epsilon \eta' (1 - i(k_1 + k_2)\tau - k_1 k_2 \tau^2) k_3^2 e^{iK\tau} \\ & \quad + \text{two perm.} + \text{c.c.} \end{aligned} \quad (142)$$

In this case, we are interested in the region  $|K\tau| \gg 1$  in order to have resonance. So the last term dominates as long as the momentum triangle is not too squeezed so one of the  $k_i$ s becomes  $< 1/\tau$  at the resonance point. The oscillatory coupling is dominantly contributed by  $\eta'$ . The integral is proportional to

$$\int d\tau \tau \sin(\omega\tau) \exp(iK\tau). \quad (143)$$

This integral can be done analytically using the relation  $t \approx -H^{-1} \ln(-H\tau)$ . But its most important properties can be understood as follows in terms of the physical picture that we described.

First, let us look at its oscillatory running in  $K$ -space. The phase of the background repeats itself after  $\Delta N_e = 2\pi H/\omega$  e-fold, during which the wave-number  $K$  changes by  $-K\Delta N_e$ . So the running of the non-Gaussianity in  $K$ -space is also oscillatory with the period given by

$$\Delta K = K\Delta N_e = \frac{2\pi KH}{\omega}. \quad (144)$$

Note that this period is changing with  $K$  in a specific way that we will see more clearly in a moment.

Next, let us look at the size of the non-Gaussianity. Each  $K$ -mode briefly resonates with the oscillatory coupling when its frequency sweeps through the resonance frequency  $\omega$ . Once its frequency differs from  $\omega$  by  $\Delta\omega$ , the integration in the 3pt starts to cancel if is performed over  $\Delta t_1 \sim \pi/\Delta\omega$ . In the meanwhile it takes  $\Delta t_2 \sim \Delta\omega/(\omega H)$  to stretch the mode and change its frequency from  $\omega$  to  $\omega - \Delta\omega$ . Equating  $\Delta t_1$  and  $\Delta t_2$  gives the time period over which the resonance occurs for this mode,

$$\Delta t \sim \sqrt{\frac{\pi}{\omega H}}. \quad (145)$$

This corresponds to the number of oscillation periods

$$\frac{\omega \Delta t}{2\pi} \sim \sqrt{\frac{\omega}{4\pi H}} \quad (146)$$

that we need to integrate over to estimate the resonance contribution. Note that one period in the integral (143) for  $K/a = \omega$  contributes  $\pi\tau_*/K$ , where  $\tau_*$  is evaluated at the resonant point. Multiplying the total number of the

resonant periods (146), using the definition (103) and  $\tilde{P}_\zeta \approx H^2/(8\pi^2\epsilon)$ , we see that the amplitude of  $S(k_1, k_2, k_3)$  is

$$f_{\text{NL}}^{\text{res}} \sim \frac{\sqrt{\pi}}{16} \eta_A \tau_* \sqrt{\frac{\omega}{H}} \sim \frac{\sqrt{\pi}}{8\sqrt{2}} \frac{\omega^{1/2} \dot{\eta}_A}{H^{3/2}}. \quad (147)$$

Slow-roll parameters acquire small oscillatory components, and here  $\eta_A$  denotes the amplitude of such an oscillation. Other prefactors of  $k_i$  are cancelled according the definition of  $S$  and the  $S$  turns out to be a function of  $K$  only. In the last step of (147), we have listed the accurate numerical number, which differs from the estimate by a factor of  $\sqrt{2}$ .

Summarizing both the running behavior and the amplitude, we get the following ansatz for the bispectrum:

$$S_{\text{ansatz}}^{\text{res}} = f_{\text{NL}}^{\text{res}} \sin\left(C \ln\left(\frac{K}{k_*}\right)\right), \quad (148)$$

where

$$C = \frac{2\pi K}{\Delta K} = \frac{\omega}{H} \quad (149)$$

and  $k_*$  gives a phase. The argument  $C \ln K$  in (148) appears because of (144). This gives a scale dependent oscillatory frequency in the  $K$ -space. In fact, this kind of dependence makes the density perturbations in the resonance model semiscale-invariant. We call it *periodic-scale-invariant*—they are invariant under a discrete subgroup of rescaling. Namely, the ansatz (148) is invariant if we rescale all  $k_i$  by  $n\Delta K/K = 2\pi nH/\omega$  e-fold, where  $n$  is an integer. Other rescaling causes a phase shift. This property is a direct consequence of the symmetry of the Lagrangian. It is periodic, so invariant under a discrete shift of the inflaton field. This periodic-scale-invariance should also be respected by the full-bispectrum results, as well as other correlation functions.

As mentioned, we have derived this ansatz from the last term in (142). Other terms will become important in the squeezed limit. The full integration (142) has been worked out in [111], and the leading order results are

$$\begin{aligned} S^{\text{res}} = f_{\text{NL}}^{\text{res}} & \left[ \sin\left(C \ln\left(\frac{K}{k_*}\right)\right) \right. \\ & \left. + \frac{1}{C} \sum_{i \neq j} \frac{k_i}{k_j} \cos\left(C \ln\left(\frac{K}{k_*}\right)\right) + \mathcal{O}\left(\frac{1}{C^2}\right) \right], \end{aligned} \quad (150)$$

where  $\mathcal{O}(1/C^2)$  terms are neglected because we need  $1/C = H/\omega \ll 1$  for large resonance. The numerical coefficient in (147) turns out to be  $\sqrt{\pi}/(8\sqrt{2})$ . As we can see, the extra terms satisfy the symmetry we mentioned and indeed give large corrections in the very squeezed limit, for example,  $k_3 < k_1 H/\omega$ . These terms also ensure a consistency condition that we will study in Section 9.2. An example is plotted in Figure 8. The spike at the very squeezed limit is due to the second term in (150). Overall, we see that the leading *shape* of this bispectrum is quite trivial, being almost a function of  $K$  only, until it gets to the very squeezed limit. The most distinctive feature of this type of non-Gaussianities is

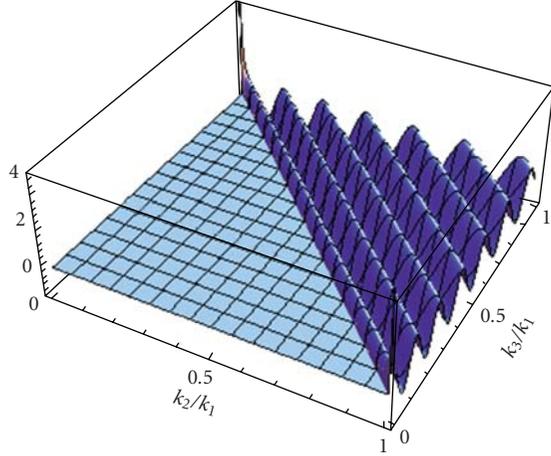


FIGURE 8: The running and shape of the resonance bispectrum (150) with  $C = 100$ .

the *running* behavior captured in (148). Unfortunately, this ansatz is not factorizable if the  $K$ -range is too large.

More arbitrary scale-dependence can be introduced if the features are applied over a finite range, or with varying periodicity and amplitude.

As a useful comparison, the resonant running here and sinusoidal running that we studied in the last subsection are clearly distinguishable from each other observationally. The resonant running oscillates with periods that are always much smaller than the local scale,  $\Delta K \ll K$ ; the frequency has a specific scale-dependence,  $\Delta K/K = \text{const.}$ ; and the frequency in the power spectrum ( $\sim \sin(C \ln k)$  in  $k$ -space) is exactly the same as that in the bispectrum ( $\sim \sin(C \ln K)$  in  $K$ -space). In contrast, the bispectrum of the sinusoidal running oscillates with a fixed period that approximately equals to the scale at the location of the sharp feature,  $\Delta K \sim k_*$ ; the frequency is scale-independent; and the power spectrum ( $\sim \sin(2k/k_*)$  in  $k$ -space) has twice an oscillatory frequency of the bispectrum ( $\sim \sin(K/k_*)$  in  $K$ -space).

As an illustration, we look at an example,

$$V(\phi) = \frac{1}{2} m^2 \phi^2 \left[ 1 + c \sin\left(\frac{\phi}{\Lambda}\right) \right]. \quad (151)$$

In this example, the inflaton is rolling over the small but periodic ripple laid on the potential. This induces an oscillatory component in the slow-roll parameters with an amplitude  $\eta_A \approx \sqrt{6} c m \phi / \Lambda^2$  and a frequency  $\omega \approx \dot{\phi} / \Lambda \approx 2m / (\sqrt{6} \Lambda)$ . So we have

$$\begin{aligned} f_{\text{NL}}^{\text{res}} &\sim \frac{c M_{\text{p}}^3}{\Lambda^{5/2} \phi^{1/2}}, \\ C &\approx \frac{2}{(\phi \Lambda)}. \end{aligned} \quad (152)$$

A numerical example is shown in Figure 9. As we can see, the ansatz (148) gives a very accurate fit to the actual running behavior. The mode function and power spectrum are the superposition of the usual unperturbed solution

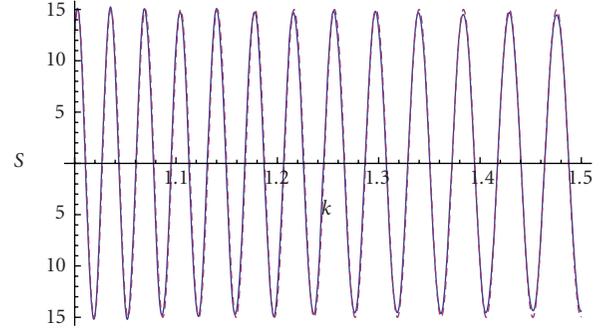


FIGURE 9: Numerical result (solid line) of the bispectrum running for the example (151) ( $m = 3 \times 10^{-6} M_{\text{p}}$ ,  $c = 5 \times 10^{-7}$ ,  $\Lambda = 0.0007 M_{\text{p}}$ ,  $\phi \approx 15 M_{\text{p}}$ ), compared with the ansatz (148) (dashed line).

and a small oscillatory component [64, 112, 113]. We can choose parameters so that the size of the ripples on the power spectrum is small, but bispectrum is made large. This is because the non-Gaussianities rise more quickly if we increase the frequency, while the mode function has difficulty responding efficiently when the external source oscillates too fast. This mechanism may be realized in terms of brane inflation [107] where the periodic feature comes from duality cascade in warped throat [114], or the monodromy inflation [115, 116] where the periodic feature comes from instanton effects [112, 113].

**6.4. Folded Shape: A Nonstandard Vacuum.** In this subsection, we study the effect of nonstandard vacuum on the primordial non-Gaussianities. We consider a different wavefunction from the Bunch-Davies vacuum when modes are well within the horizon. To start, let us first discuss several motivations for this case.

- (i) A non-Bunch-Davies vacuum can actually occur much more simply than it might sound like. Any deviation from the attractor solution of the inflaton generically generates a component of non-Bunch-Davies vacuum. This is because a general mode function is a superposition of two components,  $c_1(\mathbf{k})u(\mathbf{k}, t) + c_2(\mathbf{k})u^*(\mathbf{k}, t)$ , and in attractor solution we choose one of the component asymptotic to the Bunch-Davies vacuum. A disturbance will generically introduce a mixture with the other component. In this sense, we have already encountered such a case when we studied the effect of a sharp feature in Section 6.2. Indeed, after the inflaton crosses the sharp feature, the oscillatory behavior in the power spectrum is precisely due to the superposition of the second non-Bunch-Davies component for some finite  $k$ -range. For an *infinitely sharp* change, such a disturbance with a small amplitude extends to all  $k$  that have not exited the horizon at the time of sharp feature. An analytical illustration can be found in Section 5.3 of [107]. The location of the sharp feature can become superhorizon at the present

time, but its influence has extended to much smaller scales and becomes observable. The resonance case in Section 6.3 is another type of example. An analytical illustration can be found in Section 3.3 of [112]. For non-Gaussianities studied in Sections 6.2 and 6.3, we only concentrated on the effects caused by the features in slow-roll parameters. In analytical analyses, we approximated the mode function by the Bunch-Davies component and ignored the disturbance. The study of this subsection can be regarded as the complementary analyses on the effect of a different mode component.

- (ii) In inflationary background, modes can be quantized in terms of time-dependent creation and annihilation operators,  $a_{\mathbf{k}}(t)$  and  $a_{-\mathbf{k}}^\dagger(t)$ . The Bunch-Davies vacuum is defined as the vacuum annihilated by  $a_{\mathbf{k}}(t)$  as  $t \rightarrow -\infty$ . If a different adiabatic vacuum is defined which is annihilated by  $a_{\mathbf{k}}(t_0)$  at a finite  $t_0$ , we introduce a non-Bunch-Davies component. For example, see [117, 118]. The origin and magnitude of such a component have been debated and studied by many papers, often under the name of the “trans-Planckian effect”; see [119, 120] for summary and references.
- (iii) There are inflation models where the scale of new physics can be very low. In particular, in warped space it is proportional to the exponentially small warp factor. In some DBI inflation models [86, 92], the speed limit of the inflaton and the scale of new physics are both related to the warp factor in such a way that the local warped new scale can drop near or even below the Hubble energy scale in certain epoch of inflation. Clearly the simple scalar field Bunch-Davies vacuum is no longer sufficient. Such models further open up the possibilities of vacuum choices.

After these discussions, let us now focus on a specific simple problem [50]. We modify the wave function of the Bunch-Davies vacuum by a small second component and examine its consequence for the three-point function calculated in Section 6.1. We consider the general single field inflation with a small sound speed  $c_s$  or a large  $\lambda/\Sigma$  [50, 121].

So the mode function is

$$u_{\mathbf{k}}(\tau) = \frac{iH}{\sqrt{4\epsilon c_s k^3}} \left[ (1 + ikc_s\tau)e^{-ikc_s\tau} + C_-(1 - ikc_s\tau)e^{ikc_s\tau} \right], \quad (153)$$

where  $|C_-| \ll 1$  and can be  $k$ -dependent. In the first example above, the extra component starts at a specific time in the past. In the second class of examples, it may start either at a specific time or specific energy scale. To see a common feature without addressing these model-dependent issues, we look at the simple limit where the  $\tau$  in (153) can go all the way to  $-\infty$ . The computation of the correlation function is essentially the same as in Section 6.1. The leading order correction to the bispectra is obtained by replacing one of the three  $u_{\mathbf{k}}(\tau)$  in the integrand by its  $C_-$  component. So

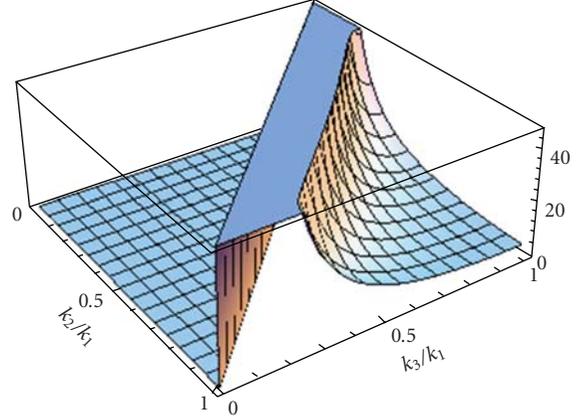


FIGURE 10: Shape of  $\tilde{S}_\lambda$  (truncated).

it simply replaces one of the  $k_i$ s in the shapes with  $-k_i$ . For example, the correction to  $S_\lambda$  is

$$\begin{aligned} \tilde{S}_\lambda = & \left( \frac{1}{c_s^2} - 1 - \frac{2\lambda}{\Sigma} \right) \frac{3k_1 k_2 k_3}{2} \\ & \times \left( \frac{\text{Re}(C_-(k_3))}{(k_1 + k_2 - k_3)^3} \right. \\ & \left. + \frac{\text{Re}(C_-(k_2))}{(k_1 - k_2 + k_3)^3} + \frac{\text{Re}(C_-(k_1))}{(-k_1 + k_2 + k_3)^3} \right). \end{aligned} \quad (154)$$

The shape of  $\tilde{S}_\lambda$  is shown in Figure 10. The most important feature of this shape is the enhancement at the folded triangle limit, for example,  $k_1 + k_2 - k_3 = 0$ . The detailed form of enhancement is model dependent. For example, it is different for another shape  $S_c$ . The divergence in this folded limit occurs due to our simple limit of taking  $\tau$  to  $-\infty$ . Imposing some kind of cutoff at the lower limit of  $\tau$  will eliminate this divergence, although as mentioned the detailed modification will be highly model dependent. For example, a simple constant cutoff  $\tau_c$  will introduce a factor of  $1 + ((1/2)x_c^2 - 1)\cos x_c - x_c \sin x_c$  for each of the three terms in (154), where  $x_c \equiv (k_1 + k_2 - k_3)c_s\tau_c$  or its cyclic. Very close to the folded limit,  $\tilde{K}_{c_s}|\tau_c| \ll 1$  ( $\tilde{K} = k_1 + k_2 - k_3$  or its cyclic), this regulates away the divergence; away from the folded limit,  $\tilde{K}_{c_s}|\tau_c| \gg 1$ , these extra factors are unity on average but with oscillations. These oscillation can be either physical, or regarded to be zero if  $x_c$  is within a regulation scale which exists since the non-Bunch-Davies component is present for a finite time in the past.

The case for slow-roll inflation is qualitatively similar, and more examples of the bispectra shapes and the observational prospects are discussed in [122, 123]. In this case, the proportional parameter for the bispectra amplitude is no longer enhanced by  $1/c_s^2$  or  $\lambda/\Sigma$ , but  $< \mathcal{O}(1)$ .

In order to facilitate the data analyses, a simple ansatz has been proposed in [123],

$$S_{\text{ansatz},1}^{\text{fold}} = 6 \left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right) - 6 \left( \frac{k_1}{k_2} + 5 \text{ perm.} \right) + 18, \quad (155)$$

which represents certain important features of this kind of bispectra. It has a smooth rising behavior in the folded limit. This ansatz is plotted in Figure 11(a). Since the real shape has a model dependent cutoff, it remains open questions how sensitive this is to data analyses and how well the ansatz (155) represents it. We can also write down an ansatz which is more directly motivated from the example (154) and the comments after that equation,

$$S_{\text{ansatz},2}^{\text{fold}} = k_1 k_2 k_3 \frac{k_1 + k_2 - k_3}{(k_c + k_1 + k_2 - k_3)^4} + 2 \text{ perm.}, \quad (156)$$

where the cutoff scale  $k_c = 1/(c_s \tau_c)$  is a parameter. For  $k_1 + k_2 - k_3 \gg k_c$  and cyclic, we have neglected the oscillatory part and only taken the average. In this ansatz, we can change the powers in the numerator and denominator to model model-dependent variations. The relation  $(k_c + k_1 + k_2 - k_3)^{-n} = (\Gamma(n))^{-1} \int_0^\infty dt t^{n-1} e^{-(k_c + k_1 + k_2 - k_3)t}$  may be used to factorize the ansatz. This ansatz is plotted in Figure 11(b).

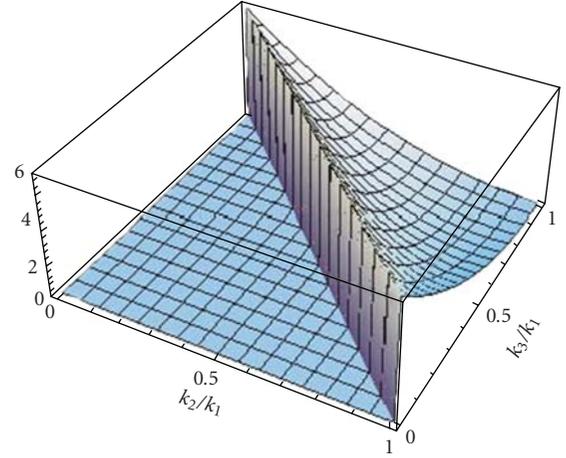
Another type of non-Bunch-Davies vacuum, namely, an  $n$ -particle state built on the normal Bunch-Davies vacuum, was studied in [124, 125] and the non-Gaussianities were found to be unobservable.

## 7. Quasisingle Field Inflation

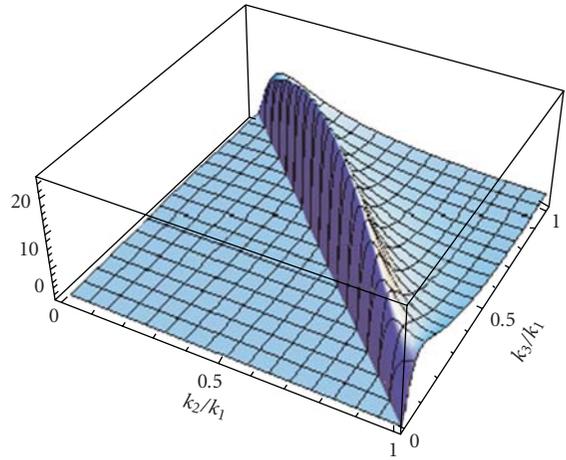
Having considered single field inflation, we now relax the condition on the number of fields. At least during inflation, we only need to consider quantum fluctuations of light fields, since if the mass of fields are very heavy, (here the relevant scale is  $m \gg H$ ), they contribute only classically and determine the classical inflaton trajectory. Multiple light fields can arise naturally if we consider the inflation models as the consequence of a UV completed framework. However, as discussed in Section 5.1, due to the back-reaction from the inflationary background, the mass of light fields are naturally of order  $H$ . The potential with such a shape is too steep for slow-roll inflation.

Therefore, as a natural step beyond the single field, let us consider slow-roll models with one inflationary direction, and one or more other directions that have mass neither much heavier nor much lighter than  $H$ . We will call the quanta in the inflationary direction as the inflaton and its mode the curvature mode, and the others isocurvature and isocurvature modes. We call these models the *quasisingle field inflation models* [46, 126].

Note that the thematic order in this paper is not chronological. The non-Gaussianities in this type of models were not computed until very recently for a couple of reasons. If the mass of particles is of order  $\mathcal{O}(H)$  or larger, the amplitude of these fields decay exponentially in time after horizon-exit. So they would not seem to



(a)



(b)

FIGURE 11: Two ansatz for the folded shape. (a) Equation (155) and (b) Equation (156) with  $k_c/k_1 = 0.1$ .

be important for superhorizon perturbations even if they couple to the curvature mode. As we will see, however, their amplitudes at or near the horizon-exit are enough to make them interesting. What really suppresses their contribution is the fast oscillation behavior present for  $m \gg H$ . Methodologically, isocurvature-to-curvature transition for non-Gaussianities was studied restrictively in the regime of superhorizon classical evolution in multi-field space [127–134], which we will explain in more details in the next section. However, for quasisingle field inflation models, a full quantum computation in the in-in formalism is necessary to properly include the contributions from both the horizon exit and the superhorizon evolution.

**7.1. Intermediate Shapes: Massive Isocurvatons.** There are potentially different ways massive isocurvatons can be coupled to the inflaton. We currently do not have a general approach in terms of model building. So what we will do is to first study this problem through a simple example, and

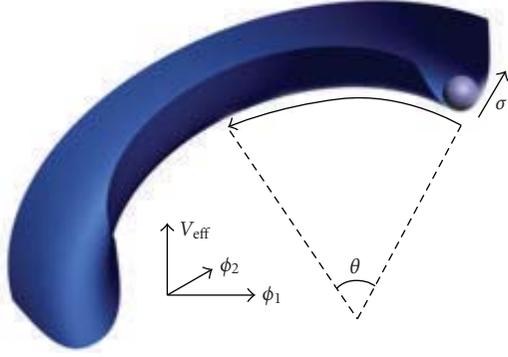


FIGURE 12: Quasi-single field inflation with turning trajectory. The field  $\theta$  and  $\sigma$  are in the polar coordinates. The  $\theta$  is the inflationary direction with a slow-roll potential. The  $\sigma$  is the isocurvature direction, which typically has mass of order  $H$ .

then discuss the features of the results that can be regarded as generic signatures of this class of models [46, 126].

We consider the case where the inflaton is turning constantly by going around (a fraction of) a circle with radius  $R$  in the angular  $\theta$  direction. See Figure 12. All the parameters, such as  $R$  and couplings, are assumed to be constant during the turning. We call this assumption the constant turn case. In the  $\theta$  direction the potential is the usual slow-roll potential  $V_{\text{sr}}(\theta)$ . The field in the radial direction is denoted as  $\sigma$  and has mass of order  $H$ , and lifted by the potential  $V(\sigma)$ . For such a turning trajectory, it is convenient to write the action in terms of fields in the polar coordinates,  $\theta$  and  $\sigma$ , instead of in the Cartesian coordinates,

$$S_m = \int d^4x \sqrt{-g} \left[ -\frac{1}{2}(R + \sigma)^2 g^{\mu\nu} \partial_\mu \theta \partial_\nu \theta - \frac{1}{2} g^{\mu\nu} \partial_\mu \sigma \partial_\nu \sigma - V_{\text{sr}}(\theta) - V(\sigma) \right]. \quad (157)$$

The potential  $V(\sigma)$  balances off the centrifugal force necessary for the turning and traps the field at the bottom of the effective potential,  $V_{\text{eff}}(\sigma) = -(1/2)\dot{\theta}_0^2(R + \sigma)^2 + V(\sigma)$ . We define the minimum of this effective potential to be  $\sigma = 0$ . We expand the effective potential as

$$V_{\text{eff}} = \text{const.} + \frac{1}{2}(V''' - \dot{\theta}_0^2)\sigma^2 + \frac{1}{6}V'''\sigma^3 + \dots, \quad (158)$$

where  $\dot{\theta}_0$  is the turning angular velocity and the primes on  $V$  denote derivatives with respect to  $\sigma$ .

To study the perturbation theory, we perturb the fields in the spatially flat gauge,

$$\theta(\mathbf{x}, t) = \theta_0(t) + \delta\theta(\mathbf{x}, t), \quad \sigma(\mathbf{x}, t) = \delta\sigma(\mathbf{x}, t), \quad (159)$$

and obtain the following Hamiltonian,

$$\mathcal{H}_0 = a^3 \left[ \frac{1}{2} R^2 \delta\dot{\theta}_I^2 + \frac{R^2}{2a^2} (\partial_i \delta\theta_I)^2 + \frac{1}{2} \delta\dot{\sigma}_I^2 + \frac{1}{2a^2} (\partial_i \delta\sigma_I)^2 + \frac{1}{2} m^2 \delta\sigma_I^2 \right], \quad (160)$$

$$\mathcal{H}_2^I = -c_2 a^3 \delta\sigma_I \delta\dot{\theta}_I, \quad (161)$$

$$\mathcal{H}_3^I = c_3 a^3 \delta\sigma_I^3, \quad (162)$$

where

$$c_2 = 2R\dot{\theta}_0, \quad c_3 = \frac{1}{6}V''', \quad m^2 = V'' + 7\dot{\theta}_0^2 \quad (163)$$

are all constants. Terms suppressed by  $\mathcal{O}(\epsilon)$  have been ignored in this gauge. The curvature perturbation  $\zeta$  is most transparent in another gauge, the uniform inflaton gauge, where

$$\theta(\mathbf{x}, t) = \theta_0(t), \quad \sigma(\mathbf{x}, t) = \sigma_0(t) + \delta\sigma(\mathbf{x}, t), \quad (164)$$

and the spatial metric is

$$h_{ij}(\mathbf{x}, t) = a^2(t) e^{2\zeta(\mathbf{x}, t)} \delta_{ij}. \quad (165)$$

In this gauge,  $\zeta$  appears in the metric as the space-dependent rescale factor and the fluctuations in the inflaton is shifted away. The relation between  $\zeta$  and  $\delta\theta$  is the gauge transformation. At the leading order this is

$$\zeta \approx -\frac{H}{\dot{\theta}_0} \delta\theta. \quad (166)$$

We will calculate the correlation functions in terms of  $\delta\theta$  and then use this relation to convert them to those of  $\zeta$ . The full-perturbation theory that one obtains in the uniform inflaton gauge justifies the above omission of several  $\mathcal{O}(\epsilon)$  terms in the spatially flat gauge [46].

There are several important points for this Hamiltonian.

First, the kinematic Hamiltonian (160) describes two free fields in the inflationary background. One is massless and has the familiar mode function,

$$u_{\mathbf{k}} = \frac{H}{R\sqrt{2k^3}} (1 + ik\tau) e^{-ik\tau}. \quad (167)$$

Another is massive and the mode function is

$$v_{\mathbf{k}} = -ie^{i(\nu+(1/2))(\pi/2)} \frac{\sqrt{\pi}}{2} H(-\tau)^{3/2} H_\nu^{(1)}(-k\tau), \quad (168)$$

where

$$\nu = \sqrt{\frac{9}{4} - \frac{m^2}{H^2}}. \quad (169)$$

For  $0 \leq m/H \leq 3/2$ , the amplitude of the mode  $v_{\mathbf{k}}$  decays as  $(-\tau)^{-\nu+3/2}$  after horizon-exit  $k\tau \rightarrow 0$ . The lighter the isocurvature is, the slower it decays. At the  $\nu \rightarrow 3/2$  (i.e.,  $m/H \rightarrow 0$ ) limit, the amplitude is frozen. For  $m/H > 3/2$ ,  $\nu$  becomes imaginary, the mode  $v_{\mathbf{k}}$  not only contains a

decay factor  $(-\tau)^{3/2}$  but also an oscillation factor  $\tau^\nu$ . This oscillation is marginal for  $m \sim H$ , but if  $m \gg H$ , it causes cancellation in the integrals of the correlation function and is equivalent to factors of Boltzmann-like suppression  $\sim e^{-m/H}$ . We will consider the case  $0 \leq \nu \leq 3/2$ .

Second, there is a sharp contrast between the  $V_{\text{sr}}'''$  for the slow-roll inflaton field and the  $V'''$  for the massive field  $\sigma$  in the noninflationary direction. The former has to be very small,  $\sim \mathcal{O}(\epsilon^2)P_\zeta^{1/2}H$ , in order to maintain the smallness of the slow-roll parameters. (Here we use  $\epsilon$  to denote collectively all slow-roll parameters,  $\epsilon \equiv -\dot{H}/H^2$ ,  $\eta \equiv \dot{\epsilon}/\epsilon H$ , and  $\xi \equiv \dot{\eta}/\eta H$ .) Consequently it contributes  $\mathcal{O}(\epsilon^2)$  to the  $f_{\text{NL}}$  of bispectrum in slow-roll inflation, generally smaller than the  $\mathcal{O}(\epsilon)$  contributions from the other terms in the same model. However, for quasisingle field inflation, the direction orthogonal to slow-roll does not have to satisfy the slow-roll conditions, and  $V'''$  is almost unconstrained. For example, in the inflationary background, it can be of order  $H$ ; and similarly,  $V''''$  can be of order one, and so forth. This isocurvaton self-interaction (162) becomes the source of large non-Gaussianities.

Third, the coupling between the isocurvaton and inflaton appears as a form of a two-point vertex operator in (161). We treat this term as part of the interaction Hamiltonian, and it is represented by the transfer vertex in Figure 13(a). The strength of the coupling is determined by the turning angular velocity  $\dot{\theta}_0$  in this model. This coupling is responsible for the transformation of the isocurvature perturbations, in particular their large non-Gaussianities, to the curvature perturbation.

We calculate correlation functions corresponding to the Feynman diagrams Figure 13 in terms of the in-in formalism, which we reviewed in Section 3.1. As an illuminating example to illustrate the different advantages of the three forms of the in-in formalism, we recall from Section 3.1 that the three-point function can be written in the following forms. The original definitions (72) and (68), which we refer to as the factorized form, lead to

$$\begin{aligned}
\langle \delta\theta^3 \rangle &= -12c_2^3 c_3 u_{p_1}^*(0) u_{p_2}(0) u_{p_3}(0) \\
&\times \text{Re} \left[ \int_{-\infty}^0 d\tilde{\tau}_1 a^3(\tilde{\tau}_1) v_{p_1}^*(\tilde{\tau}_1) u'_{p_1}(\tilde{\tau}_1) \right. \\
&\quad \times \int_{-\infty}^{\tilde{\tau}_1} d\tilde{\tau}_2 a^4(\tilde{\tau}_2) v_{p_1}(\tilde{\tau}_2) v_{p_2}(\tilde{\tau}_2) v_{p_3}(\tilde{\tau}_2) \\
&\quad \times \int_{-\infty}^0 d\tau_1 a^3(\tau_1) v_{p_2}^*(\tau_1) u'_{p_2}(\tau_1) \\
&\quad \left. \times \int_{-\infty}^{\tau_1} d\tau_2 a^3(\tau_2) v_{p_3}^*(\tau_2) u'_{p_3}(\tau_2) \right] \\
&\times (2\pi)^3 \delta^3 \left( \sum_i \mathbf{p}_i \right) + 9 \text{ other similar terms} \\
&+ 5 \text{ permutations of } \mathbf{p}_i.
\end{aligned} \tag{170}$$

The perturbation theory here starts from the fourth order. The reorganized commutator form (69) leads to

$$\begin{aligned}
\langle \delta\theta^3 \rangle &= 12c_2^3 c_3 u_{p_1}(0) u_{p_2}(0) u_{p_3}(0) \\
&\times \text{Re} \left[ \int_{-\infty}^0 d\tau_1 \int_{-\infty}^{\tau_1} d\tau_2 \int_{-\infty}^{\tau_2} d\tau_3 \right. \\
&\quad \times \int_{-\infty}^{\tau_3} d\tau_4 \prod_{i=1}^4 (a^3(\tau_i)) \\
&\quad \times a(\tau_2) (u'_{p_1}(\tau_1) - \text{c.c.}) \\
&\quad \times (v_{p_1}(\tau_1) v_{p_1}^*(\tau_2) - \text{c.c.}) \\
&\quad \times (v_{p_3}(\tau_2) v_{p_3}^*(\tau_4) u'_{p_3}(\tau_4) - \text{c.c.}) \\
&\quad \left. \times v_{p_2}(\tau_2) v_{p_2}^*(\tau_3) u'_{p_2}(\tau_3) \right] \\
&\times (2\pi)^3 \delta^3 \left( \sum_i \mathbf{p}_i \right) + 2 \text{ other similar terms} \\
&+ 5 \text{ permutations of } \mathbf{p}_i.
\end{aligned} \tag{171}$$

In the IR ( $\tau \rightarrow 0$ ), each of the ten terms in the factorized form diverge as  $\tau^{3-6\nu}$  for  $3/2 > \nu > 1/2$  ( $0 < m < \sqrt{2}H$ ); while in the commutator form, various subtractions off the complex conjugates and the requirement that the final result has to be real makes such divergence explicitly disappear.

In the UV ( $\tau \rightarrow -\infty$ ), each factor of the multiple integral that integrates from  $-\infty$  to 0 has a definite convergent direction if we choose one of the two contour tilts,  $\tau_i \rightarrow -\infty(1 \pm i\epsilon)$ , accordingly. Or more efficiently, by a Wick rotation  $\tau_i \rightarrow \pm iz_i$ . This would have been the case for the commutator form if we can break up the integrand into individual terms. However in order to achieve the explicit IR convergence, as we saw above, these terms have to be grouped; but then they have contradicting convergence directions.

To take advantage of both forms, we introduce a cutoff  $\tau_c$ , and write the IR part ( $\tau_c < \tau \leq 0$ ) of the integrals in terms of the commutator form, and the UV part ( $\tau < \tau_c$ ) in terms of the factorized form, in the following mixed form:

$$\begin{aligned}
&\sum_i \int_{\tau_c}^0 d\tau_1 \cdots \int_{\tau_c}^{\tau_{i-1}} d\tau_i \{ \text{commutator form} \} \\
&\times \int_{-\infty}^{\tau_c} d\tau_{i+1} \cdots \int_{-\infty}^{\tau_n} d\tau_n \{ \text{factorized form} \}.
\end{aligned} \tag{172}$$

This shows explicitly both convergence behavior of the correlation function. Combining with Wick-rotations of the integration contours in the UV, this form provides an efficient way to evaluate the correlation functions numerically. The shapes of bispectra are presented in Figure 14 for  $\nu = 0, 0.3, 0.5, 1$ .

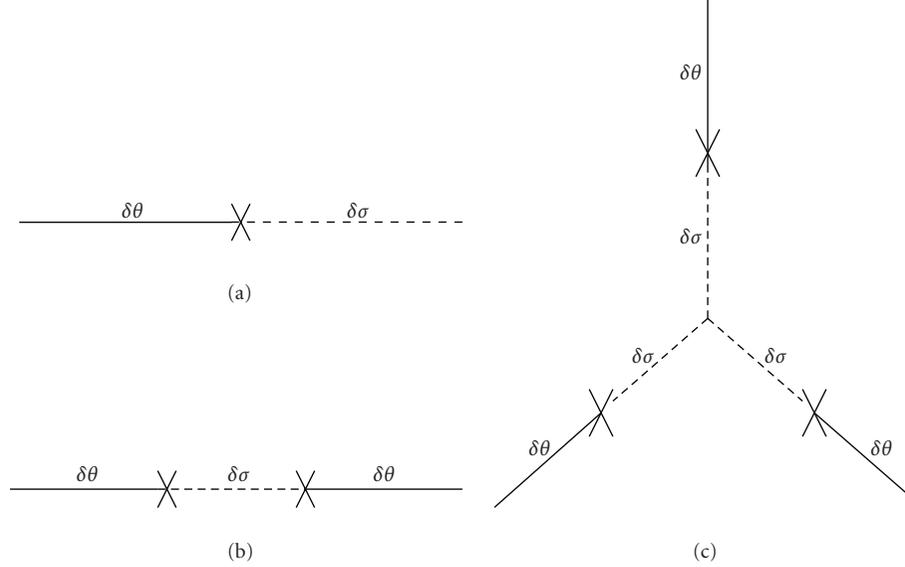


FIGURE 13: Feynman diagrams for the transfer vertex (a), corrections to the power spectrum from isocurvature modes (b), and the leading bispectrum (c).

To better understand the shapes analytically, we can work out the squeezed limit ( $p_3 \ll p_1 = p_2$ ) of the three-point function,

$$\langle \delta\theta^3 \rangle \rightarrow \frac{c_2^3 c_3}{HR^6} \frac{1}{p_1^{(7/2)-\nu} p_2 p_3^{(3/2)+\nu}} s(\nu) (2\pi)^3 \delta^3 \left( \sum_i \mathbf{p}_i \right), \quad (173)$$

where  $s(\nu)$  is a  $\nu$ -dependent numerical number.

Recall that the squeezed limit of  $S$  for the equilateral shape goes as  $p_3/p_1$ , while for the local shape  $(p_3/p_1)^{-1}$ . Both the numerical results in Figure 14 and the analytical results in (173) show that here we have a one-parameter family of shapes,  $\sim (p_3/p_1)^{1/2-\nu}$ , lie between the two. We call them the *intermediate shapes*.

The physical origin of such shapes can be understood as follows, and should be a generic signature for the quasilinear field inflation models. As we have seen, the large equilateral non-Gaussianity arises because the interacting modes cross the horizon around the same time. The shape of bispectrum peaks at the equilateral limit where the modes all have comparable wavelengths. As we will see in Section 8, the large local non-Gaussianity arises due to the classical nonlinear evolution of superhorizon modes in the multifield space; so the interactions are causally disconnected and behave local in position space. This is nonlocal from the momentum space point of view. So the shape of bispectrum peaks at the squeezed limit. Now for quasilinear field inflation, the large non-Gaussianities come from the massive isocurvature. Depending on the mass, these modes either decay right away after they exit the horizon (for  $m > \sqrt{2}H$ ), or survive for a long time at the superhorizon scales (for  $m < \sqrt{2}H$ ). In the former case, the generation and transfer of non-Gaussianities maximize for modes that are exiting the horizon around the same time, resulting in quasiequilateral shapes; in the

latter case, the generation and transfer of non-Gaussianities happen in a superhorizon fashion, resulting in quasilinear shapes. In this regard, let us look more closely at the special limit  $m/H \rightarrow 0$  ( $\nu \rightarrow 3/2$ ).

In this massless limit, an infrared cutoff to the integrals are necessary. Otherwise the transfer will last forever for the constant turn case. The cutoff corresponds to the ending of the turning. Let us discuss the following two cases. First, we still keep  $V'''$  large. Our analyses still apply in this case. Interestingly, the shape of the bispectrum goes to that of the local form in this limit. As we will explain in Section 8.1, this is a generic signature of a massless isocurvature. The infrared e-fold cutoff will introduce some running in the  $f_{\text{NL}}^{\text{int}}$  because different modes experience different turning e-folds. Second, we would like to make the isocurvature directions flat so this becomes a two-field slow-roll inflation models. Such models were intensively studied and it is known that the isocurvature modes can be transferred to the curvature mode by turning. However, since  $V''' \sim \mathcal{O}(\epsilon^{3/2})H^2/M_{\text{P}}$  is required to maintain the small slow-roll parameters, the contribution we computed here generates too small non-Gaussianity. We expect contributions from other terms are small as well. So it is much more difficult to generate large non-Gaussianities in such models, essentially because imposing the slow-roll conditions in all directions are too restrictive.

To connect with data analyses, guided by the numerical results and analytical squeezed limit, we can use the following ansatz to describe the full family of shapes:

$$S_{\text{ansatz}}^{\text{int}} = \frac{3^{(9/2)-3\nu} f_{\text{NL}}^{\text{int}} (p_1^2 + p_2^2 + p_3^2) (p_1 p_2 p_3)^{(1/2)-\nu}}{10 (p_1 + p_2 + p_3)^{(7/2)-3\nu}}. \quad (174)$$

These shapes are shown in Figure 15. Comparing with Figure 14, we can see that they match quite well except near

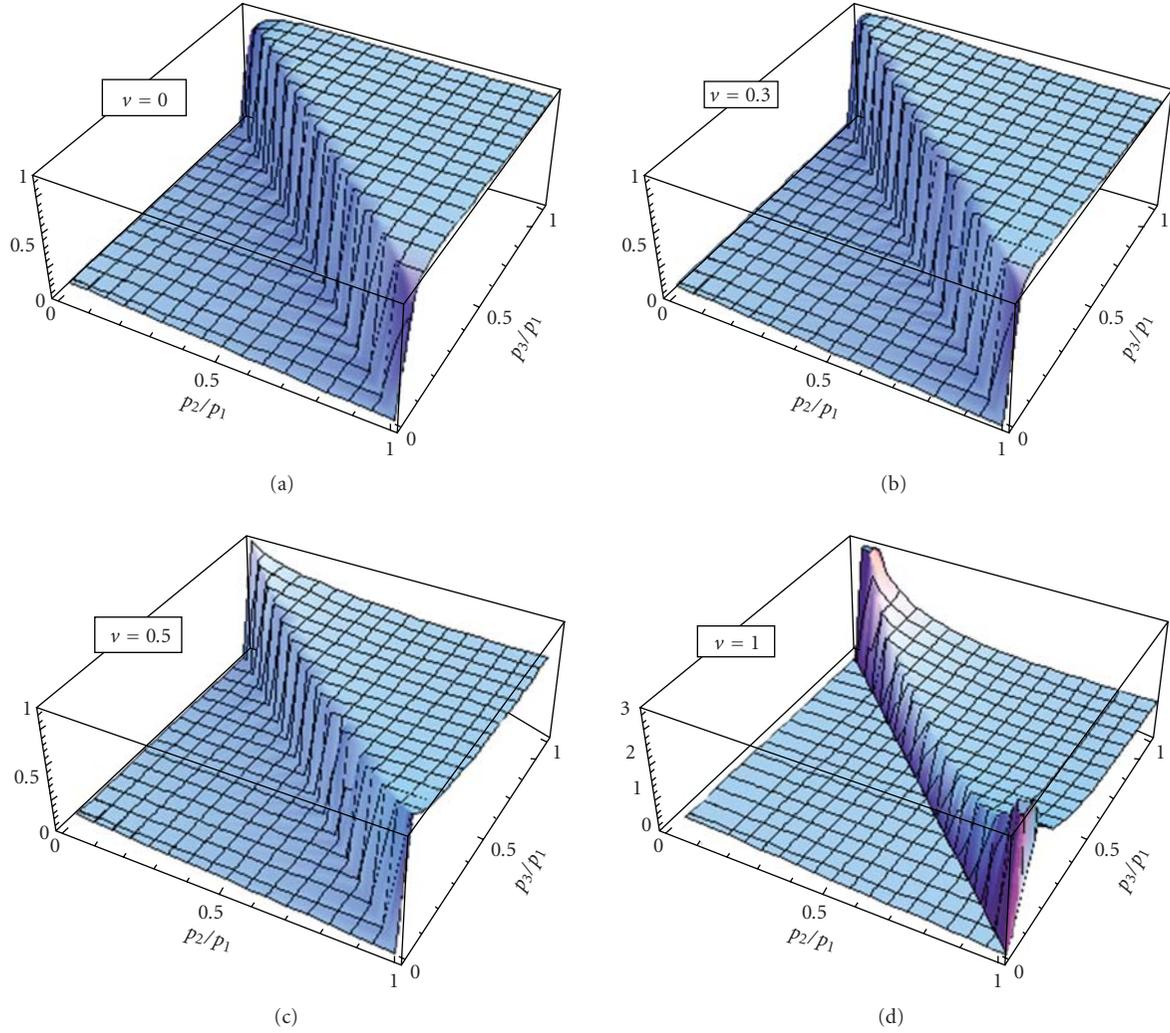


FIGURE 14: Numerical results for the shapes of bispectra with intermediate forms. We plot  $S$  with  $\nu = 0, 0.3, 0.5, 1$ . The plot is normalized such that  $S = 1$  for  $p_1 = p_2 = p_3$ .

$\nu = 0$ , around which it is better represented by another ansatz in [46]. The size of this bispectrum is

$$f_{\text{NL}}^{\text{int}} = \alpha(\nu) P_\zeta^{-1/2} \left( \frac{\dot{\theta}_0}{H} \right)^3 \left( -\frac{V''''}{H} \right), \quad (175)$$

where  $P_\zeta \approx 6.1 \times 10^{-9}$  and  $\dot{\theta}_0$  is the turning angular velocity. The  $\alpha(\nu)$  is a positive numerical number which, depending on  $\nu$ , can give an additional enhancement factor of order  $N_f$  ( $N_f$  is the turning e-folds). Since  $(\dot{\theta}_0/H)^2$  and  $V''''/H$  are the expansion parameters in the perturbation theory, they have to be small to trust our calculation. Nonetheless this is not the model-building requirement.

The fluctuations of more massive ( $m > \mathcal{O}(H)$ ) fields may become important if they play a role later in the reheating [135, 136]. Such cases typically require some tunings for special conditions, so that the highly suppressed fluctuation amplitude can become important.

## 8. Multifold Inflation

As we have seen in Section 7.1, if we take the isocurvaton mass to zero in quasisingle field inflation while keep the nonlinear self-couplings of the isocurvaton  $V''''$  large, the shape of the large bispectrum in the squeezed limit approaches the local form. The local form is in fact the earliest and most well-studied example of non-Gaussianities [66, 127, 137, 138], although it was first found to be small as we have seen in Section 4. As we will explain in this section, a large local form is a signature of massless isocurvatons that have large nonlinear evolution in multifield space. We have arrived this shape from the in-in formalism by taking the massless limit. But if we stay in this limit, there is an easier formalism, the  $\delta N$  formalism [139–141], in which the underlying physics of the local shape becomes transparent.

**8.1. Local Shape: Massless Isocurvatons.** We recall that, in single field inflation, if we use the uniform inflaton gauge where there are no fluctuations in the inflaton field,

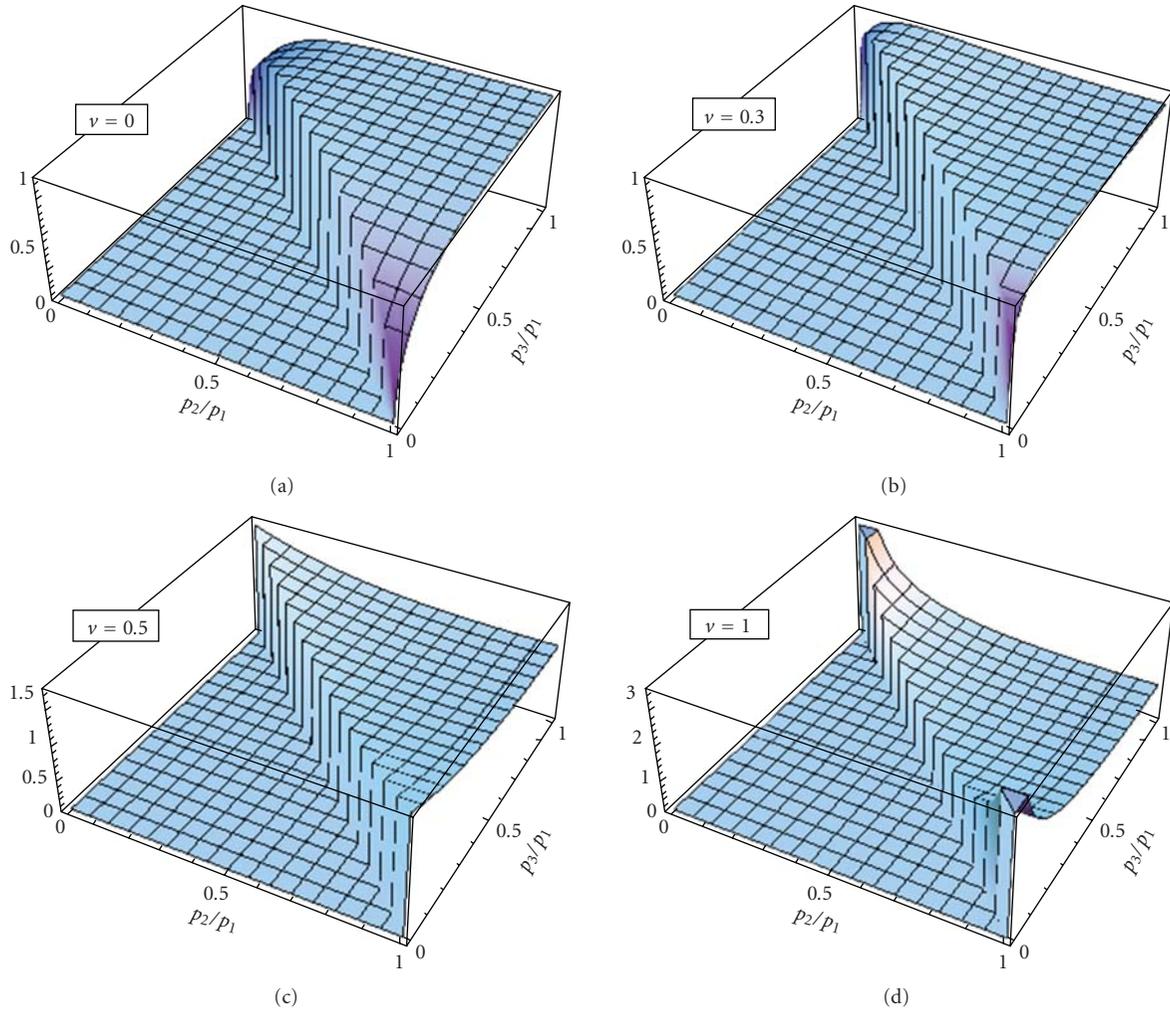


FIGURE 15: Shape ansatz (174) for the intermediate forms.

the scalar perturbation  $\zeta$  enters in the scale factor as  $a^2 e^{2\zeta}$ . For superhorizon modes,  $\zeta$  is frozen. If we look at the different comoving superhorizon patches, they are causally disconnected from each other. So they evolve independently and locally in space. In such a gauge, the only difference is a space-dependent scale factor. This is also called the separate universe picture. The primordial curvature perturbations manifest themselves as the different number of expansion e-fold,  $\delta N$ , at different positions.

We would like to generalize this picture to the multifield case in the following  $\delta N$  formalism. We will resort to a simple version of  $\delta N$  formalism stated below, which is of course a consequence of the in-in formalism, but formulated from a simple perspective which clearly illustrates the points in this section. Otherwise, as we will explain, in the most general sense the  $\delta N$  formalism should be defined as the in-in formalism written in terms of specified gauges.

- (i) We consider a set of scalars  $\phi_i$  during inflation. Inflaton is one of them but can be different linear combinations of  $\phi_i$ s as a function of time, and the other orthogonal fields are called the isocurvatons.

All the modes that we are eventually interested in should all have become *superhorizon* when the initial slice (specified below) is chosen. We look at different horizon-size patches and label them with the coarse-grained comoving coordinate  $\mathbf{x}$ . In the in-in formalism, the superhorizon modes behave as the c-number time-dependent background for each comoving patch. So we evolve these patches independently and classically.

- (ii) We pick an initial spatially flat slice, on which there is no scalar fluctuations in the metric and all the fluctuations are in the scalar fields  $\phi_{0i} + \delta\phi_i(\mathbf{x})$ . We *assume* that we know the statistics of such fluctuations.
- (iii) We pick the final uniform density slices. Relative to the unperturbed and perturbed initial spatially flat slices, we have, respectively, the unperturbed and perturbed final uniform density slices. For single field inflation, these two final surfaces are the same. For multifield models, they are generally different. Such

final slices have the properties that the universe has the same energy densities and field configurations everywhere on them. They can be chosen during either the inflation or the reheating. After that, every separated universe will have the same evolution. The only difference is the scale factor. This is the analogy of the uniform inflaton gauge in single field inflation. We study the cases where such slices exist.

- (iv) We evolve the unperturbed  $\phi_{0i}$  in the initial slice classically to the unperturbed final slice, and denote the number of e-folds as  $N_0(\phi_{0i})$ . This is the unperturbed e-fold number. We evolve the perturbed  $\phi_{0i} + \delta\phi_i(\mathbf{x})$  in the initial slice classically to the perturbed final slice, and denote the number of e-folds as  $N(\phi_{0i} + \delta\phi_i(\mathbf{x}))$ . The difference between them

$$\delta N = N(\phi_{0i} + \delta\phi_i(\mathbf{x})) - N_0(\phi_{0i}), \quad (176)$$

is the curvature perturbation  $\zeta$ . Here,  $N_0$  is a constant that can be shifted to make  $\langle \delta N \rangle$  zero.

- (v) We expand

$$\delta N = N_i \delta\phi_i + \frac{1}{2} N_{ij} \delta\phi_i \delta\phi_j + \dots, \quad (177)$$

where the subscripts on  $N$  denote the partial derivatives with respect to  $\phi_i$ . For example,  $N_{ij} = (\partial N / \partial \phi_i)(\partial N / \partial \phi_j)$ . Repeated indices are summed over. The correlation functions of  $\zeta$  can then be computed as the classical averages of the products, such as

$$\begin{aligned} \langle \zeta(\mathbf{x}_1) \zeta(\mathbf{x}_2) \rangle &= N_i N_j \langle \delta\phi_i(\mathbf{x}_1) \delta\phi_j(\mathbf{x}_2) \rangle, \\ \langle \zeta(\mathbf{x}_1) \zeta(\mathbf{x}_2) \zeta(\mathbf{x}_3) \rangle &= \frac{1}{2} N_{ij} N_k N_l \\ &\quad \times \langle \delta\phi_i(\mathbf{x}_1) \delta\phi_j(\mathbf{x}_1) \delta\phi_k(\mathbf{x}_2) \delta\phi_l(\mathbf{x}_3) \rangle \\ &\quad + 2 \text{ perm.} \end{aligned} \quad (178)$$

- (vi) We have assumed that the statistics of the  $\delta\phi_i(\mathbf{x})$  are known on the initial slice. But this is not always easy to get. So we will consider the simple cases where this statistics can be approximated as Gaussian. Otherwise, calculating such initial statistics requires using the full quantum mechanical in-in formalism.

Most generally, one identifies  $\delta N$  with the scalar curvature  $\zeta$  in the uniform inflaton gauge; and the relation between  $\delta N$  and  $\delta\phi(\mathbf{x})$  in the  $\delta N$  formalism is the gauge transformation between the uniform inflaton gauge and the spatially flat gauge. Calculating the correlation functions for  $\zeta$  becomes calculating those for  $\delta\phi(\mathbf{x})$  using the in-in formalism. An example is the one we have seen in Section 7.1.

- (vii) So far we have not used the condition that the isocurvatons are massless ( $m \ll H$ ). If they are massive, after horizon exit, the modes decay. So in the superhorizon classical regime, where the  $\delta N$  formalism is supposed to be useful, we are back in the single field inflation. Sub- and near-horizon perturbations should be computed by the full in-in formalism. Therefore, having massless isocurvatons opens up classical multifield space in which we can have sizable  $\delta N$  defined in (176).

Now let us consider the Gaussian fluctuations  $\delta\phi_i$ . From Section 2, we know that for massless scalars,

$$\langle \delta\phi_i(\mathbf{k}_1) \delta\phi_j(\mathbf{k}_2) \rangle = \frac{H_*^2}{2k_1^3} (2\pi)^3 \delta^3(\mathbf{k}_1 + \mathbf{k}_2) \delta_{ij}, \quad (179)$$

where  $H_*$  is the Hubble parameter when the corresponding mode exits the horizon. If the scalars are not exactly massless,  $H$  will have a running dependence on  $k_1$  caused by the decay of the amplitude. Using

$$\delta N(\mathbf{k}) = N_i \delta\phi_i(\mathbf{k}) + \frac{1}{2} N_{ij} \int \frac{d^3 \mathbf{k}'}{(2\pi)^3} \delta\phi_i(\mathbf{k} - \mathbf{k}') \delta\phi_j(\mathbf{k}'), \quad (180)$$

we get the power spectrum

$$\langle \zeta(\mathbf{k}_1) \zeta(\mathbf{k}_2) \rangle = (2\pi)^5 \frac{P_\zeta}{2k_1^3} \delta^3(\mathbf{k}_1 + \mathbf{k}_2), \quad (181)$$

where

$$P_\zeta = \left( \frac{H_*}{2\pi} \right)^2 N_i^2, \quad (182)$$

and the bispectrum

$$\begin{aligned} \langle \zeta(\mathbf{k}_1) \zeta(\mathbf{k}_2) \zeta(\mathbf{k}_3) \rangle \\ = N_{ij} N_i N_j \frac{H_*^4}{4} \left( \frac{1}{k_1^3 k_2^3} + 2 \text{ perm.} \right) (2\pi)^3 \delta^3(\sum \mathbf{k}_i). \end{aligned} \quad (183)$$

According to the definition (103), the shape function is

$$S^{\text{loc}} = \frac{3}{10} f_{\text{NL}}^{\text{loc}} \left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right), \quad (184)$$

where

$$f_{\text{NL}}^{\text{loc}} = \frac{5 N_{ij} N_i N_j}{6 (N_i^2)^2}. \quad (185)$$

As usual, we have ignored a mild running from the power spectrum  $P_\zeta$ . The shape (184) is called the local shape, which we plot in Figure 16. This form is already factorizable.

The physics of this shape can be understood from the derivation above. As explicitly demonstrated in (176)–(178), this non-Gaussianity is generated locally in position space for superhorizon modes. After Fourier transform, it becomes nonlocal in momentum space. That is the reason that the shape peaks at the squeezed limit.

If the perturbation  $\delta\phi(\mathbf{x}, t)$  on the initial spatially flat slice cannot be approximated as Gaussian, the shapes of final bispectra can be more complicated.

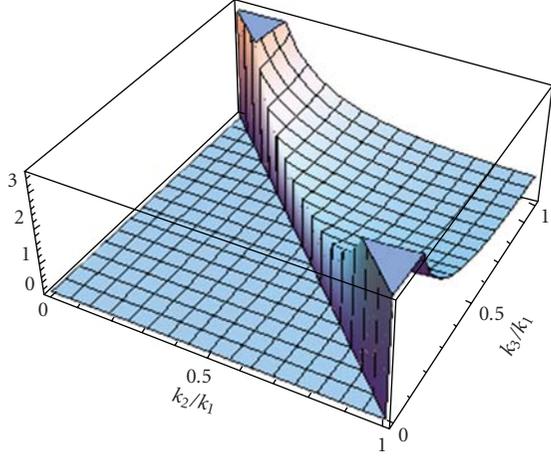


FIGURE 16: Shape of the local form (184).

(i) *An Example: The Curvaton Model.* We use the curvaton model [142–151] as an example to illustrate the generation of large local non-Gaussianity. We also use it to demonstrate the  $\delta N$  formalism.

In this model, we assume that during inflation there is another light field  $\sigma$  with the potential

$$V(\sigma) = \frac{1}{2} m_\sigma^2 \sigma^2, \quad (186)$$

and  $m_\sigma \ll H$ . This field is called the curvaton field for reasons that will become clear shortly. The energy density of the curvaton field is negligible initially. During inflation, it fluctuates and obtains the primordial amplitude  $\sigma_* = \sigma_0 + \delta\sigma(\mathbf{x})$ , where  $*$  denotes its value at the horizon-exit and after that the amplitude is approximately frozen. These perturbations are Gaussian for the potential (186) with the canonical kinetic term, but can be more complicated otherwise. Here, we study the simple Gaussian case. After inflation, these  $\sigma$ -modes remain frozen until the Hubble parameter drops below  $m_\sigma$ . Then the  $\sigma$ -field starts to oscillate around the bottom of the potential and behavior as matter. The Universe is still radiation dominated. The fraction of the matter energy density quickly grows, because the matter dilutes as  $a^{-3}$  while radiation  $a^{-4}$ . The  $\sigma$ -field decays to radiation when it reaches its lifetime.

Another assumption of the curvaton model is that the primordial fluctuations in the inflaton field is much smaller than what is needed to achieve  $\zeta \sim \mathcal{O}(10^{-5})$ , although their total energy density may still be the dominant one. So the primordial curvature perturbation is contributed by the fluctuations in the  $\sigma$  field, hence the name curvaton field.

At the initial spatially flat slice  $t_0$ , we denote the radiation and curvaton density as  $\rho_{r0}$  and  $\rho_{m0}$ , respectively, and the scale factor as  $a_0$ . Both components initially redshift as radiation. This lasts until the Hubble parameter reaches  $m_\sigma$  at  $t_1$ . We denote the scale factor at  $t_1$  as  $a_1$ . The Friedman equation at  $t_1$  is

$$3M_{\text{Pl}}^2 m_\sigma^2 = \left(\frac{a_0}{a_1}\right)^4 (\rho_{r0} + \rho_{m0}) \approx \left(\frac{a_0}{a_1}\right)^4 \rho_{r0}. \quad (187)$$

After this the curvaton starts to oscillate and behave as matter. Denote the decay rate of the curvaton as  $\Gamma$ . We use the sudden decay approximation and assume that they decay instantaneously at the epoch  $H = \Gamma$ , because a process that has the time scale  $T$  falls into the Hubble expansion epoch with  $H = 1/T$ . We denote this instant as  $t_2$ . The Friedman equation at  $t_2$  is

$$3M_{\text{Pl}}^2 \Gamma^2 = \left(\frac{a_0}{a_2}\right)^4 \rho_{r0} + \left(\frac{a_0}{a_1}\right)^4 \left(\frac{a_1}{a_2}\right)^3 \rho_{m0}. \quad (188)$$

Because at  $t_2$  the universe has the same Hubble parameter (hence the same energy density) everywhere, this is the final uniform density slice. After that, both components become radiation and the evolution everywhere is the same. So we want to work out the expansion e-folds  $N$  from  $t_0$  to  $t_2$  as a function of the initial field value  $\sigma$ . From (188), we get

$$e^{-4N} + e^{-3N} \alpha = \text{const.}, \quad (189)$$

where  $\alpha \equiv (a_0/a_1)(\rho_{m0}/\rho_{r0})$ . From (187) we can solve for  $a_0/a_1$  which is independent of  $\sigma$  at the leading order, and from (186) we know that  $\rho_{m0}$  is proportional to  $\sigma^2$ . Therefore,  $\alpha$  is proportional to  $\sigma^2$ . Also note that  $r_\rho \equiv (a_2/a_0)\alpha = e^N \alpha = \rho_m/\rho_r|_{t_2}$  is the ratio of the energy density between the curvaton and the rest of the radiation at  $t_2$ . Using these simple facts, we can differentiate (189) with respect to  $\sigma$  once and twice, and get

$$f_{\text{NL}}^{\text{loc}} = \frac{5}{6} \frac{N_{\sigma\sigma}}{N_\sigma^2} = \frac{5}{3r_\rho} - \frac{5(4+9r_\rho)}{12(4+3r_\rho)}. \quad (190)$$

In terms of the definition  $r \equiv 3\rho_m/(4\rho_r + 3\rho_m)|_{t_2} = 3r_\rho/(4+3r_\rho)$  often used in the literature,

$$f_{\text{NL}}^{\text{loc}} = \frac{5}{4r} - \frac{5}{3} - \frac{5r}{6}. \quad (191)$$

So Large local non-Gaussianity arises if  $r \ll 1$ . Note that although (189) only depends on  $\sigma$ , this is a multifield model because the curvaton takes effects during the reheating. In some simple models in which the curvaton leads to nonadiabatic perturbations between dark matter and photons,  $r$  is tightly constrained by observations [152].

The large local form has been studied most extensively in the past. Variety of possibilities exist. They all share the common feature that non-Gaussianities are generated by some massless isocurvaton fields which acquire the superhorizon evolution during the inflation. For example, in multifield slow-roll inflation a turning trajectory [153] can transfer non-Gaussianities from other directions to the inflaton direction [127–134]. But it is found to be very difficult to make non-Gaussianities large essentially because the very restrictive slow-roll conditions in all directions. In modulated reheating [154, 155] or preheating [156–161] scenarios, the role of isocurvaton is played by the massless fields which control the couplings during the reheating or preheating. Thus they create a large local non-Gaussianity in a similar fashion as the curvaton model.

Local form is also found in different contexts, such as models with special types of massive gauge fields that acquire superhorizon evolution [162–164], some nonlocal theories of inflation [165, 166], and certain cyclic universe scenario as alternatives to inflation [167–171].

The current CMB constraint on the local bispectrum is  $-10 < f_{\text{NL}}^{\text{local}} < 74$  [1]. Current constraint from large scale structure gives  $-29 < f_{\text{NL}}^{\text{local}} < 70$  [172]. Variety of methods have been invented to measure the local and other different forms of non-Gaussianities [173–186].

## 9. Summary and Discussions

*9.1. Summary.* In this subsection, we summarize the main results of Sections 6–8. Non-Gaussianities, conceptually being the expectation values of perturbations in a time-dependent background, are defined by the first-principle in-in formalism. Physically, having large primordial non-Gaussianities means that there are large nonlinear interactions of some sort determined by certain dynamics during inflation. Measuring them tells us the nature of the dynamics.

(i) *Equilateral Shape and Higher Derivative Kinetic Terms.* In single field inflation, the long wavelength modes that already exited the horizon are frozen. They cannot have large interactions with short wavelength modes that are still within the horizon. When modes are well within the horizon, they oscillate and the contributions to non-Gaussianities average out. Therefore the only chance to have large interaction is when all modes have similar wavelengths and exit the horizon at about the same time. Theories with higher derivative kinetic terms provide such interaction terms. This is why the resulting bispectrum shape peaks at the equilateral limit in the momentum space. It drops to zero at the squeezed limit  $k_3 \ll k_1 = k_2$  as  $k_3/k_1$ . It happens that, when these higher derivative terms become important enough so that the inflationary mechanism is no longer slow-roll, these large non-Gaussianities become observable. The forms of the bispectra are given in (121) and (122), and the shapes are plotted in Figures 3 and 4. The factorizable ansatz that is used to represent them in data analyses is given in (127) and plotted in Figure 5.

(ii) *Sinusoidal Running and Sharp Feature.* A sharp feature, in a potential or internal field space, introduces a sharp change in the slow-roll parameters, or the generalized slow-variation parameters. This can boost the magnitudes of time-derivatives of some parameters by several orders of magnitude while still keep the power spectrum viable. These time-derivatives act as couplings in the interaction terms. So they enhance the non-Gaussianities among the modes which are near the horizon-exit. How deep they affect the modes inside the horizon depends on how sharp the changes are.

The changes in these parameters can be roughly approximated as delta-functions in time. Correlation functions involve integrations of products of the slow-variation parameters and the mode functions. The latter contain oscillatory behavior  $\sim e^{-iK\tau}$ , where the comoving momentum  $K$  is  $k_1 +$

$k_2 + k_3$  for bispectra and  $\tau$  is the conformal time. The delta-function specifies a scale  $\tau_*$ . This is why after integration the bispectrum contains a sinusoidal factor  $\sim \sin(K/k_*)$ , where  $k_* = -1/\tau_*$  is the momentum of the mode that is near the horizon-exit at the time of the feature. So the most important property of this type of non-Gaussianity is this characteristic running. A numerical result of the running behavior is plotted in Figure 7. An ansatz is given in (137) and (139).

(iii) *Resonant Running and Periodic Features.* The periodic features do not have to be sharp. They introduce a small background oscillatory component in the slow-variation parameters. On the other hand, the mode functions are also oscillatory before they exit the horizon. Their frequencies are high when they lie deep inside the horizon and become lower as their wavelengths get stretched by the inflation. They are frozen after the wavelengths become comparable with the horizon size  $H^{-1}$ . This means that their frequencies continuously scan through the range from  $M_{\text{P}}$  (or some other large fundamental scale) to  $H$ . Therefore as long as the background oscillatory frequency  $\omega$  satisfies  $H < \omega < M_{\text{P}}$ , at some point during the evolution the small oscillatory component in the slow-variation parameters will resonant with the oscillatory behavior of the mode functions, and cause a large constructive contribution to the integration.

The periodicity of the features leads to a periodic-scale-invariance in density perturbations. Namely, they are scale invariant if we rescale all momenta by a discrete e-fold  $2\pi nH/\omega$ , where  $n$  is an integer. This is why the most important feature of this non-Gaussianity is a running behavior  $\sim \sin(C \ln K + \text{phase})$ , where  $C = \omega/H$ . This leads to the ansatz (148). The full expression is given in (150) and plotted in Figure 8. A numerical result is plotted in Figure 9.

(iv) *Folded Shape and Non-Bunch-Davies Vacuum.* The usual mode function of the Bunch-Davies vacuum has the positive energy mode  $\sim e^{-ik\tau}$ . Now we consider a non-Bunch-Davies vacuum by adding a small component of negative energy mode  $\sim e^{ik\tau}$ . The three-point function involves an integration of the product of three mode functions with momentum  $k_1$ ,  $k_2$  and  $k_3$ . So the leading correction to the Bunch-Davies results is to replace one of the  $k_i$ s with  $-k_i$ . The usual  $K = k_1 + k_2 + k_3$  in  $e^{-iK\tau}$  becomes  $\tilde{K} = k_1 + k_2 - k_3$ , and its cyclic. This effect is most important if factors of  $\tilde{K}$  appear in the denominators after the  $\tau$ -integration. Hence, the most important feature of this type of modification is to enhance the non-Gaussianity in the folded triangle limit. An example of these bispectra is given in (154) and plotted in Figure 10. Ansatz that partially capture this feature are given in (155) and (156), and plotted in Figure 11.

(v) *Intermediate Shapes and Massive Isocurvatons.* All mechanisms discussed so far for single field inflation apply to multifield inflation. We now consider new effects caused by introducing more fields to inflation models. These extra fields are called isocurvatons.

Since light fields typically acquire a mass of order  $H$ , the Hubble parameter, we first consider the quasisingle field inflation models where there is one massless inflaton while the isocurvatons have mass of order  $H$  instead of massless.

Unlike multifield slow-roll inflation, where each flat direction only has small nonlinear terms in order to satisfy the slow-roll conditions, massive directions are not inflationary direction and are free to have large nonlinear self-interactions. These nonlinear interactions can be transferred to the curvature mode through couplings and source the large non-Gaussianity.

The massive isocurvaton eventually decays after horizon exit simply because they are diluted by the expansion. How fast it decays depends on its mass. If the mass is heavier,  $m > \sqrt{2}H$ , it decays faster. So the interactions can only happen when all modes are all closer to the horizon exit. This is closer to the case of the equilateral shape that we encountered above, and results in bispectra with quasiequilateral shapes. If the mass is lighter,  $m < \sqrt{2}H$ , it decays slower. More non-Gaussianity is generated in the superhorizon scales. This is closer to the case of the local shape that we will come to below, and results in bispectra with quasilocal shapes. Overall, at the squeezed limit  $k_3 \ll k_1 = k_2$ , the bispectrum shapes behave as  $(k_3/k_1)^{1/2-\nu}$ , where  $\nu$  goes from 0 to 3/2 (corresponding to  $m$  from  $3H/2$  to 0) in the example we studied. In particular, if we take the massless limit while keeping the cubic self-interactions of isocurvaton large, we get a large bispectrum that has the same squeezed limit shape as the local one. Therefore, we have a one-parameter family of shapes that lie between the local and equilateral shape.

The numerical results of these shapes are presented in Figure 14. A simple ansatz is given in (174) that represents this family of shapes quite well, and is plotted in Figure 15.

(vi) *Local Shape and Massless Isocurvatons.* The fluctuation amplitudes of massless scalars do not decay after the horizon exit, and therefore this opens up a multifield space for the superhorizon evolution. For superhorizon modes, we can use the separate universe picture and study the classical behavior of different patches of universe. These patches are separated by horizons and evolve independently of each other. So the evolution is local in space.

Non-Gaussianities are generated when this multifield evolution is nonlinear, and any nonlinearity arising in the separate universe picture should also be local in space. A locality in position space translates to a nonlocality in momentum space. This is why the resulted local shape bispectrum peaks at the squeezed limit. The behavior is  $(k_3/k_1)^{-1}$  for  $k_3 \ll k_1 = k_2$ . This bispectrum is given in (184) and the shape is plotted in Figure 16.

In all cases, the power spectra are either approximately scale-invariant so indistinguishable from the simplest slow-roll models, or modified with features that can be made small enough to satisfy the current observational constraints.

Large bispectra generically implies large trispectra, that is, the four-point correlation functions. But trispectra contain more information and can be large even if bispectra are small. Experimentally, trispectra are more difficult to detect, but contain much more shape configurations. Each category above should have interesting extensions to trispectra. See [187–193] for the equilateral case and [194–199] for the local case.

It is certain that this list will grow in future works, providing more refined and diverse connections between theories and experiments.

9.2. *A Consistency Condition.* As we have seen, in single field inflation, the mode that has exited the horizon is frozen. This is characterized by a constant  $\zeta$  over a horizon size patch. The physical meaning of the constant  $\zeta$  is a small rescaling of the scale factor. This is the only effect that the superhorizon mode has on modes with much shorter wavelength. This fact is used by Maldacena to derive a consistency condition [47] for the three-point correlation functions in the squeezed limit for single field inflation.

(i) *Consistency Condition.* In the squeezed limit  $k_3 \ll k_1 = k_2$ ,  $k_3$  is the superhorizon mode that exited the horizon and acts as a zero-mode modulation to the two remaining modes. The correlation  $\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \rangle$  is an average of the following quantity

$$\langle \zeta_{k_1} \zeta_{k_2} \rangle_{\zeta_{k_3}} \zeta_{k_3} \quad (192)$$

over different  $\zeta_{k_3}$ . We will shift the average  $\langle \zeta_{k_3} \rangle$  to zero by definition. Here the two-point average  $\langle \zeta_{k_1} \zeta_{k_2} \rangle_{\zeta_{k_3}}$  is evaluated with different local scalings determined by the shift  $\zeta_{k_3}$ . If the two-point function is exactly scale-invariant,  $\langle \zeta_{k_1} \zeta_{k_2} \rangle_{\zeta_{k_3}}$  is just a constant. So the 3 pt vanishes because  $\langle \zeta_{k_3} \rangle = 0$ . The nonzero contribution comes from the breaking of the scale invariance. To see this, we expand the two-point average in terms of a long wavelength mode  $\zeta_{k_4}$  near the scale  $\langle \zeta_{k_4} \rangle = 0$ ,

$$\begin{aligned} \langle \zeta_{k_1} \zeta_{k_2} \rangle_{\zeta_{k_4}} &= \langle \zeta_{k_1} \zeta_{k_2} \rangle_0 + \left. \frac{d \langle \zeta_{k_1} \zeta_{k_2} \rangle}{d \zeta_{k_4}} \right|_0 \zeta_{k_4} \\ &+ \frac{1}{2} \left. \frac{d^2 \langle \zeta_{k_1} \zeta_{k_2} \rangle}{d \zeta_{k_4}^2} \right|_0 \zeta_{k_4}^2 + \dots \end{aligned} \quad (193)$$

Multiply this with  $\zeta_{k_3}$  and average over  $\zeta_{k_3}$ . The first term contributes zero since this is the scale-invariant component. The second term gives

$$\left. \frac{d \langle \zeta_{k_1} \zeta_{k_2} \rangle}{d \zeta_{k_4}} \right|_0 \langle \zeta_{k_3} \zeta_{k_4} \rangle. \quad (194)$$

To get nonzero average,  $\mathbf{k}_3 + \mathbf{k}_4 = 0$  is needed. Using the relation  $d \zeta_{k_4} = -d \ln k$ , we get

$$\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \rangle \longrightarrow - \left. \frac{d \langle \zeta_{k_1} \zeta_{k_2} \rangle}{d \ln k} \right|_0 \langle \zeta_{k_3} \zeta_{k_3} \rangle. \quad (195)$$

The higher-order terms in (193) give

$$\frac{1}{2} \frac{d^2 \langle \zeta_{k_1} \zeta_{k_2} \rangle}{(d \ln k)^2} \Big|_0 \langle \zeta_{k_3} \zeta_{k_4}^2 \rangle - \frac{1}{6} \frac{d^3 \langle \zeta_{k_1} \zeta_{k_2} \rangle}{(d \ln k)^3} \Big|_0 \langle \zeta_{k_3} \zeta_{k_4}^3 \rangle + \dots, \quad (196)$$

where  $\mathbf{k}_3 + 2\mathbf{k}_4 = 0$ ,  $\mathbf{k}_3 + 3\mathbf{k}_4 = 0$ , and so on have to be satisfied, respectively, for each term to get nonzero average. If we only consider the tree-level three-point function, these higher-order terms can be truncated since they involve more factors of  $P_\zeta$  and should be related to the loop diagram contributions to  $\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \rangle$ . The tree diagram is  $\mathcal{O}(P_\zeta^2)$ .

To connect the averages we used here with the correlation functions that we defined in previous sections, we need to restore the phase factors. Here the two-point average  $\langle \zeta_{k_1} \zeta_{k_2} \rangle_{\text{here}}$  is performed with the special point  $k_1 = k_2$  in the phase space. To connect this with the previous definition of  $\langle \zeta_{k_1} \zeta_{k_2} \rangle$ , we need to include the phase space in the neighborhood. Namely,  $\langle \zeta_{k_1} \zeta_{k_2} \rangle = \langle \zeta_{k_1} \zeta_{k_2} \rangle_{\text{here}} (2\pi)^3 \delta^3(\mathbf{k}_1 + \mathbf{k}_2)$ , so from the definition (26) we have  $\langle \zeta_{k_1} \zeta_{k_2} \rangle_{\text{here}} = (2\pi)^2 P_\zeta(k_1)/2k_1^3$ . Similarly,  $\langle \zeta^3 \rangle = \langle \zeta^3 \rangle_{\text{here}} (2\pi)^3 \delta^3(\sum \mathbf{k}_i)$ . With the usual definition of the spectrum index  $n_s - 1 \equiv d \ln P_\zeta / d \ln k$ , from (195), we get the following consistency condition [47]:

$$\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \rangle \rightarrow -(n_s - 1) \frac{1}{4k_1^3 k_3^3} P_\zeta(k_1) P_\zeta(k_3) (2\pi)^7 \delta^3 \left( \sum_i \mathbf{k}_i \right). \quad (197)$$

Although originally derived for slow-roll inflation, the only assumption is the single field. So this applies to any single field inflation models and has important physical implications that we discuss shortly [200]. Note that the derivation of this relation (197) does not rely on the smallness of the slow-variation parameters either. For the general single field inflation models that we studied in Section 6.1, at tree level this has been checked with explicit results to three different orders [50, 201] including the slow-roll limit [47]. For resonance models, this is checked to the leading order [111].

There are three types of interesting corrections to the condition (197).

Firstly, as mentioned, the right-hand side of (197) should receive corrections from loop contributions. These loop contributions are associated with higher derivatives of the two-point function. The terms (196), together with (195), provide the corresponding consistency conditions at different orders of  $P_\zeta$ . Note that for such orders, the correlation functions such as  $\langle \zeta_{k_1} \zeta_{k_2} \rangle$  on the right-hand side should also include loop corrections.

Secondly, when we assume that the only effect of the frozen superhorizon mode on the much shorter scale is a constant background rescaling, we assume that there is no interaction when these modes are all within the horizon (I would like to thank Yi Wang for helpful discussions on this point). However, large subhorizon interaction is possible in some cases, such as in Sections 6.3 and 6.4. Such interactions disappear below a new length scale at subhorizon, only then

the above assumption becomes valid. For example in the resonance model, for  $H/\omega < k_3/k_1 \ll 1$ , the modes  $\mathbf{k}_1$ ,  $\mathbf{k}_2$  and  $\mathbf{k}_3$  are guaranteed to resonant with the oscillatory background at some point when all of them are within the horizon. Only if  $k_3/k_1 \ll H/\omega$ , such resonance will not happen. This is why the consistency condition is satisfied only in the *very* squeezed limit ( $k_3 \ll k_1/C$  with  $C = \omega/H \gg 1$ ). For the squeezed region  $k_1/C < k_3 \ll k_1$ , the left-hand side of the condition is larger than the right-hand side by a factor of  $Ck_3/2k_1$ . For the non-Bunch-Davies vacuum case, a similar scale is determined by the UV cutoff  $\tau_c$ , from which the non-Bunch-Davies vacuum starts to take effect.

Thirdly, even after the long wavelength mode exits the horizon, as long as  $k_3/k_1$  is not infinitely small, there is still dependence of the two-point function on the derivative of this long wavelength mode, in addition to the overall constant shift. This introduces a different type of finite  $k_3/k_1$  corrections. They start from the second-order in  $k_3/k_1$ , because the first-order corresponds to the first spatial derivative of the long wavelength mode and should vanish due to isotropy [200]. These corrections will be amplified by the associated amplitude  $f_{\text{NL}}^{\text{non-loc}}$ , and give an additive correction  $f_{\text{NL}}^{\text{non-loc}}(k_3/k_1)^2$  to the  $n_s - 1$  on the right-hand side of the condition. For a large  $f_{\text{NL}}^{\text{non-loc}}$ , therefore, the condition needs to be satisfied in a *very* squeezed limit. The equilateral bispectra (121) and (122) are this type of examples.

The consistency condition (197) can be straightforwardly generalized to higher-order correlation functions [189, 194]. We emphasize that this condition only applies to single field inflation models. For inflation models involving more than one field, as we have seen, non-Gaussianities can be transferred from the isocurvature directions which do not respect this relation.

(ii) *Physical Implication.* Besides providing consistency checks for analytical computations, the condition also has interesting physical implications. In the following, we discuss the scale invariant cases [200], as well as the feature cases and loop corrections, ending with some cautionary remarks.

This consistency relation implies that the tree-level bispectrum in the squeezed limit is determined by the power spectrum and spectral index. We distinguish the following two cases. For the scale-invariant case,  $n_s - 1$  is of order  $\mathcal{O}(\epsilon)$  and the right-hand side of (197) takes the local form. Indeed, as we have seen, for the single field inflation models where the non-Gaussianities are large, they take the equilateral forms which vanish in the infinitely squeezed limit. For the non-scale-invariant case, especially the highly oscillatory case such as the resonance model, the power spectrum can be highly oscillatory and  $n_s - 1$  becomes large. This can still be consistent with observations since the large  $n_s - 1$  is also highly oscillatory and therefore may escape a detection so far. But such a running non-Gaussianity is orthogonal to the scale-invariant forms.

For the loop diagrams, in the scale-invariant case, these terms are suppressed by higher-orders of slow-variation parameters from, for example,  $d^2 \langle \zeta_{k_1} \zeta_{k_2} \rangle / (d \ln k)^2$ , and

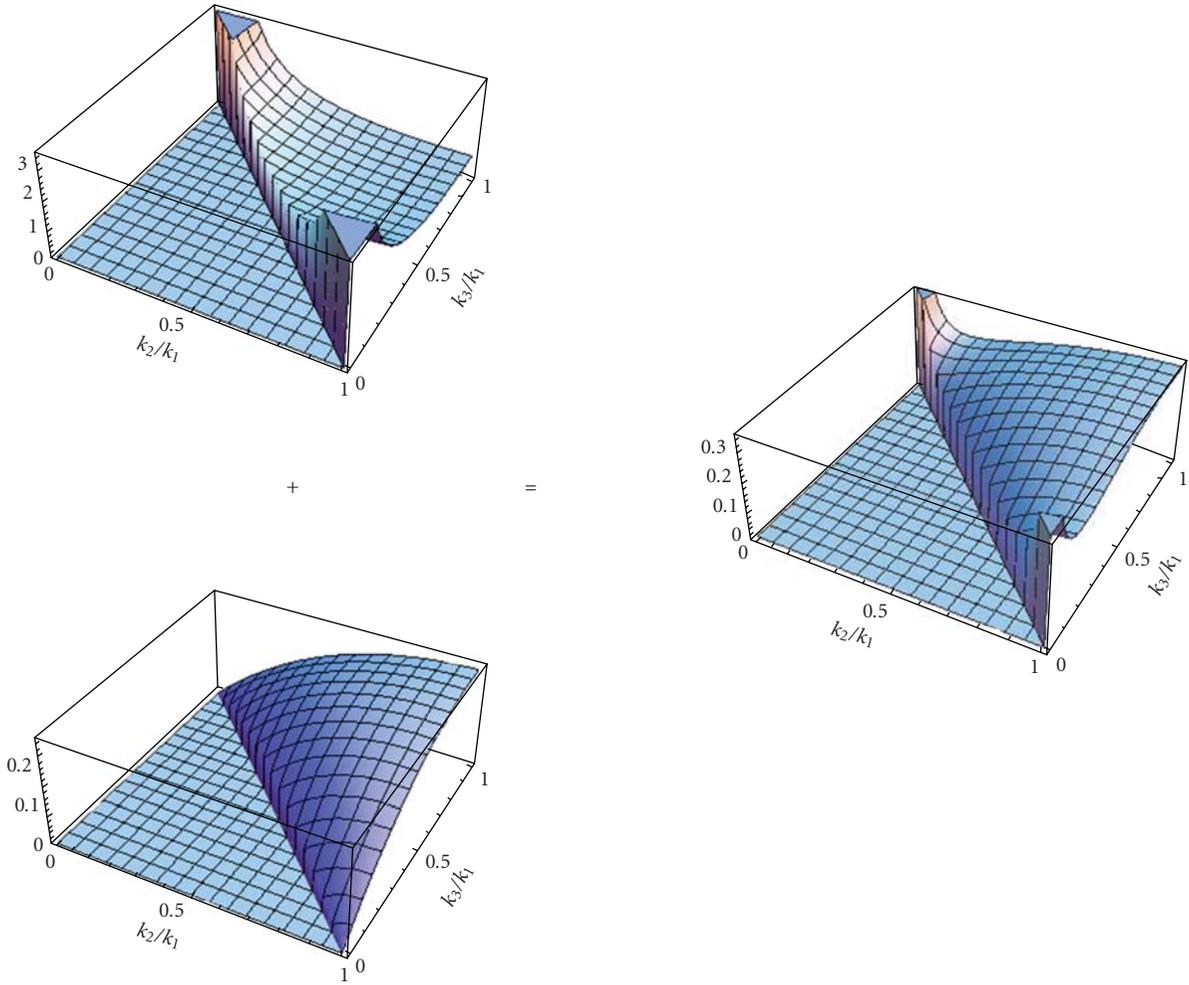


FIGURE 17: A mixing of the equilateral (Figure 4) and local shape (Figure 16).

higher-orders of  $\zeta$  from, for example,  $\langle \zeta_{k_3}^3 \rangle$ ; in the non-scale-invariant example, the extra terms are still highly oscillatory.

In summary, a detection of an approximately scale-invariant local non-Gaussianity in the infinitely squeezed triangle limit with  $f_{\text{NL}}^{\text{loc}} > \mathcal{O}(\epsilon)$  can rule out all single field inflation models.

In experiments, however, the triangle cannot be perfectly squeezed. So it is an important question how squeezed it should be to achieve the above goal. For example, in the third type of corrections we discussed previously in this subsection, we need  $f_{\text{NL}}^{\text{non-loc}}(k_3/k_1)^2$  to be smaller than  $n_s - 1$  for the consistency condition to hold, so that the contaminations from whatever nonlocal  $f_{\text{NL}}$  is small. Assuming the primordial local form is practically detectable only if  $f_{\text{NL}}^{\text{loc}} > \mathcal{O}(1)$ , we at least need  $f_{\text{NL}}^{\text{non-loc}}(k_3/k_1)^2 < \mathcal{O}(1)$ . For the class of the general single field models we studied in Section 6, if the other forms of non-Gaussianities, such as the equilateral one, can be constrained below  $f_{\text{NL}}^{\text{non-loc}} \sim \mathcal{O}(10)$ , a squeezed configuration with  $k_3/k_1 < 0.1$  will be enough for our purpose. However, a completely model-independent statement is much trickier, because there may be bispectra

with very large amplitude but orthogonal to any known bispectra that have been constrained experimentally. Besides that, in the second type of corrections, large finite- $k_3/k_1$  corrections can also arise due to subhorizon interactions. Therefore, as a cautionary remark, if we would like to rule out all single field inflation models in a rigorous model-independent fashion with a detection of scale-invariant local non-Gaussianity, we have to keep in mind the caveat that there may be single field models which only respect the consistency condition in a very squeezed region beyond the experimental reach.

**9.3. Superpositions.** Different shapes and runnings of non-Gaussianities can be superimposed in inflation models. For example, consider the following.

(i) *Mixing Shapes.* It is possible that different non-Gaussianity generation mechanisms are from different components in a model, or at different stages during inflation. So two or more different shapes can get mixed, and the final shape can be rather different. For example, in Figure 17, we plot a mixed shape between the local and equilateral shape.

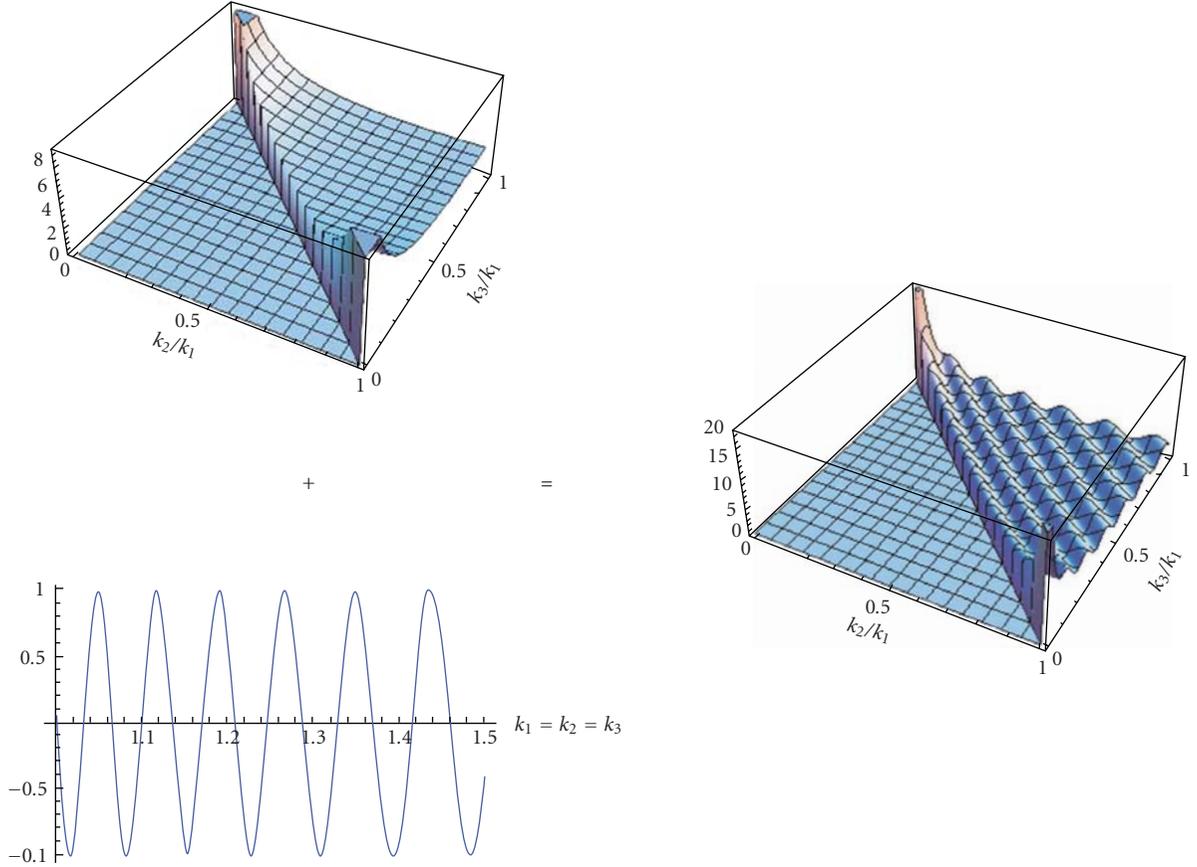


FIGURE 18: A mixing of an intermediate shape [ $\nu = 7/6$  in (174)] and a resonant running (148).

Notice that this is different from the intermediate shapes, since obviously the squeezed limit is always dominated by the local form. Examples of such models are discussed in [202].

(ii) *Mixing Shape and Running.* The shapes can also be mixed with runnings. Same as the power spectrum, the non-Gaussianities generically have some mild scale dependence. But a more dramatic case is the superposition with a strong running, such as the sinusoidal or resonant running. For example, an inflaton passing through features frequently and turning constantly at the same time on a potential landscape can generate a bispectrum which is a superposition of the resonant running and intermediate shape, as we illustrate in Figure 18. Clearly, these two signals are orthogonal to each other very well, and have to be picked up separately through different methods in data analyses.

(iii) *Orthogonalization.* If a non-Gaussianity is the linear superposition of several base components, one can generally perform a change of bases to make the new bases orthogonalized. For example, as we have seen in Section 6.1, the leading large bispectrum has two components,  $S_\lambda$  and  $S_c$ . The two shapes are very similar, and represented by the equilateral ansatz in data analyses. However since they do have small difference, one can subtract their similarities

and get a new orthogonalized base component [175]. The orthogonalization is defined by the shape correlator such as (106). Using this definition, the new bases can be chosen as

$$S_1 \approx S_\lambda + 0.22S_c \quad (198)$$

and  $S_2 = S_c$ . [Note that the  $S_\lambda$  and  $S_c$  used here do not include the prefactors  $(1/c_s^2 - 1 - 2\lambda/\Sigma)$  and  $(1/c_s^2 - 1)$  in (121) and (122)]. Their shapes are shown in Figure 19. Notice that  $S_1$  is half positive and half negative. Because  $S_1$  is not of the simplest factorizable type, the following simple ansatz has been proposed to represent  $S_1$  in data analyses [175],

$$S_{\text{ansatz}}^{\text{orth}} = -18 \left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right) + 18 \left( \frac{k_1}{k_2} + 5 \text{ perm.} \right) - 48. \quad (199)$$

We plot the shape of this ansatz in Figure 20. The current CMB constraint on this orthogonal ansatz is  $-410 < f_{\text{NL}}^{\text{orth}} < 6$  [1].

For known examples of general single field inflation, such as the DBI and k-inflation, we generically get equilateral shapes. This is also clear from their physical origin that we have emphasized. The orthogonal shape relies on a delicate cancellation between the two generic shapes. In principle, one can do this since the required parameter space is allowed in our effective field theory of general single field inflation in

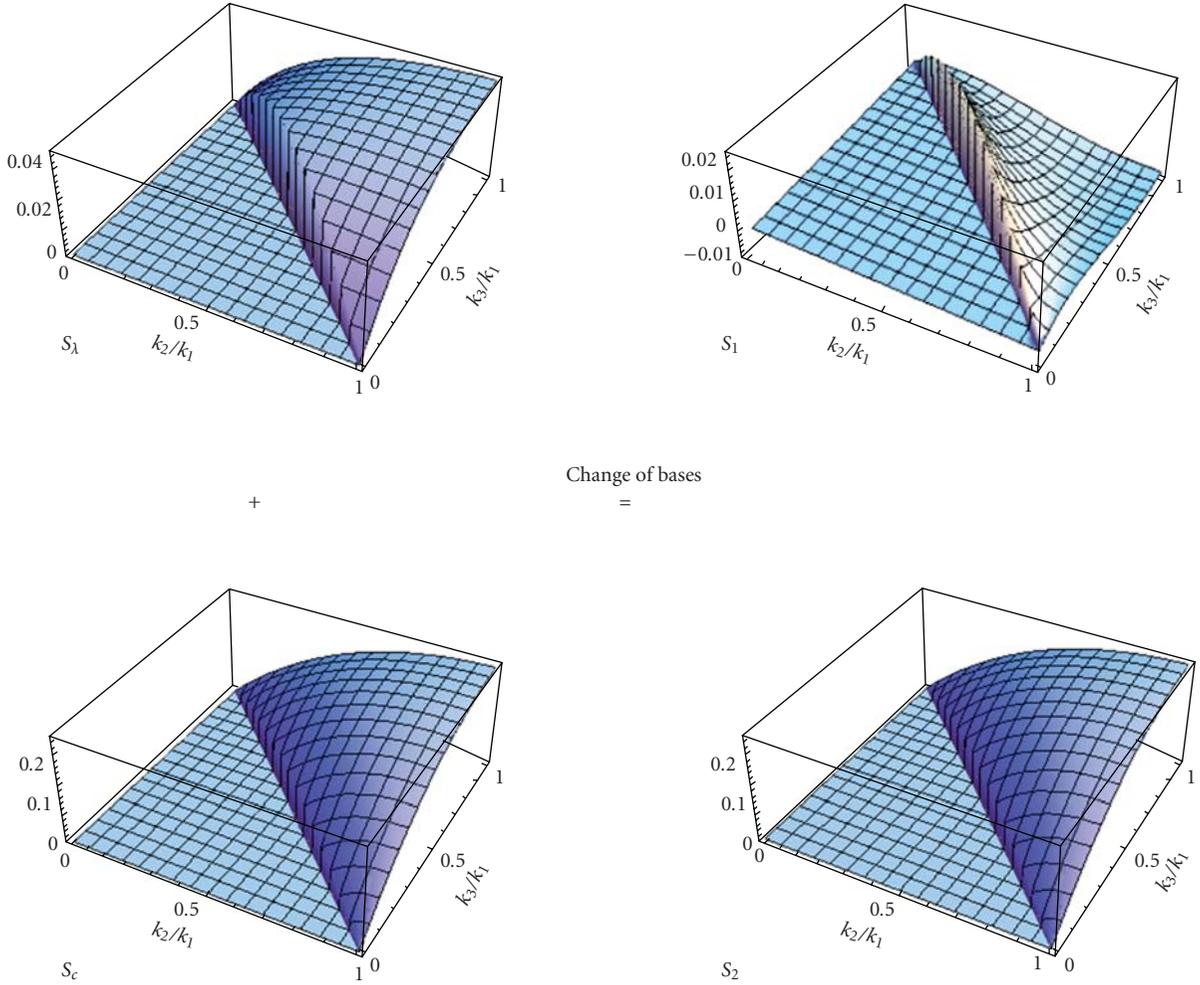


FIGURE 19: Orthogonalization of two shapes in Section 6.1 (Figures 3 and 4) through a change of bases,  $c_\lambda S_\lambda + c_c S_c = c_1 S_1 + c_2 S_2$ .

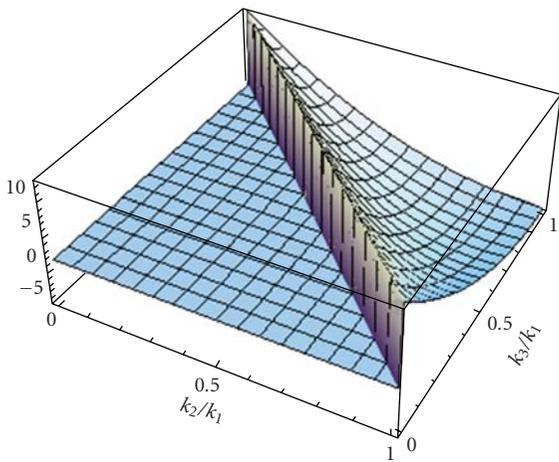


FIGURE 20: An ansatz  $-S_{\text{ansatz}}^{\text{orth}}$  in (199) for the orthogonal shape  $S_1$  in (198). Note we added a minus sign in this plot.

Section 6.1, and this may provide guidance to future model building. For example, one may fine-tune the parameters

in the k-inflation models [93, 94]. Therefore, unlike the previous cases, the direct motivation here is more oriented to data analyses. The advantage of this operation is that it makes full use of data, which impose constraints on both components. In addition, as a bonus, the ansatz for the equilateral (127), folded (155) and orthogonal (199) shapes are not linearly independent. As we can see, they all happen to be the equilateral ansatz shifted by a constant shape ansatz ( $S = \text{const.}$ ) [65]. Constraining two orthogonal bases provide efficient constraints on all three of them.

Let us do a more data-analysis-oriented exercise. We would like to construct an ansatz that is orthogonal to both local and equilateral ansatz, since both were well constrained by data. (Note that  $S_{\text{ansatz}}^{\text{orth}}$  in (199) is not quite orthogonal to the local ansatz, with a correlation  $\sim -0.48$ .) To do this we start with a trial shape  $S_{\text{trial}}$ , and demand the new orthogonal shape

$$S_{\text{ansatz}}^{\text{orth},2} = S_{\text{trial}} + c_1 S^{\text{loc}} + c_2 S_{\text{ansatz}}^{\text{eq}} \quad (200)$$

be orthogonal to both the local and equilateral ansatz,

$$S_{\text{ansatz}}^{\text{orth},2} \cdot S^{\text{loc}} = 0, \quad S_{\text{ansatz}}^{\text{orth},2} \cdot S_{\text{ansatz}}^{\text{eq}} = 0. \quad (201)$$

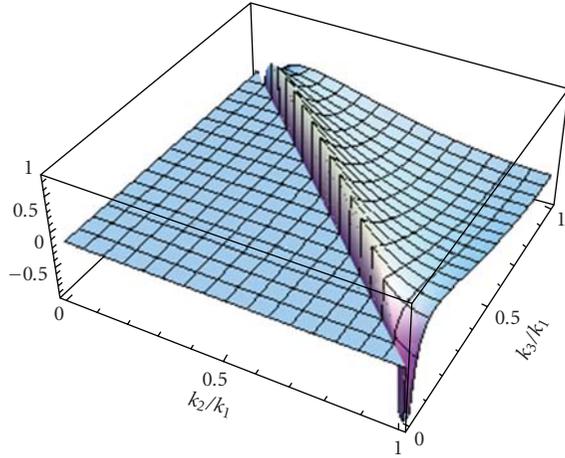


FIGURE 21: Another factorizable orthogonal ansatz (202).

The simplest factorizable trial shapes can be either the constant shape or the local-like shape  $k_1/k_2 + 5$  perm., and both give the same result. Let us use the constant shape  $S_{\text{trial}} = 1$  as an example. Solving the conditions (201) gives  $c_1 = -0.0953$  and  $c_2 = -0.204$ . So the new orthogonal ansatz is

$$S_{\text{ansatz}}^{\text{orth},2} = 1.19 \left( \frac{k_1^2}{k_2 k_3} + 2 \text{ perm.} \right) - 1.22 \left( \frac{k_1}{k_2} + 5 \text{ perm.} \right) + 3.44. \quad (202)$$

The numerical details may change slightly depending on the detailed definition and computation of the inner product (105). The shape is plotted in Figure 21. It is somewhat exotic but the ansatz is simple. By construction, this ansatz is much more orthogonal to the local form (with correlation  $\sim 0$ ) than the  $S_{\text{ansatz}}^{\text{orth}}$  currently used in [1, 175]. It also happens to have reasonably large correlation ( $\sim 0.86$ ) with the orthogonal shape in single field inflation (the  $S_1$  shown in Figure 19), similar to that ( $\sim -0.91$ ) between  $S_1$  and  $S_{\text{ansatz}}^{\text{orth}}$ . Obviously, other choices of trial shapes can result in more exotic orthogonal shapes.

One can perform a similar orthogonalization for the two shapes in (94), now they are both local to start with. More generally, if a non-Gaussianity has more base components, we can orthonormalize all of them one by one, in the sense of the Gram-Schmidt process.

**9.4. Conclusion.** The field of primordial non-Gaussianity is growing rapidly in recent years, with simultaneous progress from the experimental results, data analyses methods, nonlinear cosmology theories, physical model buildings, computational techniques, and theoretical formalisms. The progress that we have seen so far is no doubt just a beginning.

In this paper, we have studied the primordial non-Gaussianities coming from the inflation models, especially various mechanisms that can produce observable large non-Gaussianities with viable power spectra. We emphasized the

fingerprints that different underlying physics leave on non-Gaussian profiles, which break the degeneracy of model building. We described the physical pictures and presented their effective Lagrangians to the extent that they can be recognized when encountered in the inflation model building in a more fundamental theory. We also derived the resulting bispectra and represented them in terms of simple ansatz to the extent that they can be useful to data analyses. With the current rapid progress, we anticipate much more future developments along these lines through refinements and discoveries in both theories and experiments.

The standard model of cosmology—the Big Bang theory with  $\Lambda$ CDM—is now established better than ever, with the precision data coming from the cosmic microwave background and large scale structures. New data will continue to flow from many ongoing and forthcoming experiments. Although Nature does not seem to be obligated to provide us any more information beyond the standard model, exciting possibilities exist that would help us to understand the origin of the Big Bang. These include the more detailed deviations from the scale-invariance of the power spectrum, the primordial gravitational waves that we may detect from the CMB polarization, the isocurvature perturbations between matter and radiation, and the primordial non-Gaussianities. Without these types of data, the number of theoretical models with degenerate observational consequences proliferate with time and it will be hard to understand the microscopic nature of the inflation beyond our current knowledge, as well as to distinguish inflation from other possible alternatives. As we have reviewed, primordial non-Gaussianity—the collider in the very early universe—is one of the few hopes. It is becoming a target of many modern experiments. We do not know which cards Nature is hiding from us, but we are hoping and preparing for the best.

## Acknowledgments

The author would like to thank Rachel Bean, Richard Easther, Girma Hailu, Bin Hu, Min-xin Huang, Shamit Kachru, Eugene Lim, Hiranya Peiris, Sash Sarangi, Gary Shiu, Henry Tye, Yi Wang, and Jiajun Xu for valuable collaborations and sharing their insights on the works reviewed here. The author would also like to thank James Fergusson, Michele Liguori, David Lyth, David Seery, Paul Shellard, and David Wands for very helpful discussions. The author was supported by the Stephen Hawking advanced fellowship.

## References

- [1] E. Komatsu, et al., “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation,” <http://arxiv.org/abs/1001.4538>.
- [2] B. Gold, et al., “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Galactic Foreground Emission,” <http://arxiv.org/abs/1001.4555>.
- [3] D. Larson, et al., “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Power Spectra and WMAP-Derived Parameters,” <http://arxiv.org/abs/1001.4635>.

- [4] J. L. Weiland, et al., “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Planets and Celestial Calibration Sources,” <http://arxiv.org/abs/1001.4731>.
- [5] N. Jarosik, et al., “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results,” <http://arxiv.org/abs/1001.4744>.
- [6] C. L. Bennett, et al., “Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Are There Cosmic Microwave Background Anomalies?” <http://arxiv.org/abs/1001.4758>.
- [7] A. H. Guth, “Inflationary universe: a possible solution to the horizon and flatness problems,” *Physical Review D*, vol. 23, no. 2, pp. 347–356, 1981.
- [8] A. D. Linde, “A new inflationary universe scenario: a possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems,” *Physics Letters B*, vol. 108, no. 6, pp. 389–393, 1982.
- [9] A. Albrecht and P. J. Steinhardt, “Cosmology for grand unified theories with radiatively induced symmetry breaking,” *Physical Review Letters*, vol. 48, no. 17, pp. 1220–1223, 1982.
- [10] V. F. Mukhanov and G. V. Chibisov, “Quantum fluctuation and nonsingular Universe,” *JETP Letters*, vol. 33, pp. 532–535, 1981 (Russian).
- [11] S. W. Hawking, “The development of irregularities in a single bubble inflationary universe,” *Physics Letters B*, vol. 115, no. 4, pp. 295–297, 1982.
- [12] A. A. Starobinsky, “Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations,” *Physics Letters B*, vol. 117, no. 3–4, pp. 175–178, 1982.
- [13] A. H. Guth and S.-Y. Pi, “Fluctuations in the new inflationary universe,” *Physical Review Letters*, vol. 49, no. 15, pp. 1110–1113, 1982.
- [14] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, “Spontaneous creation of almost scale-free density perturbations in an inflationary universe,” *Physical Review D*, vol. 28, no. 4, pp. 679–693, 1983.
- [15] E. W. Kolb and M. S. Turner, *The Early Universe*, Addison-Wesley, Redwood city, Calif, USA, 1988.
- [16] A. Linde, “Axions in inflationary cosmology,” *Physics Letters B*, vol. 259, no. 1–2, pp. 38–47, 1991.
- [17] A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large-Scale Structure*, Cambridge University Press, Cambridge, UK, 2000.
- [18] S. Dodelson, *Modern Cosmology*, Academic Press, Amsterdam, The Netherlands, 2003.
- [19] V. Mukhanov, *Physical Foundations of Cosmology*, Cambridge University Press, Cambridge, UK, 2005.
- [20] S. Weinberg, *Cosmology*, Oxford University Press, Oxford, UK, 2008.
- [21] D. Wands, “Multiple field inflation,” *Lecture Notes in Physics*, vol. 738, pp. 275–304, 2008.
- [22] K. A. Malik and D. Wands, “Cosmological perturbations,” *Physics Reports*, vol. 475, no. 1–4, pp. 1–51, 2009.
- [23] W. H. Kinney, “TASI lectures on inflation,” <http://arxiv.org/abs/0902.1529>.
- [24] D. Baumann, “TASI Lectures on Inflation,” <http://arxiv.org/abs/0907.5424>.
- [25] D. Langlois, “Lectures on inflation and cosmological perturbations,” <http://arxiv.org/abs/1001.5259>.
- [26] D. H. Lyth and A. Riotto, “Particle physics models of inflation and the cosmological density perturbation,” *Physics Report*, vol. 314, no. 1–2, pp. 1–146, 1999.
- [27] F. Quevedo, “Lectures on string/brane cosmology,” *Classical and Quantum Gravity*, vol. 19, no. 22, pp. 5721–5779, 2002.
- [28] S.-H. H. Tye, “Brane inflation: string theory viewed from the cosmos,” *Lecture Notes in Physics*, vol. 737, pp. 949–974, 2008.
- [29] J. M. Cline, “String cosmology,” <http://arxiv.org/abs/hep-th/0612129>.
- [30] C. P. Burgess, “Lectures on cosmic inflation and its potential stringy realizations,” *Classical and Quantum Gravity*, vol. 24, no. 21, pp. S795–S852, 2007.
- [31] L. McAllister and E. Silverstein, “String cosmology: a review,” *General Relativity and Gravitation*, vol. 40, no. 2–3, pp. 565–605, 2008.
- [32] A. Mazumdar and J. Rocher, “Particle physics models of inflation and curvaton scenarios,” <http://arxiv.org/abs/1001.0993>.
- [33] E. Komatsu, “The pursuit of non-Gaussian fluctuations in the cosmic microwave background,” <http://arxiv.org/abs/astro-ph/0206039>.
- [34] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, “Non-Gaussianity from inflation: theory and observations,” *Physics Reports*, vol. 402, no. 3–4, pp. 103–266, 2004.
- [35] M. Liguori, E. Sefusatti, J. R. Fergusson, and E. P. S. Shellard, “Primordial non-Gaussianity and bispectrum measurements in the cosmic microwave background and large-scale structure,” <http://arxiv.org/abs/1001.4707>.
- [36] L. Verde, “Non-Gaussianity from large-scale structure surveys,” *Advances in Astronomy*, vol. 2010, Article ID 768675, 15 pages, 2010.
- [37] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity of large-scale cosmic microwave background anisotropies beyond perturbation theory,” *Journal of Cosmology and Astroparticle Physics*, no. 8, pp. 177–196, 2005.
- [38] J. L. Lehners, “Ekpyrotic non-Gaussianity—a review,” <http://arxiv.org/abs/1001.3125>.
- [39] R. K. Sachs and A. M. Wolfe, “Perturbations of a cosmological model and angular variations of the microwave background,” *Astrophysical Journal*, vol. 147, p. 73, 1967.
- [40] J. Schwinger, “The special canonical group,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 46, no. 10, pp. 1401–1415, 1960.
- [41] J. Schwinger, “Brownian motion of a quantum oscillator,” *Journal of Mathematical Physics*, vol. 2, no. 3, pp. 407–432, 1961.
- [42] P. M. Bakshi and K. T. Mahanthappa, “Expectation value formalism in quantum field theory. I,” *Journal of Mathematical Physics*, vol. 4, no. 1, pp. 1–11, 1963.
- [43] L. V. Keldysh, “Diagram technique for nonequilibrium processes,” *JETP Letters*, vol. 20, p. 1018, 1965.
- [44] E. Calzetta and B. L. Hu, “Closed-time-path functional formalism in curved spacetime: application to cosmological back-reaction problems,” *Physical Review D*, vol. 35, no. 2, pp. 495–509, 1987.
- [45] S. Weinberg, “Quantum contributions to cosmological correlations,” *Physical Review D*, vol. 72, no. 4, Article ID 043514, pp. 1–19, 2005.
- [46] X. Chen and Y. Wang, “Quasi-single field inflation and non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 4, article 027, 2010.
- [47] J. Maldacena, “Non-gaussian features of primordial fluctuations in single field inflationary models,” *Journal of High Energy Physics*, vol. 7, no. 5, pp. 233–264, 2003.
- [48] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant second-order perturbations and non-Gaussianity from inflation,” *Nuclear Physics B*, vol. 667, no. 1–2, pp. 119–148, 2003.

- [49] D. Seery and J. E. Lidsey, "Primordial non-Gaussianities in single-field inflation," *Journal of Cosmology and Astroparticle Physics*, no. 6, pp. 45–70, 2005.
- [50] X. Chen, M.-X. Huang, S. Kachru, and G. Shiu, "Observational signatures and non-Gaussianities of general single-field inflation," *Journal of Cosmology and Astroparticle Physics*, no. 1, article 002, 2007.
- [51] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, "Theory of cosmological perturbations," *Physics Report*, vol. 215, no. 5-6, pp. 203–333, 1992.
- [52] E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, "False vacuum inflation with Einstein gravity," *Physical Review D*, vol. 49, no. 12, pp. 6410–6433, 1994.
- [53] X. Chen, "Fine-tuning in DBI inflationary mechanism," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 12, 2008.
- [54] E. Silverstein and D. Tong, "Scalar speed limits and cosmology: acceleration from D-cceleration," *Physical Review D*, vol. 70, no. 10, Article ID 103505, 18 pages, 2004.
- [55] L. Randall and R. Sundrum, "Large mass hierarchy from a small extra dimension," *Physical Review Letters*, vol. 83, no. 17, pp. 3370–3373, 1999.
- [56] S. B. Giddings, S. Kachru, and J. Polchinski, "Hierarchies from fluxes in string compactifications," *Physical Review D*, vol. 66, no. 10, Article ID 106006, 2002.
- [57] X. Chen, S. Sarangi, S.-H. H. Tye, and J. Xu, "Is brane inflation eternal?" *Journal of Cosmology and Astroparticle Physics*, no. 11, article 015, 2006.
- [58] D. Baumann and L. McAllister, "A microscopic limit on gravitational waves from D-brane inflation," *Physical Review D*, vol. 75, no. 12, Article ID 123508, 2007.
- [59] D. H. Lyth, "What would we learn by detecting a gravitational wave signal in the cosmic microwavebackground anisotropy?" *Physical Review Letters*, vol. 78, no. 10, pp. 1861–1863, 1997.
- [60] D. Babich, P. Creminelli, and M. Zaldarriaga, "The shape of non-Gaussianities," *Journal of Cosmology and Astroparticle Physics*, no. 8, pp. 199–217, 2004.
- [61] X. Chen, "Running non-Gaussianities in Dirac-Born-Infeld inflation," *Physical Review D*, vol. 72, no. 12, Article ID 123518, 7 pages, 2005.
- [62] C. T. Byrnes, S. Nurmi, G. Tasinato, and D. Wands, "Scale dependence of local  $f_{NL}$ ," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, article 034, 2010.
- [63] X. Chen, R. Easther, and E. A. Lim, "Large non-Gaussianities in single-field inflation," *Journal of Cosmology and Astroparticle Physics*, no. 6, article no. 023, 2007.
- [64] X. Chen, R. Easther, and E. A. Lim, "Generation and characterization of large non-Gaussianities in single field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, article 010, 2008.
- [65] J. R. Fergusson and E. P. S. Shellard, "Shape of primordial non-Gaussianity and the CMB bispectrum," *Physical Review D*, vol. 80, no. 4, Article ID 043510, 26 pages, 2009.
- [66] E. Komatsu and D. N. Spergel, "Acoustic signatures in the primary microwave background bispectrum," *Physical Review D*, vol. 63, no. 6, Article ID 063002, 2001.
- [67] E. Komatsu, D. N. Spergel, and B. D. Wandelt, "Measuring primordial non-Gaussianity in the cosmic microwave background," *Astrophysical Journal*, vol. 634, no. 1, pp. 14–19, 2005.
- [68] P. Creminelli, L. Senatore, and M. Zaldarriaga, "Estimators for local non-Gaussianities," *Journal of Cosmology and Astroparticle Physics*, no. 3, article no. 019, 2007.
- [69] K. M. Smith and M. Zaldarriaga, "Algorithms for bispectra: forecasting, optimal analysis, and simulation," <http://arxiv.org/abs/astro-ph/0612571>.
- [70] A. P. S. Yadav, E. Komatsu, B. D. Wandelt, M. Liguori, F. K. Hansen, and S. Matarrese, "Fast estimator of primordial non-Gaussianity from temperature and polarization anisotropies in the cosmic microwave background. II. Partial sky coverage and inhomogeneous noise," *Astrophysical Journal*, vol. 678, no. 2, pp. 578–582, 2008.
- [71] J. R. Fergusson, M. Liguori, and E. P. S. Shellard, "General CMB and primordial bispectrum estimation: mode expansion, map-making and measures of  $f_{NL}$ ," <http://arxiv.org/abs/0912.5516>.
- [72] J. Garriga and V. F. Mukhanov, "Perturbations in k-inflation," *Physics Letters B*, vol. 458, no. 2-3, pp. 219–225, 1999.
- [73] M. Alishahiha, E. Silverstein, and D. Tong, "DBI in the sky: non-Gaussianity from inflation with a speed limit," *Physical Review D*, vol. 70, no. 12, Article ID 123505, 2004.
- [74] A. Gruzinov, "Consistency relation for single scalar inflation," *Physical Review D*, vol. 71, no. 2, pp. 027301–2, 2005.
- [75] C. Cheung, A. L. Fitzpatrick, J. Kaplan, L. Senatore, and P. Creminelli, "The effective field theory of inflation," *Journal of High Energy Physics*, vol. 2008, no. 3, article 014, 2008.
- [76] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, "Ghost inflation," *Journal of Cosmology and Astroparticle Physics*, no. 4, pp. 1–18, 2004.
- [77] S. Weinberg, "Effective field theory for inflation," *Physical Review D*, vol. 77, no. 12, Article ID 123541, 2008.
- [78] F. J. de Urries and J. Julve, "Ostrogradski formalism for higher-derivative scalar field theories," *Journal of Physics A*, vol. 31, no. 33, pp. 6949–6964, 1998.
- [79] R. Woodard, "Avoiding dark energy with 1/R modifications of gravity," *Lecture Notes in Physics*, vol. 720, pp. 403–433, 2007.
- [80] P. Creminelli, "On non-Gaussianities in single-field inflation," *Journal of Cosmology and Astroparticle Physics*, no. 10, pp. 53–62, 2003.
- [81] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, "Limits on non-Gaussianities from WMAP data," *Journal of Cosmology and Astroparticle Physics*, no. 5, article no. 004, 2006.
- [82] B. Chen, Y. Wang, and W. Xue, "Inflationary non-Gaussianity from thermal fluctuations," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 5, article 014, 2008.
- [83] D. Green, B. Horn, L. Senatore, and E. Silverstein, "Trapped inflation," *Physical Review D*, vol. 80, no. 6, Article ID 063533, 2009.
- [84] I. G. Moss and C. Xiong, "Non-Gaussianity in fluctuations from warm inflation," *Journal of Cosmology and Astroparticle Physics*, no. 4, article no. 007, 2007.
- [85] X. Chen, "Multithroat brane inflation," *Physical Review D*, vol. 71, no. 6, Article ID 063506, 5 pages, 2005.
- [86] X. Chen, "Inflation from warped space," *Journal of High Energy Physics*, no. 8, pp. 1197–1231, 2005.
- [87] S. E. Shandera and S.-H. H. Tye, "Observing brane inflation," *Journal of Cosmology and Astroparticle Physics*, no. 5, article no. 007, 2006.
- [88] S. Kecksmeti, J. Maiden, G. Shiu, and B. Underwood, "DBI inflation in the tip region of a warped throat," *Journal of High Energy Physics*, vol. 2006, no. 9, article 076, 2006.
- [89] G. Shiu and B. Underwood, "Observing the geometry of warped compactification via cosmic inflation," *Physical Review Letters*, vol. 98, no. 5, Article ID 051301, 2007.

- [90] S. Thomas and J. Ward, "IR inflation from multiple branes," *Physical Review D*, vol. 76, no. 2, Article ID 023509, 2007.
- [91] R. Bean, S. E. Shandera, S.-H. Henry Tye, and J. Xu, "Comparing brane inflation to WMAP," *Journal of Cosmology and Astroparticle Physics*, no. 5, article no. 004, 2007.
- [92] R. Bean, X. Chen, H. Peiris, and J. Xu, "Comparing infrared Dirac-Born-Infeld brane inflation to observations," *Physical Review D*, vol. 77, no. 2, Article ID 023527, 28 pages, 2008.
- [93] C. Armendáriz-Picón, T. Damour, and V. Mukhanov, "k-inflation," *Physics Letters B*, vol. 458, no. 2-3, pp. 209–218, 1999.
- [94] M. Li, T. Wang, and Y. Wang, "General single field inflation with large positive non-Gaussianity," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 3, article 028, 2008.
- [95] A. J. Tolley and M. Wyman, "Equilateral non-Gaussianity from multifield dynamics," *Physical Review D*, vol. 81, no. 4, Article ID 043502, 5 pages, 2010.
- [96] D. Langlois, S. Renaux-Petel, D. A. Steer, and T. Tanaka, "Primordial fluctuations and non-gaussianities in multifield Dirac-Born-Infeld inflation," *Physical Review Letters*, vol. 101, no. 6, Article ID 061301, 2008.
- [97] D. Langlois, S. Renaux-Petel, D. A. Steer, and T. Tanaka, "Primordial perturbations and non-Gaussianities in DBI and general multi-field inflation," *Physical Review D*, vol. 78, Article ID 063523, 15 pages, 2008.
- [98] F. Arroja, S. Mizuno, and K. Koyama, "Non-Gaussianity from the bispectrum in general multiple field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 015, 2008.
- [99] D. Langlois, S. Renaux-Petel, and D. A. Steer, "Multi-field DBI inflation: introducing bulk forms and revisiting the gravitational wave constraints," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 4, article 021, 2009.
- [100] S. Mizuno, F. Arroja, K. Koyama, and T. Tanaka, "Lorentz boost and non-Gaussianity in multifield DBI inflation," *Physical Review D*, vol. 80, no. 2, Article ID 023530, 10 pages, 2009.
- [101] Y.-F. Cai and H.-Y. Xia, "Inflation with multiple sound speeds: a model of multiple DBI type actions and non-Gaussianities," *Physics Letters B*, vol. 677, no. 5, pp. 226–234, 2009.
- [102] X. Gao, "On cross-correlations between curvature and isocurvature perturbations during inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 2, 2010.
- [103] A. A. Starobinsky, "Spectrum of adiabatic perturbations in the Universe when there are singularities in the inflation potential," *JETP Letters*, vol. 55, p. 489, 1992.
- [104] J. Adams, B. Cresswell, and R. Easther, "Inflationary perturbations from a potential with a step," *Physical Review D*, vol. 64, no. 12, Article ID 123514, 2001.
- [105] J. Hamann, L. Covi, A. Melchiorri, and A. Slosar, "New constraints on oscillations in the primordial spectrum of inflationary perturbations," *Physical Review D*, vol. 76, no. 2, Article ID 023503, 14 pages, 2007.
- [106] M. Joy, V. Sahni, and A. A. Starobinsky, "New universal local feature in the inflationary perturbation spectrum," *Physical Review D*, vol. 77, no. 2, Article ID 023514, 10 pages, 2008.
- [107] R. Bean, X. Chen, G. Hailu, S.-H. Henry Tye, and J. Xu, "Duality cascade in brane inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 3, article 026, 2008.
- [108] L. A. Kofman, "Primordial perturbations from inflation," *Physica Scripta*, vol. T36, article 108, 1991.
- [109] L. Wang and M. Kamionkowski, "Cosmic microwave background bispectrum and inflation," *Physical Review D*, vol. 61, no. 6, Article ID 063504, pp. 1–6, 2000.
- [110] E. Komatsu, A. Kogut, M. R. Nolte et al., "First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: tests of Gaussianity," *Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 119–134, 2003.
- [111] R. Flauger and E. Pajer, "Resonant Non-Gaussianity," <http://arxiv.org/abs/1002.0833>.
- [112] S. Hannestad, T. Haugbølle, P. R. Jarnhus, and M. S. Sloth, "Non-gaussianity from axion monodromy inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 6, 2010.
- [113] S. Hannestad, T. Haugbølle, P. R. Jarnhus, and M. S. Sloth, "Non-gaussianity from axion monodromy inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 6, 2010.
- [114] G. Hailu and S.-H. Henry Tye, "Structures in the gauge/gravity duality cascade," *Journal of High Energy Physics*, vol. 2007, no. 8, 2007.
- [115] E. Silverstein and A. Westphal, "Monodromy in the CMB: gravity waves and string inflation," *Physical Review D*, vol. 78, no. 10, Article ID 106003, 21 pages, 2008.
- [116] L. McAllister, E. Silverstein, and A. Westphal, "Gravity waves and linear inflation from Axion Monodromy," <http://arxiv.org/abs/0808.0706>.
- [117] J. Martin, "Trans-planckian problem of inflationary cosmology," *Physical Review D*, vol. 63, no. 12, Article ID 123501, 16 pages, 2001.
- [118] U. H. Danielsson, "Note on inflation and trans-Planckian physics," *Physical Review D*, vol. 66, no. 2, Article ID 023511, 5 pages, 2002.
- [119] R. H. Brandenberger, "Theory of cosmological perturbations and applications to superstring cosmology," in *Proceedings of the NATO Advanced Study Institute on String Theory: From Gauge Interactions to Cosmology Cargèse*, pp. 79–116, June 2004.
- [120] B. Greene, K. Schalm, J. P. van der Schaar, and G. Shiu, "Extracting new physics from the CMB," in *Proceedings of 22nd Texas Symposium on Relativistic Astrophysics*, Stanford, Calif, USA, December 2004.
- [121] P. D. Meerburg, J. P. Van Der Schaar, and M. G. Jackson, "Bispectrum signatures of a modified vacuum in single field inflation with a small speed of sound," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 2, 2010.
- [122] R. Holman and A. J. Tolley, "Enhanced non-Gaussianity from excited initial states," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 5, article 001, 2008.
- [123] P. D. Meerburg, J. P. van der Schaar, and P. S. Corasaniti, "Signatures of initial state modifications on bispectrum statistics," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 018, 2009.
- [124] J. Martin, A. Riazuelo, and M. Sakellariadou, "Nonvacuum initial states for cosmological perturbations of quantum-mechanical origin," *Physical Review D*, vol. 61, no. 8, Article ID 083518, 15 pages, 2000.
- [125] A. Gangui, J. Martin, and M. Sakellariadou, "Single field inflation and non-Gaussianity," *Physical Review D*, vol. 66, no. 8, Article ID 083502, 3 pages, 2002.
- [126] X. Chen and Y. Wang, "Large non-Gaussianities with intermediate shapes from quasi-single-field inflation," *Physical Review D*, vol. 81, no. 6, Article ID 063511, 5 pages, 2010.
- [127] D. S. Salopek and J. R. Bond, "Nonlinear evolution of long-wavelength metric fluctuations in inflationary models," *Physical Review D*, vol. 42, no. 12, pp. 3936–3962, 1990.

- [128] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity from inflation,” *Physical Review D*, vol. 65, no. 10, Article ID 103505, 2002.
- [129] F. Bernardeau and J.-P. Uzan, “Non-Gaussianity in multifield inflation,” *Physical Review D*, vol. 66, no. 10, 14 pages, 2002.
- [130] F. Bernardeau and J.-P. Uzan, “Inflationary models inducing non-Gaussian metric fluctuations,” *Physical Review D*, vol. 67, no. 12, Article ID 121301, 5 pages, 2003.
- [131] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. Van Tent, “Nonlinear perturbations in multiple-field inflation,” *Physical Review D*, vol. 73, no. 8, Article ID 083521, pp. 1–11, 2006.
- [132] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, “Large non-Gaussianity in multiple-field inflation,” *Physical Review D*, vol. 73, no. 8, Article ID 083522, pp. 1–19, 2006.
- [133] D. Seery and J. E. Lidsey, “Primordial non-Gaussianities from multiple-field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 9, pp. 187–210, 2005.
- [134] F. Vernizzi and D. Wands, “Non-Gaussianities in two-field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 019, 2006.
- [135] D. Mulryne, D. Seery, and D. Wesley, “Non-Gaussianity constrains hybrid inflation,” <http://arxiv.org/abs/0911.3550>.
- [136] J. O. Gong, C. Lin, and Y. Wang, “Non-Gaussianity from false vacuum inflation: old curvaton scenario,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 3, article 004, 2010.
- [137] A. Gangui, F. Lucchin, S. Matarrese, and S. Mollerach, “The three-point correlation function of the cosmic microwave background in inflationary models,” *Astrophysical Journal*, vol. 430, no. 2, pp. 447–457, 1994.
- [138] L. Verde, L. Wang, A. F. Heavens, and M. Kamionkowski, “Large-scale structure, the cosmic microwave background and primordial non-Gaussianity,” *Monthly Notices of the Royal Astronomical Society*, vol. 313, no. 1, pp. 141–147, 2000.
- [139] A. A. Starobinsky, “Multicomponent de Sitter (Inflationary) stages and the generation of perturbations,” *JETP Letters*, vol. 42, p. 152, 1985.
- [140] M. Sasaki and E. D. Stewart, “A general analytic formula for the spectral index of the density perturbations produced during inflation,” *Progress of Theoretical Physics*, vol. 95, no. 1, pp. 71–78, 1996.
- [141] D. H. Lyth and Y. Rodríguez, “Inflationary prediction for primordial non-gaussianity,” *Physical Review Letters*, vol. 95, no. 12, Article ID 121302, 4 pages, 2005.
- [142] A. Linde and V. Mukhanov, “Non-Gaussian isocurvature perturbations from inflation,” *Physical Review D*, vol. 56, no. 2, pp. R535–R539, 1997.
- [143] D. H. Lyth and D. Wands, “Generating the curvature perturbation without an inflaton,” *Physics Letters B*, vol. 524, no. 1-2, pp. 5–14, 2002.
- [144] T. Moroi and T. Takahashi, “Effects of cosmological moduli fields on cosmic microwave background,” *Physics Letters B*, vol. 522, no. 3-4, pp. 215–221, 2001.
- [145] T. Moroi and T. Takahashi, “Erratum: effects of cosmological moduli fields on cosmic microwave background (Physics Letters, Section B: Nuclear, Elementary Particle (2001) 522 (215) PII: S0370269301012953),” *Physics Letters B*, vol. 539, no. 3-4, p. 303, 2002.
- [146] D. H. Lyth, C. Ungarelli, and D. Wands, “Primordial density perturbation in the curvaton scenario,” *Physical Review D*, vol. 67, no. 2, Article ID 023503, 2003.
- [147] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity in the curvaton scenario,” *Physical Review D*, vol. 69, no. 4, Article ID 043503, 2004.
- [148] Q.-G. Huang, “Large non-Gaussianity implication for curvaton scenario,” *Physics Letters B*, vol. 669, no. 5, pp. 260–265, 2008.
- [149] M. Li, C. Lin, T. Wang, and Y. Wang, “Non-Gaussianity, isocurvature perturbation, gravitational waves, and a no-go theorem for the isocurvaton,” *Physical Review D*, vol. 79, no. 6, Article ID 063526, 10 pages, 2009.
- [150] K. Enqvist and T. Takahashi, “Signatures of non-Gaussianity in the curvaton model,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 9, article 012, 2008.
- [151] Q.-G. Huang and Y. Wang, “Curvaton dynamics and the non-linearity parameters in the curvaton model,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 9, article 025, 2008.
- [152] E. Komatsu, J. Dunkley, M. R. Nolte et al., “Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [153] C. Gordon, D. Wands, B. A. Bassett, and R. Maartens, “Adiabatic and entropy perturbations from inflation,” *Physical Review D*, vol. 63, no. 2, Article ID 023506, 11 pages, 2001.
- [154] G. Dvali, A. Gruzinov, and M. Zaldarriaga, “New mechanism for generating density perturbations from inflation,” *Physical Review D*, vol. 69, no. 2, Article ID 023505, 6 pages, 2004.
- [155] L. Kofman, “Probing string theory with modulated cosmological fluctuations,” <http://arxiv.org/abs/astro-ph/0303614>.
- [156] K. Enqvist, A. Jokinen, A. Mazumdar, T. Multamäki, and A. Väihkönen, “Non-gaussianity from preheating,” *Physical Review Letters*, vol. 94, no. 16, Article ID 161301, 4 pages, 2005.
- [157] A. Jokinen and A. Mazumdar, “Very large primordial non-Gaussianity from multiple fields: application to massless preheating,” *Journal of Cosmology and Astroparticle Physics*, no. 4, pp. 25–45, 2006.
- [158] N. Barnaby and J. M. Cline, “Non-Gaussianity from tachyonic preheating in hybrid inflation,” *Physical Review D*, vol. 75, no. 8, Article ID 086004, 14 pages, 2007.
- [159] A. Chambers and A. Rajantie, “Lattice calculation of non-gaussian density perturbations from the massless preheating inflationary model,” *Physical Review Letters*, vol. 100, no. 4, Article ID 041302, 4 pages, 2008.
- [160] A. Chambers and A. Rajantie, “Erratum: lattice calculation of non-Gaussian density perturbations from the massless preheating inflationary model (Physical Review Letters (2008) 100 (041302)),” *Physical Review Letters*, vol. 101, no. 14, Article ID 149903, 1 pages, 2008.
- [161] J. R. Bond, A. V. Frolov, Z. Huang, and L. Kofman, “Non-gaussian curvature spikes from chaotic billiards in inflation preheating,” *Physical Review Letters*, vol. 103, no. 7, Article ID 071301, 4 pages, 2009.
- [162] N. Bartolo, E. Dimastrogiovanni, S. Matarrese, and A. Riotto, “Anisotropic bispectrum of curvature perturbations from primordial non-abelian vector fields,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 10, article 015, 2009.
- [163] C.A. Valenzuela-Toledo, Y. Rodríguez, and D. H. Lyth, “Non-Gaussianity at tree and one-loop levels from vector field perturbations,” *Physical Review D*, vol. 80, no. 10, Article ID 103519, 7 pages, 2009.

- [164] E. Dimastrogiovanni, N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity and statistical anisotropy from vector field populated inflationary models,” <http://arxiv.org/abs/1001.4049>.
- [165] N. Barnaby and J. M. Cline, “Large non-Gaussianity from non-local inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 7, article 017, 2007.
- [166] N. Barnaby and J. M. Cline, “Predictions for non-Gaussianity from non-local inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 6, article 030, 2008.
- [167] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, “Non-Gaussianities in new ekpyrotic cosmology,” *Physical Review Letters*, vol. 100, no. 17, Article ID 171302, 4 pages, 2008.
- [168] K. Koyama, S. Mizuno, F. Vernizzi, and D. Wands, “Non-Gaussianities from ekpyrotic collapse with multiple fields,” *Journal of Cosmology and Astroparticle Physics*, no. 11, pp. 1–21, 2007.
- [169] J. Khoury and F. Piazza, “Rapidly-varying speed of sound, scale invariance and non-Gaussian signatures,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 7, article 026, 2009.
- [170] J.-L. Lehners and S. Renaux-Petel, “Multifield cosmological perturbations at third order and the ekpyrotic trispectrum,” *Physical Review D*, vol. 80, no. 6, Article ID 063503, 18 pages, 2009.
- [171] J.-L. Lehners and P. J. Steinhardt, “Non-Gaussianity generated by the entropic mechanism in bouncing cosmologies made simple,” *Physical Review D*, vol. 80, no. 10, Article ID 103520, 5 pages, 2009.
- [172] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, “Constraints on local primordial non-Gaussianity from large scale structure,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, 2008.
- [173] A. P. S. Yadav and B. D. Wandelt, “Evidence of primordial non-gaussianity ( $f_{NL}$ ) in the wilkinson microwave anisotropy probe 3-year data at  $2.8\sigma$ ,” *Physical Review Letters*, vol. 100, no. 18, Article ID 181301, 4 pages, 2008.
- [174] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $f_{NL}^{local}$  from WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, 2009.
- [175] L. Senatore, K. M. Smith, and M. Zaldarriaga, “Non-Gaussianities in single field inflation and their optimal limits from the WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 1, 2010.
- [176] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, “Imprints of primordial non-Gaussianities on large-scale structure: scale-dependent bias and abundance of virialized objects,” *Physical Review D*, vol. 77, no. 12, Article ID 123514, 17 pages, 2008.
- [177] V. Desjacques and U. Seljak, “Signature of primordial non-Gaussianity of  $\varphi^3$  type in the mass function and bias of dark matter haloes,” *Physical Review D*, vol. 81, no. 2, 2010.
- [178] L. Verde and S. Matarrese, “Detectability of the effect of inflationary non-gaussianity on halo bias,” *Astrophysical Journal*, vol. 706, no. 1, part 2, pp. L91–L95, 2009.
- [179] M. Roncarelli, L. Moscardini, E. Branchini et al., “Imprints of primordial non-Gaussianities in X-ray and SZ signals from galaxy clusters,” *Monthly Notices of the Royal Astronomical Society*, vol. 402, no. 2, pp. 923–933, 2010.
- [180] Ø. Rudjord, F. K. Hansen, X. Lan, M. Liguori, D. Marinucci, and S. Matarrese, “An estimate of the primordial non-gaussianity parameter  $f_{NL}$  using the needlet bispectrum from wmap,” *Astrophysical Journal*, vol. 701, no. 1, pp. 369–376, 2009.
- [181] Ø. Rudjord, F. K. Hansen, X. Lan, M. Liguori, D. Marinucci, and S. Matarrese, “Directional variations of the non-gaussianity parameter  $f_{NL}$ ,” *Astrophysical Journal*, vol. 708, no. 2, pp. 1321–1325, 2010.
- [182] J. Smidt, A. Amblard, P. Serra, and A. Cooray, “Measurement of primordial non-Gaussianity using the WMAP 5-year temperature skewness power spectrum,” *Physical Review D*, vol. 80, no. 12, 19 pages, 2009.
- [183] P. Cabella, et al., “Foreground influence on primordial non-Gaussianity estimates: needlet analysis of WMAP 5-year data,” <http://arxiv.org/abs/0910.4362>.
- [184] M. Bucher, B. Van Tent, and C. S. Carvalho, “Detecting bispectral acoustic oscillations from inflation using a new flexible estimator,” <http://arxiv.org/abs/0911.1642>.
- [185] M. LoVerde, A. Miller, S. Shandera, and L. Verde, “Effects of scale-dependent non-Gaussianity on cosmological structures,” *Journal of Cosmology and Astroparticle Physics*, vol. 4, article 014, 2008.
- [186] E. Sefusatti, M. Liguori, A. P. S. Yadav, M. G. Jackson, and E. Pajer, “Constraining running non-gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, 2009.
- [187] X. Chen, B. Hu, M.-X. Huang, G. Shiu, and Y. Wang, “Large primordial trispectra in general single field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, 2009.
- [188] F. Arroja, S. Mizuno, K. Koyama, and T. Tanaka, “Full trispectrum in single field DBI inflation,” *Physical Review D*, vol. 80, no. 4, Article ID 043527, 14 pages, 2009.
- [189] M.-X. Huang and G. Shiu, “Inflationary trispectrum for models with large non-Gaussianities,” *Physical Review D*, vol. 74, no. 12, 5 pages, 2006.
- [190] F. Arroja and K. Koyama, “Non-Gaussianity from the trispectrum in general single field inflation,” *Physical Review D*, vol. 77, no. 8, Article ID 083517, 14 pages, 2008.
- [191] X. Gao and B. Hu, “Primordial trispectrum from entropy perturbations in multifield DBI model,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, 2009.
- [192] G. Xian, L. Miao, and L. Chunshan, “Primordial non-Gaussianities from the trispectra in multiple field inflationary models,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 11, 2009.
- [193] S. Mizuno, F. Arroja, and K. Koyama, “Full quantum trispectrum in multifield DBI inflation,” *Physical Review D*, vol. 80, no. 8, Article ID 083517, 18 pages, 2009.
- [194] D. Seery, J. E. Lidsey, and M. S. Sloth, “The inflationary trispectrum,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 1, 2007.
- [195] D. Seery and J. E. Lidsey, “Non-Gaussianity from the inflationary trispectrum,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 1, 2007.
- [196] D. Seery, M. S. Sloth, and F. Vernizzi, “Inflationary trispectrum from graviton exchange,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 3, 2009.
- [197] P. Adshead, R. Easther, and E. A. Lim, ““in-in” formalism and cosmological perturbations,” *Physical Review D*, vol. 80, no. 8, 2009.
- [198] N. Bartolo, E. Dimastrogiovanni, S. Matarrese, and A. Riotto, “Anisotropic trispectrum of curvature perturbations induced by primordial non-Abelian vector fields,” *Journal of Cosmology and Astroparticle Physics*, vol. 11, article 028, 2009.
- [199] C.A. Valenzuela-Toledo and Y. Rodríguez, “Non-gaussianity from the trispectrum and vector field perturbations,” *Physics Letters B*, vol. 685, no. 2-3, pp. 120–127, 2010.

- [200] P. Creminelli and M. Zaldarriaga, “A single-field consistency relation for the three-point function,” *Journal of Cosmology and Astroparticle Physics*, no. 10, pp. 101–107, 2004.
- [201] C. Cheung, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, “On the consistency relation of the three-point function in single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 2, 2008.
- [202] S. Renaux-Petel, “Combined local and equilateral non-Gaussianities from multifield DBI inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 10, 2009.

## Review Article

# Review of Local Non-Gaussianity from Multifield Inflation

Christian T. Byrnes<sup>1</sup> and Ki-Young Choi<sup>2</sup>

<sup>1</sup> Fakultät für Physik, Universität Bielefeld, Postfach 100131, 33501 Bielefeld, Germany

<sup>2</sup> Department of Physics, Pusan National University, Busan 609-735, Republic of Korea

Correspondence should be addressed to Ki-Young Choi, kiyoung.choi@pusan.ac.kr

Received 11 January 2010; Revised 5 May 2010; Accepted 9 June 2010

Academic Editor: Sarah Shandera

Copyright © 2010 C. T. Byrnes and K.-Y. Choi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We review models which generate a large non-Gaussianity of the local form. We first briefly consider three models which generate the non-Gaussianity either at or after the end of inflation; the curvaton scenario, modulated (p)reheating, and an inhomogeneous end of inflation. We then focus on ways of generating the non-Gaussianity during inflation. We derive general conditions which a product or sum separable potential must satisfy in order to generate a large local bispectrum during slow-roll inflation. As an application, we consider two-field hybrid inflation. We then derive a formalism not based on slow roll which can be applied to models in which the slow-roll parameters become large before inflation ends. An exactly soluble two-field model is given in which this happens. Finally, we also consider further non-Gaussian observables, a scale dependence of  $f_{\text{NL}}$  and the trispectrum.

## 1. Introduction

There are many models of the universe which can predict a large non-Gaussianity. However the predicted amplitude and the shape of the non-Gaussianity are different among different classes of models. One category is those which generate the non-Gaussianity due to nontrivial classical dynamics on superhorizon scales. These models predict the shape of the bispectrum to be of the so-called “local type”, which can be expressed as an expansion of the Bardeen potential [1]

$$\Phi(\mathbf{x}) = \Phi_L(\mathbf{x}) + f_{\text{NL}}(\Phi_L^2(\mathbf{x}) - \langle \Phi_L^2(\mathbf{x}) \rangle), \quad (1)$$

where  $\Phi$  is the curvature perturbation on a Newtonian slice and  $\Phi_L$  is its linear and Gaussian part.  $\langle \Phi_L^2(\mathbf{x}) \rangle$  denotes the ensemble average in a statistically homogeneous distribution. The current limit on the local type of the nonlinearity parameter  $f_{\text{NL}}$  from seven years of WMAP data [2] is  $-10 < f_{\text{NL}} < 74$  at the 95% confidence level. Constraints are expected to improve rapidly and significantly, first with Planck data and later using large scale structure data, see the recent reviews [3–5]. The Bardeen potential is related to the

primordial curvature perturbation of  $\zeta$  on large scales and in the matter dominated era by  $\Phi = (3/5)\zeta$ .

The curvature perturbation at horizon exit is determined by the classical perturbations of the scalar fields,  $\delta\phi_i(\mathbf{x})$ . The subsequent evolution of  $\zeta$  can be conveniently described by the  $\delta N$  formalism [6–11]. The curvature perturbation is given by up to quadratic terms [11]

$$\zeta = \delta N = \sum_I N_{,I} \delta\varphi_{I*} + \frac{1}{2} \sum_{IJ} N_{,IJ} \delta\varphi_{I*} \delta\varphi_{J*} + \dots, \quad (2)$$

where  $N(\mathbf{x}, t)$  is the e-folding number evaluated in an unperturbed Universe, from the epoch of horizon exit to later epoch of uniform energy density hypersurface (for an extension to include gradient terms, see [12]). The power spectrum  $\mathcal{P}_\zeta$  and the bispectrum  $B_\zeta$  are defined by

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle \equiv (2\pi)^3 \delta^3(\mathbf{k}_1 + \mathbf{k}_2) \frac{2\pi^2}{k_1^3} \mathcal{P}_\zeta(k_1), \quad (3)$$

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle \equiv (2\pi)^3 \delta^3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_\zeta(k_1, k_2, k_3).$$

From this, we can define the observable quantities, the spectral index, the tensor-to-scalar ratio, and the nonlinearity parameter

$$n_\zeta - 1 \equiv \frac{\partial \log \mathcal{P}_\zeta}{\partial \log k}, \quad (4)$$

$$r = \frac{\mathcal{P}_T}{\mathcal{P}_\zeta} = \frac{8\mathcal{P}_*}{M_p^2 \mathcal{P}_\zeta}, \quad (5)$$

$$f_{\text{NL}} = \frac{5}{6} \frac{k_1^3 k_2^3 k_3^3}{k_1^3 + k_2^3 + k_3^3} \frac{B_\zeta(k_1, k_2, k_3)}{4\pi^4 \mathcal{P}_\zeta^2}, \quad (6)$$

where  $\mathcal{P}_T = 8\mathcal{P}_*/M_p^2 = 8H_*^2/(4\pi^2 M_p^2)$  is the power spectrum of the tensor metric fluctuations. It is well known that single-field inflation does not lead to a detectably large non-Gaussianity, in fact  $f_{\text{NL}}$  is suppressed by slow-roll parameters [13]. Observably large non-Gaussianity can be obtained by breaking the slow-roll conditions during inflation [14], using extended kinetic terms [15–25], see also the reviews [26, 27], or going beyond models of single-field inflation [11, 28–34]

It is natural to consider multiple scalar field since they are ubiquitous in many beyond the standard model of particle physics, such as supersymmetry and string theory. These scalar fields generate nonadiabatic perturbations during inflation and change the evolution of the curvature perturbation after horizon exit. The residual isocurvature perturbation may be present in the primordial density fluctuation and can be correlated with the curvature perturbation or may be responsible for an observably large non-Gaussianity in the cosmic microwave background and large scale structure, for observational limits on isocurvature perturbations see [1, 35, 36]. In this paper, we will only consider models with adiabatic primordial perturbations, in which the isocurvature perturbation present during inflation is converted into an adiabatic perturbation. We also neglect the secondary non-Gaussianities generated at later times, for example see [37, 38].

There are popular multifield models [39] which may generate observably large non-Gaussianity. These include the curvaton scenario, modulated (p)reheating, and an inhomogeneous end of inflation, see Section 2. In these scenarios, large non-Gaussianity is generated either by the means of ending inflation, or after inflation. It was shown recently that it also possible to generate large non-Gaussianity during the evolution of slow-roll multifield inflation, see Section 3.

All of these models generate the large non-Gaussianity after horizon exit, such as after reheating, at the end of inflation, at the phase transition, or during inflation after horizon exit and involves the perturbation of the nonadiabatic mode. Therefore, the non-Gaussianity of these models is of the local type which is distinguishable from other shapes of non-Gaussianity (for a list of possibilities see, e.g., [40–42]), in which the non-Gaussianity is generated intrinsically from the quantum fluctuations, or during horizon exit.

In Section 2, we summarise the aforementioned three models, which are popular methods of generating a large non-Gaussianity. In Section 3, we review the possibility of

generating a large non-Gaussianity from multifield slow-roll inflation, and in Section 4, we consider hybrid inflation with two inflaton fields as an application. Then in Section 5, we discuss multifield models of inflation without assuming the slow-roll conditions and present an exact solution. Non-Gaussian observables beyond  $f_{\text{NL}}$ , such as its scale dependence and the trispectrum are introduced in Section 6. Finally, we conclude in Section 7.

## 2. Summary of Popular Models Generating a Large Local Non-Gaussianity

Multifield inflationary models [39] can lead to an observable signature of non-Gaussianity. These include models in which the large non-Gaussianity is generated either by the means of ending inflation, or after inflation. We review the three most popular of these models in this section, for a discussion on how these models are related see [43]. In the next section, we review the possibility of generating large non-Gaussianity during multifield slow-roll inflation. Later in Section 5, we discuss multifield models of inflation without assuming the slow-roll conditions. This does not exhaust all of the possible ways of generating a large local non-Gaussianity from inflation, see also the reviews [44–47] and the papers [48, 49]. Non-Gaussianity of the local form can also be generated in the ekpyrotic scenario [50, 51], although in the simplest case  $f_{\text{NL}}$  is large and negative [52], which is observationally ruled out.

*2.1. Curvaton Scenario.* In this scenario, there is a light weakly-interacting inhomogeneous scalar field, the curvaton  $\chi$ , in addition to the inflaton field,  $\phi$ . The additional scalar field is completely subdominant during inflation however, it can dominate the energy density of the Universe later since the time-averaged equation of state becomes that of the pressureless matter ( $P = 0$ ), whose energy density decreases more slowly than that of the radiation which the inflaton field has decayed into. After decay, the inhomogeneity of the scalar field leads to the density perturbation of radiation [53, 54]. This is the basis of the curvaton scenario [55–59] which liberates the inflaton field from being required to generate the observed spectrum of perturbations (although there is instead an upper bound on their allowed magnitude).

The significant non-Gaussianity of the local type can be generated in the curvaton model [60]. A study of non-Gaussianity in the curvaton scenario using second-order cosmological perturbation theory was done by Bartolo et al. [31] with a sudden decay approximation and also by Malik and Lyth [32] using a fully numeric approach, which accounted for both the sudden decay and nonsudden decay approximations. Lyth and Rodríguez [11] used the  $\delta N$  formalism including the quadratic terms (second order terms in (2)) to calculate the non-Gaussianity. The nonlinear generalisation of the curvature perturbation and a numerical study was done to give a full probability distribution function in [61]. This simple curvaton model was extended into the mixed curvaton-inflaton scenario in which fluctuations of both the inflaton and a curvaton contribute to the primordial

density perturbation studying the isocurvature perturbation in [62]. The multiple curvaton fields which contribute to the primordial density perturbation was studied in [63–65]. Observational constraints on curvaton models with isocurvature perturbation and non-Gaussianity can be found in [66–68].

The generation of large non-Gaussianity in the curvaton scenario can be easily understood using the result of the nonlinear  $\delta N$  formalism [61, 62]. Assuming the sudden decay of the curvaton field on the  $H = \Gamma$  uniform energy-density hypersurface, that is, when the local Hubble rate equals the decay rate for the curvaton, leads to a nonlinear relation between the local curvaton density and the radiation density before and after the decay

$$\Omega_r e^{4(\zeta_{\text{inf}} - \zeta)} + \Omega_\chi e^{3(\zeta_\chi - \zeta)} = 1, \quad (7)$$

where  $\Omega_r + \Omega_\chi = 1$ . Here,  $\zeta$  is the primordial curvature perturbation which remains constant on large scales in the radiation-dominated era after the curvaton decays and  $\zeta_\chi$  is the curvaton perturbation. Expanding this equation order by order yields up to second order [61]

$$\begin{aligned} \zeta = & (1 - r_\chi) \zeta_{\text{inf}} + r_\chi \zeta_\chi + \frac{r_\chi (1 - r_\chi) (3 + r_\chi)}{2} (\zeta_\chi - \zeta_{\text{inf}})^2 \\ & + \text{higher-order terms,} \end{aligned} \quad (8)$$

where  $r_\chi \equiv 3\Omega_\chi/(4 - \Omega_\chi)$  at the time of the curvaton decay. The curvature perturbation of curvaton field [61]

$$\zeta_\chi = \frac{2}{3} \frac{\delta_1 \chi}{\chi_0} - \frac{1}{3} \left( \frac{\delta_1 \chi}{\chi_0} \right)^2 + \frac{2}{9} \left( \frac{\delta_1 \chi}{\chi_0} \right)^3, \quad (9)$$

is almost constant before the curvaton decays from the start of curvaton oscillation. Note that the intrinsic non-Gaussianity in the curvaton field during oscillation is  $f_{\text{NL},\chi} = -5/4$ , which comes from the nonlinear relation between  $\delta_1 \chi$  and  $\zeta_\chi$ . After the curvaton decays into radiation, the curvature perturbation of radiation has the following non-Gaussianity parameters: [11, 31, 61]

$$f_{\text{NL}} = \frac{\tilde{r}^2}{(1 + \tilde{r})^2} \left[ \frac{5}{4r_\chi} - \frac{5}{3} - \frac{5r_\chi}{6} \right], \quad (10)$$

where  $r_\chi$  is evaluated when the curvaton decays based on the sudden decay approximation but gives good agreement with a full numerical study [32, 61], and  $\tilde{r}$  is defined as

$$\tilde{r} \equiv \frac{|r_\chi \zeta_\chi|^2}{|(1 - r_\chi) \zeta_{\text{inf}}|^2}. \quad (11)$$

In the limit of the pure curvaton scenario,  $\zeta_{\text{inf}} \rightarrow 0$ , it recovers the usual normalization for  $f_{\text{NL}}$ , that is,  $\tilde{r}^2/(1 + \tilde{r})^2 \rightarrow 1$ . Therefore, the large  $f_{\text{NL}}$  is obtained for small  $r_\chi$ . This large  $f_{\text{NL}}$ , in spite of the order of unity non-Gaussianity in  $\zeta_\chi$ , has the origin in the nonlinear relation between  $\zeta_\chi$  (or  $\delta_1 \chi$ ) and  $\zeta$  in (8).

In fact (7), and thus (8), can be used “at any time” before the curvaton decay with  $r_\chi$  evaluated at that time to find the total curvature perturbation at that moment, as well as at the epoch of curvaton decay. This enables us to understand the evolution of the curvature perturbation  $\zeta$  on the uniform energy density hypersurface before the curvaton decay. Initially when the primordial radiation was generated, for example, from the inflaton decay,  $r_\chi$  is negligibly small so we cannot ignore the perturbations of the inflaton field. At this early time,  $\zeta = \zeta_{\text{inf}}$  and  $f_{\text{NL}} = 0$  since  $\tilde{r} \propto r_\chi^2$  which makes  $f_{\text{NL}}$  negligible in (10). As time goes on,  $r_\chi$  increases while  $\zeta_{\text{inf}}$  and  $\zeta_\chi$  are almost constant since there is no energy flow between two fluids satisfying adiabatic condition except around the decay time, thus the curvature perturbation evolves *gradually*. If the curvaton survives long enough,  $r_\chi$  reaches a value with which the  $r_\chi \zeta_\chi$  dominates the first term,  $\zeta_{\text{inf}}$ , and the curvaton perturbation is then responsible for the primordial inhomogeneity of the Universe. The large non-Gaussianity is possible only when  $r_\chi$  at the decay time is small that is, the energy density of curvaton is subdominant compared to that of radiation. This is because when  $r_\chi = 1$ , the curvaton energy density dominates over the radiation component before decay, then  $\zeta$  is effectively the curvature perturbation of a single field (curvaton),  $\zeta_\chi$ , and therefore non-Gaussianity  $f_{\text{NL}} = f_{\text{NL},\chi} = -5/4$ . We will also see this phenomena in the multifield slow-roll inflation later, in Section 3.

**2.2. Modulated (P)reheating.** After inflation, the energy density in the inflaton field must be transferred into radiation. In the simplest case of adiabatic perturbations, this process does not affect the primordial curvature perturbation on scales which are observable today, because these scale were much larger than the horizon at the time of reheating. However, in a spirit similar to the curvaton model, there may be a subdominant light scalar field present during inflation which modulates the efficiency of reheating. This makes the efficiency of reheating a spatially dependent process. The quasiscale invariant perturbations in this field, which during inflation are an isocurvature perturbation, may be converted into the primordial curvature perturbation during this process. For a review of reheating after inflation, see for example [69, 70].

As a simple illustration of this, in the “old” models of perturbative reheating, the decay rate of the inflaton is given by  $\Gamma \sim \lambda^2 m$ , where  $\lambda$  is the strength of the coupling between the inflaton and reheating fields and  $m$  is the mass of the inflaton. Provided that  $\Gamma$  is much less than the Hubble rate at the end of inflation, so that reheating takes place slowly, the reheating temperature is given by

$$T_{\text{reh}} \sim \sqrt{\Gamma M_P} \sim \lambda \sqrt{m M_P}. \quad (12)$$

If the coupling strength is a function of the local value of a second scalar field  $\chi$  then this will give rise to a perturbation in the decay rate of the inflaton field, and thus in the reheating temperature which is responsible for the density

perturbation after reheating. Provided that the  $\chi$  field is perturbed, the curvature perturbation can be written as

$$\zeta = \alpha \frac{\delta\Gamma}{\Gamma}, \quad (13)$$

where  $\alpha$  is a parameter depending on the ratio of  $\Gamma$  to the Hubble expansion at the end of inflation, whose value increases with decreasing  $\Gamma/H$  and leads to  $1/6$  in the limit  $\Gamma \ll H_{\text{inf}}$  [33, 71]. Hence, perturbations of the light scalar field  $\chi$  can be imprinted into the radiation temperature through the reheating process [71, 72]. In a similar way, the mass of the decaying particle may be modulated [73]. The above equation can be extended to second order, which is required to calculate  $f_{\text{NL}}$ . In the simple quadratic case that  $\Gamma = \Gamma_0 + \Gamma_1(\chi/\bar{\chi})^2 = \Gamma_0 + \Gamma_1(1 + \delta\chi/\bar{\chi})^2$ , where  $\bar{\chi}$  and  $\delta\chi$  are the homogeneous background part and the perturbation of the field  $\chi$ , respectively, the result is [33, 74–76]

$$f_{\text{NL}} = \frac{5}{12} \frac{\Gamma_0 + \Gamma_1}{\Gamma_1 \alpha}. \quad (14)$$

Hence, we can see that the non-Gaussianity can easily be larger than unity in this model by two sources. One is due to the small value of  $\alpha$  which occurs when the decay rate of the inflation is not much smaller than the Hubble constant at the end of inflation. The other is when  $\Gamma_1$  is small in which  $\chi$  field only controls part of the decay channels of inflaton field.

It is now generally considered that a period of rapid and highly nonlinear preheating preceded the perturbative reheating phase [69]. In this case, one can instead consider modulated preheating, see, for example, [77–79]. A particularly rapid form of preheating, known as instant preheating [80], has attracted more attention in this area, as it is possible to find analytic estimates [81–85]. It is possible to generate the primordial curvature perturbation in these models by having either a weak [81] or a strong [83] symmetry breaking of the two-field potential during inflation which affects the efficiency of instant preheating, or by having two inflaton fields which are coupled with different strengths to the preheat field [82]. However, all of the above models of modulated instant preheating are only allowed, even at the linear level in perturbations, for a small corner of parameter space and even then only for a special choice of initial conditions [85].

**2.3. Inhomogeneous End of Inflation.** An inhomogeneous phase transition can occur also at the end of inflation between inflationary and non-inflationary expansion due to the condition of how inflation ends [34, 86]. In the case of multicomponent inflation, there is a family of inflationary trajectories. If the family of trajectories is not perpendicular to the surface on which inflation ends then there is a possibility that inflation ends on a slice which is not of uniform density and the generation of the curvature perturbation at the end of inflation is expected [34] (This was also discussed in [87–89] using the linear relation between two fields at the end of inflation and the non-Gaussianity is transferred from that of the light field which is generated from its nonlinear self-coupling). In single field inflation, the

energy density is determined solely by the single field  $\phi$ , and the end of inflation is parameterised by the unique value  $\phi_e$ . If there is another field  $\sigma$ , then  $\phi_e(\sigma)$  will depend on position through the perturbation  $\delta\sigma(\mathbf{x})$ . As a result, the change of e-folding number from a spacetime slice of uniform energy density just before the end of inflation to a spacetime slice of uniform density just after the end of inflation will have a perturbation  $\delta N_e = \zeta_e$ . This  $\zeta_e$  is the curvature perturbation generated by the end of inflation and can dominate over the contribution from inflation [34].

Using the perturbation of  $\phi_e(\sigma)$

$$\delta\phi_e = \phi'_e \delta\sigma + \frac{1}{2} \phi_e'' (\delta\sigma)^2, \quad (15)$$

the curvature perturbation generated at the end of inflation is [34]

$$\begin{aligned} \zeta_e &= N'_e \delta\phi_e + \frac{1}{2} N_e'' (\delta\phi_e)^2 \\ &= N'_e \phi'_e \delta\sigma + \frac{1}{2} [2N_e'' \phi_e'^2 + n'_e \phi_e''] (\delta\sigma)^2. \end{aligned} \quad (16)$$

When  $\zeta_e$  dominates the perturbation generated during the inflation, the power spectrum is given by

$$\mathcal{P}_e = \frac{\phi_e'^2}{2\epsilon_e} \left( \frac{H^*}{2\pi} \right)^2 \quad (17)$$

The non-Gaussianity can also be generated at the end of inflation since the perturbation at the end of inflation is not required to be Gaussian. Thus, we expect that the observable large non-Gaussianity can be generated in this process. From (16) the non-Gaussianity is [34]

$$f_{\text{NL}} = -\frac{5}{3} \sqrt{\frac{\epsilon_e}{2}} \frac{\phi_e''}{(\phi_e')^2}. \quad (18)$$

This was applied to the simple example of extended hybrid inflation using sudden end approximation [90, 91].

Recently Sasaki [92], Naruko and Sasaki [93] calculated analytically the curvature perturbation and the non-Gaussianity in the multi-brid inflation models from horizon exit to the end of inflation considering the general couplings to the waterfall field. In general, we cannot just ignore the curvature perturbation generated during inflation. Furthermore, that is also affected by the end of inflation. A deeper understanding of non-Gaussianity both from the evolution during multifield hybrid inflation and effects from the end of inflation is considered by Byrnes et al. [94, 95] which is summarised in Section 3.

### 3. Non-Gaussianity during Slow-Roll Inflation

It is well known that the single field slow-roll inflation with canonical kinetic terms generates a nonlinear parameter of the order of the spectral tilt of the spectrum and thus too small to be observed [13]. On the other hand, in multifield inflationary models the nonadiabatic perturbation may change this conclusion. Furthermore, even within slow-roll inflation an observably large non-Gaussianity can be generated if certain conditions are satisfied.

Rigopoulos et al. introduced a formalism [96, 97] to deal with nonlinearity based on a system of fully nonlinear equations for long wavelengths and applied it to a two-field inflation model. In the case of two-field quadratic inflation with two massive fields, an analytic calculation is possible with slow-roll conditions, and this shows that the nonlinear parameter is small [98, 99]. The numerical calculation confirms the analytical prediction that it is possible to generate a narrow spike of large non-Gaussianity while the inflaton trajectory turns a corner, but the non-Gaussianity decays quickly after the corner [99, 100]. This spike is due to the temporary jump in the slow-roll parameters [101, 102].

Recently, an analysis based on the general analytic formula was done by Byrnes et al. [94, 95], and they showed that it is possible to generate observable large non-Gaussianity even during slow-roll multifield inflation models and explicitly wrote the conditions as well as concrete examples (There is a claim [103] that the sizeable value of  $f_{\text{NL}}$  can arise from a loop corrections). In this section, we summarise the general conditions for two-field slow-roll inflation to generate a large non-Gaussianity and briefly give a specific example. In Section 4 we consider the multifield hybrid inflation model which has been studied in depth, and we include the effects from the end of inflation.

**3.1. General Formulas.** In the case of a potential which is either sum or product separable, it is possible to compute the curvature perturbation using the  $\delta N$ -formalism and slow-roll conditions, and an analytic formula for the nonlinearity parameter can be obtained. The general formula for the nonlinearity parameter  $f_{\text{NL}}$  was calculated in the case of a separable potential by sum [99], and this was later extended to the non-canonical kinetic terms as well as to the separable potential by product [104]. This has been further generalized to the arbitrary number of fields [105] and to the trispectrum [106].

For a product separable potential,  $W(\varphi, \chi) = U(\varphi)V(\chi)$ , we can find the number of  $e$ -foldings analytically in the slow-roll limit [107],

$$N(\varphi_*, \chi_*) = -\frac{1}{M_P^2} \int_*^e \frac{U}{U} d\varphi = -\frac{1}{M_P^2} \int_*^e \frac{V}{V} d\chi. \quad (19)$$

Then, the power spectrum and the nonlinear parameters can be calculated straightforwardly using the derivatives of  $e$ -folding number [99, 104, 107]. In the case of two inflatons with canonical kinetic terms, the power spectrum and spectral index are found to be [104]

$$\begin{aligned} \mathcal{P}_\zeta &= \frac{W_*}{24\pi^2 M_P^4} \left( \frac{u^2}{\epsilon_\varphi^*} + \frac{v^2}{\epsilon_\chi^*} \right), \\ n_\zeta - 1 &= -2\epsilon^* - 4 \frac{u^2(1 - \eta_{\varphi\varphi}^*/2\epsilon_\varphi^*) + v^2(1 - \eta_{\chi\chi}^*/2\epsilon_\chi^*)}{u^2/\epsilon_\varphi^* + v^2/\epsilon_\chi^*}, \\ r &= 16 \left( \frac{u^2}{\epsilon_\varphi^*} + \frac{v^2}{\epsilon_\chi^*} \right)^{-1}, \end{aligned} \quad (20)$$

where the slow-roll parameters are

$$\epsilon_\varphi = \frac{M_P^2}{2} \left( \frac{U_\varphi}{U} \right)^2 = \epsilon \cos^2 \theta, \quad \epsilon_\chi = \frac{M_P^2}{2} \left( \frac{V_\chi}{V} \right)^2 = \epsilon \sin^2 \theta, \quad (21)$$

where  $\theta$  is the angle between the adiabatic perturbation and one of the fields and

$$\eta_{\varphi\varphi} = M_P^2 \frac{V_{\varphi\varphi}}{V}, \quad \eta_{\varphi\chi} = M_P^2 \frac{U_\varphi V_\chi}{W}, \quad \eta_{\chi\chi} = M_P^2 \frac{V_{\chi\chi}}{V}. \quad (22)$$

Note that the superscript “\*” and “ $e$ ” denotes that the values are evaluated at horizon exit and at end of inflation, respectively. In the above,  $u$  and  $v$  are evaluated at the end of inflation as

$$u \equiv \frac{\epsilon_\varphi^e}{\epsilon^e} = \cos^2 \theta^e, \quad v \equiv \frac{\epsilon_\chi^e}{\epsilon^e} = \sin^2 \theta^e. \quad (23)$$

The nonlinearity parameter  $f_{\text{NL}}$  becomes [104]

$$\begin{aligned} f_{\text{NL}} &= \frac{5}{6} \frac{2}{(u^2/\epsilon_\varphi^* + v^2/\epsilon_\chi^*)^2} \\ &\times \left[ \frac{u^3}{\epsilon_\varphi^*} \left( 1 - \frac{\eta_{\varphi\varphi}^*}{2\epsilon_\varphi^*} \right) + \frac{v^3}{\epsilon_\chi^*} \left( 1 - \frac{\eta_{\chi\chi}^*}{2\epsilon_\chi^*} \right) \right. \\ &\quad \left. - \left( \frac{u}{\epsilon_\varphi^*} - \frac{v}{\epsilon_\chi^*} \right)^2 \mathcal{A}_P \right] \end{aligned} \quad (24)$$

where

$$\begin{aligned} \mathcal{A}_P &\equiv -\frac{\epsilon_\varphi^e \epsilon_\chi^e}{(\epsilon^e)^2} \left[ \hat{\eta}^e - 4 \frac{\epsilon_\varphi^e \epsilon_\chi^e}{\epsilon^e} \right], \\ \hat{\eta} &\equiv \frac{\epsilon_\chi \eta_{\varphi\varphi} + \epsilon_\varphi \eta_{\chi\chi}}{\epsilon}. \end{aligned} \quad (25)$$

Similar formulas are found in the case of a sum separable potential [99, 104].

**3.2. Conditions for Generating a Large  $f_{\text{NL}}$ .** Using an analytic formula for the nonlinear parameter  $f_{\text{NL}}$  in the case of a sum or product separable potential, it is possible to generate significant non-Gaussianity even during slow-roll inflation with Gaussian perturbations at Hubble exit and the general conditions for it can be written explicitly [94].

For the separable potential by product large nonlinear parameter  $|f_{\text{NL}}| \gtrsim 1$  is possible when (there is a symmetric region of large non-Gaussianity under the exchange of the two fields).

$$\begin{aligned} \sin^2 \theta^* &\lesssim \sin^4 \theta^e \left( \frac{1}{\sqrt{\sin^2 \theta^e \mathcal{G}_P}} - 1 \right), \\ \mathcal{G}_P &= \frac{6}{5} \left| -\eta_{\chi\chi}^* + 2\eta_{\chi\chi}^e \right|^{-1}. \end{aligned} \quad (26)$$

with this condition, the large  $f_{\text{NL}}$  is given by

$$f_{\text{NL}} \simeq \frac{5}{6} \frac{\sin^6 \theta^e}{(\sin^2 \theta^* + \sin^4 \theta^e)^2} \left[ -\eta_{\chi\chi}^* + 2\eta_{\chi\chi}^e \right]. \quad (27)$$

For large  $f_{\text{NL}}$ , the substantial increase in  $\sin \theta$  is necessary between horizon exit and the end of inflation, which corresponds to the curve of the trajectory in the field space during the slow-roll inflation. However, the background trajectory must be almost entirely in the one-field direction.

One specific toy model is the quadratic times exponential potential [94], where the potential is given by

$$W(\varphi, \chi) = \frac{1}{2} e^{-\lambda \varphi^2 / M_{\text{P}}^2} m^2 \chi^2. \quad (28)$$

With expansion of exponential, this potential can be understood as the quadratic chaotic inflation model which has a coupling to another light scalar field. For this potential, considering the condition for large non-Gaussianity, (26), large  $f_{\text{NL}}$  is generated when the  $\varphi$  at horizon exit is small. One example is found to give  $|f_{\text{NL}}| \sim 30$  with  $\varphi^* \sim 10^{-3} M_{\text{P}}$  for  $\lambda = 0.05$  and  $\chi^* = 16 M_{\text{P}}$ .

For a sum separable potential, we can find similar conditions for generating a large  $f_{\text{NL}}$  [94].

#### 4. Hybrid Inflation with Two Inflaton Fields

We consider a model of two field hybrid inflation, whose potential is given by

$$\begin{aligned} W(\varphi, \chi) &= W_0 \exp\left(\frac{1}{2} \eta_{\varphi\varphi} \frac{\varphi^2}{M_{\text{P}}^2}\right) \exp\left(\frac{1}{2} \eta_{\chi\chi} \frac{\chi^2}{M_{\text{P}}^2}\right) \\ &\simeq W_0 \left(1 + \frac{1}{2} \eta_{\varphi\varphi} \frac{\varphi^2}{M_{\text{P}}^2} + \frac{1}{2} \eta_{\chi\chi} \frac{\chi^2}{M_{\text{P}}^2}\right), \end{aligned} \quad (29)$$

which is vacuum dominated, that is, which satisfies  $|\eta_{\varphi\varphi} \varphi^2| \ll M_{\text{P}}^2$  and  $|\eta_{\chi\chi} \chi^2| \ll M_{\text{P}}^2$ . It is in this regime that the two ways of writing the potential given above are equivalent (at leading order in slow roll), so we can treat the potential as being both sum and product separable. We assume that inflation ends abruptly by a waterfall field which is heavy during inflation and hence does not affect the dynamics during inflation. First, we calculate observables during slow-roll inflation. We will consider the full potential including the waterfall field in Section 4.2, which incorporate the effects from the end of inflation considered in Section 2.3. We will see that this can lead to a change in observables on the surface where the waterfall field is destabilised.

In the vacuum-dominated regime, the slow-roll solutions are

$$\varphi(N) = \varphi_* e^{-\eta_{\varphi\varphi} N}, \quad \chi(N) = \chi_* e^{-\eta_{\chi\chi} N}, \quad (30)$$

where “\*” denotes the value at the horizon exit. Throughout this section, whenever we write a quantity without making it explicit at which time it should be evaluated, we mean the equation to be valid at any time  $N$   $e$ -foldings after Hubble exit and while slow roll is valid. Generally, we will

be interested in quantities at the end of inflation, in which case we take  $N = 60$ .

The slow-roll parameters are

$$\epsilon_{\varphi} = \frac{1}{2} \eta_{\varphi\varphi}^2 \frac{\varphi^2}{M_{\text{P}}^2}, \quad \epsilon_{\chi} = \frac{1}{2} \eta_{\chi\chi}^2 \frac{\chi^2}{M_{\text{P}}^2}, \quad \epsilon = \epsilon_{\varphi} + \epsilon_{\chi}. \quad (31)$$

We note that the dominant slow-roll parameters  $\eta_{\varphi\varphi}$  and  $\eta_{\chi\chi}$  are constants during inflation in the vacuum dominated regime and that they are much larger than the slow-roll parameters  $\epsilon_{\varphi}$  and  $\epsilon_{\chi}$  throughout inflation.

From the previous section, large non-Gaussianity can be realised in either of two regions

$$\begin{aligned} \cos^2 \theta &\equiv \frac{\dot{\varphi}^2}{\dot{\varphi}^2 + \dot{\chi}^2} \simeq \frac{\epsilon_{\varphi}}{\epsilon_{\varphi} + \epsilon_{\chi}} \ll 1, \quad \text{or} \\ \sin^2 \theta &\equiv \frac{\dot{\chi}^2}{\dot{\varphi}^2 + \dot{\chi}^2} \simeq \frac{\epsilon_{\chi}}{\epsilon_{\varphi} + \epsilon_{\chi}} \ll 1. \end{aligned} \quad (32)$$

Since the two regions are symmetrical [94] (before specifying the values of  $\eta_{\varphi\varphi}$  and  $\eta_{\chi\chi}$ ), in the rest of this section, we will focus on the second region. In this region where  $\epsilon_{\varphi} \gg \epsilon_{\chi}$ ,  $|f_{\text{NL}}| > 1$  is fulfilled by the condition

$$\sin^2 \theta^* \lesssim \sin^4 \theta \left( \sqrt{\frac{5 |\eta_{\chi\chi}|}{6 \sin^2 \theta}} - 1 \right), \quad (33)$$

in other words,

$$|\eta_{\chi\chi}|^{-1} e^{-4(\eta_{\varphi\varphi} - \eta_{\chi\chi})N} \lesssim \sin^2 \theta \simeq \frac{\epsilon_{\chi}}{\epsilon_{\varphi}} \lesssim |\eta_{\chi\chi}|. \quad (34)$$

This implies three inequalities on  $\theta$

$$\begin{aligned} \sin^2 \theta^* &< \frac{1}{3} \left(\frac{5}{6}\right)^2 \left(\frac{3}{4}\right)^4 |\eta_{\chi\chi}|^2, \\ \sin^2 \theta &< \frac{5}{6} |\eta_{\chi\chi}|, \quad \frac{\sin^2 \theta}{\sin^2 \theta^*} > \frac{24}{5} \frac{1}{|\eta_{\chi\chi}|}. \end{aligned} \quad (35)$$

Note that in this region  $\sin \theta \simeq \eta_{\chi\chi} \chi / (\eta_{\varphi\varphi} \varphi)$  from (30), we require  $N(\eta_{\varphi\varphi} - \eta_{\chi\chi}) > 1$  so that  $\sin^2 \theta$  grows significantly during inflation.

*4.1. Simplified Formula for the Observables When  $f_{\text{NL}}$  Is Large.* We can substantially simplify all of the above formula in the case where  $f_{\text{NL}}$  is large. We define the quantity

$$\tilde{r} \equiv \frac{(\partial N / \partial \chi_*)^2}{(\partial N / \partial \varphi_*)^2} = \frac{\epsilon_{\chi}}{\epsilon_{\varphi}} e^{2(\eta_{\varphi\varphi} - \eta_{\chi\chi})N}. \quad (36)$$

In the region, we are considering where  $f_{\text{NL}}$  is large, this is approximately given by the initial and final angles of the background trajectory with different exponents

$$\tilde{r} \simeq \frac{\sin^4 \theta}{\sin^2 \theta^*}. \quad (37)$$

In the case of large non-Gaussianity it follows that

$$\mathcal{P}_\zeta \simeq \frac{W_*}{24\pi^2 M_p^4 \epsilon_*} \left( 1 + \frac{\epsilon_\chi}{\epsilon_\varphi} e^{2(\eta_{\varphi\varphi} - \eta_{\chi\chi})N} \right) = \frac{8}{r} \left( \frac{H_*}{2\pi} \right)^2, \quad (38)$$

$$n_\zeta - 1 \simeq 2 \frac{\eta_{\varphi\varphi} + \tilde{r}\eta_{\chi\chi}}{1 + \tilde{r}}, \quad (39)$$

$$r \simeq \frac{16\epsilon_*}{1 + \tilde{r}}, \quad (40)$$

$$f_{\text{NL}} \simeq \frac{5}{6} \frac{\sin^6 \theta^e}{(\sin^2 \theta^* + \sin^4 \theta^e)^2} \eta_{\chi\chi} = \frac{5}{6} \frac{\tilde{r}}{(1 + \tilde{r})^2} \eta_{\chi\chi} e^{2(\eta_{\varphi\varphi} - \eta_{\chi\chi})N}. \quad (41)$$

The first condition in (35) implies that

$$\chi_* \ll \varphi_*. \quad (42)$$

We, therefore, require a very small value of  $\chi_*$  in order to have a large non-Gaussianity. While this may be a fine tuning, we comment that requiring a large ratio of the initial field values is common in models where the isocurvature perturbation generates a significant part of the primordial adiabatic perturbation, see the discussion in the conclusion of [85]. If the inflaton perturbations are neglected a priori, such as in the pure (i.e., not mixed) curvaton scenario, this remains implicitly true. For an explicit discussion of this, in the simple case that both the inflaton and curvaton fields have a quadratic potential, see [108].

The sign of  $f_{\text{NL}}$  is determined by the sign of  $\eta_{\chi\chi}$ . The amplitude of  $f_{\text{NL}}$  depends exponentially on the difference of the slow-roll parameters,  $\eta_{\varphi\varphi} - \eta_{\chi\chi}$ , which we require to be positive to be in the branch of large non-Gaussianity where  $\sin^2 \theta \ll 1$ , while the spectral index depends on a weighted sum of the slow-roll parameters, so it is possible to have a large non-Gaussianity and a scale invariant spectrum. However, it is not possible to have a large and positive  $f_{\text{NL}}$  and a red spectrum of perturbations. We will see in Section 4.2.2 that by including the effect of the waterfall field this conclusion may change, depending on the values of the coupling constants between the two inflaton fields and the waterfall field.

In Table 1, we give some explicit examples of values of  $\eta_{\varphi\varphi}$ ,  $\eta_{\chi\chi}$ ,  $\varphi_*$  and  $\chi_*$  which lead to a large non-Gaussianity. Using (4), we also calculate the spectral index. The first example in Table 1 shows that it is possible to have  $|f_{\text{NL}}| \simeq 100$  and a scale invariant spectrum. We also see that it is possible to generate a large non-Gaussianity during slow roll with  $\eta_{\varphi\varphi}$  and  $\eta_{\chi\chi}$  both positive or both negative, or when one is positive and the other negative corresponding to a saddle point. The results for this model were verified using a novel calculational method of momentum transport equations in [109].

**4.2. Effect of the Waterfall Field and Further Evolution after Inflation.** In this section, we include the effects of the waterfall field  $\rho$  which is required to end hybrid inflation. Inflation ends when the waterfall field is destabilised that is, when its effective mass becomes negative. During inflation,

the waterfall field is heavy, and it is trapped with a vacuum expectation value of zero, so we can neglect it during inflation. The end of inflation occurs when the effective mass of the waterfall field is zero, which occurs on a hypersurface defined in general by [92, 93]

$$\sigma^2 = G(\varphi, \chi) \equiv g_1^2 \varphi^2 + g_2^2 \chi^2, \quad (43)$$

which is realised by the potential  $W(\varphi, \chi)$ , defined by (29), where  $W_0$  is given by

$$W_0 = \frac{1}{2} G(\varphi, \chi) \rho^2 + \frac{\lambda}{4} \left( \rho^2 - \frac{\sigma^2}{\lambda} \right)^2. \quad (44)$$

Here,  $g_1$  ( $g_2$ ) is the coupling between the  $\varphi$  ( $\chi$ ) field and  $\rho$  is the waterfall field with self-coupling  $\lambda$ . In general, the hypersurface defined by this end condition is not a surface of uniform energy density.

As discussed earlier, this is an example of a model with an inhomogeneous end of inflation, that is, where inflation ends at slightly different times in different places. It has also been shown for the hybrid potential we are considering that this can be used to generate a large amount of non-Gaussianity, for certain parameters values and fine tuning of the parameters [90, 93]. However, these papers concern the large non-Gaussianity generated at the end of inflation rather than during slow-roll inflation, by having a very large ratio of couplings  $g_1/g_2 \ll 1$ . Here, we consider the case where  $g_1$  and  $g_2$  have the same order of magnitude with  $g_1^2/g_2^2 = \eta_{\varphi\varphi}/\eta_{\chi\chi}$  in Section 4.2.1 and with  $g_1^2 = g_2^2$  in Section 4.2.2.

**4.2.1.  $g_1^2/g_2^2 = \eta_{\varphi\varphi}/\eta_{\chi\chi}$ .** In this case, we have chosen the coupling constants (which can satisfy  $g_1^2 < 0$  and/or  $g_2^2 < 0$ ) such that the surface where the waterfall field is destabilised corresponds to a surface of uniform energy density. This is because the end condition can be rewritten as  $\sigma^2 = 2g_2^2/\eta_{\chi\chi}((1/2)\eta_{\varphi\varphi}\varphi^2 + (1/2)\eta_{\chi\chi}\chi^2) \propto W$ . In this case, the value of all observable quantities such as the power spectrum and non-Gaussianity are the same as those we calculated previously which were valid at the final hypersurface of uniform energy density during inflation.

**4.2.2.  $g_1^2 = g_2^2$ .** In this case, the end of inflation given by the condition in (43) does not occur on a uniform energy density hypersurface [93]. We will show how the non-Gaussianity is modified by the condition at the end of inflation in this example. In general, we expect there to be some modification to non-Gaussianity from the end of inflation, except in the special case we considered in Section 4.2.1. We have checked in [95] that the correction from the extra expansion which occurs from the surface on which inflation ends up to a surface of uniform energy density is small, and therefore one can still use the  $\delta N$  formalism in this case.

TABLE 1: Table showing some initial conditions for the hybrid inflation model that lead to large levels of non-Gaussianity. The table shows the parameter  $\tilde{r}$ , the bispectrum and the trispectrum nonlinearity parameters, the spectral index and tensor-to-scalar ratio. They are evaluated when the number of  $e$ -foldings from the end of inflation is  $N_k = 60$ . The trispectrum parameters  $\tau_{\text{NL}}$  and  $g_{\text{NL}}$  are defined and calculated in Section 6.1.

$\eta_{\varphi\varphi}$	$\eta_{\chi\chi}$	$\varphi^*$	$\chi^*$	$\tilde{r}$	$f_{\text{NL}}$	$\tau_{\text{NL}}$	$g_{\text{NL}}$	$n_\zeta - 1$	$r$
0.04	-0.04	1	$6.8 \times 10^{-5}$	1	-123	$4.4 \times 10^4$	-33	0	0.006
0.04	-0.04	1	$1.5 \times 10^{-4}$	5	-68	$8 \times 10^3$	-24	-0.05	0.002
0.08	0.01	1	0.0018	1	9.27	247	0.77	0.09	0.026
0.02	-0.04	1	0.00037	1	-11.1	357	-2.6	-0.02	0.002
-0.01	-0.09	1	$3 \times 10^{-6}$	0.16	-132	$1.8 \times 10^5$	-44	-0.04	0.0007
0.06	-0.01	1	$4.3 \times 10^{-4}$	0.1	-3	148	-0.2	0.11	0.026
0.01	-0.06	1	$7.5 \times 10^{-6}$	0.04	-8	$2.5 \times 10^3$	-2	0.01	0.0008

In this case, the power spectrum and  $f_{\text{NL}}$  in the limit of large non-Gaussianity are [93]

$$\mathcal{P}_\zeta \simeq \frac{W_*}{24\pi^2 M_P^4 \epsilon_*} \left( 1 + \frac{\eta_{\varphi\varphi}^2 \epsilon_\chi}{\eta_{\chi\chi}^2 \epsilon_\varphi} e^{2(\eta_{\varphi\varphi} - \eta_{\chi\chi})N} \right) = \frac{8}{\tilde{r}} \left( \frac{H_*}{2\pi} \right)^2, \quad (45)$$

$$n_\zeta - 1 \simeq 2 \frac{\eta_{\varphi\varphi} + (\eta_{\varphi\varphi}^2/\eta_{\chi\chi}^2) \tilde{r} \eta_{\chi\chi}}{1 + (\eta_{\varphi\varphi}^2/\eta_{\chi\chi}^2) \tilde{r}}, \quad (46)$$

$$r \simeq 16\epsilon_* \left( 1 + \frac{\eta_{\varphi\varphi}^2 \tilde{r}}{\eta_{\chi\chi}^2} \right)^{-1}, \quad (47)$$

$$f_{\text{NL}} \simeq \frac{5}{6} \frac{(\eta_{\varphi\varphi}^2/\eta_{\chi\chi}^2) \tilde{r}}{\left( 1 + (\eta_{\varphi\varphi}^2/\eta_{\chi\chi}^2) \tilde{r} \right)^2} \eta_{\varphi\varphi} e^{2(\eta_{\varphi\varphi} - \eta_{\chi\chi})N}. \quad (48)$$

We can see that the observables at the end of inflation are changed by the additional ratio of  $\eta_{\varphi\varphi}^2/\eta_{\chi\chi}^2$  before  $\tilde{r}$ . Furthermore,  $f_{\text{NL}}$  has an additional factor of  $\eta_{\varphi\varphi}/\eta_{\chi\chi}$ .

In Table 2, we show the values of  $f_{\text{NL}}$ ,  $n_\zeta - 1$  and  $r$  with end condition  $g_1^2 = g_2^2$  for the same parameter values that we used in Table 1, where the inflation ends on the uniform energy density hypersurface. The first two examples show that if  $\eta_{\varphi\varphi} = -\eta_{\chi\chi}$ , then the observables are unchanged for two different end conditions except that the sign of  $f_{\text{NL}}$  is switched. The second example in the table shows that in this case it is possible to have a red spectral index and a positive value of  $f_{\text{NL}}$ . For many values of the initial parameters when  $|\eta_{\varphi\varphi}/\eta_{\chi\chi}| \neq 1$ , the magnitude of  $f_{\text{NL}}$  decreases compared to Table 1, but the reverse can also happen, an example of this where  $|f_{\text{NL}}|$  grows by more than an order of magnitude is shown in the final row of Tables 1 and 2.

**4.2.3. Further Evolution after Inflation.** So far, we have assumed a quick transition to the radiation epoch at the end of inflation, thereby neglecting the dynamics of the waterfall field. However, if we consider the role of the waterfall field, then after the waterfall field is destabilised, there may be a further evolution of the primordial curvature perturbation, which will lead to a change of the observable parameters. This applies to any model with an inhomogeneous end of inflation, since there are isocurvature perturbations still

present after the waterfall field is destabilised and inflation has ended. Further evolution will depend on the details of reheating in a model dependent way. To the best of our knowledge, this issue has not been considered in depth in any paper. If we assume an instantaneous transition to radiation domination (so a completely efficient and immediate decay of the waterfall and inflaton fields), then there will be no further change to the observables as we have argued in the previous section. However, this is clearly an idealised case.

In the special case where the waterfall field is also light during inflation, Barnaby and Cline [110] have shown there is the possibility of generating a large non-Gaussianity during preheating for certain parameter values. This is possible even if there is only one inflaton field and the waterfall field present. However, in this case, inflation does not end abruptly when the waterfall field is destabilised so this is not the scenario we have considered in this paper. A recent claim that even a heavy waterfall field may lead to a scale invariant local non-Gaussianity has been withdrawn [111].

## 5. Multiple-Field Inflation without Slow Roll

In this section, we demonstrate a new method that provides, for certain classes of models, analytical expressions for  $f_{\text{NL}}$  valid in regimes *beyond* a slow-roll approximation [112]. This method is based on the first-order Hamilton-Jacobi formalism developed by Salopek and Bond [113], which allows us to express inflationary observables in the multifield case, without having to focus on a slow-roll regime (see [114] for a similar application of this formalism to the single-field case).

The main idea on which our arguments are based is to focus on inflationary multifield trajectories in which the *Hubble rate*, and not the potential, is separable. This approach allows us to analytically study cases where the non-Gaussian parameter  $f_{\text{NL}}$  becomes large, in regimes in which the slow-roll parameters are enhanced after Hubble exit, but before inflation ends. In particular, we provide analytical equations that express the nonlinearity parameter  $f_{\text{NL}}$ , in terms of quantities that generalise the slow-roll parameters during inflation. Using this method, we demonstrate in the next subsection an exact solution of multifield inflation

TABLE 2: Same as Table 1, but with different end condition,  $g_1^2 = g_2^2$  as used in Section 4.2.2.

$\eta_{\varphi\varphi}$	$\eta_{\chi\chi}$	$\varphi_*$	$\chi_*$	$\tilde{r}$	$f_{\text{NL}}$	$n_\zeta - 1$	$r$
0.04	-0.04	1	$6.8 \times 10^{-5}$	1	123	0	0.006
0.04	-0.04	1	$1.5 \times 10^{-4}$	5	68	-0.05	0.002
0.08	0.01	1	0.0018	1	4.59	0.02	0.0008
0.02	-0.04	1	0.00037	1	3.5	0.02	0.026
-0.01	-0.09	1	$3 \times 10^{-6}$	0.16	-0.2	-0.02	0.0008
0.06	-0.01	1	$4.3 \times 10^{-4}$	0.1	38	0.01	0.006

which can give rise to a large non-Gaussianity due to a strong break down in slow roll shortly before inflation ends.

The first-order Hamilton-Jacobi equations of motion are

$$H^2 = \frac{1}{3}W(\phi, \chi) + \frac{2}{3} \left[ \left( \frac{\partial H}{\partial \phi} \right)^2 + \left( \frac{\partial H}{\partial \chi} \right)^2 \right], \quad (49)$$

$$\dot{\phi} = -2 \frac{\partial H}{\partial \phi}, \quad \dot{\chi} = -2 \frac{\partial H}{\partial \chi}.$$

We concentrate on solutions satisfying the following Ansatz for the Hubble parameter  $H(\phi, \chi)$

$$H(\phi, \chi) = H^{(1)}(\phi) + H^{(2)}(\chi), \quad (50)$$

that is, we demand that it can be split as a sum (hence, we follow an analysis related to [99]. In an analogous way, one could also consider a situation in which the Hubble parameter splits as a *product* of single-field pieces [104]. The formula's in this section were extended to an arbitrary number of fields with a sum separable Hubble parameter in [115]) of two pieces, each one depending on a single field. Notice that this is different with respect to the requirement of separable potentials usually done in the literature and earlier in this paper.

It is convenient to define the following quantities:

$$\delta^\phi = \left( \frac{H_{,\phi}^{(1)}}{H} \right)^2, \quad \delta^\chi = \left( \frac{H_{,\chi}^{(2)}}{H} \right)^2, \quad (51)$$

$$\gamma^\phi = \frac{H_{,\phi\phi}^{(1)}}{H}, \quad \gamma^\chi = \frac{H_{,\chi\chi}^{(2)}}{H},$$

and  $\delta = \delta^\phi + \delta^\chi$ . Although their definition resembles the corresponding one for the usual slow-roll parameters  $\epsilon$  and  $\eta$ , they do *not* coincide with them when taking a slow roll limit (it is nevertheless simple to work out the relation with the slow roll parameters, defined by (21) and (22)). In a slow-roll regime, one finds  $\delta^\phi \simeq \epsilon^\phi/2$ ,  $\gamma^\phi \simeq (\eta^{\phi\phi} - \epsilon^\phi)/2$ . We assume that the quantities  $\delta$  and  $\gamma$  are much smaller than unity at Hubble exit  $t = t_*$ , in order that we can use the  $\delta N$  formalism. During inflation, by definition, we have to ensure that the quantity  $\epsilon_H \equiv -\dot{H}/H^2 = 2\delta < 1$ . The quantities  $\gamma$  can, however, become much larger than unity during inflation.

Using similar techniques to those applied to a separable potential, one can calculate the spectral index and  $f_{\text{NL}}$ , the details were given in [112] and the results are

$$n_\zeta - 1 = -4\delta_* - 4 \frac{u_H (1 - (\gamma_*^\phi/\delta_*^\phi) u_H)}{u_H^2/\delta_*^\phi + v_H^2/\delta_*^\chi} + (-4) \frac{v_H (1 - (\gamma_*^\chi/\delta_*^\chi) v_H)}{u_H^2/\delta_*^\phi + v_H^2/\delta_*^\chi} \quad (52)$$

$$\frac{6}{5} f_{\text{NL}} = 2 \frac{(u_H^2/\delta_*^\phi) (1 - (\gamma_*^\phi/\delta_*^\phi) u_H)}{(u_H^2/\delta_*^\phi + v_H^2/\delta_*^\chi)^2} + 2 \frac{(v_H^2/\delta_*^\chi) (1 - (\gamma_*^\chi/\delta_*^\chi) v_H)}{(u_H^2/\delta_*^\phi + v_H^2/\delta_*^\chi)^2} + 2 \frac{(u_H/\delta_*^\phi - v_H/\delta_*^\chi)^2 \mathcal{A}_H}{(u_H^2/\delta_*^\phi + v_H^2/\delta_*^\chi)^2}, \quad (53)$$

where the new symbols are defined by

$$u_H \equiv \frac{H_*^{(1)} + Z_e}{H_*}, \quad v_H \equiv \frac{H_*^{(2)} - Z_e}{H_*},$$

$$Z_e = \frac{(H_e^{(2)} \delta_e^\phi - H_e^{(1)} \delta_e^\chi)}{\delta_e}, \quad (54)$$

$$\mathcal{A}_H = -\frac{H_e^2}{H_*^2} \frac{\delta_e^\phi \delta_e^\chi}{\delta_e} \left( \frac{1}{2} - \frac{\gamma_e^{ss}}{\delta_e} \right),$$

$$\gamma^{ss} = \frac{(\delta^\chi \gamma^\phi + \delta^\phi \gamma^\chi)}{\delta}.$$

We reiterate that the formula for  $f_{\text{NL}}$  is exact (apart from neglecting a slow-roll suppressed contribution due to the non-Gaussianity of the fields at Hubble exit) and is not based on a slow-roll expansion. Although  $\delta_*$  and  $\gamma_*$  are necessarily small, the quantity  $\mathcal{A}$  is not suppressed by these parameters and can assume large values during inflation, enhancing  $f_{\text{NL}}$ . We note in agreement with [99] that if one of the fields has reached a minimum so  $\dot{\phi} = 0$  or  $\dot{\chi} = 0$  at the end of inflation then  $\mathcal{A}_H = 0$  and the non-Gaussianity at the end of inflation will be very small. Therefore, any model of inflation with a separable potential or separable Hubble factor with a large non-Gaussianity present at the end of inflation must

have both fields still evolving, and therefore the presence of isocurvature modes. It would, therefore, also be interesting to study the evolution of the perturbations after inflation, to see if this leads to an important change in the observables [116].

*5.1. Exact Solution.* Very few exact solutions in multiple-field inflation are known [117]. Here, we present an exact solution which can give rise to a very large non-Gaussianity at the end of inflation. A potential with a similar form may be motivated in string theory, in the context of Kähler moduli inflation [118–125].

The potential we consider is [112]

$$W(\phi, \chi) = U_0 \left( -A_1 e^{-\alpha\phi} + A_2 e^{-2\alpha\phi} - B_1 e^{-\beta\chi} + B_2 e^{-2\beta\chi} + \frac{A_1 B_1}{2} e^{-\alpha\phi - \beta\chi} \right), \quad (55)$$

where the parameters  $\alpha$  and  $\beta$  satisfy

$$\alpha^2 = \frac{3}{2} - \frac{6A_2}{A_1^2}, \quad \beta^2 = \frac{3}{2} - \frac{6B_2}{B_1^2}. \quad (56)$$

The Hamilton-Jacobi equations have the following exact solution for  $H$  and the fields:

$$\begin{aligned} H &= H_0 \left( 1 - \frac{A_1}{2} e^{-\alpha\phi} - \frac{B_1}{2} e^{-\beta\chi} \right), \\ \phi &= \frac{1}{\alpha} \ln \left[ e^{\alpha\phi_*} - A_1 \alpha^2 H_0 t \right], \\ \chi &= \frac{1}{\beta} \ln \left[ e^{\beta\chi_*} - B_1 \beta^2 H_0 t \right], \end{aligned} \quad (57)$$

where we have defined  $H_0 = \sqrt{U_0/3}$ ,  $t$  is cosmic time, and we have set  $t_* = 0$ . The overall factor to the potential  $U_0$  can be freely chosen so that the amplitude of the scalar power spectrum matches the observed amplitude of perturbations in the CMB. The scale factor results

$$a(t) = a_0 \left( e^{\alpha\phi_*} - \alpha^2 A_1 H_0 t \right)^{1/2\alpha^2} \left( e^{\beta\chi_*} - \beta^2 B_1 H_0 t \right)^{1/2\beta^2} e^{H_0 t}. \quad (58)$$

Notice that the solution becomes singular at late times, when the scale factor vanishes and the field values diverge. This singularity occurs well after inflation ends, and we will not need to discuss it in our analysis. From the relations

$$\gamma^\phi = -\alpha\sqrt{\delta^\phi}, \quad \gamma^\chi = -\beta\sqrt{\delta^\chi}, \quad (59)$$

we notice that in this example  $\gamma^\phi$  and  $\gamma^\chi$  can become much larger than unity at the end of inflation if  $|\alpha|$  and/or  $|\beta|$  are much greater than one. This corresponds to a break down in slow roll, although  $\epsilon_H$  remains smaller than unity during the inflationary era by definition.

As a simple, concrete example of parameter choice which gives rise to a large non-Gaussianity, we choose a regime in which the parameters  $|\alpha|$  and  $|\beta|$  are both large, let us say larger than some quantity  $R \gg 1$ . We note from (56) that

we are, therefore, required to have  $A_2 < 0$  and  $B_2 < 0$ . In this case, the potential does not have a minimum, but inflation still ends through  $\epsilon_H$  growing larger than unity, and we can trust our results in this regime. The potential would need modifications which apply after inflation in order for reheating to take place and these modifications may provide a minimum for the potential. We parameterise the values of the quantities  $\delta_e^\phi$  and  $\delta_e^\chi$  at the end of inflation as

$$\delta_e^\phi = \frac{1}{2m}, \quad \delta_e^\chi = \frac{1}{2n} \quad \text{such that} \quad \frac{1}{m} + \frac{1}{n} = 1, \quad (60)$$

where the final equality follows from  $\epsilon_H = 1$ . We assume that the quantities  $m$  and  $n$  are not too big. Namely, they satisfy the inequality  $m, n \ll R^2$ . This implies that  $H_0 \simeq H_* \simeq H_e$ . We choose  $|\gamma_*^\phi| = 1/40p$ ,  $|\gamma_*^\chi| = 1/40q$ , such that we can write

$$N_{\text{tot}} = 60 \simeq 10(p + q). \quad (61)$$

This fixes the initial values for the fields  $\phi_*$  and  $\chi_*$ , to the values

$$\begin{aligned} e^{\alpha\phi_*} &= 20p\alpha^2 A_1, \\ e^{\beta\chi_*} &= 20q\beta^2 B_1. \end{aligned} \quad (62)$$

The conditions (60) imply

$$\begin{aligned} H_0 t_e &= 20p - \sqrt{\frac{m}{2\alpha^2}}, \\ H_0 t_e &= 20q - \sqrt{\frac{n}{2\beta^2}}. \end{aligned} \quad (63)$$

Since  $m$  and  $n$  are much smaller than  $R^2$ , we expect  $p \simeq q$ . We hence have the simple relations

$$\delta_*^\phi \simeq \frac{1}{4N_{\text{tot}}^2 \alpha^2}, \quad \delta_*^\chi \simeq \frac{1}{4N_{\text{tot}}^2 \beta^2}, \quad \gamma_*^\phi \simeq \gamma_*^\chi \simeq -\frac{1}{2N_{\text{tot}}}. \quad (64)$$

From (52), we find

$$n_\zeta - 1 \simeq -\frac{2}{N_{\text{tot}}} \simeq -0.04. \quad (65)$$

So, we have a red spectral index in agreement with present day observations [126] and in [112] it was also shown that the tensor-to-scalar ratio for this model is negligible.

When evaluated at the end of inflation, we find

$$\frac{6}{5} f_{\text{NL}} \simeq -2\sqrt{2mn} \left[ \frac{\alpha^2 n - \beta^2 m}{\alpha^2 n^2 + \beta^2 m^2} \right]^2 \left( \frac{\alpha}{\sqrt{n}} + \frac{\beta}{\sqrt{m}} \right). \quad (66)$$

We stress that this formula provide only the dominant contributions to  $f_{\text{NL}}$ . It is valid in the case  $|\alpha|, |\beta| > R \gg 1$ , and  $m, n < R^2$ .

As an explicit case, we take  $\alpha = -100$ ,  $\beta = 20$ ,  $m = 6$ ,  $n = 6/5$  and the initial conditions to satisfy (62) with  $N_{\text{tot}} = 60$ .

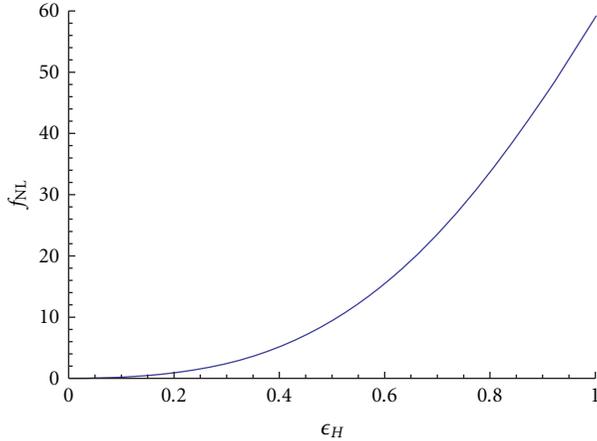


FIGURE 1: Plot showing  $f_{\text{NL}}$  as a function of  $\epsilon_H$  towards the end of inflation, for the values of the parameters given in the text. Inflation ends when  $\epsilon_H = 1$ ; for this example  $f_{\text{NL}} \approx 59$  at that time.

Then, independently of the values of  $A_1$  and  $B_1$ , we find  $f_{\text{NL}} \approx 58$  from the simplified formula (66), and  $f_{\text{NL}} \approx 53$  from the full formula (53). See Figures 1 and 2. Notice that the results for  $f_{\text{NL}}$  differ from each other by quantities of order  $1/R = 1/\beta$  as expected by the approximations we have made. At the end of inflation, the parameters are  $\gamma^\phi = 29$ ,  $\gamma^\chi = -13$  and this shows that the slow-roll approximation has been significantly broken, (it is nevertheless simple to work out the relation with the slow roll parameters, defined by (21) and (22). In a slow-roll regime, one finds  $\delta^\phi \simeq \epsilon^\phi/2$ ,  $\gamma^\phi \simeq (\eta^{\phi\phi} - \epsilon^\phi)/2$ ). From Figure 2, one can see that the trajectory is straight for much of inflation but it turns near the end of inflation (during the last  $e$ -folding before inflation ends), and it is during this time that  $f_{\text{NL}}$  grows larger than unity.

Notice that our requirements of final values for the quantities  $\delta_e^\phi$  and  $\delta_e^\chi$  (both much bigger than  $1/R$ ) imposes fine-tuning constraints on the initial conditions, since at leading order in  $1/R$  the values for  $\gamma_*^\phi$  and  $\gamma_*^\chi$  must coincide (see (62) and recall that  $p \simeq q$ ).

We stress that the enhancement of non-Gaussianity occurs towards the end of inflation. What happens just *after* inflation is a model-dependent issue; we cannot address this question within the approximations used in this concrete model. For our choice of parameters, we notice that  $f_{\text{NL}}$  is still increasing at the end of inflation, see Figure 1, but for other choices it may start to decrease before inflation ends, see [115]. We stress that in the regime where  $|\alpha|, |\beta| \gg 1$  the slow-roll parameters will necessarily become much greater than unity by the end of inflation, which may correspond to  $|\gamma^{\text{ss}}| \gg 1$ : this is one reason by which  $f_{\text{NL}}$  can become large. Our formalism is at least in principle suitable to study the evolution of non-Gaussianity after inflation, when the parameter  $\epsilon_H$  becomes larger than unity (although we restate that the specific potential being considered here anyway needs to be modified after inflation ends).

## 6. Higher Order Non-Gaussian Observables

Non-Gaussianity is most popularly parameterised in terms of one nonlinearity parameter,  $f_{\text{NL}}$ . There are several different  $f_{\text{NL}}$ 's used in the literature which parameterise different shape dependences of the bispectrum [41, 42], the most popular being the local model discussed in this paper and the so-called equilateral model which can arise from models of inflation with noncanonical kinetic terms (a popular example is DBI inflation [15–25]). But can we learn more than one number from an observation of non-Gaussianity? And if  $f_{\text{NL}}^{\text{local}}$  is detected how can we hope to distinguish between the models which can generate this? Fortunately, observations of non-Gaussianity have the possibility of teaching us much more than the value of one parameter. Two examples of further non-Gaussian observables are a scale dependence of  $f_{\text{NL}}$  and the trispectrum.

**6.1. Trispectrum.** The four point function of the primordial curvature perturbation is defined by

$$\begin{aligned} & \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \zeta_{\mathbf{k}_4} \rangle_e \\ & \equiv T_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) (2\pi)^3 \delta^3(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 + \mathbf{k}_4), \end{aligned} \quad (67)$$

which using the  $\delta N$  formalism, and assuming that the fields have a Gaussian distribution at Hubble exit, is given by

$$\begin{aligned} & T_\zeta(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \mathbf{k}_4) \\ & = \tau_{\text{NL}} [P_\zeta(k_{13})P_\zeta(k_3)P_\zeta(k_4) + (11 \text{ perms})] \\ & \quad + \frac{54}{25} g_{\text{NL}} [P_\zeta(k_2)P_\zeta(k_3)P_\zeta(k_4) + (3 \text{ perms})], \end{aligned} \quad (68)$$

where  $k_{13} \equiv |\mathbf{k}_1 + \mathbf{k}_3|$ , and the trispectrum nonlinearity parameters can be calculated using [98, 106, 127]

$$\tau_{\text{NL}} = \frac{N_{AB}N^ACN^BN^C}{(N_D N^D)^3}, \quad (69)$$

$$g_{\text{NL}} = \frac{25 N_{ABC}N^AN^BN^C}{54 (N_D N^D)^3}.$$

Hence, we see that the trispectrum depends on two nonlinearity parameters (as opposed to one,  $f_{\text{NL}}$ , for the bispectrum), and they may be observationally distinguishable since they are prefactors of terms with different shape dependences (68). The current observational bound on the local type of the bispectrum from seven years of WMAP data is  $-10 < f_{\text{NL}} < 74$  at the  $2\sigma$  level [2]. Recently, there has been a first constraint on both  $\tau_{\text{NL}}$  and  $g_{\text{NL}}$ , which came from using WMAP5 data, with the bounds  $-7.4 < g_{\text{NL}}/10^5 < 8.2$  and  $-0.6 < \tau_{\text{NL}}/10^4 < 3.3$  at 95% confidence [128]. There have been two other observational constraints on the trispectrum through  $g_{\text{NL}}$ , setting  $\tau_{\text{NL}} = 0$ . The bounds, which in both cases are roughly  $|g_{\text{NL}}| \lesssim 10^5$ – $10^6$ , come from large scale structure [129], and the CMB [130]. If there is no detection of non-Gaussianity it is expected that with Planck data the bounds will be reduced to about  $|f_{\text{NL}}| \lesssim 10$ ,  $\tau_{\text{NL}} \lesssim 10^3$  and  $g_{\text{NL}} \lesssim 10^5$  at the  $2\sigma$  level and future observations

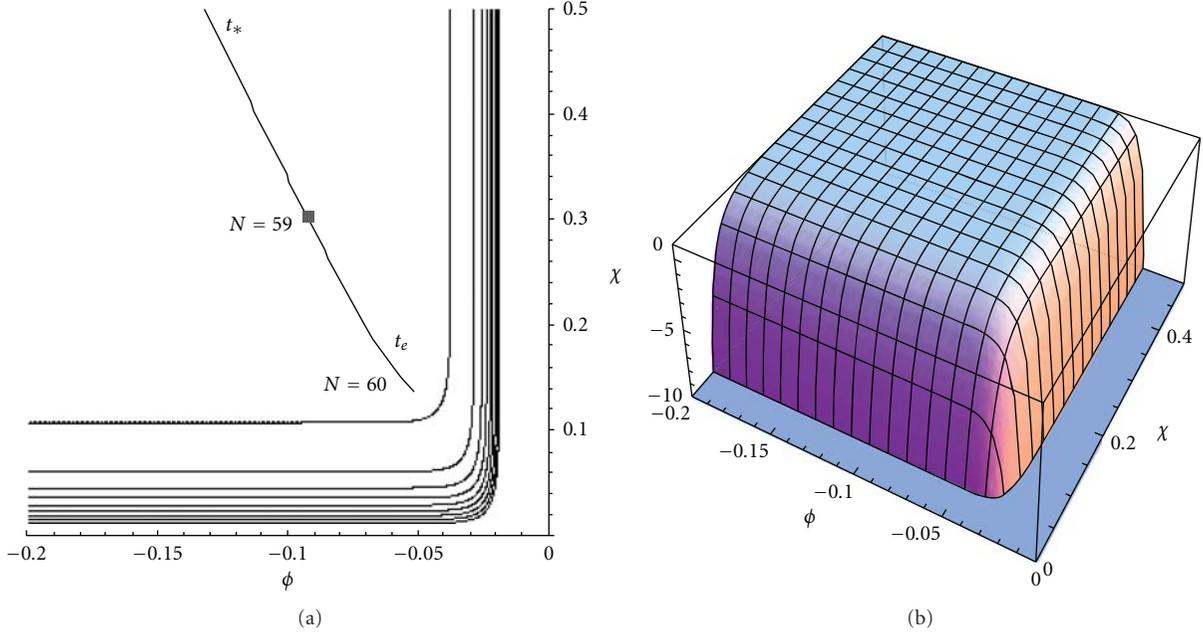


FIGURE 2: (a) shows the trajectory considered for the parameters given after (66) superimposed on a contour plot of the potential. The trajectory is shown from the time of horizon crossing at  $t_*$  and runs until the end of inflation at  $N = 60$ . The square on the trajectory, labelled with  $N = 59$ , indicates a point along the trajectory one  $e$ -folding before inflation ends as  $\phi$  and  $\chi$  roll towards zero. This shows that the fields roll much more quickly during the final stage of inflation, and the trajectory curves near the end. (b) shows the potential for the same parameter values. Notice that inflation ends on the plateau long before the potential becomes negative.

may become even tighter [131–135]. Note that the expected future constraint on  $g_{\text{NL}}$  is about two orders of magnitude weaker than that on  $\tau_{\text{NL}}$  [135]. We will see that it is also possible for some models and parameter ranges that the first observational signature will come through the trispectrum rather than the bispectrum.

**6.1.1. Trispectrum in Two-Field Hybrid Inflation.** In the regime where  $|g_{\text{NL}}|$  and  $\tau_{\text{NL}}$  are greater than unity, they are given by

$$g_{\text{NL}} = \frac{10}{3} \frac{\tilde{r}(\eta_{\phi\phi} - 2\eta_{\chi\chi}) - \eta_{\chi\chi}}{1 + \tilde{r}} f_{\text{NL}}, \quad (70)$$

$$\tau_{\text{NL}} = \frac{\tilde{r}}{(1 + \tilde{r})^3} \eta_{\chi\chi}^2 e^{4N(\eta_{\phi\phi} - \eta_{\chi\chi})} = \frac{1 + \tilde{r}}{\tilde{r}} \left( \frac{6}{5} f_{\text{NL}} \right)^2. \quad (71)$$

We see that  $g_{\text{NL}}$  is subdominant to  $f_{\text{NL}}$  and hence will not provide a competitive observational signature. It follows from (71) that  $\tau_{\text{NL}} > (6f_{\text{NL}}/5)^2$ , so  $\tau_{\text{NL}}$  may be large and provide an extra observable parameter for this model. This inequality between  $\tau_{\text{NL}}$  and  $f_{\text{NL}}$  is true in general [75], and equality is reached whenever a single field direction during inflation generates the primordial curvature perturbation. However, it is usually assumed that  $\tau_{\text{NL}} \sim f_{\text{NL}}^2$  since both arise from second derivatives in the  $\delta N$  formalism. In fact, for our model it is possible to have a small  $f_{\text{NL}}$  (and hence also a small  $g_{\text{NL}}$ ) but a large and potentially observable  $\tau_{\text{NL}}$ . For this, we require that  $\tilde{r} \ll 1$ , although in practice if we make it too

small it may no longer be possible to satisfy a constraint on a minimum possible field velocity consistent with a classical slow-roll trajectory, as discussed in [95] (see also [136] and a discussion which reaches a different conclusion is given in [137]). In the final example in Table 1, we give an explicit example of parameter values which give rise to an  $f_{\text{NL}}$  which is probably too small to be detected with Planck but with a very large trispectrum through  $\tau_{\text{NL}} > 10^3$  that should be detectable at a high significance. For another example with  $f_{\text{NL}}, g_{\text{NL}} \lesssim O(1)$  but  $\tau_{\text{NL}} \gg 1$ , see [76]. In contrast, it has been shown in several papers [127, 138–142] that in the curvaton scenario where the curvaton has a nonquadratic potential it is possible to realise  $|g_{\text{NL}}| \gg 1$  while  $\tau_{\text{NL}} = (6f_{\text{NL}}/5)^2$  is small with some tuning of parameters. This is also possible in the exact solution which we presented in Section 5 and we consider this next.

**6.1.2. Trispectrum in the Exact Solution.** In this regime, where  $|\alpha| \gg |\beta| > R$  and  $m, n > 1/R^2$ , it is also possible to give compact expressions for the trispectrum (4-point function) nonlinearity parameters, in terms of  $f_{\text{NL}}^2$  as

$$\tau_{\text{NL}} \simeq \left( \frac{6}{5} f_{\text{NL}} \right)^2, \quad \frac{54}{25} g_{\text{NL}} \simeq -\frac{3}{2} \frac{n - m}{m} \left( \frac{6}{5} f_{\text{NL}} \right)^2. \quad (72)$$

The complete result was given in [115]. Hence, both of the trispectrum nonlinearity parameters are generally large whenever  $f_{\text{NL}}$  is. We note that if  $m \simeq 1$  then from (60)  $n \gg 1$ , so the trispectrum through  $g_{\text{NL}}$  will give the dominant signal of non-Gaussianity through a large, negative  $g_{\text{NL}}$ .

6.2. *Scale Dependence of  $f_{\text{NL}}$ .* In its simplest form, the local form of  $f_{\text{NL}}$  defined by (1) is a constant parameter, independent of both position and scale. However, in realistic models  $f_{\text{NL}}$  is likely to be mildly scale-dependent. This may happen in two ways. In general, single-field models (such as the curvaton scenario) where the scalar field which generates the primordial curvature perturbation has a nonquadratic potential the non-linearities this generates will give rise to scale dependence of  $f_{\text{NL}}$ . Alternatively, even in models where all of the fields have a quadratic potential; if the primordial curvature perturbation has contributions from more than one field, and the fields do not all have the same mass, then  $f_{\text{NL}}$  will again have a scale dependence. This is because the correlation between the first and second order terms of  $\zeta$ , which the bispectrum depends on, will become scale-dependent. This is indeed what happens in the two-field hybrid model presented earlier. For details about a scale dependence of local  $f_{\text{NL}}$ , see Byrnes et al. [143].

We define a second observable parameter derived from the bispectrum

$$n_{f_{\text{NL}}} = \frac{d \ln |f_{\text{NL}}|}{d \ln k}, \quad (73)$$

which is analogous to the scale dependence of the power spectrum. There is a subtlety here, which is that in general  $f_{\text{NL}}$  may depend on three independent parameters,  $k_1, k_2$ , and  $k_3$ . However, it was shown in [143] that provided one takes the derivative while keeping the ratio of the three  $k$  vectors fixed then  $n_{f_{\text{NL}}}$  is independent of the shape of the triangle described by the three  $k$  vectors, which makes this a well defined quantity. Observational prospects for this quantity were considered in [144], who showed that the Planck satellite is sensitive to  $n_{f_{\text{NL}}} \simeq 0.1$ , assuming a fiducial value of  $f_{\text{NL}} = 50$ . This observational sensitivity to  $n_{f_{\text{NL}}}$  is about a factor of two larger than the current preferred value of the power spectrum's spectral index and may therefore provide an interesting extra constraint on non-Gaussian models. The scale dependence of an equilateral form of  $f_{\text{NL}}$  has also been considered from both a theoretical and observational perspective, see for example [144–147].

In the example of two-field hybrid inflation,  $f_{\text{NL}}$  has a scale dependence both because of the exponential term in  $f_{\text{NL}}$ , (41), and because  $\tilde{r}$  will vary through the change of the initial value of  $\sin^2 \theta^*$ . We find

$$\frac{\partial \ln \tilde{r}}{\partial \ln k} = \frac{\partial \ln e^{2N(\eta_{\phi\phi} - \eta_{\chi\chi})}}{\partial \ln k} = -2(\eta_{\phi\phi} - \eta_{\chi\chi}). \quad (74)$$

Using this we find from (41) that

$$n_{f_{\text{NL}}} \equiv \frac{d \log f_{\text{NL}}}{d \log k} = -4 \frac{\eta_{\phi\phi} - \eta_{\chi\chi}}{1 + \tilde{r}}. \quad (75)$$

In the case that we include the effect from the surface where the waterfall field is destabilised and  $g_1^2 = g_2^2$ , we find from (48) that

$$n_{f_{\text{NL}}} = -4 \frac{\eta_{\phi\phi} - \eta_{\chi\chi}}{1 + (\eta_{\phi\phi}/\eta_{\chi\chi})^2 \tilde{r}}. \quad (76)$$

For both cases, the spectral index of  $f_{\text{NL}}$  satisfies

$$-4(\eta_{\phi\phi} - \eta_{\chi\chi}) < n_{f_{\text{NL}}} < 0, \quad (77)$$

for any value of  $\tilde{r}$  and hence  $f_{\text{NL}}$  will be smaller on small scales.

Because we require a relatively large value of  $\eta_{\phi\phi} - \eta_{\chi\chi} > 1/N$  for our model to generate a large non-Gaussianity it is quite possible for our model to generate a relatively significant scale dependence of  $f_{\text{NL}}$ . However, the amount also depends on  $\tilde{r}$ , and when this is large, then the  $\chi$  field is almost solely responsible for generating  $\zeta$  at both first and second order and  $n_{f_{\text{NL}}}$  is suppressed. We can also see in agreement with the statement at the beginning of this subsection that in this case, if the mass of the two fields are equal, then the two fields have the same scale dependence and  $n_{f_{\text{NL}}} = 0$ .

We note that this is in contrast to the large non-Gaussianity from an inhomogeneous end of inflation found in [93]. In the specific cases, they considered to generate a large non-Gaussianity the non-Gaussianity was generated purely at the end of inflation and  $f_{\text{NL}}$  is scale-independent. In detail, we see from (4.4) and (4.24) in [93] that their formulas for  $f_{\text{NL}}$  does not depend on  $N$  or on any quantities evaluated at Hubble exit. This is also in contrast to the exact solution considered in Section 5.1. Our expression for  $f_{\text{NL}}$  in (66) depends on the initial values through  $p$  and  $q$ , and this dependence drops out at leading order in  $\alpha$ . Hence,  $f_{\text{NL}}$  is *independent* of the number of  $e$ -foldings and scale-independent.

## 7. Conclusions

We have reviewed various models which can generate a large local non-Gaussianity. A feature shared by all of these models is that they have more than one light scalar field present during inflation. This extra degree of freedom generates an isocurvature perturbation which is at least partially converted into the primordial curvature perturbation after horizon exit of the modes which are observable today. In the curvaton and modulated reheating scenarios, this conversion occurs after the end of inflation, while in the inhomogeneous end of inflation scenario, this conversion occurs on the non-uniform energy density hypersurface on which inflation ends. For these three scenarios, the light field which generates the primordial curvature perturbation after or at the end of inflation can be treated as a test field which does not affect the inflationary dynamics.

Our main focus has been on models in which a large non-Gaussianity is generated during inflation. This can occur even within slow-roll inflation for certain potentials and certain trajectories. We have shown, at least in the case of a separable potential, that the trajectory is required to be almost entirely along the direction of one field but that the orthogonal field must become more important towards the end of inflation, and hence the inflationary trajectory must curve. In absolute terms, the change to the angle of the

background trajectory is small (compared to a trajectory which turns by a right angle during inflation), but in relative terms it must grow by at least an order of magnitude. This is in contrast to the previous three scenarios.

We have reviewed two-field hybrid inflation as an explicit model of a separable potential where the conditions required to generate a large non-Gaussianity can be satisfied. The conditions can be satisfied for any possible combination of positive and negative  $\eta$  parameters, so the potential can be bowl-shaped, a hill-top, or have a saddle point. The main conditions which must be satisfied is that the difference of the two  $\eta$  parameters must not be too small,  $\eta_{\varphi\varphi} - \eta_{\chi\chi} \sim 0.1$ , and the value of the  $\chi$  field must be very subdominant to that of the  $\varphi$  field initially (or vice versa). In general, but depending on the coupling constants between the two inflaton fields and the waterfall field, there is a change to observables at the end of inflation, due to the fact that the surface on which the waterfall field is destabilised and inflation ends might not be a surface of uniform energy density. This effect is responsible for the inhomogeneous end of inflation scenario. It is then a model, dependent question whether there will be further evolution to the observables during reheating in this model, this deserves further attention.

One similarity that this hybrid inflation model has together with the quadratic curvaton scenario is that in both cases the initially subdominant (approximately isocurvature) field  $\chi$  has the ratio  $\delta\chi/\chi \sim \zeta_\chi$  approximately constant and the field fluctuations do not become more non-Gaussian with time. However, the effect of this light field on the primordial curvature perturbation grows, during inflation in the hybrid scenario and before the curvaton decay in the curvaton scenario. It is this nonlinear transfer between the field fluctuation and  $\zeta$ , described by the  $\delta N$  formalism, which can generate a large non-Gaussianity. Therefore, the non-Gaussianity in the hybrid scenario which we have studied is generated on super horizon scales during slow-roll inflation, in a similar way to which non-Gaussianity is generated over time in the curvaton scenario before the decay of the curvaton. The evolution of  $f_{\text{NL}}$  during inflation is explicitly calculated and plotted in [95]. This conclusion is somewhat different from that in [47], and we plan to elaborate on this point in a future work. For more discussion on the distinction between non-Gaussianity generated by a non-Gaussian field perturbation, and non-Gaussianity generated by a nonlinear transfer between a Gaussian field perturbation and  $\zeta$ , see [109]. An example where the subdominant fields fluctuations can become non-Gaussian due to a large self interaction was discussed by Bernardeau [45].

In order to study models where slow-roll breaks down before the end of inflation, it is clearly necessary to go beyond a formalism based on the slow-roll approximation. We have shown how this can be done in the context a separable Hubble parameter instead of a separable potential and this leads to an exact expression for (the local part of)  $f_{\text{NL}}$  in these models. As an explicit example, an exact two-field solution with an exponential potential was given. For some parameter choices, this leads to a strong break down of slow roll before the end of inflation, which may give rise to a large non-Gaussianity. Further work is also required for this model to

understand how the potential may be modified after the end of inflation in order that reheating occurs.

Non-Gaussianity is a topical field, in which observations have improved greatly over the last decade through both studies of the CMB and large-scale structure. Observations so far have heavily focused on constraining the bispectrum nonlinearity parameter  $f_{\text{NL}}$ . Currently, the tightest constraint comes from the WMAP satellite, assuming the local model of non-Gaussianity this constrains the amplitude of the non-Gaussian part of the primordial curvature perturbation to be less than about one thousandth the amplitude of the Gaussian perturbation. This constraint is likely to be tightened considerably by the Planck satellite, which is currently taking data, or instead there might be a detection. A detection of  $f_{\text{NL}}$  at this level would rule out simplest models of inflation, which are single field with a canonical kinetic term. Clearly, this would be an extremely exciting result.

However, even if we are in the fortunate position of having a detection of  $f_{\text{NL}}$  as well as improved constraints/detection of the scalar-to-tensor ratio and the spectral index, there will probably still be several viable scenarios, as detailed in this paper, which for suitable parameter choices and initial conditions can match the observations. Fortunately, non-Gaussianity is about much more than one number. The trispectrum (four-point function) depends on two nonlinearity parameters. In general,  $\tau_{\text{NL}} \geq (6f_{\text{NL}}/5)^2$ . If the current observational hints (which are not statistically significant) that  $f_{\text{NL}} \sim 40$  turn out to be true, then both the bispectrum and the trispectrum should be large enough for Planck to detect. Even if the bispectrum turns out to be much smaller, although for many models  $\tau_{\text{NL}}$  is close to the lower bound, we have seen that in the model of hybrid inflation it is possible to have  $\tau_{\text{NL}} \gg f_{\text{NL}}^2$ , so the trispectrum might even be the first observational signature of non-Gaussianity. Alternatively, the trispectrum through a large  $g_{\text{NL}}$  might give the first observational signature, as is possible in self-interacting curvaton models or the exact solution with an exponential potential. If  $f_{\text{NL}}$  is detected, it will also be possible to either constrain or detect a scale dependence of this parameter. Although it is often assumed to be constant, this is only true for certain simple models, and for example in the two-field hybrid inflation model it generally has a significant scale dependence. We have, therefore, seen that non-Gaussianity is an important and powerful method of constraining and distinguishing between the many models of inflation.

## Acknowledgments

The authors are extremely grateful to their collaborators with whom they have worked on numerous projects related to non-Gaussianity, which part of this paper is based on. The authors thank Takahiro Tanaka, Teruaki Suyama, and Shuichiro Yokoyama for discussions. K.Y. Choi was partly supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-341-C00008) and by the second stage of Brain Korea 21 Project in 2006.

## References

- [1] E. Komatsu and D. N. Spergel, "Acoustic signatures in the primary microwave background bispectrum," *Physical Review D*, vol. 63, no. 6, Article ID 063002, 2001.
- [2] E. Komatsu, K. M. Smith, J. Dunkley, et al., "Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation," <http://arxiv.org/abs/1001.4538>.
- [3] E. Komatsu, "Hunting for primordial non-Gaussianity in the cosmic microwave background," *Classical and Quantum Gravity*, vol. 27, no. 12, Article ID 124010, 2010.
- [4] M. Liguori, E. Sefusatti, J. R. Fergusson, and E. P. S. Shellard, "Primordial non-Gaussianity and bispectrum measurements in the cosmic microwave background and large-scale structure," <http://arxiv.org/abs/1001.4707>.
- [5] L. Verde, "Non-Gaussianity from large-scale structure surveys," *Advances in Astronomy, Special issue "Testing the Gaussianity and Statistical Isotropy of the Universe"*, vol. 2010, Article ID 768675, 15 pages, 2010, <http://arxiv.org/abs/1001.5217>.
- [6] A. A. Starobinsky, "Multicomponent de Sitter (inflationary) stages and the generation of perturbations," *JETP Letters*, vol. 42, no. 3, pp. 152–155, 1985.
- [7] A. A. Starobinsky, "Mnogokomponentnye de-sitterovskie (inflyatsionnye) stadii i generatsiya vozmushchenii," *Pisma Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*, vol. 42, no. 3, pp. 124–127, 1985.
- [8] M. Sasaki and E. D. Stewart, "A general analytic formula for the spectral index of the density perturbations produced during inflation," *Progress of Theoretical Physics*, vol. 95, no. 1, pp. 71–78, 1996.
- [9] M. Sasaki and T. Tanaka, "Super-horizon scale dynamics of multi-scalar inflation," *Progress of Theoretical Physics*, vol. 99, no. 5, pp. 763–781, 1998.
- [10] D. H. Lyth, K. A. Malik, and M. Sasaki, "A general proof of the conservation of the curvature perturbation," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 5, article 004, 2005.
- [11] D. H. Lyth and Y. Rodríguez, "Inflationary prediction for primordial non-Gaussianity," *Physical Review Letters*, vol. 95, no. 12, Article ID 121302, 4 pages, 2005.
- [12] Y.-I. Takamizu, S. Mukohyama, M. Sasaki, and Y. Tanaka, "Non-Gaussianity of superhorizon curvature perturbations beyond  $\delta N$  formalism," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 6, article 019, 2010.
- [13] J. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *Journal of High Energy Physics*, vol. 2003, no. 5, article 013, 2003.
- [14] X. Chen, R. Easther, and E. A. Lim, "Large non-Gaussianities in single-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 6, article 023, 2007.
- [15] E. Silverstein and D. Tong, "Scalar speed limits and cosmology: acceleration from D-cceleration," *Physical Review D*, vol. 70, no. 10, Article ID 103505, 18 pages, 2004.
- [16] M. Alishahiha, E. Silverstein, and D. Tong, "DBI in the sky: non-Gaussianity from inflation with a speed limit," *Physical Review D*, vol. 70, no. 12, Article ID 123505, 15 pages, 2004.
- [17] X. Chen, M.-X. Huang, S. Kachru, and G. Shiu, "Observational signatures and non-Gaussianities of general single-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 1, article 002, 2007.
- [18] M.-X. Huang and G. Shiu, "Inflationary trispectrum for models with large non-Gaussianities," *Physical Review D*, vol. 74, no. 12, Article ID 121301, 2006.
- [19] D. Langlois, S. Renaux-Petel, D. A. Steer, and T. Tanaka, "Primordial perturbations and non-Gaussianities in DBI and general multi-field inflation," *Physical Review D*, vol. 78, Article ID 063523, 15 pages, 2008.
- [20] F. Arroja, S. Mizuno, and K. Koyama, "Non-Gaussianity from the bispectrum in general multiple field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 015, 2008.
- [21] X. Gao and B. Hu, "Primordial trispectrum from entropy perturbations in multifield DBI model," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 012, 2009.
- [22] Y.-F. Cai and H.-Y. Xia, "Inflation with multiple sound speeds: a model of multiple DBI type actions and non-Gaussianities," *Physics Letters B*, vol. 677, no. 5, pp. 226–234, 2009.
- [23] X. Chen, B. Hu, M.-X. Huang, G. Shiu, and Y. Wang, "Large primordial trispectra in general single field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 008, 2009.
- [24] F. Arroja, S. Mizuno, K. Koyama, and T. Tanaka, "Full trispectrum in single field DBI inflation," *Physical Review D*, vol. 80, no. 4, Article ID 043527, 2009.
- [25] S. Mizuno, F. Arroja, K. Koyama, and T. Tanaka, "Lorentz boost and non-Gaussianity in multifield DBI inflation," *Physical Review D*, vol. 80, no. 2, Article ID 023530, 2009.
- [26] K. Koyama, "Non-Gaussianity of quantum fields during inflation," *Class.Quant.Grav.*, vol. 27, Article ID 124001, 2010, <http://arxiv.org/abs/1002.0600>.
- [27] X. Chen, "Primordial non-Gaussianities from inflation models," <http://arxiv.org/abs/1002.1416>.
- [28] K. Enqvist, A. Jokinen, A. Mazumdar, T. Multamäki, and A. Väihkönen, "Non-Gaussianity from preheating," *Physical Review Letters*, vol. 94, no. 16, Article ID 161301, 2005.
- [29] K. Enqvist, A. Jokinen, A. Mazumdar, T. Multamäki, and A. Väihkönen, "Non-Gaussianity from instant and tachyonic preheating," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 3, article 010, 2005.
- [30] A. Jokinen and A. Mazumdar, "Very large primordial non-Gaussianity from multiple fields: application to massless preheating," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 4, article 003, 2006.
- [31] N. Bartolo, S. Matarrese, and A. Riotto, "Non-Gaussianity in the curvaton scenario," *Physical Review D*, vol. 69, no. 4, Article ID 043503, 2004.
- [32] K. A. Malik and D. H. Lyth, "A numerical study of non-Gaussianity in the curvaton scenario," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 9, article 008, 2006.
- [33] M. Zaldarriaga, "Non-Gaussianities in models with a varying inflaton decay rate," *Physical Review D*, vol. 69, no. 4, Article ID 043508, 2004.
- [34] D. H. Lyth, "Generating the curvature perturbation at the end of inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 11, article 006, 2005.
- [35] J. Väliviita and T. Giannantonio, "Constraints on primordial isocurvature perturbations and spatial curvature by Bayesian model selection," *Physical Review D*, vol. 80, no. 12, Article ID 123516, 2009.
- [36] C. Hikage, D. Munshi, A. Heavens, and P. Coles, "Adiabatic versus isocurvature non-Gaussianity," *Monthly Notices of the Royal Astronomical Society*, vol. 404, no. 3, pp. 1505–1511, 2010.

- [37] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity and the cosmic microwave background anisotropies,” <http://arxiv.org/abs/1001.3957>.
- [38] C. Pitrou, J.-P. Uzan, and F. Bernardeau, “The cosmic microwave background bispectrum from the non-linear evolution of the cosmological perturbations,” *JCAP*, no. 1007, article 003, 2010, <http://arxiv.org/abs/1003.0481>.
- [39] D. Wands, “Multiple field inflation,” *Lecture Notes in Physics*, vol. 738, pp. 275–304, 2008.
- [40] E. Komatsu, N. Afshordi, N. Bartolo, et al., “Non-Gaussianity as a probe of the physics of the primordial universe and the astrophysics of the low redshift universe,” <http://arxiv.org/abs/0902.4759>.
- [41] D. Babich, P. Creminelli, and M. Zaldarriaga, “The shape of non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, no. 8, article 009, 2004.
- [42] J. R. Fergusson and E. P. S. Shellard, “Shape of primordial non-Gaussianity and the CMB bispectrum,” *Physical Review D*, vol. 80, no. 4, Article ID 043510, 2009.
- [43] L. Alabidi, K. A. Malik, C. T. Byrnes, and K.-Y. Choi, “How the curvaton scenario, modulated reheating and an inhomogeneous end of inflation are related,” <http://arxiv.org/abs/1002.1700>.
- [44] A. Mazumdar and J. Rocher, “Particle physics models of inflation and curvaton scenarios,” <http://arxiv.org/abs/1001.0993>.
- [45] F. Bernardeau, “Gravity and non-gravity mediated couplings in multiple-field inflation,” *Class.Quant.Grav.*, vol. 27, Article ID 124004, 2010, <http://arxiv.org/abs/1003.2869>.
- [46] D. Wands, “Local non-Gaussianity from inflation,” *Class.Quant.Grav.*, vol. 27, Article ID 124002, 2010, <http://arxiv.org/abs/1004.0818>.
- [47] T. Tanaka, T. Suyama, and S. Yokoyama, “Use of delta N formalism—difficulties in generating large local-type non-Gaussianity during inflation,” *Class.Quant.Grav.*, vol. 27, Article ID 124003, 2010, <http://arxiv.org/abs/1003.5057>.
- [48] N. Barnaby and J. M. Cline, “Predictions for non-Gaussianity from non-local inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 6, article 030, 2008.
- [49] D. Langlois and L. Sorbo, “Primordial perturbations and non-Gaussianities from modulated trapping,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 014, 2009.
- [50] J.-L. Lehners and P. J. Steinhardt, “Non-Gaussianity generated by the entropic mechanism in bouncing cosmologies made simple,” *Physical Review D*, vol. 80, no. 10, Article ID 103520, 2009.
- [51] J.-L. Lehners, “Ekpyrotic non-Gaussianity—a review,” *Advances in Astronomy, Special issue “Testing the Gaussianity and Statistical Isotropy of the Universe”*, vol. 2010, Article ID 903907, 19 pages, 2010, <http://arxiv.org/abs/1001.3125>.
- [52] K. Koyama, S. Mizuno, F. Vernizzi, and D. Wands, “Non-Gaussianities from ekpyrotic collapse with multiple fields,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, article 024, 2007.
- [53] S. Mollerach, “Isocurvature baryon perturbations and inflation,” *Physical Review D*, vol. 42, no. 2, pp. 313–325, 1990.
- [54] A. Linde and V. Mukhanov, “Non-Gaussian isocurvature perturbations from inflation,” *Physical Review D*, vol. 56, no. 2, pp. R535–R539, 1997.
- [55] K. Enqvist and M. S. Sloth, “Adiabatic CMB perturbations in pre-Big-Bang string cosmology,” *Nuclear Physics B*, vol. 626, no. 1-2, pp. 395–409, 2002.
- [56] D. H. Lyth and D. Wands, “Generating the curvature perturbation without an inflaton,” *Physics Letters B*, vol. 524, no. 1-2, pp. 5–14, 2002.
- [57] T. Moroi and T. Takahashi, “Effects of cosmological moduli fields on cosmic microwave background,” *Physics Letters B*, vol. 522, no. 3-4, pp. 215–221, 2001.
- [58] T. Moroi and T. Takahashi, “Erratum: effects of cosmological moduli fields on cosmic microwave background,” *Physics Letters B*, vol. 539, no. 3-4, p. 303, 2002.
- [59] A. Linde and V. Mukhanov, “The curvaton web,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 4, article 009, 2006.
- [60] D. H. Lyth, C. Ungarelli, and D. Wands, “Primordial density perturbation in the curvaton scenario,” *Physical Review D*, vol. 67, no. 2, Article ID 023503, 2003.
- [61] M. Sasaki, J. Väiviita, and D. Wands, “Non-Gaussianity of the primordial perturbation in the curvaton model,” *Physical Review D*, vol. 74, no. 10, Article ID 103003, 2006.
- [62] D. Langlois, F. Vernizzi, and D. Wands, “Non-linear isocurvature perturbations and non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 12, article 004, 2008.
- [63] K.-Y. Choi and J.-O. Gong, “Multiple scalar particle decays and perturbation generation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 6, article 007, 2007.
- [64] H. Assadullahi, J. Väiviita, and D. Wands, “Primordial non-Gaussianity from two curvaton decays,” *Physical Review D*, vol. 76, no. 10, Article ID 103003, 2007.
- [65] Q.-G. Huang, “The N-vaton,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 9, article 017, 2008.
- [66] C. Gordon and A. Lewis, “Observational constraints on the curvaton model of inflation,” *Physical Review D*, vol. 67, no. 12, Article ID 123513, 2003.
- [67] M. Beltrán, “Isocurvature, non-Gaussianity, and the curvaton model,” *Physical Review D*, vol. 78, no. 2, Article ID 023530, 2008.
- [68] K. Nakayama and J. Yokoyama, “Gravitational wave background and non-Gaussianity as a probe of the curvaton scenario,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 1, article 010, 2010.
- [69] B. A. Bassett, S. Tsujikawa, and D. Wands, “Inflation dynamics and reheating,” *Reviews of Modern Physics*, vol. 78, no. 2, pp. 537–589, 2006.
- [70] A. V. Frolov, “Non-linear dynamics and primordial curvature perturbations from preheating,” *Class.Quant.Grav.*, vol. 27, Article ID 124006, 2010, <http://arxiv.org/abs/1004.3559>.
- [71] G. Dvali, A. Gruzinov, and M. Zaldarriaga, “New mechanism for generating density perturbations from inflation,” *Physical Review D*, vol. 69, no. 2, Article ID 023505, 2004.
- [72] L. Kofman, “Probing string theory with modulated cosmological fluctuations,” <http://arxiv.org/abs/astro-ph/0303614>.
- [73] G. Dvali, A. Gruzinov, and M. Zaldarriaga, “Cosmological perturbations from inhomogeneous reheating, freeze-out, and mass domination,” *Physical Review D*, vol. 69, no. 8, Article ID 083505, 5 pages, 2004.
- [74] C. W. Bauer, M. L. Graesser, and M. P. Salem, “Fluctuating annihilation cross sections and the generation of density perturbations,” *Physical Review D*, vol. 72, no. 2, Article ID 023512, 7 pages, 2005.
- [75] T. Suyama and M. Yamaguchi, “Non-Gaussianity in the modulated reheating scenario,” *Physical Review D*, vol. 77, no. 2, Article ID 023505, 2008.

- [76] K. Ichikawa, T. Suyama, T. Takahashi, and M. Yamaguchi, "Primordial curvature fluctuation and its non-Gaussianity in models with modulated reheating," *Physical Review D*, vol. 78, no. 6, Article ID 063545, 2008.
- [77] M. Bastero-Gil, V. Di Clemente, and S. F. King, "Preheating curvature perturbations with a coupled curvaton," *Physical Review D*, vol. 70, no. 2, Article ID 023501, 2004.
- [78] J. R. Bond, A. V. Frolov, Z. Huang, and L. Kofman, "Non-Gaussian curvature spikes from chaotic billiards in inflation preheating," *Physical Review Letters*, vol. 103, no. 7, Article ID 071301, 2009.
- [79] K. Kohri, D. H. Lyth, and C.A. Valenzuela-Toledo, "On the generation of a non-Gaussian curvature perturbation during preheating," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 2, article 023, 2010.
- [80] G. Felder, L. Kofman, and A. Linde, "Instant preheating," *Physical Review D*, vol. 59, no. 12, Article ID 123523, 7 pages, 1999.
- [81] E. W. Kolb, A. Riotto, and A. Vallinotto, "Curvature perturbations from broken symmetries," *Physical Review D*, vol. 71, no. 4, Article ID 043513, 10 pages, 2005.
- [82] C. T. Byrnes and D. Wands, "Scale-invariant perturbations from chaotic inflation," *Physical Review D*, vol. 73, no. 6, Article ID 063509, 8 pages, 2006.
- [83] T. Matsuda, "Generating the curvature perturbation with instant preheating," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 3, article 003, 2007.
- [84] T. Battefeld, "Modulated perturbations from instant preheating after new ekpyrosis," *Physical Review D*, vol. 77, no. 6, Article ID 063503, 2008.
- [85] C. T. Byrnes, "Constraints on generating the primordial curvature perturbation and non-Gaussianity from instant preheating," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 1, article 011, 2009.
- [86] Q.-G. Huang, "A geometric description of the non-Gaussianity generated at the end of multi-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 6, article 035, 2009.
- [87] F. Bernardeau and J.-P. Uzan, "Non-Gaussianity in multifield inflation," *Physical Review D*, vol. 66, no. 10, Article ID 103506, 2002.
- [88] F. Bernardeau and J.-P. Uzan, "Inflationary models inducing non-Gaussian metric fluctuations," *Physical Review D*, vol. 67, no. 12, Article ID 121301, 2003.
- [89] F. Bernardeau and T. Brunier, "Non-Gaussianities in extended D-term inflation," *Physical Review D*, vol. 76, no. 4, Article ID 043526, 2007.
- [90] L. Alabidi and D. Lyth, "Curvature perturbation from symmetry breaking the end of inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 8, article 006, 2006.
- [91] M. P. Salem, "Generation of density perturbations at the end of inflation," *Physical Review D*, vol. 72, no. 12, Article ID 123516, 7 pages, 2005.
- [92] M. Sasaki, "Multi-brid inflation and non-Gaussianity," *Progress of Theoretical Physics*, vol. 120, no. 1, pp. 159–174, 2008.
- [93] A. Naruko and M. Sasaki, "Large non-Gaussianity from multi-brid inflation," *Progress of Theoretical Physics*, vol. 121, no. 1, pp. 193–210, 2009.
- [94] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall, "Conditions for large non-Gaussianity in two-field slow-roll inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 10, article 008, 2008.
- [95] C. T. Byrnes, K. Y. Choi, and L. M. H. Hall, "Large non-Gaussianity from two-component hybrid inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 2, article 017, 2009.
- [96] G. I. Rigopoulos and E. P. S. Shellard, "Non-linear inflationary perturbations," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 10, article 006, 2005.
- [97] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, "Nonlinear perturbations in multiple-field inflation," *Physical Review D*, vol. 73, no. 8, Article ID 083521, 11 pages, 2006.
- [98] L. Alabidi and D. H. Lyth, "Inflation models and observation," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 016, 2006.
- [99] F. Vernizzi and D. Wands, "Non-Gaussianities in two-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 019, Article ID 019, 2006.
- [100] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, "Quantitative bispectra from multifield inflation," *Physical Review D*, vol. 76, no. 8, Article ID 083512, 2007.
- [101] S. Yokoyama, T. Suyama, and T. Tanaka, "Primordial non-Gaussianity in multiscalar inflation," *Physical Review D*, vol. 77, no. 8, Article ID 083511, 2008.
- [102] S. Yokoyama, T. Suyama, and T. Tanaka, "Primordial non-Gaussianity in multiscalar inflation," *Physical Review D*, vol. 77, no. 8, Article ID 083511, 2008.
- [103] H. R. S. Cogollo, Y. Rodríguez, and C. A. Valenzuela-Toledo, "On the issue of the  $\zeta$  series convergence and loop corrections in the generation of observable primordial non-Gaussianity in slow-roll inflation: I. The bispectrum," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 029, 2008.
- [104] K.-Y. Choi, L. M. H. Hall, and C. van de Bruck, "Spectral running and non-Gaussianity from slow-roll inflation in generalized two-field models," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 2, article 029, 2007.
- [105] T. Battefeld and R. Easther, "Non-Gaussianities in multi-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 3, article 020, 2007.
- [106] D. Seery and J. E. Lidsey, "Non-Gaussianity from the inflationary trispectrum," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 1, article 008, 2007.
- [107] J. García-Bellido and D. Wands, "Metric perturbations in two-field inflation," *Physical Review D*, vol. 53, no. 10, pp. 5437–5445, 1996.
- [108] N. Bartolo and A. R. Liddle, "Simplest curvaton model," *Physical Review D*, vol. 65, no. 12, Article ID 121301, 2002.
- [109] D. J. Mulryne, D. Seery, and D. Wesley, "Moment transport equations for non-Gaussianity," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 1, article 024, 2010.
- [110] N. Barnaby and J. M. Cline, "Non-Gaussian and nonscale-invariant perturbations from tachyonic preheating in hybrid inflation," *Physical Review D*, vol. 73, no. 10, Article ID 106012, 2006.
- [111] D. Mulryne, D. Seery, and D. Wesley, "Non-Gaussianity constrains hybrid inflation," <http://arxiv.org/abs/0911.3550>.
- [112] C. T. Byrnes and G. Tasinato, "Non-Gaussianity beyond slow roll in multi-field inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 016, 2009.
- [113] D. S. Salopek and J. R. Bond, "Nonlinear evolution of long-wavelength metric fluctuations in inflationary models," *Physical Review D*, vol. 42, no. 12, pp. 3936–3962, 1990.

- [114] W. H. Kinney, “Hamilton-Jacobi approach to non-slow-roll inflation,” *Physical Review D*, vol. 56, no. 4, pp. 2002–2009, 1997.
- [115] D. Battefeld and T. Battefeld, “On non-Gaussianities in multi-field inflation (N fields): bi- and tri-spectra beyond slow-roll,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 11, article 010, 2009.
- [116] K.-Y. Choi, J.-O. Gong, and D. Jeong, “Evolution of the curvature perturbation during and after multi-field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 2, article 032, 2009.
- [117] M. Sasaki, “A note on nonlinear curvature perturbations in an exactly soluble model of multi-component slow-roll inflation,” *Classical and Quantum Gravity*, vol. 24, no. 9, pp. 2433–2437, 2007.
- [118] J. P. Conlon and F. Quevedo, “Kähler moduli inflation,” *Journal of High Energy Physics*, vol. 2006, no. 1, article 146, 2006.
- [119] J. R. Bond, L. Kofman, S. Prokushkin, and P. M. Vaudrevange, “Roulette inflation with Kähler moduli and their axions,” *Physical Review D*, vol. 75, no. 12, Article ID 123511, 2007.
- [120] A. Misra and P. Shukla, “Large volume axionic Swiss cheese inflation,” *Nuclear Physics B*, vol. 800, no. 1-2, pp. 384–400, 2008.
- [121] M. Badziak and M. Olechowski, “Volume modulus inflation and a low scale of SUSY breaking,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 7, article 021, 2008.
- [122] E. Palti, G. Tasinato, and J. Ward, “Weakly-coupled IIA flux compactifications,” *Journal of High Energy Physics*, vol. 2008, no. 6, article 084, 2008.
- [123] H.-X. Yang and H.-L. Ma, “Two-field Kähler moduli inflation in large volume moduli stabilization,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 024, 2008.
- [124] J. J. Blanco-Pillado, D. Buck, E. J. Copeland, M. Gomez-Reino, and N. J. Nunes, “Kahler moduli inflation revisited,” *JHEP*, vol. 2010, no. 1, article 081, 2010, <http://arxiv.org/abs/0906.3711>.
- [125] N. Barnaby, J. R. Bond, Z. Huang, and L. Kofman, “Pre-heating after modular inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article 021, 2009.
- [126] E. Komatsu, J. Dunkley, M. R. Nolte et al., “Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal*, vol. 180, no. 2, pp. 330–376, 2009.
- [127] C. T. Byrnes, M. Sasaki, and D. Wands, “Primordial trispectrum from inflation,” *Physical Review D*, vol. 74, no. 12, Article ID 123519, 2006.
- [128] J. Smidt, A. Amblard, A. Cooray, A. Heavens, D. Munshi, and P. Serra, “A measurement of cubic-order primordial non-Gaussianity ( $g_{NL}$  and  $\tau_{NL}$ ) With WMAP 5-Year Data,” <http://arxiv.org/abs/1001.5026>.
- [129] V. Desjacques and U. Seljak, “Signature of primordial non-Gaussianity of  $\phi^3$  type in the mass function and bias of dark matter haloes,” *Physical Review D*, vol. 81, no. 2, Article ID 023006, 2010.
- [130] P. Vielva and J. L. Sanz, “Constraints on  $f_{NL}$  and  $g_{NL}$  from the analysis of the N-pdf of the CMB large-scale anisotropies,” *Monthly Notices of the Royal Astronomical Society*, vol. 404, no. 2, pp. 895–907, 2010.
- [131] T. Okamoto and W. Hu, “Angular trispectra of CMB temperature and polarization,” *Physical Review D*, vol. 66, no. 6, Article ID 063008, 2002.
- [132] N. Kogo and E. Komatsu, “Angular trispectrum of CMB temperature anisotropy from primordial non-Gaussianity with the full radiation transfer function,” *Physical Review D*, vol. 73, no. 8, Article ID 083007, 2006.
- [133] A. Cooray, C. Li, and A. Melchiorri, “Trispectrum of 21-cm background anisotropies as a probe of primordial non-Gaussianity,” *Physical Review D*, vol. 77, no. 10, Article ID 103506, 2008.
- [134] D. Jeong and E. Komatsu, “Primordial non-Gaussianity, scale-dependent bias, and the bispectrum of galaxies,” *Astrophysical Journal*, vol. 703, no. 2, pp. 1230–1248, 2009.
- [135] J. Smidt, A. Amblard, C. T. Byrnes, A. Cooray, and D. Munshi, “CMB constraints on primordial non-Gaussianity from the bispectrum and trispectrum and a consistency test of single-field inflation,” *Phys.Rev.*, Article ID 123007, 2010, <http://arxiv.org/abs/1004.1409>.
- [136] P. Creminelli, S. Dubovsky, A. Nicolis, L. Senatore, and M. Zaldarriaga, “The phase transition to eternal inflation,” *Journal of High Energy Physics*, vol. 2008, no. 9, article 036, 2008.
- [137] Y. Rodríguez and C. A. Valenzuela-Toledo, “On the Issue of the  $\zeta$  series convergence and loop corrections in the generation of observable primordial non-Gaussianity in slow-roll inflation. II. the trispectrum,” *Physical Review D*, vol. 81, no. 2, Article ID 023531, 2010.
- [138] K. Enqvist and S. Nurmi, “Non-Gaussianity in curvaton models with nearly quadratic potentials,” *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 10, article 013, 2005.
- [139] Q.-G. Huang and Y. Wang, “Curvaton dynamics and the non-linearity parameters in the curvaton model,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 9, article 025, 2008.
- [140] Q.-G. Huang, “A curvaton with a polynomial potential,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 11, article 005, 2008.
- [141] K. Enqvist and T. Takahashi, “Signatures of non-Gaussianity in the curvaton model,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 9, article 012, 2008.
- [142] K. Enqvist, S. Nurmi, O. Taanila, and T. Takahashi, “Non-Gaussian fingerprints of self-interacting curvaton,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 4, article 009, 2010.
- [143] C. T. Byrnes, S. Nurmi, G. Tasinato, and D. Wands, “Scale dependence of local  $f_{NL}$ ,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 2, article 034, 2010.
- [144] E. Sefusatti, M. Liguori, A. P. S. Yadav, M. G. Jackson, and E. Pajer, “Constraining running non-Gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article 022, 2009.
- [145] M. LoVerde, A. Miller, S. Shandera, and L. Verde, “Effects of scale-dependent non-Gaussianity on cosmological structures,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article 014, 2008.
- [146] X. Chen, “Running non-Gaussianities in Dirac-Born-Infeld inflation,” *Physical Review D*, vol. 72, no. 12, Article ID 123518, 7 pages, 2005.
- [147] J. Khoury and F. Piazza, “Rapidly-varying speed of sound, scale invariance and non-Gaussian signatures,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 7, article 026, 2009.

## Review Article

# Non-Gaussianity from Particle Production during Inflation

**Neil Barnaby**

*McLennan Physical Laboratories, Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George Street, Toronto, ON, Canada M5S 3H8*

Correspondence should be addressed to Neil Barnaby, barnaby@cita.utoronto.ca

Received 15 January 2010; Accepted 11 June 2010

Academic Editor: Sarah Shandera

Copyright © 2010 Neil Barnaby. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In a variety of models the motion of the inflaton may trigger the production of some non-inflaton particles during inflation, for example via parametric resonance or a phase transition. Such models have attracted interest recently for a variety of reasons, including the possibility of slowing the motion of the inflaton on a steep potential. In this review we show that interactions between the produced particles and the inflaton condensate can lead to a qualitatively new mechanism for generating cosmological fluctuations from inflation. We illustrate this effect using a simple prototype model  $g^2(\phi - \phi_0)^2\chi^2$  for the interaction between the inflaton,  $\phi$ , and iso-inflaton,  $\chi$ . Such interactions are quite natural in a variety of inflation models from supersymmetry and string theory. Using both lattice field theory and analytical calculations, we study the production of  $\chi$  particles and their subsequent rescatterings off the condensate  $\phi(t)$ , which generates bremsstrahlung radiation of light inflaton fluctuations  $\delta\phi$ . This mechanism leads to observable features in the primordial power spectrum. We derive observational constraints on such features and discuss their implications for popular models of inflation. Inflationary particle production also leads to a very novel kind of nongaussian signature which may be observable in future missions.

## 1. Introduction

In recent years the inflationary paradigm has become a cornerstone of modern cosmology. In the simplest scenario the observed cosmological perturbations are seeded by the quantum vacuum fluctuations of the inflaton field [1–5]. This mechanism predicts a nearly scale invariant spectrum of adiabatic primordial fluctuations, consistent with recent observational data [6]. In addition to this standard mechanism, there are also several alternatives for generating cosmological perturbations from inflation; examples include modulated fluctuations [7–10] and the curvaton mechanism [11]. These various scenarios all lead to similar predictions for the power spectrum. On the other hand, nongaussian statistics (such as the bispectrum) provide a powerful tool to observationally discriminate between different mechanisms for generating the curvature perturbation. In this paper, which is based on [12–15], we will present a *qualitatively new* mechanism for generating cosmological perturbations during inflation. We discuss in detail the predictions of this new scenario for both the spectrum and nongaussianity of the primordial curvature fluctuations, showing how this

new mechanism may be observationally distinguished from previous approaches.

*1.1. Non-Gaussianity from Inflation.* The possibility to discriminate between various inflationary scenarios has led to a recent surge of interest in computing and measuring nongaussian statistics. Although single field, slow roll models are known to produce negligible nongaussianity [16–18], there are now a variety of scenarios available in the literature which may predict an observable signature. Departures from gaussianity are often parametrized in the following form:

$$\zeta(x) = \zeta_g(x) + \frac{3}{5}f_{NL}[\zeta_g^2(x) - \langle\zeta_g^2(x)\rangle], \quad (1)$$

where  $\zeta(x)$  is the primordial curvature perturbation,  $\zeta_g(x)$  is a Gaussian random field, and  $f_{NL}$  characterizes the degree of nongaussianity. The ansatz (1) is known as the “local” form of nongaussianity.

Although the local ansatz (1) has received significant attention, it is certainly not the only well-motivated model for a nongaussian curvature perturbation. For example, the

nongaussian part of  $\zeta(x)$  need not be correlated with the gaussian part. Consider a primordial curvature perturbation of the form

$$\zeta(x) = \zeta_g(x) + F_{NL}[\chi_g(x)], \quad (2)$$

where  $F_{NL}$  is some nonlinear (not necessarily quadratic) function, and  $\chi_g(x)$  is a gaussian field which is uncorrelated with  $\zeta_g(x)$ . Both (1) and (2) are local in position space; however, these two types of nongaussianity will have very different observational implications. The uncorrelated ansatz (2) for the primordial curvature perturbation can arise, for example, in models with preheating into light fields [19–21]. (See also [22, 23] for more discussion of nongaussianity from preheating and [24, 25] for another model where nongaussianity is generated at the end of inflation.)

A useful quantity to consider is the *bispectrum*,  $B(k_1, k_2, k_3)$ , which is the 3-point correlation function of the Fourier transform of the primordial curvature perturbation

$$\langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B(k_i), \quad (3)$$

where  $k_i \equiv |\mathbf{k}_i|$ . The delta function appearing in (3) reflects translational invariance and ensures that  $B(k_i)$  depends on three momenta  $\mathbf{k}_i$  which form a triangle:  $\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0$ . Rotational invariance implies that  $B(k_i)$  is symmetric in its arguments.

If we assume the ansatz (1) for the primordial curvature perturbation, then  $B(k_i)$  has a very particular dependence on momenta; it peaks in the squeezed limit where one of the wavenumbers is much smaller than the remaining two (e.g.,  $k_1 \ll k_2, k_3$ ). Such a bispectrum is referred to as having a *squeezed* shape. However, other shapes of bispectrum are worth considering. A bispectrum is referred to as “equilateral” if it peaks when  $k_1 = k_2 = k_3$  and “flattened” if it peaks when one of the wavenumbers is half the size of the remaining two (e.g.,  $2k_1 = k_2 = k_3$ ).

Without assuming any specific form for the primordial curvature perturbation, such as (1) or (2), one may characterize an arbitrary bispectrum (3) by specifying its shape, running, and size [26]. As discussed above, the shape refers to the configuration of triangle on which  $B(k_i)$  is maximal (squeeze, equilateral, or flattened). The running of the bispectrum refers to how the magnitude of  $B(k_i)$  depends on the overall size of the triangle. For example, in the case of scale invariant fluctuations, the bispectrum must scale as  $B(\lambda k_1, \lambda k_2, \lambda k_3) = \lambda^{-6} B(k_1, k_2, k_3)$ . Finally, the overall size of the bispectrum is often quantified by evaluating the magnitude of  $B(k_i)$  on some fixed equilateral triangle. However, the skewness of the probability density function (defined later) might provide a better measure of the size of nongaussianity.

Different types of nongaussian signatures are correlated with properties of the underlying inflation model. Let us first consider some examples with small running (by “small” here we refer to any model where the running of the bispectrum is proportional to slow-variation parameters or arises due to loop effects. This does *not* necessarily mean that such running cannot lead to interesting observational signatures, see; [27–30]).

- (1) A large bispectrum of local shape, along with iso-curvature effects, is associated with models where multiple fields are light (or otherwise dynamically important) during inflation. Examples include the curvaton mechanism [31–34] or models with turning points along the inflationary trajectory [35–39]. The observational bound on local type nongaussianity, coming from the WMAP7 [6] data, is  $-10 < f_{NL}^{\text{local}} < 74$  [40] at 95% confidence level. When combined with large scale structure (LSS) data the bound becomes somewhat stronger:  $-1 < f_{NL}^{\text{local}} < 65$  [41].
- (2) A large local bispectrum *without* any isocurvature fluctuations can *only* be produced by nonlocal inflation models [42–44]. For any single-field inflation model described by a local low-energy effective field theory, the results of [45] imply that the ratio of the 3-point correlation function to the square of the 2-point function must be of order of the spectral tilt, in the squeezed limit. Hence, it has been argued that a large squeezed bispectrum must be associated with the presence of multiple light degrees of freedom and hence iso-curvature effects. However, in [42–44] it was shown that single field nonlocal inflation models can produce a large squeezed bispectrum in the regime where the underlying scale of nonlocality is much larger than the Hubble scale during inflation. Such constructions evade the no-go theorem of [45] precisely because they violate the usual assumption of cluster decomposition. Moreover, models of this type are *not* subsumed by the general analysis of [46] since nonlocal field theories with infinitely many derivatives cannot be obtained in the regime of low-energy effective field theory. It is nevertheless sensible to study such constructions since they may be derived from ultra-violet (UV) complete frameworks, such as string field theory or  $p$ -adic string theory. See [47–49] for details concerning the underlying consistency of nonlocal field theories, and see [50] for a succinct review of nonlocal cosmology.
- (3) A large equilateral bispectrum is typically associated with a small sound speed for the inflaton perturbations [26], such as in Dirac-Born-Infeld (DBI) inflation models [51, 52]. However, such a signature may also be obtained in multifield gelaton [53] or trapped inflation [54] models. The observational bound on equilateral type nongaussianity is  $-125 < f_{NL}^{\text{equil}} < 435$  at 95% confidence level [55].
- (4) A large flattened bispectrum is associated with nonvacuum initial conditions [26, 56–58]. (To our knowledge there is no explicit computation of the observational bound on flattened nongaussianity. In [56], a template (the enfolded model) was proposed. The analysis of [55] is sufficiently general to study this shape; however, they do not explicitly place bounds on  $f_{NL}^{\text{flat}}$  but instead constrain an alternative shape (the orthogonal model) which is a superposition of flattened and equilateral shapes.)

If we relax the assumption that the bispectrum is close to scale invariant, then a much richer variety of nongaussian signatures is possible. For example, in models with sharp steps in the inflaton potential [59, 60] the bispectrum is large only for triangles with a particular characteristic size. We will refer to such a signature as a *localized nongaussian feature*. Localized nongaussianities are not well constrained by current observation but may be observable in future missions.

Given the significant role that nongaussianity may play in discriminating between different models of the early universe, it is of crucial importance to explore and classify all possible consistent signatures for the bispectrum and other nongaussian statistics. Indeed, in this paper we will describe a new kind of signature—uncorrelated nongaussian features—which is predicted in a variety of simple and well-motivated models of inflation, but which has nevertheless been overlooked in previous literature.

**1.2. Inflationary Particle Production.** Recently, a new mechanism for generating cosmological perturbations during inflation was proposed [12]. This new mechanism, dubbed *infrared (IR) cascading*, is qualitatively different from previous proposals (such as the curvaton or modulated fluctuations) in that it does not rely on the quantum vacuum fluctuations of some light scalar fields during inflation. Rather, the scenario involves the production of massive iso-curvature particles *during* inflation. These subsequently rescatter off the slow-roll condensate to generate bremsstrahlung radiation of light inflaton fluctuations (which induce curvature perturbations and temperature anisotropies in the usual manner). IR cascading can also be distinguished from previous mechanisms from the observational perspective: this new mechanism leads to novel features in both the spectrum and bispectrum.

In principle, IR cascading may occur in any model where non-inflaton (iso-curvature) particles are produced during inflation. Models of this type have attracted considerable interest recently; examples have been studied where particle production occurs via parametric resonance [12, 13, 54, 61–66], as a result of a phase transition [19, 20, 67–74] or otherwise [75]. Recent interest in inflationary particle production has been stimulated by various considerations.

- (1) Particle production arises naturally in a number of microscopically realistic models of inflation, including examples from string theory [54] and supersymmetric (SUSY) field theory [76]. In particular, inflationary particle production is a generic feature of open string inflation models [13], such as brane/axion monodromy [77–79].
- (2) The energetic cost of producing particles during inflation has a dissipative effect on the dynamics of the inflaton. Particle production may therefore slow the motion of the inflaton, even on a steep potential. This gives rise to a new inflationary mechanism, called *trapped inflation* [54, 80, 81], which may circumvent some of the fine-tuning problems associated with standard slow-roll inflation. See [54]

for an explicit string theory realization of trapped inflation and [81] for a generalization to higher-dimensional moduli spaces and enhanced symmetry loci. The idea of using dissipative dynamics to slow the motion of the inflaton is qualitatively similar to warm inflation [82] and also to the variant of natural inflation [83, 84] proposed recently by Anber and Sorbo [75].

- (3) Observable features in the primordial power spectrum, generated by particle production and IR cascading, offer a novel example of the non-decoupling of high scale physics in the Cosmic Microwave Background (CMB) [61, 85–87]. In the most interesting examples, the produced particles are extremely massive for (almost) the entire history of the universe; however, their effect cannot be integrated out due to the nonadiabatic time dependence of the iso-inflaton mode functions during particle production. In [61] particle production during large field inflation was proposed as a possible probe of Planck-scale physics.

In this paper we study in detail the impact of particle production and IR cascading on the observable primordial curvature perturbations. In order to illustrate the basic physics we focus on a very simple and general prototype model where the inflaton,  $\phi$ , and iso-inflaton,  $\chi$ , fields interact via the coupling

$$\mathcal{L}_{\text{int}} = -\frac{g^2}{2}(\phi - \phi_0)^2\chi^2. \quad (4)$$

We expect, however, that our results will generalize in a straightforward way to more complicated models, such as fermion iso-inflaton fields or gauged interactions, wherein the physics of particle production and rescattering is essentially the same. Our result may also have implications for inflationary phase transitions, because spinodal decomposition can be interpreted as a kind of particle production, and similar bilinear interactions will induce rescattering effects.

Scalar field interactions of type (4) have also been studied recently in connection with nonequilibrium Quantum Field Theory (QFT) [88–90], in particular with applications to the theory of preheating after inflation [91–95] and also moduli trapping [80, 81] at enhanced symmetry points. Although our focus is on particle production *during* inflation (as opposed to during preheating, after inflation) some of our results nevertheless have implications for preheating, moduli trapping, and also non-equilibrium QFT more generally. For example, in [12] analytical and numerical studies of rescattering and IR cascading during inflation made it possible to observe, for the first time, the dynamical approach to the turbulent scaling regime that was discovered in [96, 97].

Particle production during inflation in model (4) leads to observable features in the primordial power spectrum,  $P(k)$ . A number of recent studies have found evidence for localized features in  $P(k)$  that are incompatible with the simplest power-law model  $P(k) \sim k^{n_s-1}$  [62, 73, 98–110]. Although these observed features may simply be statistical anomalies

(see, e.g., [111]), there remains the tantalizing possibility that they represent some new physics beyond the simplest slow-roll model. Upcoming polarization data may play an important role in distinguishing these possibilities [73]. In the meantime, it is interesting to determine the extent to which such features may be explained by a simple and well-motivated model such as (4). Moreover, because (4) is a complete microscopic model (as opposed to a phenomenological modification of the power spectrum) it is possible to predict a host of correlated observables, such as features in the scalar bispectrum and tensor power spectrum. Hence, it should be possible to robustly rule out (or confirm) the possibility that some massive iso-curvature particles were produced during inflation.

If detected, features from particle production and IR cascading will provide a rare and powerful new window into the microphysics driving inflation. This scenario opens up the possibility of learning some details about how the inflaton couples to other particles in nature, as opposed to simply reconstructing the inflaton potential along the slow-roll trajectory. Moreover, due to the non-decoupling discussed above, features from particle production and IR cascading may probe new (beyond the standard model) physics at extraordinarily high energy scales.

The outline of this paper is as follows. In Section 2 we provide a brief, qualitative overview of the dynamics of particle production and IR cascading in model (4). In Section 3 we study in detail this same dynamics using fully nonlinear lattice field theory simulations. In Section 4 we provide an analytical theory of particle production and IR cascading in an expanding universe. A complimentary analytical analysis, using second-order cosmological perturbation theory, is provided in Section 5. In Section 6 we consider the observational constraints on inflationary particle production using a variety of data sets. In Section 7 we provide several explicit microscopic realizations of our scenario and study the implications of our observational constraints on models of string theory inflation, in particular brane monodromy. In Section 8 we quantify and characterize the nongaussianity generated by particle production and IR cascading. Finally, in Section 9, we conclude and discuss possible future directions.

## 2. Overview and Summary of the Mechanism

In this section we provide a brief overview of the dynamics of particle production and IR cascading in model (4) and also summarize the resulting observational signatures. In the remainder of this paper we will flesh out the details of this mechanism with analytical and numerical calculations.

We consider the following model:

$$S = \int d^4x \sqrt{-g} \left[ \frac{M_p^2}{2} R - \frac{1}{2} (\partial\phi)^2 - V(\phi) - \frac{1}{2} (\partial\chi)^2 - \frac{g^2}{2} (\phi - \phi_0)^2 \chi^2 \right], \quad (5)$$

where  $R$  is the Ricci curvature constructed from the metric  $g_{\mu\nu}$ ,  $\phi$  is the inflaton field, and  $\chi$  is the iso-inflaton. As usual, we assume a flat FRW space-time with scale factor  $a(t)$

$$ds^2 \equiv g_{\mu\nu} dx^\mu dx^\nu = -dt^2 + a^2(t) d\mathbf{x}^2 \quad (6)$$

and employ the reduced Planck mass  $M_p \cong 2.43 \times 10^{18} \text{GeV}$ . We leave the potential  $V(\phi)$  driving inflation unspecified for assuming that it is sufficiently flat in the usual sense, that is,  $\epsilon \ll 1$ ,  $|\eta| \ll 1$  where

$$\epsilon \equiv \frac{M_p^2}{2} \left( \frac{V'}{V} \right)^2, \quad \eta \equiv M_p^2 \frac{V''}{V} \quad (7)$$

are the usual slow-roll parameters.

Note that one might wish to supplement (5) by its supersymmetric completion in order to protect the flatness of the inflaton potential from large radiative corrections coming from loops of the  $\chi$  field. We expect that our results will carry over in a straightforward way to SUSY models and also to more complicated scenarios such as higher spin iso-inflaton fields and (possibly) inflationary phase transitions.

The coupling  $(g^2/2)(\phi - \phi_0)^2 \chi^2$  in (5) is introduced to ensure that the iso-inflaton field can become instantaneously massless at some point  $\phi = \phi_0$  along the inflaton trajectory (which we assume occurs during the observable range of  $e$ -foldings of inflation). At this moment  $\chi$  particles will be produced by quantum effects.

Let us first consider the homogeneous dynamics of the inflaton field,  $\phi(t)$ . Near the point  $\phi = \phi_0$  we can generically expand

$$\phi(t) \cong \phi_0 + vt, \quad (8)$$

where  $v \equiv \dot{\phi}(0)$ , and we have arbitrarily set the origin of time so that  $t = 0$  corresponds to the moment when  $\phi = \phi_0$ . (We are, of course, assuming that  $\dot{\phi}(0) \neq 0$ .) The interaction (4) induces an effective (time varying) mass for the  $\chi$  particles of the form

$$m_\chi^2 = g^2 (\phi - \phi_0)^2 \cong k_\star^2 t^2, \quad (9)$$

where we have defined the characteristic scale

$$k_\star = \sqrt{g|v|}. \quad (10)$$

It is straightforward to verify that the simple expression (9) will be a good approximation for  $(H|t|)^{-1} \lesssim \mathcal{O}(\epsilon, \eta)$  which, in most models, will be true for the entire observable 60  $e$ -foldings of inflation.

Note that, without needing to specify the background inflationary potential  $V(\phi)$ , we can write the ratio  $k_\star/H$  as

$$\frac{k_\star}{H} = \sqrt{\frac{g}{2\pi\mathcal{P}_\zeta^{1/2}}}, \quad (11)$$

where  $\mathcal{P}_\zeta^{1/2} = 5 \times 10^{-5}$  is the usual amplitude of the vacuum fluctuations from inflation. In this work we assume that  $k_\star > H$  which is easily satisfied for reasonable values of the coupling  $g^2 > 10^{-7}$ . In particular, for  $g^2 \sim 0.1$  we have  $k_\star/H \sim 30$ .

The scenario we have in mind is the following. Inflation starts at some field value  $\phi > \phi_0$  and the inflaton rolls toward the point  $\phi = \phi_0$ . Initially, the iso-inflaton field is extremely massive  $m_\chi \gg H$  and hence it stays pinned in the vacuum,  $\chi = 0$ , and does not contribute to superhorizon curvature fluctuations. Eventually, at  $t = 0$ , the inflaton rolls through the point  $\phi = \phi_0$  where  $m_\chi = 0$  and  $\chi$  particles are produced. To describe this burst of particle production one must solve for the following equation for the  $\chi$  particle mode functions in an expanding universe:

$$\ddot{\chi}_k + 3H\dot{\chi}_k + \left[ \frac{k^2}{a^2} + k_\star^4 t^2 \right] \chi_k = 0. \quad (12)$$

Equations of this type are well-studied in the context of preheating after inflation [92] and moduli trapping [80]. The initial conditions for (12) should be chosen to ensure that the q-number field  $\chi$  is in the adiabatic vacuum in the asymptotic past (see Sections 3 and 4 for more details). In the regime  $k_\star > H$  particle production is fast compared to the expansion time and one can solve (12) very accurately for the occupation number of the created  $\chi$  particles

$$n_k = e^{-\pi k^2/k_\star^2}. \quad (13)$$

Very quickly after the moment  $t = 0$ , within a time  $\Delta t \sim k_\star^{-1} \ll H^{-1}$ , these produced  $\chi$  particles become nonrelativistic ( $m_\chi > H$ ), and their number density starts to dilute as  $a^{-3}$ .

Following the initial burst of particle production there are two distinct physical effects which take place. First, the energetic cost of producing the gas of massive out-of-equilibrium  $\chi$  particles drains energy from the inflaton condensate, forcing  $\dot{\phi}$  to drop abruptly. This velocity dip is the result of the backreaction of the produced  $\chi$  fluctuations on homogeneous condensate  $\phi(t)$ . The second physical effect is that the produced massive  $\chi$  particles rescatter off the condensate via the diagram in Figure 1 and emit bremsstrahlung radiation of light inflaton fluctuations (particles).

Backreaction and rescattering leave distinct imprints in the observable cosmological perturbations. Let us first discuss the impact of backreaction. In Figure 2 we plot the velocity dip resulting from the backreaction of the produced  $\chi$  particle on the homogeneous inflaton condensate  $\phi(t)$ . From this figure we see that the quantity  $\ddot{\phi}/(H\dot{\phi})$  becomes large in the dip. This violation of slow roll is a transient effect; at late times the produced  $\chi$  particles become extremely massive and their number density dilutes as  $a^{-3}$ .

One can understand the temporary slowing down of the inflaton from an analytical perspective. Backreaction is taken into account using the mean-field equation

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} + g^2(\phi - \phi_0)\langle\chi^2\rangle = 0, \quad (14)$$

where the vacuum average is computed following [80, 92]

$$\langle\chi^2\rangle \cong \frac{n_\chi a^{-3}}{g|\phi - \phi_0|}. \quad (15)$$

In (14) we have implicitly assumed that the usual Coleman-Weinberg corrections to the inflaton potential have *already*

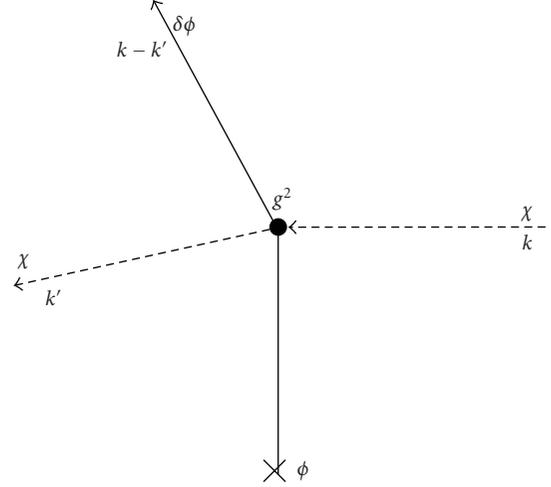


FIGURE 1: Rescattering diagram.

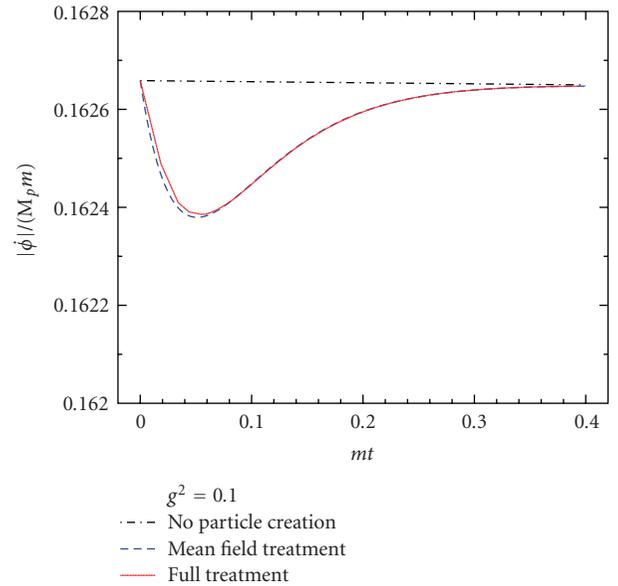


FIGURE 2:  $|\ddot{\phi}|/(M_p m)$  plotted against  $mt$  for  $g^2 = 0.1$  (where  $m = V_{,\phi\phi}$  is the effective inflaton mass). Time  $t = 0$  corresponds to the moment when  $\phi = \phi_0$  and  $\chi$  particles are produced copiously. The solid red line is the lattice field theory result taking into account the full dynamics of rescattering and IR cascading while the dashed blue line is the result of a mean-field theory treatment which ignores rescattering [64]. The dot-dashed black line is the inflationary trajectory in the absence of particle creation.

been absorbed into  $V(\phi)$ , hence the vacuum average  $\langle\chi^2\rangle$  should include *only* the effects of nonadiabatic particle production. (Here  $n_\chi = \int (d^3k/(2\pi)^3) n_k \sim k_\star^3$  is the total number density of produced  $\chi$  particles, and the factor  $a^{-3}$  reflects the usual volume dilution of non-relativistic matter.) In Figure 2 we have plotted the solution of (14) along with the exact result obtained from lattice field theory simulations, illustrating the accuracy of this simple treatment.

Using the mean-field approach, one finds that the transient violation of slow roll leads to a “ringing pattern” (damped oscillations) in the power spectrum  $P_\phi(k) = (k^3/2\pi^2)|\delta\phi_k|^2$  of inflaton fluctuations [64]. This ringing pattern is localized around wavenumbers which left the horizon at the moment when particle production occurred. The effect is very much analogous to Fresnel diffraction at a sharp edge.

The second physical effect, rescattering, was considered for the first time in the context of inflationary particle production in [12]. Figure 1 illustrates the dominant process: bremsstrahlung emission of long-wavelength  $\delta\phi$  fluctuations from rescattering of the produced  $\chi$  particles off the condensate. The time scale for such processes is set by the microscopic scale,  $k_\star^{-1}$ , and is thus very short compared to the expansion time,  $H^{-1}$ . Moreover, the production of inflaton fluctuations  $\delta\phi$  deep in the infrared (IR) is extremely energetically inexpensive, since the inflaton is very nearly massless. The combination of the short time scale for rescattering and the energetic cheapness of radiating IR  $\delta\phi$  leads to a rapid build-up of power in long wavelength inflaton modes: IR cascading. This effect leads to a bump-like feature in the power spectrum of inflaton fluctuations, very different from the ringing pattern associated with backreaction. The bump-like feature from rescattering dominates over the ringing pattern from backreaction for all values of parameters.

In [12] model (5) was studied using lattice field theory simulations, without neglecting any physical processes (that is to say that full nonlinear structure of the theory, including backreaction and rescattering effects, was accounted for consistently). However, this same dynamics can be understood analytically by solving the equation for the inflaton fluctuations  $\delta\phi$  in the approximation that all interactions are neglected, except for the diagram in Figure 1. The appropriate equation is

$$\delta\ddot{\phi} + 3H\delta\dot{\phi} - \frac{\bar{\nabla}^2}{a^2}\delta\phi + V_{,\phi\phi}\delta\phi \cong -g^2[\phi(t) - \phi_0]\chi^2. \quad (16)$$

See [14] for a detailed analytical theory. The solution of (16) may be split into two parts: the solution of the homogeneous equation and the particular solution which is due to the source term. The former simply corresponds to the usual scale invariant quantum vacuum fluctuations from inflation. The particular solution, on the other hand, corresponds to inflaton fluctuations generated by rescattering. The abrupt growth of  $\chi$  inhomogeneities at  $t = 0$  sources the particular solution and generates inflaton fluctuations which subsequently cross the horizon and freeze in.

As mentioned earlier, rescattering generates a bump-like contribution to the primordial power spectrum of the curvature perturbations. To good approximation this may be described by a simple semi analytic fitting function

$$P(k) = A_s \left(\frac{k}{k_0}\right)^{n_s-1} + A_{\text{IR}} \left(\frac{\pi e}{3}\right)^{3/2} \left(\frac{k}{k_{\text{IR}}}\right)^3 e^{-(\pi/2)(k/k_{\text{IR}})^2}, \quad (17)$$

where the first term corresponds to the usual vacuum fluctuations from inflation (with amplitude  $A_s$  and spectral

index  $n_s$ ) while the second term corresponds to the bump-like feature from particle production and IR cascading. The amplitude of this feature ( $A_{\text{IR}}$ ) depends on  $g^2$  while the location ( $k_{\text{IR}}$ ) depends on  $\phi_0$ .

In [13] the simple fitting function (17) was used to place observational constraints on inflationary particle production using a variety of cosmological data sets. Current data are consistent with rather large spectral distortions of the type (17). Features as large as  $\mathcal{O}(10\%)$  of the usual scale-invariant fluctuations from inflation are allowed, in the case that  $k_{\text{IR}}$  falls within the range of scales relevant for CMB experiments. Such a feature corresponds to a realistic coupling  $g^2 \sim 0.01$ . Even larger values of  $g^2$  are allowed if the feature is localized on smaller scales. In Figure 3 we have illustrated the primordial power spectrum in model (5) for a representative choice of parameters. We also plot the CMB angular Temperature-Temperature (TT) power spectrum for the same parameters.

The prototype model (5) may be realized microscopically in a variety of different particle physics frameworks. In particular, particle production is a rather generic feature of open string inflation models [13] where the inflaton,  $\phi$ , has a geometrical interpretation as the position of some mobile D-brane. In this context the iso-inflaton,  $\chi$ , corresponds to a low-lying open string excitation which is stretched between the mobile inflationary brane and any other (spectator) branes which inhabit the compactification volume. If the inflationary and spectator branes become coincident during inflation, then the symmetry of the system is enhanced [80] and some low-lying stretched string states will become instantaneously massless, mimicking interaction (4) (see also [54]). An explicit realization of this scenario is provided by brane/axion monodromy models [77–79]. Our observational constraints on inflationary particle production may be used to place bounds on parameters of the underlying string model [13].

The bump-like feature in  $P(k)$ , illustrated in Figure 3, must be associated with a nongaussian feature in the bispectrum [12, 14]. Indeed, it is evident already from inspection of (16) that the inflaton fluctuations generated by rescattering are significantly nongaussian; the particular solution of (16) is bi-linear in the gaussian field  $\chi$ . The nongaussian signature from IR cascading is rather novel. The nongaussian part of  $\zeta$  is uncorrelated with the gaussian part. Moreover, the bispectrum  $B(k_i)$  is very far from scale invariant; it peaks strongly for triangles with a characteristic size  $\sim k_{\text{IR}}$ , corresponding to the location of the bump in the power spectrum (17). The shape of the bispectrum therefore depends sensitively on the size of the triangle and is not well described by any of the templates that have been proposed in the literature to date.

The magnitude of this new kind of nongaussianity may be quite large. To quantify the effect it is useful to introduce the probability density function (PDF),  $P(\zeta)$ , which is the probability that the curvature perturbation has a fluctuation of size  $\zeta$ . If we define the central moments of the PDF as

$$\langle \zeta^n \rangle = \int \zeta^n P(\zeta) d\zeta, \quad (18)$$

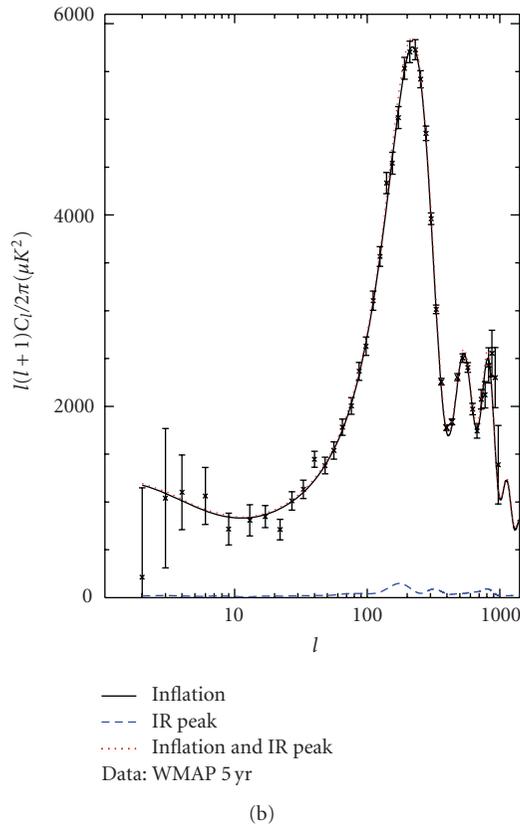
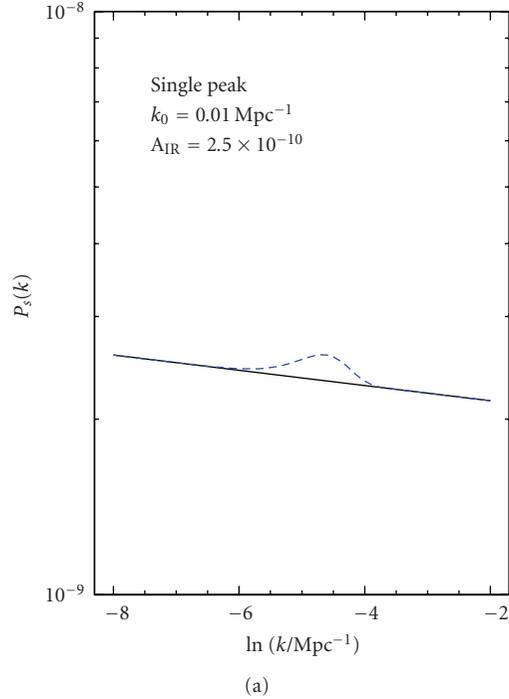


FIGURE 3: (a) shows a sample bump in the power spectrum with amplitude  $A_{\text{IR}} = 2.5 \times 10^{-10}$  which corresponds to a coupling  $g^2 \sim 0.01$ . The feature is located at  $k_{\text{IR}} = 0.01 \text{ Mpc}^{-1}$ . This example represents a distortion of  $\mathcal{O}(10\%)$  as compared to the usual vacuum fluctuations and is consistent with the data at  $2\sigma$ . (b) shows the CMB angular TT power spectrum for this example, illustrating that the distortion shows up mostly in the first peak.

then a useful measure of nongaussianity is the dimensionless skewness of the PDF, defined by

$$\hat{S}_3 \equiv \frac{\langle \zeta^3 \rangle_c}{\langle \zeta^2 \rangle^{3/2}}, \quad (19)$$

where the subscript  $c$  indicates that only the connected part of the correlator should be included. The skewness  $\hat{S}_3$  encodes information about the bispectrum  $B(k_i)$  integrated over all size and shape configurations and thus provides a meaningful single number to compare the nongaussianity of inflation models which may have very different shapes or running [27].

If we choose  $g^2 \sim 0.01$  (which is compatible with observation for all values of  $\phi_0$ ), then model (5) produces the same value of  $\hat{S}_3$  as a local model (1) with  $|f_{\text{NL}}| \sim 10^2$ . This large value suggests that nongaussianity from particle production during inflation may be observable in future missions.

Depending on model parameters, the nongaussian features predicted by model (5) may lead to a rich variety of observable consequences for the CMB or Large Scale Structure (LSS). The phenomenology of this model is quite different from other constructions that have been proposed to obtain large nongaussianity from inflation. However, the underlying microscopic description (5) is extremely simple and, indeed, rather generic from the low-energy perspective. Explicit realizations of interaction (4) have been obtained from string theory and SUSY. Moreover, in order to obtain an observable signature it was not necessary to fine-tune the inflationary trajectory or appeal to re-summation of an infinite series of high-dimension operators.

### 3. Numerical Study of Rescattering and IR Cascading

**3.1. HLattice Simulations.** In this section we study numerically the creation of  $\delta\phi$  fluctuations by rescattering of the produced  $\chi$  particles off the condensate  $\phi(t)$  in model (5). To this end, we have written a new lattice field theory code, HLattice [112], for simulating the interactions of scalar fields in a cosmological setting. HLattice can be used to simulate the dynamics of any number of interacting scalar fields with arbitrary scalar potential and metric on field space [113]. We solve the Klein-Gordon equations for the scalar field dynamics in an expanding FRW space-time and also solve the Friedmann equation self-consistently for the scale factor,  $a(t)$ . Since the production of long wavelength  $\delta\phi$  modes is so energetically inexpensive, a major requirement for successfully capturing this effect is respecting energy conservation to very high accuracy. HLattice conserves energy with an accuracy of order  $\sim 10^{-8}$ , as compared to  $10^{-3}$ – $10^{-5}$ , which has been obtained using previous codes such as DEFROST [114] or LATTICEASY [115]. A minimum accuracy of order  $10^{-4}$  is required for the problem at hand.

The box size of our  $512^3$  simulations corresponds to a comoving scale which is initially  $(20/2\pi) \sim 3$  times the horizon size  $H^{-1}$ , while  $k_* \cong 60\sqrt{g}H$ . We run our

simulations for roughly 3  $e$ -foldings from the initial moment  $t = 0$  when the  $\chi$  particles are produced although a single  $e$ -folding would have been sufficient to capture the effect. For the sake of illustration, we have chosen the standard chaotic inflation potential  $V = m^2\phi^2/2$  with  $m = 10^{-6}\sqrt{8\pi}M_p$  for our numerical analysis. However, our results do not depend sensitively on the choice of background inflation model. (The model independence of our result arises simply because all the dynamics of rescattering and IR cascading occurs within a single  $e$ -folding from the moment when  $\phi = \phi_0$ . Over such a short time it will always be a good approximation to expand  $\phi(t) \cong \phi_0 + vt$ . Hence the dependence on the background dynamics arises only through  $v = \dot{\phi}(0)$  which is determined by the Hubble scale and the observed amplitude of curvature perturbations. This claim of model independence is born out by explicit analytical calculations in the next section.) We have considered both  $\phi_0 = 2\sqrt{8\pi}M_p$  and  $\phi_0 = 3.2\sqrt{8\pi}M_p$  and also three different values of the coupling constant:  $g^2 = 0.01, 0.1, 1$ . As expected, the coupling  $g^2$  determines the magnitude of the effect while  $\phi_0$  simply shifts the location of the power spectrum feature. For this choice of inflationary potential, the choice  $\phi_0 = 3.2\sqrt{8\pi}M_p$  corresponds to putting the feature on scale slightly smaller than today's horizon. On the other hand,  $\phi_0 = 2\sqrt{8\pi}M_p$  corresponds to placing the feature on scales much smaller than those probed by the CMB (we considered this case in order to be able to directly contrast our results with [64]).

In order to capture the quantum production of  $\chi$  particles using classical lattice simulations, we start our numerical evolution very shortly *after* particle production has occurred, when the  $\chi_k$  modes are nearly adiabatic, but before any significant inflaton fluctuations have been produced. In practice, this corresponds to initializing the simulation at a time  $t_{\text{initial}} = \mathcal{O}(k_\star^{-1})$ . The initial conditions for the modes  $\chi_k(t)$  are given by the usual Bogoliubov computation [80, 92]. These are chosen to reproduce the occupation number  $n_k = e^{-\pi k^2/k_\star^2}$ , while ensuring that the source term for the  $\delta\phi$  fluctuations is turned on smoothly at the initial time. As long as the initial conditions are chosen appropriately, our results are not sensitive to the choice of  $t_{\text{initial}}$ .

At the initial time, the occupation numbers in the inflaton and iso-inflaton fluctuations are small. However, very quickly the massive  $\chi$  particles are diluted away by the expansion of the universe, and the occupation number of the produced IR  $\delta\phi$  fluctuations grows large compared to unity. Thus, classical lattice field theory simulations are sufficient to capture the late-time dynamics. (In the next section we will provide a quantum mechanical treatment of the dynamics of particle production and IR cascading during inflation, which will serve as an a posteriori justification for our classical lattice calculation.)

Our approach is very similar to the methodology that has been employed successfully in studies of preheating after inflation for many years [114, 115]. In that case the initial fluctuations of the fields are chosen to reproduce the exact behaviour of the quantum correlation functions. The occupation numbers of the fields are small at the initial time. However, these grow rapidly as a result of the preheating

instability, and classical simulations are sufficient to capture the late-time dynamics.

**3.2. Numerical Results.** We have studied the fully nonlinear dynamics of  $\chi$  particle production and the subsequent interactions of the produced  $\chi$  with the inflaton field in model (5), as described above. We are interested in the power spectrum of the inflaton fluctuations

$$P_\phi(k) = \frac{k^3}{2\pi^2} |\delta\phi_k|^2. \quad (20)$$

This contains a contribution coming from the usual quantum vacuum fluctuations from inflation that is close to the usual power-law form  $k^{n_s-1}$  on large scales. Such a contribution would be present even in the absence of particle production and is not particularly interesting for us. In order to isolate the effects of rescattering we have subtracted off this component in Figures 4, 5, and 6. In all cases we have normalized  $P_\phi$  to the amplitude of the usual vacuum fluctuations from inflation,  $H^2/(2\pi)^2$ .

Figure 4 shows time evolution of the power in the inflaton fluctuations generated by rescattering, for three different time steps early in the evolution. This figure illustrates how multiple rescatterings lead to a dynamical cascading of power into the IR. To illustrate the magnitude of this effect, the horizontal yellow line corresponds to the amplitude of the usual vacuum fluctuations from inflation. For  $g^2 \gtrsim 0.06$ , the fluctuations from rescattering come to dominate over the vacuum fluctuations within a single  $e$ -folding. In Figure 5 we illustrate how the magnitude of the spectral distortion depends on the coupling,  $g^2$ . (The apparent change in the location of the feature for different values of  $g^2$  arises because we are plotting the power spectrum as a function of  $\ln(k/k_\star)$  and  $k_\star$  depends on  $g$ .)

At late times, the IR portion of the power spectrum illustrated in Figure 4 will remain fixed since the modes associated with these scales have gone outside the horizon and become frozen. On the other hand, the UV portion of this curve corresponds to modes that are still inside the horizon, hence we expect  $\delta\phi_k \sim a^{-1}$  and the UV tail of the power spectrum should damp as  $a^{-2}$ , due to the Hubble expansion. We observe precisely this behaviour in our lattice field theory simulations, and this is illustrated in Figure 6, which displays the dynamics of IR cascading over a much longer time scale.

Within a few  $e$ -foldings from the time of particle production, the entire bump-like feature from IR cascading becomes frozen outside the horizon. At this point the fluctuations have become classical, large-scale adiabatic density perturbations and are observable in the present epoch (presuming that  $\phi = \phi_0$  occurs during the observable range of  $e$ -foldings). In Figure 3 we have illustrated this bump-like feature in both the primordial power spectrum and angular TT spectrum, for a representative choice of parameters.

**3.3. Backreaction Effects.** As discussed previously, the production of  $\chi$  fluctuations at  $t = 0$  backreacts on the homogeneous  $\phi(t)$  causing a transient violation of slow roll.

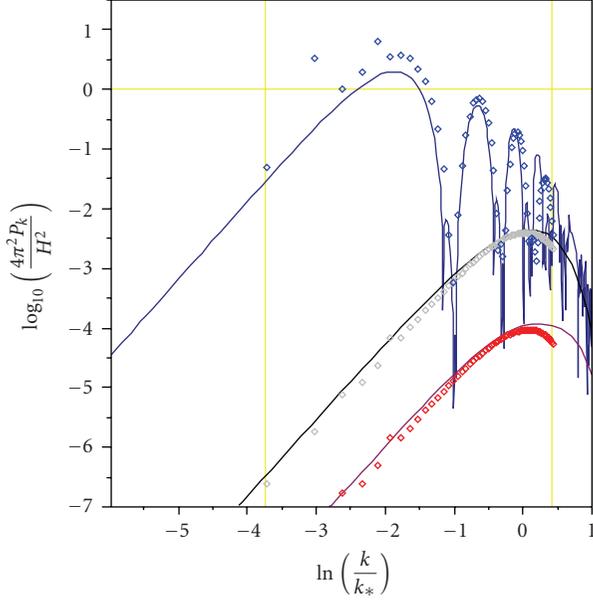


FIGURE 4: The power spectrum of inflaton modes induced by rescattering (normalized to the usual vacuum fluctuations) as a function of  $\ln(k/k_*)$ , plotted for three representative time steps in the evolution, showing the cascading of power into the IR. For each time step we plot the analytical result (the solid line) and the data points obtained using lattice field theory simulations (diamonds). The time steps correspond to the following values of the scale factor:  $a = 1.03, 1.04, 2.20$  (where  $a = 1$  at the moment when  $\phi = \phi_0$ ). By this time the amplitude of fluctuations is saturated due to the expansion of the universe. The vertical lines show the range of scales from our lattice simulation.

We can study this backreaction numerically, by averaging the inhomogeneous field  $\dot{\phi}(t, x)$  over the simulation box. The result is plotted in Figure 2. We have also plotted the analytical solution of the mean field (14), showing that this agrees with the exact numerical result.

The dynamics illustrated in Figure 2 is easy to understand physically. The production of  $\chi$  particles at  $t = 0$  drains kinetic energy from the condensate and hence  $\dot{\phi}$  must decrease abruptly. However, within a few  $e$ -foldings of the moment  $t = 0$ , the produced iso-inflaton particles become extremely massive and are diluted by the expansion as  $a^{-3}$ . At late times the inflaton velocity  $\dot{\phi}$  must tend to the slow-roll value. Notice that the velocity  $\dot{\phi}$  including backreaction effects is not changed significantly, as compared to the usual slow-roll result. This illustrates the energetic cheapness of particle production and IR cascading in model (5).

The transient violation of slow roll illustrated in Figure 2 is expected to induce a ringing pattern in the vacuum fluctuations from inflation [64]. This effect is accounted for automatically in our HLattice simulations. However, we would like to disentangle the effect of backreaction on the cosmological fluctuations from the effect of rescattering. This will be useful in order to compare the relative importance of different physical processes and also to guide our analytical efforts in the next section. To this end, we consider the

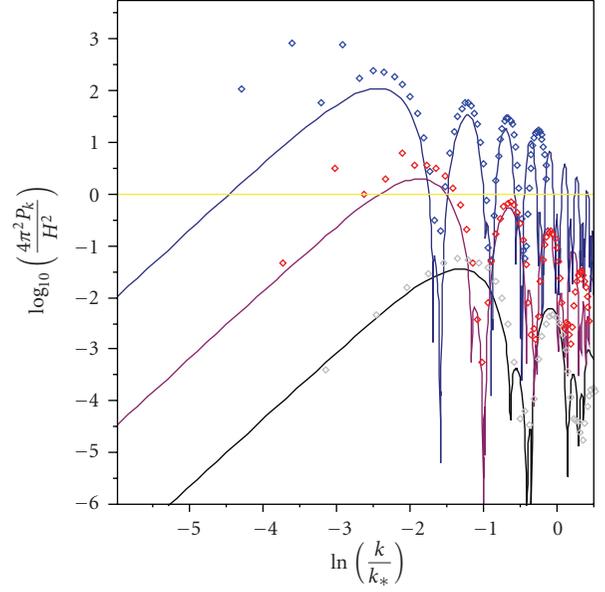


FIGURE 5: The dependence of the power spectrum  $P_\phi$  on the coupling  $g^2$ . The three curves correspond to  $P_\phi$  for  $g^2 = 0.01, 0.1, 1$ , evaluated at a fixed value of the scale factor,  $a = 2.20$ . We see that even for small values of  $g^2$  the inflaton modes induced by rescattering constitute a significant fraction of the usual vacuum fluctuations after only a single  $e$ -folding.

evolution of the curvature perturbation on comoving hypersurfaces,  $\mathcal{R}$ . In linear theory the equation for the Fourier modes  $\mathcal{R}_k$  is well known

$$\mathcal{R}_k'' + 2\frac{z'}{z}\mathcal{R}_k' + k^2\mathcal{R}_k = 0. \quad (21)$$

Here the prime denotes derivatives with respect to conformal time  $\tau = \int(dt/a)$  and  $z \equiv a\dot{\phi}/H$ . Equation (21) is only strictly valid in the absence of entropy perturbations. However, in our case the  $\chi$  field is extremely massive  $m_\chi^2 \gg H^2$  for nearly the entire duration of inflation, hence one may expect that direct iso-curvature contributions to  $\mathcal{R}$  are small. We have solved (21) numerically. In order to take backreaction effects into account we compute the dynamics of  $z(t) = a(t)\dot{\phi}(t)/H(t)$  by averaging over our HLattice simulation box. Next, we solve (21) given this background evolution and compute the power spectrum

$$P_{\mathcal{R}}(k) = \frac{k^3}{2\pi^2} |\mathcal{R}_k|^2. \quad (22)$$

The result is very close to the usual power-law form  $k^{n_s-1}$ , with small superposed oscillations resulting from the transient violation of slow roll; see Figure 7. In order to make the ringing pattern more visible, we have subtracted off the usual (nearly) scale-invariant result which would be obtained in the absence of particle production. For comparison, we also plot the bump-like feature from rescattering and IR cascading. This latter contribution was obtained using the results for  $P_\phi(k)$  from the previous subsection and the naive formula  $\mathcal{R} \sim (H/\dot{\phi})\delta\phi$  (so that  $P_{\mathcal{R}} \sim (2\epsilon M_p^2)^{-1}P_\phi$ ).

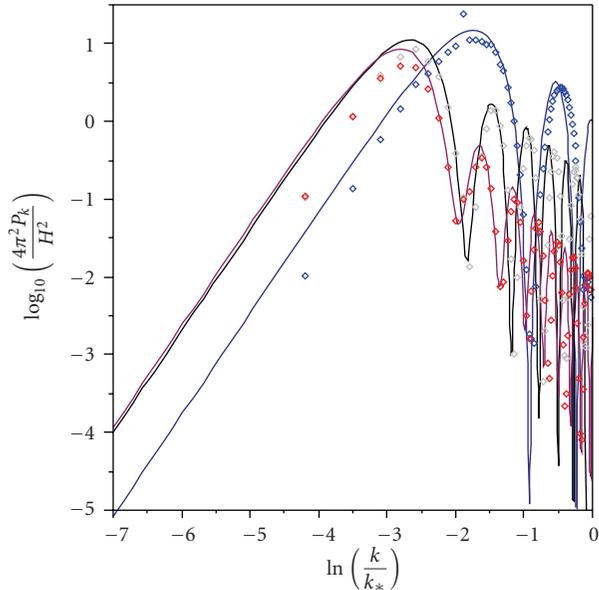


FIGURE 6: The power spectrum of inflaton modes induced by rescattering (normalized to the usual vacuum fluctuations) as a function of  $\ln(k/k_*)$ , plotted for three representative time steps in the late-time evolution. This figure illustrates the final stages of IR cascading; we see the peak of the bump-like feature slide to  $k \sim e^{-3}k_*$ , at which point the associated mode functions  $\delta\phi_k$  have crossed the horizon and become frozen. At later times in the evolution the peak of the feature and also the IR tail ( $\sim k^3$ ) remain fixed. Modes associated with the UV end of the spectrum are still inside the horizon and continue to evolve as  $\delta\phi_k \sim a^{-1}$ , which explains the damping of the  $k > e^{-2}k_*$  part of the spectrum. For each time step we plot the analytical result (the solid line) and the data points obtained using lattice field theory simulations (diamonds).

From Figure 7 we see that IR cascading has a *much* more significant impact on the observable curvature fluctuations than does backreaction. Indeed, for  $g^2 = 0.1$  the transient violation of slow roll yields an order  $10^{-2}$  correction to the vacuum fluctuations while the correction from IR cascading is of order  $10^1$ . This dominance is generic for all values of the coupling. Thus, in developing an analytical theory of particle production during inflation, it is a very good approximation to completely ignore backreaction effects.

#### 4. Analytical Formalism

In the last section, we have studied particle production, rescattering, and IR cascading using nonlinear lattice field theory simulations. In this section we will develop a detailed analytical theory, in order to understand those results from a physical perspective. These results were first presented in [14]. We consider, again, model (5). The equations of motion that we wish to solve are

$$-\square\phi + V'(\phi) + g^2(\phi - \phi_0)\chi^2 = 0, \quad (23)$$

$$-\square\chi + g^2(\phi - \phi_0)^2\chi = 0, \quad (24)$$

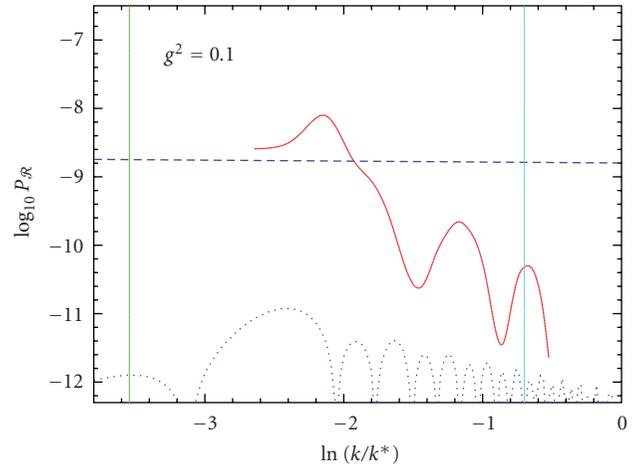


FIGURE 7: A comparison of curvature fluctuations from different physical effects. The dashed blue line is the usual (nearly) scale invariant vacuum fluctuations from inflation. The red solid line is the bump-like feature induced by rescattering and IR cascading. The dotted blue line is the ringing pattern resulting from the momentary slowing-down of the inflaton (computed using the mean field approach of [64]). The vertical lines show  $aH$  at the beginning of particle production and after  $\sim 3e$ -foldings. This figure clearly illustrates the dominance of IR cascading over backreaction effects. For illustration we have taken  $g^2 = 0.1$ , but the dominance is generic for all values of the coupling.

where  $\square = g_{\mu\nu}\nabla^\mu\nabla^\nu$  is the covariant d'Alembertian. It will be useful to work with conformal time  $\tau$ , related to cosmic time  $t$  via  $a d\tau = dt$ . In terms of conformal time the metric takes the form

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t)d\mathbf{x} \cdot d\mathbf{x} \\ &= a^2(\tau)[-d\tau^2 + d\mathbf{x} \cdot d\mathbf{x}]. \end{aligned} \quad (25)$$

We denote derivatives with respect to cosmic time as  $\dot{f} \equiv \partial_t f$  and with respect to conformal time as  $f' \equiv \partial_\tau f$ . The Hubble parameter  $H = \dot{a}/a$  has conformal time analogue  $\mathcal{H} = a'/a$ . For an inflationary (quasi-de Sitter) phase ( $H \cong \text{const}$ ) one has

$$a = -\frac{1}{H\tau} \frac{1}{1-\epsilon}, \quad \mathcal{H} = -\frac{1}{\tau} \frac{1}{1-\epsilon} \quad (26)$$

to leading order in the slow roll parameter  $\epsilon \ll 1$ .

As discussed in Section 2, the motion of the homogeneous inflaton  $\phi(t)$  leads to the production of a gas of  $\chi$  particles at the moment  $t = 0$  when  $\phi = \phi_0$ . The first step in our analytical computation is to describe this burst of particle production in an expanding universe. Following the initial burst, both backreaction and rescattering effects take place. Our formalism will focus on the latter effect, which is much more important, and we provide only a cursory treatment of backreaction.

**4.1. Particle Production in an Expanding Universe.** The first step in our scenario is the quantum mechanical production of  $\chi$  particles due to the motion of  $\phi$ . To understand this

effect we must solve the equation for the  $\chi$  fluctuations in the rolling inflaton background. Approximating  $\phi \cong \phi_0 + vt$  (24) gives

$$\ddot{\chi} + 3H\dot{\chi} - \frac{\bar{\nabla}^2}{a^2}\chi + k_\star^4 t^2 \chi = 0, \quad (27)$$

where  $k_\star \equiv \sqrt{g|v|}$ . We remind the reader that  $k_\star \gg H$  for reasonable values of the coupling; see (11).

The flat space analogue of (27) is very well understood from studies of broad-band parametric resonance during preheating [92] and also moduli trapping at enhanced symmetry points [80]. One does not expect this treatment to differ significantly in our case since both the time scale for particle production  $\Delta t$  and the characteristic wavelength of the produced fluctuations  $\lambda$  are small compared to the Hubble scale:  $\Delta t \sim \lambda^{-1} \sim k_\star^{-1} \ll H^{-1}$ . Hence, we expect that the occupation number of produced  $\chi$  particles will not differ significantly from the flat-space result (13), at least on scales  $k \gtrsim H$ . Furthermore, notice that the  $\chi$  field is extremely massive for most of the inflation

$$\frac{m_\chi^2}{H^2} \cong \frac{k_\star^4 t^2}{H^2}. \quad (28)$$

Since  $k_\star \gg H$ , it follows that  $m_\chi^2 \gg H^2$ , except in a tiny interval  $H|\Delta t| \sim (H/k_\star)^2$  which amounts to roughly  $10^{-3}$  e-foldings for  $g^2 \sim 0.1$ . Therefore, we do not expect any significant fluctuations of  $\chi$  to be produced on superhorizon scales  $k \lesssim H$ .

Let us now consider the solutions of (27). We work with conformal time  $\tau$  and write the Fourier transform of the quantum field  $\chi$  as

$$\chi(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} \frac{\xi_{\mathbf{k}}^\chi(\tau)}{a(\tau)} e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (29)$$

Note the explicit factor of  $a^{-1}$  in (29) which is introduced to give  $\xi_{\mathbf{k}}^\chi$  a canonical kinetic term. The q-number valued Fourier transform  $\xi_{\mathbf{k}}^\chi(\tau)$  can be written as

$$\xi_{\mathbf{k}}^\chi(\tau) = a_{\mathbf{k}} \chi_{\mathbf{k}}(\tau) + a_{-\mathbf{k}}^\dagger \chi_{\mathbf{k}}^\star(\tau), \quad (30)$$

where the annihilation/creation operators satisfy the usual commutation relation

$$[a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = \delta^{(3)}(\mathbf{k} - \mathbf{k}'), \quad (31)$$

and the c-number valued mode functions  $\chi_{\mathbf{k}}$  obey the following oscillator-like equation:

$$\chi_{\mathbf{k}}''(\tau) + \omega_{\mathbf{k}}^2(\tau)\chi_{\mathbf{k}}(\tau) = 0. \quad (32)$$

The time-dependent frequency is

$$\begin{aligned} \omega_{\mathbf{k}}^2(\tau) &= k^2 + a^2 m_\chi^2(\tau) - \frac{a''}{a} \\ &\cong k^2 + \frac{1}{\tau^2} \left[ \frac{k_\star^4}{H^2} t^2(\tau) - 2 \right], \end{aligned} \quad (33)$$

where  $m_\chi^2(\tau) = g^2(\phi - \phi_0)^2 \cong k_\star^4 t^2(\tau)$  is the time-dependent effective mass of the  $\chi$  particles, and

$$t(\tau) = \frac{1}{H} \ln \left( \frac{-1}{H\tau} \right) \quad (34)$$

is the usual cosmic time variable. We have arbitrarily set the origin of conformal time so that  $\tau = -1/H$  corresponds to the moment when  $\phi = \phi_0$ .

In Figure 8(a) we have plotted a representative solution of (32) in order to illustrate the qualitative behaviour of the modes  $\chi_{\mathbf{k}}$ . In Figure 8(b) we plot the occupation number  $n_{\mathbf{k}}$  of particles with momentum  $\mathbf{k}$ , defined as the energy of the mode  $(1/2)|\chi_{\mathbf{k}}'|^2 + (1/2)\omega_{\mathbf{k}}^2|\chi_{\mathbf{k}}|^2$  divided by the energy  $\omega_{\mathbf{k}}$  of each particle. Explicitly, we define

$$n_{\mathbf{k}} = \frac{\omega_{\mathbf{k}}}{2} \left[ \frac{|\chi_{\mathbf{k}}'|^2}{\omega_{\mathbf{k}}^2} + |\chi_{\mathbf{k}}|^2 \right] - \frac{1}{2} \quad (35)$$

where the term  $-1/2$  comes from extracting the zero-point energy of the linear harmonic oscillator (see [92] for a review). From Figure 8(a) we see that, near the massless point  $t = 0$ , the fluctuations  $\chi_{\mathbf{k}}$  get a ‘‘kick’’, and from Figure 8(b) we see that the occupation number  $n_{\mathbf{k}}$  jumps abruptly at this same moment.

Let us now try to understand analytically the behaviour of the solutions of (32). At early times  $t \ll -k_\star^{-1}$ , the frequency  $\omega_{\mathbf{k}}$  varies adiabatically

$$\left| \frac{\omega_{\mathbf{k}}'}{\omega_{\mathbf{k}}^2} \right| \ll 1. \quad (36)$$

In this in-going adiabatic regime the modes  $\chi_{\mathbf{k}}$  are not excited and the solution of (32) is well described by the adiabatic solution  $\chi_{\mathbf{k}}(\tau) = f_{\mathbf{k}}(\tau)$  where

$$f_{\mathbf{k}}(\tau) \equiv \frac{1}{\sqrt{2\omega_{\mathbf{k}}(\tau)}} \exp \left[ -i \int^\tau d\tau' \omega_{\mathbf{k}}(\tau') \right]. \quad (37)$$

We have normalized (37) to be pure positive frequency so that the state of the iso-inflaton field at early times corresponds to the adiabatic vacuum with no  $\chi$  particles. (Inserting (37) into (35) one finds  $n_{\mathbf{k}} = 0$  for the adiabatic solution, as expected.)

The adiabatic solution (37) ceases to be a good approximation very close to the moment when  $\phi = \phi_0$ , that is at times  $|t| \lesssim k_\star^{-1}$ . In this regime the adiabaticity condition (36) is violated for modes with wavenumber  $H \lesssim k \lesssim k_\star$  and  $\chi$  particles within this momentum band are produced. During the non-adiabatic regime we can still represent the solutions of (32) in terms of the functions  $f_{\mathbf{k}}(\tau)$  as

$$\chi_{\mathbf{k}}(\tau) = \alpha_{\mathbf{k}}(\tau) f_{\mathbf{k}}(\tau) + \beta_{\mathbf{k}}(\tau) f_{\mathbf{k}}^\star(\tau). \quad (38)$$

This expression affords a solution of (32) provided the time-dependent Bogoliubov coefficients obey the following set of coupled equations:

$$\begin{aligned} \alpha_{\mathbf{k}}'(\tau) &= \frac{\omega_{\mathbf{k}}'(\tau)}{2\omega_{\mathbf{k}}(\tau)} \exp \left[ +2i \int^\tau d\tau' \omega_{\mathbf{k}}(\tau') \right] \beta_{\mathbf{k}}(\tau), \\ \beta_{\mathbf{k}}'(\tau) &= \frac{\omega_{\mathbf{k}}'(\tau)}{2\omega_{\mathbf{k}}(\tau)} \exp \left[ -2i \int^\tau d\tau' \omega_{\mathbf{k}}(\tau') \right] \alpha_{\mathbf{k}}(\tau). \end{aligned} \quad (39)$$

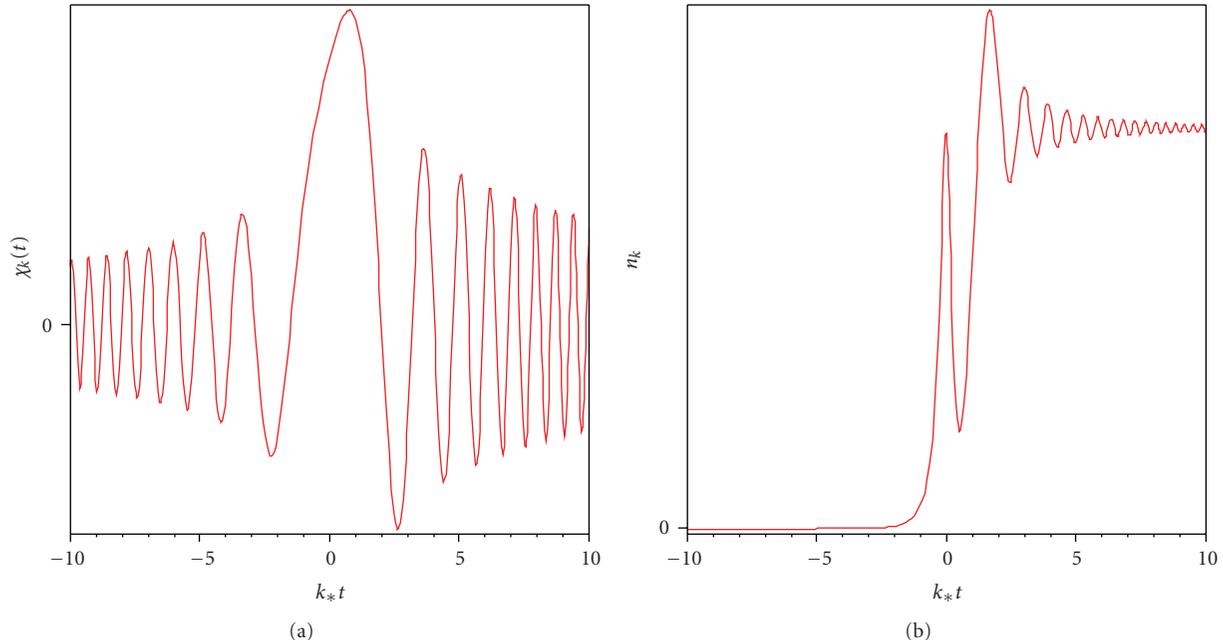


FIGURE 8: (a) illustrates the time dependence of the solutions  $\chi_k$  of (32) for a representative choice of parameters. The oscillatory behaviour at early times represents the adiabatic initial condition. At  $t = 0$  the effective frequency (33) varies nonadiabatically and the fluctuations get a “kick”. (b) plots the occupation number (35) for the same mode. No particles are present in the adiabatic in-going regime. This figure illustrates how the violations of adiabaticity at  $t = 0$  lead to production of  $\chi$  particles.

The Bogoliubov coefficients are normalized as  $|\alpha_k|^2 - |\beta_k|^2 = 1$ , and the assumption that no  $\chi$  particles are present in the asymptotic past (this assumption is justified since any initial excitation of  $\chi$  would have been damped out exponentially fast by the expansion of the universe) fixes the initial conditions  $\alpha_k = 1$ ,  $\beta_k = 0$  for  $t \rightarrow -\infty$ . This is known as the adiabatic initial condition.

From the structure of (39) it is clear that violation of condition (36) near  $t = 0$  leads to a rapid growth in the  $|\beta_k|$  coefficient. The time variation of  $\beta_k$  can be interpreted as a corresponding growth in the occupation number of the  $\chi$  particles

$$n_k = |\beta_k|^2. \quad (40)$$

At late times ( $t \gtrsim k_*^{-1}$ ) adiabaticity is restored and the growth of  $n_k = |\beta_k|^2$  must saturate. By inspection of (39) we can see that the Bogoliubov coefficients must tend to constant values in the out-going adiabatic regime. Therefore, within less than an  $e$ -folding from the moment of particle production the solution  $\chi_k$  of (32) can be represented as a simple superposition of positive frequency  $f_k$  modes and negative frequency  $f_k^*$  modes. Our goal now is to derive an analytical expression for the modes  $\chi_k$  which is valid in this out-going adiabatic region.

Let us first study the adiabatic solution  $f_k(\tau)$ . If we focus on the interesting region of phase space,  $H \lesssim k \lesssim k_*$ , then the adiabatic solution (37) is very well approximated by

$$f_k(\tau) \cong \frac{1}{a^{1/2} k_* \sqrt{2t(\tau)}} e^{-(i/2)k_*^2 t^2(\tau)}, \quad (41)$$

where  $t(\tau)$  is defined by (34). It is interesting to note that (41) is identical to the analogous flat-space result [12], except for the factor of  $a^{-1/2}$ . Taking into account also the explicit factor of  $a^{-1}$  in our definition of the Fourier transform (29), we recover the expected large-scale behaviour for a massive field in de Sitter space, that is,  $\chi \sim a^{-3/2}$ . This dependence on the scale factor is easy to understand physically, it simply reflects the volume dilution of nonrelativistic particles:  $\rho_\chi \sim m_\chi^2 \chi^2 \sim a^{-3}$ .

Next, we seek an expression for the Bogoliubov coefficients  $\alpha_k, \beta_k$  in the out-going adiabatic regime  $t \gtrsim k_*^{-1}$ . From (39) it is clear that the value of the Bogoliubov coefficients at late times can depend only on dynamics during the interval  $|t| \lesssim k_*^{-1}$  where adiabaticity condition (36) is violated. This interval is tiny compared to the expansion time, and we are justified in treating  $a(\tau)$  as a constant during this phase. Hence, it follows that the flat space computation of the Bogoliubov coefficients [80, 92] must apply, at least for scales  $k \gtrsim H$ . To a very good approximation we therefore have the well-known result

$$\alpha_k \cong \sqrt{1 + e^{-\pi k^2/k_*^2}}, \quad (42)$$

$$\beta_k \cong -ie^{-\pi k^2/(2k_*^2)} \quad (43)$$

in the out-going adiabatic regime. Equation (43) gives the usual expression (13) for the co-moving occupation number of particles produced by a single burst of broad-band parametric resonance:

$$n_k = |\beta_k|^2 = e^{-\pi k^2/k_*^2}. \quad (44)$$

Finally, we arrive at an expression for the out-going adiabatic  $\chi$  modes which is accurate for interesting scales  $k_* \lesssim k \lesssim H$ . Putting together results (41) and (38) along with well-known expressions (42) and (43) we arrive at

$$\begin{aligned} \chi_k(\tau) \cong & \sqrt{1 + e^{-\pi k^2/k_*^2}} \frac{1}{a^{1/2} k_* \sqrt{2t(\tau)}} e^{-(i/2)k_*^2 t^2(\tau)} \\ & - i e^{-\pi k^2/(2k_*^2)} \frac{1}{a^{1/2} k_* \sqrt{2t(\tau)}} e^{+(i/2)k_*^2 t^2(\tau)}, \end{aligned} \quad (45)$$

valid for  $t \gtrsim k_*^{-1}$ . Equation (45) is the main result of this subsection. We will now justify that this expression is quite sufficient for our purposes.

For modes deep in the UV,  $k \gtrsim k_*$ , our expression (45), is not accurate. (Expression (41) for the adiabatic modes  $f_k$  is not valid at high momenta where  $\omega_k \cong k$ .) However, such high momentum particles are not produced, condition (36) is always satisfied for  $k \gg k_*$ . Note that the absence of particle production deep in the UV is built into our expression (45): as  $k \rightarrow \infty$  this function tends to the vacuum solution  $\chi_k \rightarrow f_k$ .

Our expression (45) is also not valid deep in the IR, for modes  $k < H$ . To justify this neglect requires somewhat more care. Notice that, even very far from the massless point,  $t = 0$ , long wavelength modes  $k \ll H$  should not be thought of as particle like. The large-scale mode functions are not oscillatory but rather damp exponentially fast as  $\chi \sim a^{-3/2}$ . Hence, even if we started with some super-horizon fluctuations of  $\chi$  at the beginning of inflation, these would be suppressed by an exponentially small factor before the time when particle production occurs. Any super-horizon fluctuation generated near  $t = 0$  would need to be exponentially huge to overcome this damping. However, resonant particle production during inflation does *not* lead to exponential growth of mode functions. (In this regard our scenario is very different from preheating at the end of inflation. In the latter case the inflaton passes *many* times through the massless point  $m_\chi = 0$ , and there are, correspondingly, many bursts of particle production. After many oscillations of the inflaton field, the  $\chi$  particle occupation numbers build up to become exponentially large, and, averaged over many oscillations of the background, the  $\chi$  mode functions grow exponentially. However, in our case there is only a *single* burst of particle production at  $t = 0$ . The resulting occupation number (13) is always less than unity, and the solutions of (32) never display exponential growth.)

To verify explicitly that there is no significant effect for super-horizon fluctuations let us consider solving (27), neglecting gradient terms. The equation we wish to solve, then, is

$$\partial_t^2 (a^{3/2} \chi) + \left[ k_*^4 t^2 - \frac{9}{4} H^2 \right] (a^{3/2} \chi) = 0. \quad (46)$$

(For simplicity we take  $\epsilon = 0$  for this paragraph; however, this has no effect on our results.) The solution of this

equation may be written in terms of parabolic cylinder functions  $D_\nu(z)$  as

$$\begin{aligned} \chi(t, \mathbf{x}) \sim & \frac{1}{a^{3/2}} \left( C_1 D_{-1/2+(9H^2/8k_*^2)i} [(1+i)k_* t] \right. \\ & \left. + C_2 D_{-1/2-(9H^2/8k_*^2)i} [(-1+i)k_* t] \right). \end{aligned} \quad (47)$$

For our purposes the precise values of the coefficients  $C_1, C_2$  are not important. Rather, it suffices to note that for  $k_* |t| \gtrsim 1$  function (47) behaves as

$$\chi(t, \mathbf{x}) \sim |t|^{-1/2} e^{-3Ht/2} \times [\text{oscillatory}]. \quad (48)$$

This explicit large-scale asymptotics confirms our previous claims that the super-horizon fluctuations of  $\chi$  damp to zero exponentially fast, as  $a^{-3/2} \sim e^{-3Ht/2}$ . As discussed previously, this damping is easy to understand in terms of the volume dilution of non-relativistic particles. We can also understand the power-law damping that appears in (48) from a physical perspective. The properly normalized modes behave as  $a^{3/2} \chi \sim \omega_k^{-1/2}$  while on large scales we have  $\omega_k \sim |m_\chi| \sim |t|$ . Hence, the late-time damping factor  $t^{-1/2}$  which appears in (48) reflects the fact that the  $\chi$  particles become ever more massive as  $\phi$  rolls away from the point  $\phi_0$ .

Finally, it is straightforward to see that function (47) does not display any exponential growth near  $t = 0$ . Hence, we conclude that there is no significant generation of super-horizon  $\chi$  fluctuations due to particle production. (This is strictly true only in the linearized theory. It is possible that  $\chi$  particles are generated by nonlinear effects such as rescattering. However, even such second-order  $\chi$  fluctuations will be extremely massive compared to the Hubble scale and must therefore suffer exponential damping  $a^{-3/2}$  on large scales.)

In this subsection we have seen that the quantum production of  $\chi$  particles in an expanding universe proceeds very much as it does in flat space. This is reasonable since particle production occurs on a time scale short compared to the expansion time and involves modes which are inside the horizon at the time of production.

**4.2. Inflaton Fluctuations.** In Section 4.1 we studied the quantum production of  $\chi$  particles which occurs when  $\phi$  rolls past the massless point  $\phi = \phi_0$ . Subsequently, there are two distinct physical processes which take place: backreaction and rescattering. As we have argued in Section 3, the former effect has a negligible impact of the observable spectrum of cosmological perturbations. Hence, we will not study this effect in any detail (see [61, 63, 64] for analytical calculations). Instead we provide a cursory treatment of backreaction in Appendix A, in order to clear up some common misconceptions.

In this subsection we study the rescattering of produced  $\chi$  particle off the inflaton condensate. The dominant process to consider is the diagram illustrated in Figure 1, corresponding to bremsstrahlung emission of  $\delta\phi$  fluctuations (particles) in the background of the external field. (There is also a subdominant process of the type  $\chi\chi \rightarrow \delta\phi\delta\phi$  which is phase-space suppressed.) Taking into account only the rescattering

diagram illustrated in Figure 1 is equivalent to solving the following equation for the q-number inflaton fluctuation:

$$\left[ \partial_\tau^2 + 3H\partial_\tau - \frac{\bar{\nabla}^2}{a^2} + m^2 \right] \delta\phi = -g^2[\phi(t) - \phi_0]\chi^2, \quad (49)$$

where we have introduced the notation  $m^2 \equiv V_{,\phi\phi}$  for the inflaton effective mass. (Note that we are not assuming a background potential of the form  $m^2\phi^2/2$ , only that  $V_{,\phi\phi} \neq 0$  in the vicinity of the point  $\phi = \phi_0$ .)

Equation (49) may be derived by noting that (5) gives an interaction of the form  $g^2(\phi - \phi_0)\delta\phi\chi^2$  between the inflaton and iso-inflaton, in the background of the external field  $\phi(t)$ . Equivalently, one may construct this equation by a straightforward iterative solution of (23).

We work in conformal time and define the q-number Fourier transform  $\xi_{\mathbf{k}}^\phi(\tau)$  of the inflaton fluctuation analogously to (29):

$$\delta\phi(\tau, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} \frac{\xi_{\mathbf{k}}^\phi(\tau)}{a(\tau)} e^{i\mathbf{k}\cdot\mathbf{x}}. \quad (50)$$

(To avoid potential confusion we again draw the attention of the reader to the explicit factor  $a^{-1}$  in our convention for the Fourier transform.) The equation of motion (49) now takes the form

$$\begin{aligned} & \left[ \partial_\tau^2 + k^2 + a^2 m^2 - \frac{a''}{a} \right] \xi_{\mathbf{k}}^\phi(\tau) \\ & = -gk_\star^2 a(\tau) t(\tau) \int \frac{d^3k'}{(2\pi)^{3/2}} \xi_{\mathbf{k}'}^\chi \xi_{\mathbf{k}-\mathbf{k}'}^\chi(\tau). \end{aligned} \quad (51)$$

The solution of (51) consists of two parts: the solution of the homogeneous equation and the particular solution which is due to the source. The former corresponds, physically, to the usual vacuum fluctuations from inflation. On the other hand, the particular solution corresponds physically to the secondary inflaton modes which are generated by rescattering.

**4.3. Homogeneous Solution and Green Function.** We consider first the homogeneous solution of (51). Since the homogeneous solution is a Gaussian field, we may expand the q-number Fourier transform in terms of annihilation/creation operators  $b_{\mathbf{k}}, b_{\mathbf{k}}^\dagger$  and c-number mode functions  $\phi_{\mathbf{k}}(\tau)$  as

$$\xi_{\mathbf{k}}^\phi(\tau) = b_{\mathbf{k}}\phi_{\mathbf{k}}(\tau) + b_{-\mathbf{k}}^\dagger\phi_{\mathbf{k}}^*(\tau). \quad (52)$$

Here the inflaton annihilation/creation operators  $b_{\mathbf{k}}, b_{\mathbf{k}}^\dagger$  obey

$$\left[ b_{\mathbf{k}}, b_{\mathbf{k}'}^\dagger \right] = \delta^{(3)}(\mathbf{k} - \mathbf{k}') \quad (53)$$

and commute with the annihilation/creation operators of the  $\chi$ -field:

$$\left[ a_{\mathbf{k}}, b_{\mathbf{k}'}^\dagger \right] = \left[ a_{\mathbf{k}}, b_{\mathbf{k}'} \right] = 0. \quad (54)$$

Using (26) and (7) it is straightforward to see that the homogeneous inflaton mode functions obey the following equation:

$$\partial_\tau^2 \phi_{\mathbf{k}} + \left[ k^2 - \frac{1}{\tau^2} \left( \nu^2 - \frac{1}{4} \right) \right] \phi_{\mathbf{k}} = 0, \quad (55)$$

where we have defined

$$\nu \cong \frac{3}{2} - \eta + \epsilon. \quad (56)$$

The properly normalized mode function solutions are well known and may be written in terms of the Hankel function of the first kind as

$$\phi_{\mathbf{k}}(\tau) = \frac{\sqrt{\pi}}{2} \sqrt{-\tau} H_\nu^{(1)}(-k\tau). \quad (57)$$

This solution corresponds to the usual quantum vacuum fluctuations of the inflaton field during inflation.

In passing, let us compute the power spectrum of the quantum vacuum fluctuations from inflation. Using the solutions (57) we have

$$P_\phi^{\text{vac}}(k) = \frac{k^3}{2\pi^2} \left| \frac{\phi_{\mathbf{k}}}{a} \right|^2 \cong \frac{H^2}{(2\pi)^2} \left( \frac{k}{aH} \right)^{n_s-1} \quad (58)$$

on large scales  $k \ll aH$ . The explicit factor of  $a^{-2}$  in (58) appears to cancel the  $a^{-1}$  in our definition of the Fourier transform (50). The spectral index is

$$n_s - 1 = 3 - 2\nu \cong 2\eta - 2\epsilon \quad (59)$$

using (56).

Given the solution (57) of the homogeneous equation, it is now trivial to construct the retarded Green function for (51). This may be written in terms of the free theory mode functions (57) as

$$\begin{aligned} G_{\mathbf{k}}(\tau - \tau') &= i\Theta(\tau - \tau') \left[ \phi_{\mathbf{k}}(\tau)\phi_{\mathbf{k}}^*(\tau') - \phi_{\mathbf{k}}^*(\tau)\phi_{\mathbf{k}}(\tau') \right] \\ &= \frac{i\pi}{4} \Theta(\tau - \tau') \sqrt{\tau\tau'} \left[ H_\nu^{(1)}(-k\tau) H_\nu^{(1)}(-k\tau')^* \right. \\ &\quad \left. - H_\nu^{(1)}(-k\tau)^* H_\nu^{(1)}(-k\tau') \right]. \end{aligned} \quad (60)$$

**4.4. Particular Solution: Rescattering Effects.** We now consider the particular solution of (51). This is readily constructed using the Green function (60) as

$$\begin{aligned} & \xi_{\mathbf{k}}^\phi(\tau) \\ &= -\frac{gk_\star^2}{(2\pi)^{3/2}} \int d\tau' d^3k' G_{\mathbf{k}}(\tau - \tau') a(\tau') t(\tau') \xi_{\mathbf{k}'}^\chi \xi_{\mathbf{k}-\mathbf{k}'}^\chi(\tau'). \end{aligned} \quad (61)$$

Notice that this particular solution is statistically independent of the homogeneous solution (52). In other words, the particular solution can be expanded in terms of the annihilation/creation operators  $a_{\mathbf{k}}, a_{\mathbf{k}}^\dagger$  associated with the  $\chi$

field whereas the homogeneous solution is written in terms of the annihilation/creation operators  $b_{\mathbf{k}}, b_{\mathbf{k}}^\dagger$  associated with the inflaton vacuum fluctuations. These two sets of operators commute with one another.

We will ultimately be interested in computing the  $n$ -point correlation functions of the particular solution (62). For example, carefully carrying out the Wick contractions, the connected contribution to the 2-point function is

$$\begin{aligned} & \langle \xi_{\mathbf{k}_1}^\phi \xi_{\mathbf{k}_2}^\phi(\tau) \rangle \\ &= \frac{2g^2 k_*^4}{(2\pi)^3} \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2) \\ & \times \int d\tau' d\tau'' a(\tau') a(\tau'') t(\tau') t(\tau'') G_{k_1}(\tau - \tau') G_{k_2}(\tau - \tau'') \\ & \times \int d^3 k' \chi_{k_1 - k'}(\tau') \chi_{k_1 - k'}^*(\tau'') \chi_{k'}(\tau') \chi_{k'}^*(\tau''). \end{aligned} \quad (62)$$

The power spectrum of  $\delta\phi$  fluctuations generated by rescattering is then defined in terms of the 2-point function in the usual manner

$$\langle \xi_{\mathbf{k}}^\phi(t) \xi_{\mathbf{k}'}^\phi(\tau) \rangle \equiv \delta^{(3)}(\mathbf{k} + \mathbf{k}') \frac{2\pi^2}{k^3} a(\tau)^2 P_\phi^{\text{resc}}. \quad (63)$$

(The explicit factor of  $a^2$  in definition (63) appears to cancel the factor of  $a^{-1}$  in our convention for Fourier transforms (50).)

The total power spectrum is simply the sum of the contribution from the vacuum fluctuations (58) and the contribution from rescattering (63):

$$P_\phi(k) = P_\phi^{\text{vac}}(k) + P_\phi^{\text{resc}}(k). \quad (64)$$

There are no cross-terms, owing to the fact that  $a_{\mathbf{k}}$  and  $b_{\mathbf{k}}$  commute.

**4.5. Renormalization.** We now wish to evaluate the 2-point correlator (62). In principle, this is straightforward: first substitute result (45) for the  $\chi_k$  modes and result (60) for the Green function into (62), next evaluate the integrals. However, there is a subtlety. The resulting power spectrum is formally infinite. Moreover, the 2-point correlation function (62) receives contributions from two distinct effects. There is a contribution from particle production, which we are interested in. However, there is also a contribution coming from quantum vacuum fluctuations of the  $\chi$  field interacting nonlinearly with the inflaton. The latter contribution would be present even in the absence of particle production, when  $\alpha_k = 1, \beta_k = 0$ .

In order to isolate the effects of particle production on the inflaton fluctuations, we would like to subtract off the contribution to the 2-point correlation function (62) which is coming from the quantum vacuum fluctuations of  $\chi$ . This subtraction also has the effect of rendering the power spectrum (63) finite, since it extracts the usual UV divergent contribution associated with the Minkowski-space vacuum fluctuations.

As a step towards renormalizing the 2-point correlation function of inflaton fluctuations from rescattering (62), let us first consider the simpler problem of renormalizing the 2-point function of the gaussian field  $\chi$ . We defined the renormalized 2-point function in momentum space as follows:

$$\begin{aligned} \langle \xi_{k_1}^\chi(t_1) \xi_{k_2}^\chi(t_2) \rangle_{\text{ren}} &= \langle \xi_{k_1}^\chi(t_1) \xi_{k_2}^\chi(t_2) \rangle \\ & - \langle \xi_{k_1}^\chi(t_1) \xi_{k_2}^\chi(t_2) \rangle_{\text{in}}. \end{aligned} \quad (65)$$

In (65) the quantity  $\langle \xi_{k_1}^\chi(t_1) \xi_{k_2}^\chi(t_2) \rangle_{\text{in}}$  is the contribution which would be present even in the absence of particle production, computed by simply taking solution (38) with  $\alpha_k = 1, \beta_k = 0$ . Explicitly, we have

$$\langle \xi_{k_1}^\chi(t_1) \xi_{k_2}^\chi(t_2) \rangle_{\text{in}} = \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2) f_{k_1}(t_1) f_{k_2}^*(t_2), \quad (66)$$

where  $f_k$  are the adiabatic solutions (37).

To see the impact of this subtraction, let us consider the renormalized variance for the iso-inflaton field,  $\langle \chi^2 \rangle$ . Employing prescription (65) we have

$$\begin{aligned} \langle \chi^2(\tau, \mathbf{x}) \rangle_{\text{ren}} &= \int \frac{d^3 k}{(2\pi)^3 a^2(\tau)} \left[ |\chi_k(\tau)|^2 - \frac{1}{2\omega_k(\tau)} \right] \\ &= \langle \chi^2(\tau, \mathbf{x}) \rangle - \delta_M, \end{aligned} \quad (67)$$

where  $\delta_M$  is the contribution from the Coleman-Weinberg potential. This proves that our prescription reproduces the scheme advocated in [80]. The renormalized variance (67) is finite and may be computed explicitly using our solutions (45). We find

$$\langle \chi^2(t, \mathbf{x}) \rangle_{\text{ren}} \cong \frac{n_\chi a^{-3}}{g |\phi - \phi_0|}, \quad (68)$$

where

$$n_\chi \equiv \int \frac{d^3 k}{(2\pi)^3} n_k \sim k_*^3 \quad (69)$$

is the total co-moving number density of produced  $\chi$  particles. Result (68) was employed in [64] to quantify the effect of backreaction on the inflaton condensate in the mean field treatment (14). Hence, the renormalization scheme (65) was implicit in that calculation also.

At the level of the 2-point function, our renormalization scheme is tantamount to assuming that Coleman-Weinberg corrections are already absorbed into the definition of the inflaton potential,  $V(\phi)$ . In general, such corrections might steepen  $V(\phi)$  and spoil slow-roll inflation. Here, we assume that this problem has already been dealt with, either by fine-tuning the bare inflaton potential or else by including extended SUSY (which can minimize dangerous corrections). See also [80] for a related discussion.

Having established a scheme for renormalizing the 2-point function of the gaussian field  $\chi$ , it is now straightforward to consider higher-order correlation functions. We

simply rewrite the 4-point function as a product of 2-point functions using Wick's theorem. Next, each Wick contraction is renormalized as (65). Applying this prescription to (62) amounts to

$$\begin{aligned}
& \left\langle \xi_{\mathbf{k}_1}^\phi(\tau) \xi_{\mathbf{k}_2}^\phi(\tau) \right\rangle_{\text{ren}} \\
&= \frac{2g^2 k_\star^4}{(2\pi)^3} \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2) \\
&\times \int d\tau' d\tau'' t(\tau') t(\tau'') a(\tau') a(\tau'') G_{k_1}(\tau - \tau') G_{k_2}(\tau - \tau'') \\
&\times \int d^3 k' \left[ \chi_{k_1-k'}(\tau') \chi_{k_1-k'}^\star(\tau'') - f_{k_1-k'}(\tau') f_{k_1-k'}^\star(\tau'') \right] \\
&\times \left[ \chi_{k'}(\tau') \chi_{k'}^\star(\tau'') - f_{k'}(\tau') f_{k'}^\star(\tau'') \right], \tag{70}
\end{aligned}$$

where  $f_k(t)$  are the adiabatic solutions defined in (37).

**4.6. Power Spectrum.** We are now in a position to compute the renormalized power spectrum of inflation fluctuations generated by rescattering,  $P_\phi^{\text{resc}}(k)$ . We renormalize the 2-point correlator of the inflaton fluctuations generated by rescatter according to (70) and extract the power spectrum by comparison to (63). We have relegated the technical details to Appendix B and here we simply state the final result:

$$\begin{aligned}
& P_\phi^{\text{resc}}(k) \\
&= \frac{g^2 k^3 k_\star}{16\pi^5} \left[ \frac{e^{-\pi k^2/(2k_\star^2)}}{2\sqrt{2}} \left( I_2(k, \tau)^2 + |I_1(k, \tau)|^2 \right) \right. \\
&\quad \left. + \left[ e^{-\pi k^2/(4k_\star^2)} + \frac{1}{2\sqrt{2}} e^{-3\pi k^2/(8k_\star^2)} \right] \right. \\
&\quad \times \left( I_2(k, \tau)^2 - \text{Re}[I_1(k, \tau)] \right) \\
&\quad \left. + \left[ \frac{8\sqrt{2}}{3\sqrt{3}} e^{-\pi k^2/(3k_\star^2)} + \frac{4\sqrt{2}}{5\sqrt{5}} e^{-3\pi k^2/(5k_\star^2)} \right] \right. \\
&\quad \left. \times \text{Im}[I_1(k, \tau) I_2(k, \tau)] \right], \tag{71}
\end{aligned}$$

where the functions  $I_1, I_2$  are the curved space generalization of the characteristic integrals defined in [12]. Explicitly we have

$$\begin{aligned}
I_1(k, \tau) &= \frac{1}{a(\tau)} \int d\tau' G_k(\tau - \tau') e^{ik_\star^2 t^2(\tau')}, \\
I_2(k, \tau) &= \frac{1}{a(\tau)} \int d\tau' G_k(\tau - \tau'). \tag{72}
\end{aligned}$$

The characteristic integral  $I_2$  can be evaluated analytically; however, the resulting expression is not particularly enlightening. Evaluation of the integral  $I_1$  requires numerical methods. More details are in Appendix B. Equation (71) is the main result of this section.

**4.7. Comparison to Lattice Field Theory Simulations.** In Section 3 the results of our analytical formalism were plotted alongside the output of fully nonlinear HLattice simulations. It is evident from Figures 4, 5, and 6 that the agreement between these approaches is extremely good, even very late into the evolution and in the regime  $g^2 \sim 1$ . The consistency of perturbative quantum field theory analytics and nonlinear classical lattice simulations provides a highly nontrivial check on our calculation.

**4.8. The Bispectrum.** So far, we have shown how to compute analytically the power spectrum generated by particle production, rescattering, and IR cascading during inflation in model (5). We found that IR cascading leads to a bump-like contribution to the primordial power spectrum of the inflaton fluctuations. However, this same dynamics must also have a nontrivial impact on nongaussian statistics, such as the bispectrum. Indeed, it is already evident from our previous analysis that the inflaton fluctuations generated by rescattering may be significantly nongaussian. From expression (62) we see that the particular solution (due to rescattering) which is bi-linear is the gaussian field  $\chi$ .

We define the bispectrum of the inflaton field fluctuations in terms of the three-point correlation function as

$$\left\langle \xi_{\mathbf{k}_1}^\phi \xi_{\mathbf{k}_2}^\phi \xi_{\mathbf{k}_3}^\phi(\tau) \right\rangle = (2\pi)^3 a^3(\tau) \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_\phi(k_i). \tag{73}$$

The factor  $a^3$  appears in (73) to cancel the explicit factors of  $a^{-1}$  in our convention (50) for the Fourier transform. It is well known that the nongaussianity associated with the usual quantum vacuum fluctuations of the inflaton is negligible [16–18]; therefore, when evaluating the bispectrum (73) we consider only the particular solution (62) which is due to rescattering. Carefully carrying out the Wick contractions, we find the following result for the renormalized 3-point function:

$$\begin{aligned}
& \left\langle \xi_{\mathbf{k}_1}^\phi \xi_{\mathbf{k}_2}^\phi \xi_{\mathbf{k}_3}^\phi(\tau) \right\rangle_{\text{ren}} \\
&= \frac{4g^3 k_\star^6}{(2\pi)^{9/2}} \delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \prod_{i=1}^3 \int d\tau_i t(\tau_i) a(\tau_i) G_{k_i}(\tau - \tau_i) \\
&\times \int d^3 p \left[ \chi_{k_1-p}(\tau_1) \chi_{k_1-p}^\star(\tau_2) - f_{k_1-p}(\tau_1) f_{k_1-p}^\star(\tau_2) \right] \\
&\times \left[ \chi_{k_3+p}(\tau_2) \chi_{k_3+p}^\star(\tau_3) - f_{k_3+p}(\tau_2) f_{k_3+p}^\star(\tau_3) \right] \\
&\times \left[ \chi_p(\tau_1) \chi_p^\star(\tau_3) - f_p(\tau_1) f_p^\star(\tau_3) \right] + (k_2 \longleftrightarrow k_3), \tag{74}
\end{aligned}$$

where the modes  $\chi_k$  are defined by (38) and  $f_k$  are the adiabatic solutions (37). On the last line of (74) we have labeled schematically terms which are identical to the preceding three lines, only with  $k_2$  and  $k_3$  interchanged. One may verify that this expression is symmetric under interchange of the momenta  $k_i$  by changing dummy variables of integration.

It is now straightforward (but tedious) to plug expressions (41) and (45) into (74) and evaluate the integrals.

This computation is tractable analytically because the time and phase-space integrals decouple. The bispectrum is then extracted by comparison to (73). This computation is carried out in detail in [14] where we have shown that  $B_\phi(k_i)$  peaks only over triangles with a characteristic size, corresponding to the location of the bump in the power spectrum. This result is easy to understand on physical grounds, all the dynamics of rescattering and IR cascading take place over a very short time near the moment  $t = 0$ . Hence, the effect of this dynamics on the primordial fluctuations must be limited to scales leaving the horizon near the time when particle production occurs.

We will provide a cursory discussion of the bispectrum  $B_\phi(k_i)$  in Section 8 when we discuss nongaussianity from particle production during inflation.

**4.9. Inclusion of a Bare Iso-Inflaton Mass.** In passing, it may be interesting to consider how the analysis of this section is modified in the case that our prototype action (5) is supplemented by a bare mass term for the iso-inflaton field of the form  $\delta\mathcal{L} = -(1/2)\mu^2\chi^2$ . Thus, in place of (5) suppose that we consider the model

$$\begin{aligned} \mathcal{L} = & -\frac{1}{2}(\partial\phi)^2 - V(\phi) \\ & -\frac{1}{2}(\partial\chi)^2 - \frac{1}{2}\mu^2\chi^2 - \frac{g^2}{2}(\phi - \phi_0)^2\chi^2. \end{aligned} \quad (75)$$

Now the  $\chi$  particles do not become massless at the point  $\phi = \phi_0$ , but rather the effective mass-squared

$$m_\chi^2 = \mu^2 + g^2(\phi - \phi_0)^2 \quad (76)$$

reaches a minimum value  $\mu^2$  (which we assume to be positive). Such a correction may arise due to a variety of effects and will reduce the impact of particle production and IR cascading on the observable cosmological fluctuations.

Let us briefly consider how the additional bare mass term in (75) alters the dynamics of particle production. The iso-inflaton fluctuations now obey the equation

$$\ddot{\chi} + 3H\dot{\chi} - \frac{\bar{\nabla}^2}{a^2}\chi + [\mu^2 + k_\star^4 t^2]\chi = 0 \quad (77)$$

rather than (27). This equation was solved in [80] in the regime where particle production is fast compared to the expansion time. (In the opposite regime, which corresponds to a fine-tuned coupling  $g^2 \ll 10^{-7}$ , the iso-inflaton will be light for a significant fraction of inflation. In that case the theory (5) must be considered as a multifield inflation model and one can no longer consistently assume  $\langle\chi\rangle = 0$ . In other words, relaxing the assumption of fast particle production significantly changes the scenario under consideration, and we do not pursue this possibility any further.) The occupation number of produced  $\chi$  particles is

$$n_k = r e^{-\pi k^2/k_\star^2} \quad (78)$$

which differs from our previous result (13) by the suppression factor

$$r \equiv e^{-\pi\mu^2/k_\star^2} \leq 1. \quad (79)$$

Therefore, the effect of the inclusion of a bare mass for the iso-inflaton is to suppress the number density of produced  $\chi$  particles by an amount  $r$ . This suppression reflects the reduced phase space of produced particles: the adiabaticity condition  $|\omega'_k/\omega_k^2| \ll 1$  is violated only for modes with  $k < \sqrt{k_\star^2 - \mu^2}$ .

The reduction of  $n_k$  translates into a suppression for the  $n$ -point correlation functions of the iso-inflaton. For example, the renormalized variance  $\langle\chi^2\rangle_{\text{ren}} \propto \int d^3k n_k$  is suppressed by a factor of  $r$ . The power spectrum of inflaton fluctuations generated by rescattering,  $P_\phi^{\text{resc}}$ , is proportional to the 4-point correlator of  $\chi$  and hence picks up a suppression factor of order  $r^2$ . Similarly, the bispectrum  $B_\phi$  is proportional to the 6-point correlator of  $\chi$  and must be reduced by a factor of order  $r^3$ . The condition

$$\mu^2 \ll k_\star^2 \quad (80)$$

is equivalent to  $r \cong 1$  and ensures that the addition of a bare iso-inflaton mass will have a negligible impact on any observable.

For models obtained from string theory or supergravity (SUGRA), it is natural to expect  $\mu$  of the order of the Hubble scale during inflation [116–118]. In the context of SUGRA, the finite energy density driving inflation breaks SUSY and induces soft scalar potentials with curvature of order  $V''_{\text{soft}} \sim \mu^2 \sim H^2$  [117]. In the case of string theory, many scalars are conformally coupled to gravity [118] through an interaction of the form  $\delta\mathcal{L} = -(1/12)R\chi^2$  where the Ricci scalar is  $R \sim H^2$  during inflation. More generally, any nonminimal coupling  $\delta\mathcal{L} = -(\xi/2)R\chi^2$  between gravity and the iso-inflaton will induce a contribution of order  $H$  to the effective mass of  $\chi$ , as long as  $\xi = \mathcal{O}(1)$ . In all models where  $\mu^2 \sim H^2$  condition (80) is satisfied for reasonable values of the coupling  $g^2 > 10^{-7}$ , see; (11). Thus, we expect that corrections of the form  $\delta\mathcal{L} = -(1/2)\mu^2\chi^2$  will not alter our results in a wide variety of well-motivated models.

## 5. Cosmological Perturbation Theory

In Section 4 we developed an analytical theory of particle production and IR cascading during inflation which is in very good agreement with nonlinear lattice field theory simulations. However, this formalism suffers from a neglect of metric perturbations, and, consequently, we were unable to rigorously discuss the gauge invariant curvature perturbation  $\zeta$ . (This variable is related to the quantity  $\mathcal{R}$ , defined in Section 3.3 as  $\zeta \cong -\mathcal{R}$  on large scales.) Hence, the reader may be concerned about gauge ambiguities in our results. In this section we address such concerns, showing that metric perturbations may be incorporated in a straightforward manner and that their consistent inclusion does not change our results in any significant way. We will do so by showing explicitly that, with appropriate choice of gauge, (49) and (27) for the fluctuations of the inflaton and iso-inflaton still hold, to first approximation. We will also go beyond our previous analysis by explicitly showing that in this same gauge the spectrum of the curvature fluctuations,  $P_\zeta$ , is trivially related to the spectrum of inflaton fluctuations,  $P_\phi$ .

To render the analysis tractable we would like to take full advantage of the results derived in the last section. To do so, we employ the Seery et al. formalism for working directly with the field equations [119] and make considerable use of results derived by Malik in [120, 121]. (Note that our notations differ somewhat from those employed by Malik. The reader is therefore urged to take care in comparing our formulae.)

We expand the inflaton and iso-inflaton fields up to second order in perturbation theory as

$$\begin{aligned}\phi(\tau, \mathbf{x}) &= \phi(\tau) + \delta_1\phi(\tau, \mathbf{x}) + \frac{1}{2}\delta_2\phi(\tau, \mathbf{x}), \\ \chi(\tau, \mathbf{x}) &= \delta_1\chi(\tau, \mathbf{x}) + \frac{1}{2}\delta_2\chi(\tau, \mathbf{x}).\end{aligned}\quad (81)$$

The perturbations are defined to average to zero  $\langle \delta_n\phi \rangle = \langle \delta_n\chi \rangle = 0$  so that  $\langle \phi(t, \mathbf{x}) \rangle = \phi(t)$  and  $\langle \chi(t, \mathbf{x}) \rangle = 0$ . (The condition  $\langle \chi \rangle = 0$  is ensured by the fact that  $m_\chi \gg H$  for nearly the entire duration of inflation.)

We employ the flat slicing and threading throughout this section. With this gauge choice the perturbed metric takes the form

$$\begin{aligned}g_{00} &= -a^2(1 + 2\psi_1 + \psi_2), \\ g_{0i} &= a^2\partial_i \left[ B_1 + \frac{1}{2}B_2 \right], \\ g_{ij} &= a^2\delta_{ij},\end{aligned}\quad (82)$$

so that spatial hypersurfaces are flat. Note also that in this gauge the field perturbations  $\delta_n\phi$ ,  $\delta_n\chi$  coincide with the Sasaki-Mukhanov variables [122] at both first and second order.

This perturbative approach, of course, neglects the momentary slowdown of the inflaton background due to backreaction. However, we have already shown in Section 3.3 that backreaction has a tiny impact on the observable cosmological perturbations (see also [12]).

**5.1. Gaussian Perturbations.** In [120] Malik has derived closed-form evolution equations for the field perturbations  $\delta_n\phi$ ,  $\delta_n\chi$  at both first order ( $n = 1$ ) and, second ( $n = 2$ ) order in perturbation theory. Let us first study the Gaussian perturbations. The closed-form Klein-Gordon equation for  $\delta_1\phi$  derived in [120] can be written as

$$\delta_1\phi'' + 2\mathcal{H}\delta_1\phi' - \bar{\nabla}^2\delta_1\phi + \left[ a^2m^2 - 3\left(\frac{\phi'}{M_p}\right)^2 \right] \delta_1\phi = 0. \quad (83)$$

Following our previous analysis we expand the first-order perturbation in terms of annihilation/creation operators as

$$\delta_1\phi(t, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^{3/2}} \left[ b_{\mathbf{k}} \frac{\delta_1\phi_{\mathbf{k}}(\tau)}{a(\tau)} e^{i\mathbf{k}\cdot\mathbf{x}} + \text{h.c.} \right], \quad (84)$$

where h.c. denotes the Hermitian conjugate of the preceding term, and we draw the attention of the reader to the

the explicit factor of  $a^{-1}$  in our definition of the Fourier transform. Working to leading order in slow-roll parameters we have

$$\delta_1\phi_k'' + \left[ k^2 + \frac{1}{\tau^2}(-2 + 3\eta - 9\epsilon) \right] \delta_1\phi_k = 0. \quad (85)$$

This equation coincides exactly with (55), and the properly normalized solutions again take the form of (57). The only difference is that the order of the Hankel function,  $\nu$ , is now given by

$$\nu \cong \frac{3}{2} - \eta + 3\epsilon \quad (86)$$

rather than by (56). The power spectrum of the Gaussian fluctuations is, again, given by (58). The correction to the order of the Hankel function  $\nu$  translates into a correction to the spectral index: instead of (59) we now have

$$n_s - 1 = 2\eta - 6\epsilon \quad (87)$$

which is precisely the standard result [123].

Thus, as far as the quantum vacuum fluctuations of the inflaton are concerned, the only impact of consistently including metric perturbations is an  $\mathcal{O}(\epsilon)$  correction to the spectral index  $n_s$ .

Let us now turn our attention to the first-order fluctuations of the iso-inflaton. The closed-form Klein-Gordon equation for  $\delta_1\chi$  derived in [120] can be written as

$$\delta_1\chi'' + 2\mathcal{H}\delta_1\chi' - \bar{\nabla}^2\delta_1\chi + a^2k_*^4 t^2(\tau)\delta_1\phi = 0. \quad (88)$$

This coincides *exactly* with (27), which we have already solved. The fact that linear perturbations of  $\chi$  do not couple to the metric fluctuations follows from the condition  $\langle \chi \rangle = 0$ .

**5.2. Non-Gaussian Perturbations.** Now let us consider the second-order perturbation equations. The closed-form Klein-Gordon equation for  $\delta_2\phi$  derived in [120] can be written as

$$\begin{aligned}\delta_2\phi'' + 2\mathcal{H}\delta_2\phi' - \bar{\nabla}^2\delta_2\phi + \left[ a^2m^2 - 3\left(\frac{\phi'}{M_p}\right)^2 \right] \delta_2\phi \\ = J(\tau, \mathbf{x}).\end{aligned}\quad (89)$$

As usual, the left-hand side is identical to the first-order equation (83) while the source term  $J$  is constructed from a bi-linear combination of the first order quantities  $\delta_1\phi$  and  $\delta_1\chi$ . In order to solve (89) we require explicit expressions for the Green function  $G_k$  and the source term  $J$ . The Green function is trivial for the case at hand; it is still given by our previous result (56), provided that one takes into account the fact that the order of the Hankel functions  $\nu$  is now given by (86) rather than (56). In other words, the Green function for the non-Gaussian perturbations (89) differs from the result obtained neglecting metric perturbations only by  $\mathcal{O}(\epsilon)$  corrections.

Next, we would like to consider the source term,  $J$ , appearing in (89). Schematically, we can split the source into contributions bi-linear in the Gaussian inflaton fluctuation  $\delta_1\phi$  and contributions bi-linear in the iso-inflaton  $\delta_1\chi$ :

$$J = J_\phi + J_\chi. \quad (90)$$

The contribution  $J_\phi$  would be present even in the absence of the iso-inflaton. These correspond, physically, to the usual nongaussian corrections to the inflaton vacuum fluctuations coming from self-interactions. This contribution to the source is well studied in the literature and is known to contribute negligibly to the bispectrum [119]. Thus, in what follows, we will ignore  $J_\phi$ .

On the other hand, the contribution  $J_\chi$  appearing in (90) depends only on the iso-inflaton fluctuations  $\delta_1\chi$ . This contribution can be understood, physically, as generating nongaussian inflaton fluctuations  $\delta_2\phi$  by rescattering of the produced  $\chi$  particles off the condensate. Hence, the contribution  $J_\chi$  may source large nongaussianity and is most interesting for us. It is straightforward to compute  $J_\chi$  explicitly for our model using the general results of [120]. We find

$$\begin{aligned} J_\chi = & -2a^2g^2(\phi - \phi_0)(\delta_1\chi)^2 \\ & \pm \frac{\sqrt{2\epsilon}}{M_p} \left[ -a^2g^2(\phi - \phi_0)^2(\delta_1\chi)^2 - \frac{1}{2}(\bar{\nabla}\delta_1\chi)^2 \right. \\ & + \nabla^{-2} \left( \partial_i(\delta_1\chi)\bar{\nabla}^2\partial^i(\delta_1\chi) + \bar{\nabla}^2(\delta_1\chi)\bar{\nabla}^2(\delta_1\chi) \right. \\ & \left. \left. + \delta_1\chi'\bar{\nabla}^2\delta_1\chi + (\bar{\nabla}\delta_1\chi')^2 \right) \right], \end{aligned} \quad (91)$$

where the upper sign is for  $\phi' > 0$  and the lower sign is for  $\phi' < 0$ . Notice that the contributions to  $J_\chi$  on the third and fourth line of (91) contain the inverse spatial Laplacian  $\nabla^{-2}$  and are thus nonlocal. These terms all contain at least as many gradients as inverse gradients, and hence the large-scale limit is well defined. In [124] it was argued that these terms nearly always contribute negligibly to the curvature perturbation on large scales.

Let us now examine the structure of the iso-inflaton source  $J_\chi$ , (91). The first line of (91) goes like  $a^2g^2(\phi - \phi_0)(\delta_1\chi)^2$ . This coincides exactly with the source term in (49) which was already studied in Section 4. On the other hand, the terms on the second, third and fourth lines of (91) are new. These represent corrections to IR cascading which arise due to the consistent inclusion of metric perturbations. We will now argue that these ‘‘extra’’ terms are negligible as compared to the first line. If we denote the energy density in gaussian iso-inflaton fluctuations as  $\rho_\chi \sim m_\chi^2(\delta_1\chi)^2$  then, by inspection, we see that the first line of (91) is parametrically of order  $\rho_\chi/|\phi - \phi_0|$  while the remaining terms are of order  $\sqrt{\epsilon}\rho_\chi/M_p$ . Hence, we expect the first term to dominate for the field values  $\phi \cong \phi_0$  which are relevant for IR cascading. This suggests that the dominant contribution to  $J_\chi$  is the term which we have already taken into account in Section 4.

Let us now make this argument more quantitative. Inspection reveals that the only ‘‘new’’ contribution to (91) which has any chance of competing with the ‘‘old’’ term  $a^2g^2(\phi - \phi_0)(\delta_1\chi)^2$  is the one proportional to  $\sqrt{\epsilon}a^2g^2(\phi - \phi_0)^2(\delta_1\chi)^2/M_p$  (the first term on the second line). This new correction has the possibility of becoming significant because it grows after particle production, as  $\phi$  rolls away from  $\phi_0$ . This growth, which reflects the fact that the energy density in the  $\chi$  particles increases as they become more massive, cannot persist indefinitely. Within a few  $e$ -foldings of particle production the iso-inflaton source term must behave as  $J_\chi \sim a^{-3}$ , corresponding to the volume dilution of non-relativistic particles. Hence, in order to justify the analysis of Section 4 we must check that the term

$$J_{\text{new}} \sim \frac{\sqrt{\epsilon}}{M_p} a^2g^2(\phi - \phi_0)^2(\delta_1\chi)^2 \quad (92)$$

does not dominate over the term which we have already considered

$$J_{\text{old}} \sim a^2g^2(\phi - \phi_0)(\delta_1\chi)^2 \quad (93)$$

during the relevant time  $H\Delta t = \mathcal{O}(1)$  after particle production. It is straightforward to show that

$$\frac{J_{\text{old}}}{J_{\text{new}}} \sim \frac{M_p}{\sqrt{\epsilon}} \frac{1}{\phi - \phi_0} \sim \frac{M_p H}{\dot{\phi}\sqrt{\epsilon}} \frac{1}{N} \sim \frac{1}{\epsilon} \frac{1}{N}, \quad (94)$$

where  $N = Ht$  is the number of  $e$ -foldings elapsed from particle production to the time when IR cascading has completed. Hence,  $N = \mathcal{O}(1)$ , and we conclude that the second, third, and fourth lines of (91) are (at least) slow-roll suppressed as compared to the first line.

In summary, we have shown that consistent inclusion of metric perturbations yields corrections to the inflaton fluctuations  $\delta\phi$  which fall into two classes.

- (1) Slow-roll suppressed corrections to the inflaton vacuum fluctuations  $\delta_1\phi$  (these amount to changing the definition of  $\nu$  in solution (57)). These corrections have two physical effects. First, they yield an  $\mathcal{O}(\epsilon)$  correction to the spectral index. Second, they modify the propagator  $G_k$  by an  $\mathcal{O}(\epsilon)$  correction.
- (2) Corrections to the source  $J$  for the nongaussian inflaton perturbation  $\delta_2\phi$ . These corrections are the second, third, and fourth lines of (91) which, as we have seen, are slow-roll suppressed.

It should be clear that neither of these corrections alters our previous analysis in any significant way.

**5.3. Correlators.** So far, we have shown that a consistent inclusion of metric perturbations does not significantly alter our previous results for the field perturbations. Specifically,  $\delta_1\chi$  is identical to our previous solution of (27) for the iso-inflaton while  $\delta_1\phi$  coincides with the homogeneous solution of (49), up to slow-roll corrections. At second order in perturbation theory, we have seen that

$$\delta_2\phi = \int d^4x' G(x - x') J_\chi(x') + \mathcal{O}[(\delta_1\phi)^2]. \quad (95)$$

To a leading order in slow roll,  $J_\chi \cong -2a^2g^2(\phi - \phi_0)(\delta_1\chi)^2$ , and the first term coincides with our previous result for the particular solution of (49). The terms of order  $(\delta_1\phi)^2$  represent nongaussian corrections to the vacuum fluctuations from inflation (coming from self-interactions of  $\delta\phi$  and the nonlinearity of gravity). These would be present even in the absence of particle production and are known to have a negligible impact on the spectrum and bispectrum [119].

We are ultimately interested in the connected  $n$ -point correlation functions of  $\delta\phi$ . For example, the 2-point function  $\langle(\delta\phi)^2\rangle$  gets a contribution of the form  $\langle(\delta_1\phi)^2\rangle$  which gives the usual nearly scale-invariant large-scale power spectrum from inflation. The cross-term  $\langle\delta_1\phi\delta_2\phi\rangle$  is of order  $\langle(\delta_1\phi)^4\rangle$  and represents a negligible ‘‘loop’’ correction to the scale-invariant spectrum from inflation. (The cross-term does not involve the iso-inflaton since  $\delta_1\phi$  and  $\delta_1\chi$  are statistically independent.) Finally, there is a contribution  $\langle(\delta_2\phi)^2\rangle$  which involves terms of order  $\langle\chi^4\rangle$  coming from rescattering and terms of order  $\langle(\delta_1\phi)^4\rangle$  which represent (more) loop corrections to the scale-invariant spectrum from inflation. Thus, we can schematically write

$$P_\phi(k) = P_\phi^{\text{vac}}(k)[1 + (\text{loops})] + P_\phi^{\text{resc}}(k). \quad (96)$$

Here  $P_\phi^{\text{vac}} \sim k^{n_s-1}$  is the usual nearly scale-invariant spectrum from inflation and  $P_\phi^{\text{resc}}$  is the bump-like contribution from rescattering and IR cascading which we have studied in the previous section. The ‘‘loop’’ corrections to  $P_k^{\text{vac}}(k)$  have been studied in detail in the literature (see, e.g., [125, 126]) and are known to be negligible in most models.

We can also make a similar schematic decomposition of the bispectrum by considering the structure of the 3-point correlator  $\langle(\delta\phi)^3\rangle$ . Following our previous line of reasoning, it is clear that the dominant contribution comes from rescattering and is of order  $\langle\chi^6\rangle$ . The terms involving  $\langle(\delta_1\phi)^3\rangle$ , on the other hand, represent the usual nongaussianity generated during single field slow-roll inflation and are known to be small [119].

**5.4. The Curvature Perturbation.** Ultimately one wishes to compute not the field perturbations,  $\delta_n\phi$ ,  $\delta_n\chi$ , but rather the gauge invariant curvature fluctuation,  $\zeta$ . We expand this in perturbation theory in the usual manner:

$$\zeta = \zeta_1 + \frac{1}{2}\zeta_2. \quad (97)$$

In [121] Malik has derived expressions for the large scale curvature perturbation in terms of the Sasaki-Mukhanov variables at both first order and second order in perturbation theory. We remind the reader that in the flat slicing (which we employ) the Sasaki-Mukhanov variable for each field simply coincides with the field perturbation (*i.e.*,  $Q_\phi = \delta\phi$  and  $Q_\chi = \delta\chi$ ).

At first order in perturbation theory the iso-inflaton does not contribute to the curvature perturbation (since  $\langle\chi\rangle = 0$ ), and we have

$$\zeta_1 = -\frac{\mathcal{H}}{\phi'}\delta_1\phi. \quad (98)$$

At second order in perturbation theory the expression for the curvature perturbation is more involved. Using the results of [121] and working to leading order in slow-roll parameters we find

$$\begin{aligned} \zeta_2 \cong & -\frac{\mathcal{H}}{\phi'}\left[\delta_2\phi - \frac{\delta_2\phi'}{3\mathcal{H}}\right] \\ & + \frac{1}{3(\phi')^2}\left[(\delta_1\chi)^2 + a^2g^2v^2t^2(\tau)(\delta_1\chi)^2\right] \\ & + \frac{1}{3(\phi')^2}\left[(\delta_1\phi')^2 + a^2m^2(\delta_1\phi)^2\right] + 2\left(\frac{\mathcal{H}\delta_1\phi}{\phi'}\right)^2. \end{aligned} \quad (99)$$

Let us discuss the various contributions to this equation. The third line contributes to the nongaussianity of the vacuum fluctuations during inflation. (The term  $2\zeta_1^2$  on the last line (which would seem to contribute  $\mathcal{O}(1)$  to the nonlinearity parameter  $f_{NL}$ ) stems from using the Malik and Wands [127] definition of  $\zeta_2$ . It can be related to the definition of Lyth and Rodríguez [128] (which also agrees with Maldacena) as  $\zeta_2 = \zeta_2^{\text{LR}} + 2\zeta_1^2$ . Thus, in the absence of particle production (99) would predict  $f_{NL} \sim \mathcal{O}(\epsilon, \eta)$ , in agreement with [16–18]. See also [19, 20] for a related discussion.) These terms are known to be negligible [16–18, 119].

Next, we consider the second line of (99). This represents the direct contribution of the gaussian fluctuations  $\delta_1\chi$  to the curvature perturbation. This contribution is tiny since the  $\chi$  particles are extremely massive for nearly the entire duration of inflation, and hence  $\delta_1\chi \sim a^{-3/2}$  (see also [54] for a related discussion). The smallness of this contribution to  $\zeta$  can be understood physically by noting that the super-horizon iso-curvature fluctuations in our model are negligible.

Finally, let us consider the contribution on the first line of (99). This contribution is the most interesting. To make contact with observations we must compute the curvature perturbation at late times and on large scales. In Section 4 we have already shown that  $\delta_n\phi$  is constant on large scales and at late times for both  $n = 1$  and  $n = 2$ . This is the expected result: the curvature fluctuations are frozen far outside the horizon and in the absence of entropy perturbations [129]. (Note that, in some cases, the curvature fluctuations may evolve significantly after horizon exit [130, 131]. This is a concern in models where there are significant violations of slow roll. In Section 3.3 we have already shown that the transient violation of slow roll has a negligible effect on the curvature fluctuations in our model. Hence, the result  $\zeta_n \sim \delta_n\phi \sim \text{const}$  far outside the horizon is consistent with previous studies.) Hence  $\delta_2\phi'$  is completely negligible and the first term on the first line of (99) must dominate over the second term. We conclude that, at late times and on large scales, the second order curvature perturbation is very well approximated by

$$\zeta_2 \cong -\frac{\mathcal{H}}{\phi'}\delta_2\phi + \dots \quad (100)$$

In summary, we have shown that the power spectrum of curvature fluctuations from inflation in model (5) is trivially related to the power spectrum of inflaton fluctuations

$$P_\zeta(k) \cong \frac{H^2}{\dot{\phi}^2} P_\phi(k) = \frac{1}{2\epsilon M_p^2} P_\phi(k) \quad (101)$$

at both first order and second order in cosmological perturbation theory. This relation is valid at late times and for scales far outside the horizon. Curvature spectrum (101) may be written as

$$P_\zeta(k) = P_\zeta^{\text{vac}}(k)[1 + (\text{loops})] + P_\zeta^{\text{resc}}(k). \quad (102)$$

The power spectrum of the inflaton vacuum fluctuations agrees with the usual result obtained in linear theory [123]

$$P_\zeta^{\text{vac}}(k) \cong \frac{H^2}{8\pi^2 \epsilon M_p^2} \left(\frac{k}{aH}\right)^{2\eta-6\epsilon}. \quad (103)$$

In (102) we have schematically labeled the corrections arising from the second line of (99) and the source  $J_\phi$  as ‘‘loop’’. These are nongaussian corrections to the inflaton vacuum fluctuations arising from self-interactions of the inflaton and also the nonlinearity of gravity. Such corrections are negligible. The most interesting contribution to the power spectrum (102) is due to rescattering,  $P_\zeta^{\text{resc}}(k)$ . This quantity is proportional to our previous result (71).

In passing, notice that the bispectrum  $B_\phi$  (defined by (73)) of inflaton fluctuations will differ from the bispectrum  $B$  of the curvature fluctuations (defined by (3)) only by a simple rescaling:

$$B(k_i) \cong -\left(\frac{H}{\dot{\phi}}\right)^3 B_\phi(k_i) = -\frac{1}{(2\epsilon)^{3/2} M_p^3} B_\phi(k_i). \quad (104)$$

The dominant contribution to  $B_\phi$  comes from rescattering effects  $\langle \delta_2 \phi^3 \rangle \sim \langle \delta_1 \chi^6 \rangle$ .

The analysis of this section justifies our neglect of metric fluctuations in Section 4.

## 6. Observational Constraints on Particle Production during Inflation

In the previous sections of this paper, we have demonstrated that particle production and IR cascading in model (4) leads to a bump-like feature in the primordial power spectrum. We now aim to determine the extent to which such a spectral distortion is compatible with current cosmological data. The results of this section have first appeared in [13].

One motivation for this study is to determine whether features generated by particle production during inflation can explain some of the anomalies in the observed primordial power spectrum,  $P(k)$ . A number of different studies have hinted at the possible presence of some localized features in the power spectrum [62, 73, 98–110], which are not compatible with the simplest power-law  $P(k) \sim k^{n_s-1}$  model. Although such glitches may simply be statistical anomalies [111], there is also the tantalizing possibility that they represent a signature of primordial physics beyond the simplest

slow-roll inflation scenario. Forthcoming polarization data may play a crucial role in distinguishing between these possibilities [73]. However, in the meantime, it is interesting to determine the extent to which such features may be explained by microscopically realistic inflation models, such as (5).

To answer this question we provide a simple semianalytic fitting function that accurately captures the shape of the feature generated by particle production and IR cascading during inflation. Next, we confront this modified power spectrum with a variety of observational data sets. We find no evidence for a detection; however, we note that observations are consistent with relatively large spectral distortions of the type predicted by model (4). If the feature is located on scales relevant for Cosmic Microwave Background (CMB) experiments, then its amplitude may be as large as  $\mathcal{O}(10\%)$  of the usual scale-invariant fluctuations, corresponding to  $g^2 \sim 0.01$ . Our results translate into a  $\phi_0$ -dependent bound on the coupling  $g^2$  which is crucial in order to determine whether the nongaussian signal associated with particle production and IR cascading is detectable in future missions.

We also consider the more complicated features which result from multiple bursts of particle production and IR cascading. Such features are a prediction of a number of string theory inflation models, including brane/axion monodromy [77–79]. For appropriate choice of the spacing between the features, we find that the constraint on  $g^2$  in this scenario is even weaker than the single-bump case.

*6.1. A Simple Parametrization of the Power Spectrum.* In [12] it was shown that particle production and IR cascading during inflation in model (4) generates a bump-like contribution to the primordial power spectrum. As shown in Figure 9, this feature can be fit with a very simple function  $P_{\text{bump}} \sim k^3 e^{-\pi k^2/(2k_*^2)}$ . The bump-like contribution from IR cascading is complimentary to the usual (nearly) scale-invariant contribution to the primordial power spectrum  $P_{\text{vac}} \sim k^{n_s-1}$  coming from the quantum vacuum fluctuations of the inflaton. The total, observable, power spectrum in model (4) is simply the superposition of these two contributions:  $P(k) \sim k^{n_s-1} + k^3 e^{-\pi k^2/(2k_*^2)}$  (see (64)). This simple formula can be motivated from analytical considerations [12] and provides a good fit to lattice field theory simulations near the peak of the feature and also in the IR tail. (This fitting formula does *not* capture the small oscillatory structure in the UV tail of the feature (see Figure 9) which does not concern us since that region is not phenomenologically interesting.)

It is straightforward to generalize this discussion to allow for multiple bursts of particle production during inflation. Suppose there are multiple points  $\phi = \phi_i$  ( $i = 1, \dots, n$ ) along the inflationary trajectory where new degrees of freedom  $\chi_i$  become massless:

$$\mathcal{L}_{\text{int}} = -\sum_{i=0}^n \frac{g_i^2}{2} (\phi - \phi_i) \chi_i^2. \quad (105)$$

For each instant  $t_i$  when  $\phi = \phi_i$  there will be an associated burst of particle production and subsequent rescattering of

the produced massive  $\chi_i$  off the condensate  $\phi(t)$ . Each of these events proceeds as described above and leads to a new bump-like contribution to the power spectrum. These features simply superpose owing to that fact that each field  $\chi_i$  is statistically independent (so that the cross-terms involving  $\chi_i\chi_j$  with  $i \neq j$  in the computation of the two-point function must vanish). Thus, we arrive at the following parametrization of the primordial power spectrum in models with particle production during inflation:

$$P(k) = A_s \left( \frac{k}{k_0} \right)^{n_s-1} + \sum_{i=1}^n A_i \left( \frac{\pi e}{3} \right)^{3/2} \left( \frac{k}{k_i} \right)^3 e^{-(\pi/2)(k/k_i)^2}, \quad (106)$$

where  $A_s$  is the amplitude of the usual nearly scale-invariant vacuum fluctuations from inflation, and  $k_0$  is the pivot, which we choose to be  $k_0 = 0.002 \text{ Mpc}^{-1}$  following [132]. The constants  $A_i$  depend on the couplings  $g_i^2$  and measure of the size of the features from particle production. We have normalized these amplitudes so that the power in the  $i$ th bump, measured at the peak of the feature, is given by  $A_i$ . The location of each feature,  $k_i$ , is related to the number of  $e$ -foldings  $N$  from the end of inflation to the time when the  $i$ th burst of particle production occurs: roughly  $\ln(k_i/H) \sim N_i$  where  $N = N_i$  at the moment when  $\phi = \phi_i$ . From a purely phenomenological perspective the locations  $k_i$  are completely arbitrary.

We compare (106) to lattice field theory simulations in order to determine the amplitude  $A_i$  in terms of model parameters. We find

$$A_i \cong 10^{-6} g_i^{15/4}. \quad (107)$$

The power-law form of this relation was determined by inspection of numerical results; however, it can also be motivated by analytical considerations. Assuming standard chaotic inflation  $V = m^2\phi^2/2$ , we have tested this formula for  $g^2 = 1, 0.1, 0.01$ , taking both  $\phi_0 = 2\sqrt{8\pi}M_p$  and  $\phi_0 = 3.2\sqrt{8\pi}M_p$ . We found agreement up to factors order unity in all cases.

Theoretical consistency of our calculation of the shape of the feature bounds the coupling as  $10^{-7} \lesssim g_i^2 \lesssim 1$  [12]. Hence, power spectrum (106) can be obtained from sensible microphysics only when  $10^{-20} \lesssim A_i \lesssim 10^{-6}$ . This constraint still allows for a huge range of observational possibilities: near the upper bound the feature is considerably larger than the vacuum fluctuations while near the lower bound the feature is completely undetectable.

Note that for each bump in (106) the IR tail  $P_{\text{bump}} \rightarrow k^3$  as  $k \rightarrow 0$  is similar to the feature considered by Hoi et al. in [108], consistent with causality arguments about the generation of curvature perturbations by local physics.

**6.2. Data Sets and Analysis.** The primordial power spectrum for our model is parametrized as (106). Our aim is to derive observational constraints on the various model parameters  $A_s$ ,  $n_s$ ,  $k_i$ , and  $A_i$  using CMB, galaxy power spectrum,

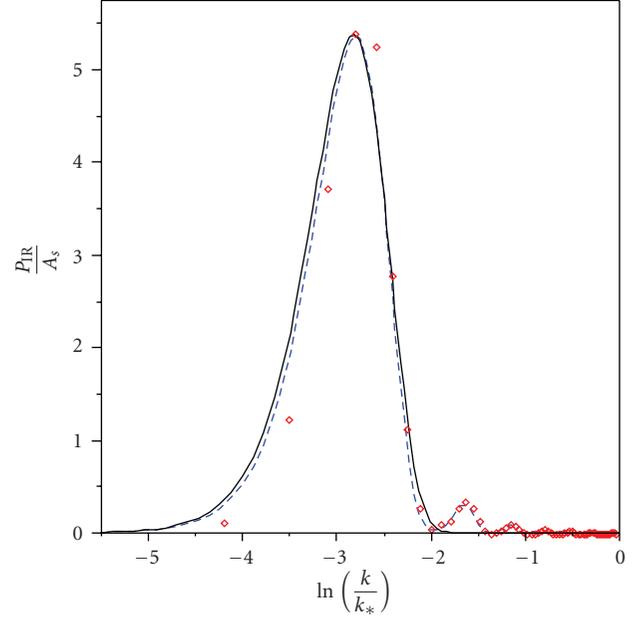


FIGURE 9: The bump-like features generated by IR cascading. We plot the feature power spectrum  $P_{\text{bump}}(k)$  obtained from fully nonlinear lattice field theory simulations (the red points) and also the result of an analytical calculation (the dashed blue curve) using the formalism described in [14]. We also superpose the fitting function  $\sim k^3 e^{-\pi k^2/(2k_*^2)}$  (the solid black curve) to illustrate the accuracy of this simple formula.

and weak lensing data. To this end we use the cosmoMC package [133] to run Markov Chain Monte Carlo (MCMC) calculations to determine the likelihood of the cosmological parameters, including our new parameters  $A_i$  and  $k_i$ . We employ the following data sets.

**Cosmic Microwave Background (CMB).** Our complete CMB data sets include WMAP-5yr [132, 134], BOOMERANG [135–137], ACBAR [138–141], CBI [142–145], VSA [146], DASI [147, 148], and MAXIMA [149]. We have included the Sunyaev-Zeldovic (SZ) secondary anisotropy [150, 151] for WMAP-5yr, ACBAR, and CBI data sets. The SZ template is obtained from hydrodynamical simulation [152]. Also included for theoretical calculation of CMB power spectra is the CMB lensing contribution.

**Type Ia Supernova (SN).** We employ the Union Supernova Ia data (307 SN Ia samples) from The Supernova Cosmology Project [153].

**Large Scale Structure (LSS).** The 2dF Galaxy Redshift Survey (2dFGRS) data [154] and Sloan Digital Sky Survey (SDSS) Luminous Red Galaxy (LRG) data release 4 [155] are utilized.

Note that we have used the likelihood code based on the non-linear modeling by Tegmark et al. [155] (marginalizing the bias  $b$  and the  $Q$  parameter). However with a large bump in the linear power spectrum, this naive treatment may be not sufficient to characterize the non-linear response to the

TABLE 1: Weak lensing data sets.

Data sets	$A_{\text{eff}}$ (deg <sup>2</sup> )	$n_{\text{eff}}$ (arcmin <sup>-2</sup> )
COSMOS [156, 157]	1.6	40
CFHTLS-wide [158, 159]	22	12
GaBODS [160, 161]	13	12.5
RCS [160, 161]	53	8
VIRMOS-DESCART [159, 162]	8.5	15

feature on small scales. Ideally, this should be obtained from N-body simulations; however, such a study is beyond the scope of the current work.

There are several other caveats on our results in the high- $k$  regime. First, we assume linear bias for the galaxies, which may not be entirely safe at sufficiently small scales. Moreover, sharp features in the matter power spectrum can cause sharp features in the bias as a function of  $k$ .

Keeping in mind these caveats our constraints on small scales  $k \gtrsim 0.1 \text{ Mpc}^{-1}$  should be taken with a grain of salt and considered as accurate only up to factors order unity.

*Weak Lensing (WL).* Five WL data sets are used in this paper. The effective survey area  $A_{\text{eff}}$  and galaxy number density  $n_{\text{eff}}$  of each survey are listed in Table 1.

For COSMOS data we use the CosmoMC plug-in written by Lesgourgues [157], modified to do numerical marginalization on three nuisance parameters in the original code.

For the other four weak lensing data sets we use the likelihood given in [163]. To calculate the likelihood we have written a CosmoMC plug-in code, with simplified marginalization on the parameters of galaxy number density function  $n(z)$ . More details about this plug-in can be found in [164].

As for the LSS data, for small scales  $k \gtrsim 0.1 \text{ Mpc}^{-1}$  there is the caveat that the nonlinear evolution of the power spectrum in the presence of bump-like distortions may not be treated accurately.

**6.3. Observational Constraints: A Single Burst of Particle Production.** We now present our results for the observational constraints on particle production during inflation, assuming two different scenarios.

The minimal scenario to consider is a single burst of particle production during inflation, which corresponds to taking  $n = 1$  in (105). The power spectrum is given by (106) with  $n = 1$ , and, with some abuse of notation, we denote  $k_1 \equiv k_{\text{IR}}$  and  $A_1 \equiv A_{\text{IR}}$ . The prior we have used for  $A_{\text{IR}}$  is  $0 \leq A_{\text{IR}} \leq 25 \times 10^{-10}$  and for  $k_{\text{IR}}$  is  $-9.5 \leq \ln[k/\text{Mpc}^{-1}] \leq 1$ . The former condition ensures that the bump-like feature from IR cascading does not dominate over the observed scale-invariant fluctuations while the latter is necessary in order to have the feature in the observable range of scales. In Figure 10 we plot the marginalized posterior likelihood for the new parameters  $A_{\text{IR}}$  and  $k_{\text{IR}}$  describing the bump while in Table 2 we give

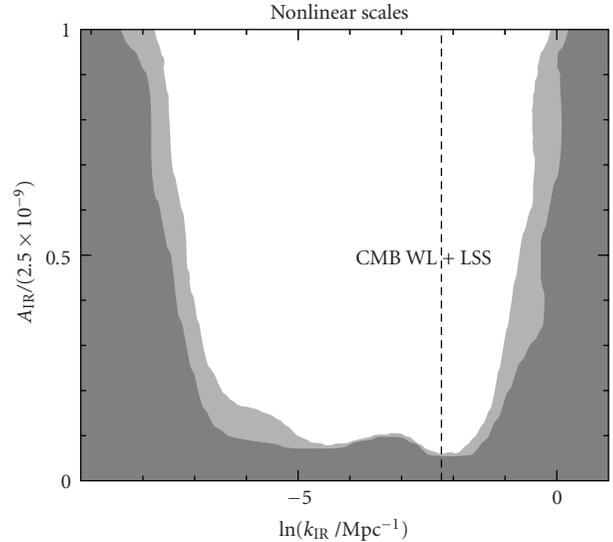


FIGURE 10: Marginalized posterior likelihood contours for the parameters  $A_{\text{IR}}$  and  $k_{\text{IR}}$  (the magnitude and position of the feature, resp.) in the single-bump model. Black and grey regions correspond to parameter values allowed at 95.4% and 99.7% confidence levels, respectively. At small scales, to the right of the dashed vertical line, our results should be taken with a grain of salt since the nonlinear evolution of the power spectrum may not be modeled correctly in the presence of bump-like distortions.

TABLE 2: Constraints on the standard “vanilla” cosmological parameters for the single-bump model. All errors are 95.4% confidence level.

$\Omega_b h^2$	$0.0227^{+0.0010}_{-0.0010}$
$\Omega_c h^2$	$0.112^{+0.0050}_{-0.0044}$
$\theta$	$1.0424^{+0.0042}_{-0.0043}$
$\tau$	$0.08^{+0.03}_{-0.03}$
$n_s$	$0.956^{+0.024}_{-0.024}$
$\ln[10^{10} A_s]$	$3.206^{+0.074}_{-0.068}$
$A_{\text{SZ}}$	$1.62^{+0.76}_{-0.74}$
$\Omega_m$	$0.264^{+0.026}_{-0.022}$
$\sigma_8$	$0.807^{+0.034}_{-0.030}$
$z_{\text{re}}$	$10.5^{+2.5}_{-2.7}$
$H_0$	$71.6^{+2.3}_{-2.4}$

the best fit values for the remaining (vanilla) cosmological parameters.

For very large scales  $\lesssim \text{Gpc}^{-1}$ , the data do not contain much information (due to cosmic variance), and hence the constraint on any modification of the power spectrum is weak. In this region the spectral distortion may be larger than 100% of the usual scale-invariant fluctuations and couplings  $g^2$  of order unity are allowed. For smaller scales  $k \gtrsim \text{Gpc}^{-1}$  the constraints are stronger, and we have, very roughly,  $A_{\text{IR}}/A_s \lesssim 0.1$  corresponding to  $g^2 \lesssim 0.01$ . For very small scales,  $k \gtrsim 0.1 \text{ Mpc}^{-1}$ , our constraints should be taken with a grain of salt since the nonlinear evolution of the power spectrum may not be modeled correctly in the presence of

bump-like distortions. At small scales nonlinear effects tend to wipe out features of this type (see, e.g., [165]), and hence observational constraints for  $k \gtrsim 0.1 \text{ Mpc}^{-1}$  may be weaker than what is presented in Figure 10. Note that in most of this nonlinear regime we find essentially no constraint on  $A_{\text{IR}}$ , which is consistent with what would be expected in a more comprehensive treatment.

The IR cascading bump in the primordial power spectrum will be accompanied by a corresponding nongaussian feature in the bispectrum, that will be discussed in more detail in the next section. From the perspective of potentially observing this signal it is most interesting if this feature is located on scales probed by CMB experiments. (There is also the fascinating possibility that the nongaussianity from IR cascading could show up in the large-scale structure as in [27, 166–168]. We leave a detailed discussion to future studies.) To get some intuition into what kinds of features in the CMB scales are still allowed by the data we focus on an example with  $A_{\text{IR}} = 2.5 \times 10^{-10}$  which, using (107), corresponds to a reasonable coupling value  $g^2 \sim 0.01$ . We take the bump to be located at  $k_{\text{IR}} = 0.01 \text{ Mpc}^{-1}$  and fix the remaining model parameters to  $A_s = 2.44 \times 10^{-9}$ ,  $n_s = 0.97$  (which are compatible with the usual values). This sample bump in the power spectrum is illustrated in Figure 3(a) and is consistent with the data at  $2\sigma$ . In Figure 3(b) we plot the associated angular CMB TT spectrum. This example represents a surprisingly large spectral distortion: the total power in the feature as compared to the scale invariant vacuum fluctuations is  $P_{\text{bump}}/P_{\text{vac}} \sim 0.1$ , evaluated at the peak of the bump. We discuss the nongaussianity associated with this feature.

**6.4. Observational Constraints: Multiple Bursts of Particle Production.** Next, we consider a slightly more complicated scenario: multiple bursts of particle production leading many localized features in the power spectrum. For simplicity we assume that all bumps have the same magnitude  $A_i \equiv A_{\text{IR}}$ , and we further suppose a fixed number of  $e$ -foldings  $\delta N$  between each burst of particle production. This implies that the location of the  $i$ th bump will be given by  $k_i = e^{(i-1)\Delta} k_1$  where  $\Delta$  is a model parameter controlling the density of features. We take the number of bursts,  $n$ , to be sufficiently large that the series of features extends over the whole observable range. In the next section we will see that these assumptions are not restrictive and that many well-motivated models are consistent with this simple setup.

Our multibump model, then, has three parameters:  $A_{\text{IR}}$ ,  $k_1$ , and  $\Delta$ . We take the prior on the amplitude to be  $A_{\text{IR}} \leq 25 \times 10^{-10}$  as in Section 6.3. If the features are very widely spaced,  $\Delta \gtrsim 1$ , then the constraint on each bump will obviously be identical to the results for the single-bump case presented in Section 6.3. Hence the most interesting case to consider is  $\Delta \lesssim 1$  so that the bumps are partially overlapping. Our prior for the density of features is therefore  $0 \leq \Delta \leq 1$ . Finally, the location of the first bump will be a historical accident in realistic models, hence we marginalize over all possible values of  $k_1$  and present our constraints and 2d likelihood plots in the space of  $A_{\text{IR}}$  and  $\Delta$ .

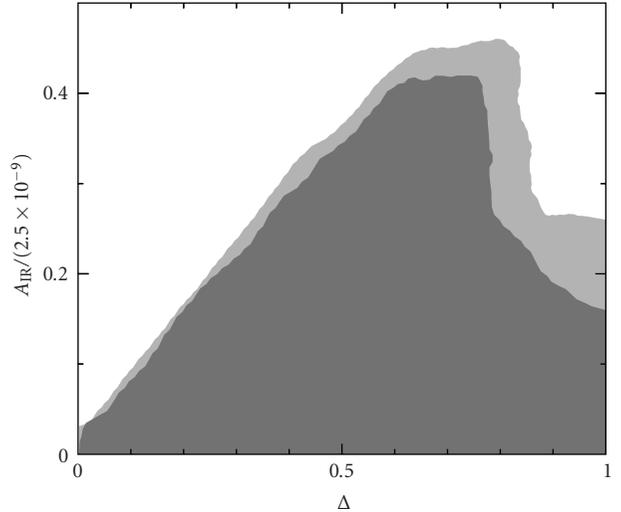


FIGURE 11: Marginalized posterior likelihood contours for the parameters  $A_{\text{IR}}$  and  $\Delta$  (the feature amplitude and spacing, resp.) of the multiple-bump model. Black and grey regions correspond to values allowed at 95.4% and 99.7% confidence levels, respectively.

TABLE 3: Constraints on the standard “vanilla” cosmological parameters for the multiple-bump model. All error bars are 95.4% confidence level.

$\Omega_b h^2$	$0.0227^{+0.0009}_{-0.0009}$
$\Omega_c h^2$	$0.1126^{+0.0049}_{-0.0044}$
$\theta$	$1.0424^{+0.0039}_{-0.0043}$
$\tau$	$0.078^{+0.031}_{-0.026}$
$n_s$	$0.93^{+0.04}_{-0.17}$
$\ln[10^{10} A_s]$	$2.8^{+0.4}_{-0.9}$
$A_{SZ}$	$1.60^{+0.77}_{-0.76}$
$\Omega_m$	$0.265^{+0.026}_{-0.021}$
$\sigma_8$	$0.807^{+0.034}_{-0.030}$
$z_{re}$	$10.3^{+2.6}_{-2.5}$
$H_0$	$71.4^{+2.2}_{-2.4}$

This marginalized likelihood plot is presented in Figure 11. In Table 3 we present the best-fit values for the vanilla cosmological parameters.

From the likelihood plot, Figure 11, there is evidently a preferred value of the feature spacing, roughly  $\Delta \sim 0.75$ , for which the constraints are the weakest. This can be understood as follows. For very high density  $\Delta \rightarrow 0$  the localized features from IR cascading smear together, and the total power spectrum (106) is  $P(k) \sim A_s (k/k_0)^{n_s-1} + C$  where the size of the constant deformation scales linearly with the density of features:  $C \propto \Delta^{-1}$ . Therefore, the upper bound on the amplitude  $A_{\text{IR}}$  should scale linearly with  $\Delta$ . Indeed, this linear trend is very evident from Figure 11 in the small- $\Delta$  regime. This linear behaviour must break down at some point since as the features become infinitely widely spaced the constraint on  $A_{\text{IR}}$  must go to zero. This explains the bump in the likelihood plot, Figure 11, near  $\Delta \sim 0.75$ .

In passing, notice that the behaviour  $P(k) \sim A_s(k/k_0)^{n_s-1} + C$  for  $\Delta \ll 1$  also explains why the best-fit  $A_s$  in Table 3 is somewhat lower than the standard value and why the spectral tilt  $n_s - 1$  is somewhat more red.

To get some intuition for the kinds of multibump distortions that are allowed by the data, we consider an example with  $A_{\text{IR}} = 1 \times 10^{-9}$ ,  $\Delta = 0.75$  and fix the vanilla parameters to  $A_s = 1.04 \times 10^{-9}$ ,  $n_s = 0.93$ . This choice of parameters is consistent with the data at  $2\sigma$  and corresponds to a reasonable coupling  $g^2 \sim 0.02$ . In Figure 12 we plot the primordial power spectrum  $P(k)$  and also the CMB TT angular power spectrum for this example.

## 7. Particle Physics Models

In previous sections of this paper we have studied in some detail the observational signatures of inflationary particle production in model (4), and we have also derived observational constraints on the model parameters  $g^2$  and  $\phi_0$ . We now aim to provide some explicit microscopic derivations of model (5) from popular theories of particle physics including string theory and SUSY. Our examples are meant to be illustrative rather than exhaustive. From the low energy perspective interactions of type (4) are completely generic, and hence we expect that many microscopic derivations may be possible.

**7.1. Open String Inflation Models.** String theory inflation models may be divided into two classes depending on the origin of the inflaton: closed string models and open string models. In the former case the inflaton is typically a geometrical modulus associated with the compactification manifold (examples include racetrack inflation [169], Kähler modulus inflation [170], Roulette inflation [171], and fibre inflation [172]). In the latter case the inflaton is typically the position modulus of some mobile D-brane (One notable exception is inflation driven by the open string tachyon, for example nonlocal string field theory models [42–44, 47–50].) living in the compactification manifold (examples include brane inflation [173, 174] such as the warped KKLM model [175], D3/D7 inflation [176], and DBI inflation [51, 52]). In open string inflation models there may be, in addition to the mobile inflationary brane, some additional “spectator” branes. If the mobile brane collides with any spectator brane during inflation, then some of the stretched string states between these branes will become massless at the moment when the branes are coincident [54, 80], precisely mimicking interaction (4). Thus, we expect particle production, IR cascading, and the bump-like features described above to be a reasonably generic prediction of open string inflation.

**7.2. String Monodromy Models.** A concrete example of the heuristic scenario discussed in the last subsection is provided by the brane monodromy and axion monodromy string theory inflation models proposed in [77–79]. In the original brane monodromy model [77] one considers type IIA string theory compactified on a nil manifold that is the product of

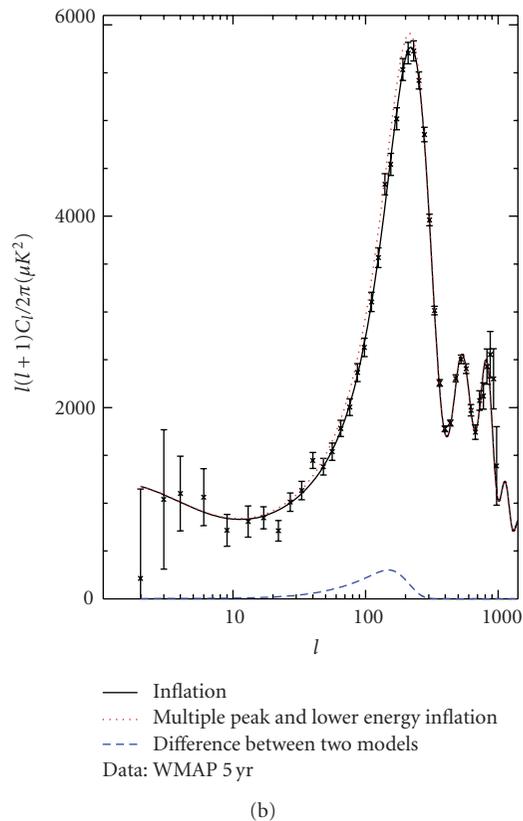
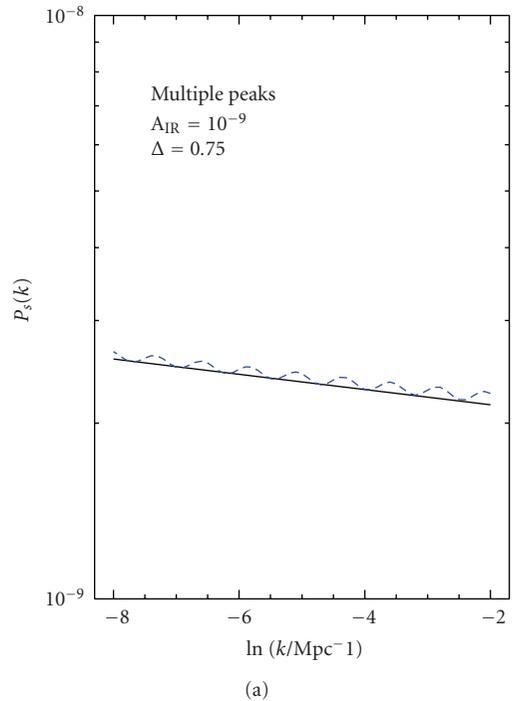


FIGURE 12: (a) shows a sample multiple-bump distortion with amplitude  $A_{\text{IR}} = 1 \times 10^{-9}$  which corresponds to  $g^2 \sim 0.02$ . The feature spacing is  $\Delta = 0.75$ . We take the vanilla parameters to be  $A_s = 1.04 \times 10^{-9}$ ,  $n_s = 0.93$  so that the scale of inflation is slightly lower than in the standard scenario and the spectral tilt is slightly redder. (b) shows the CMB angular TT power spectrum for this example.

two twisted tori. The metric on each of these twisted tori has the form

$$\frac{ds^2}{\alpha'} = L_{u_1}^2 du_1^2 + L_{u_2}^2 du_2^2 + L_x^2 (dx' + Mu_1 du_2)^2, \quad (108)$$

where  $x' = x - (M/2)u_1 u_2$ , and  $M$  is an integer flux number. The dimensionless constants  $L_{u_1}$ ,  $L_{u_2}$ , and  $L_x$  determine the size of the compactification.

Inflation is realized by the motion of a D4-brane along the direction  $u_1$  of the internal manifold. The D4 spans our large 3 dimensions and wraps 1 cycle along the direction  $u_2$  of the internal space. The size of this 1 cycle, in string units, is given by

$$L = \sqrt{L_{u_2}^2 + L_x^2 M^2 u_1^2}. \quad (109)$$

Hence, the brane prefers to minimize its world-volume by moving to the location  $u_1 = 0$  where this 1 cycle has minimal size. This preference gives a potential to the D4-brane position which goes like  $V \sim u_1$  in the large  $u_1$  regime that is relevant for large field inflation.

In [54] it was shown that this scenario allows for the inclusion of a number of spectator branes stabilized at positions  $u_1 = i/M$  (with  $i$  integer) along the inflationary trajectory. As the mobile inflationary D4 rolls through these points, particles (strings) will be produced, and the resulting distribution of features will look nearly identical to the simple multi-bump scenario studied in Section 6.4. To see this, let us now determine the distribution of features that is predicted from brane monodromy inflation. The canonical inflaton  $\phi$  can be related to the position of the mobile D4 as

$$\phi = Bu_1^{1/p}, \quad (110)$$

where  $B$  is a constant with dimensions of mass that depends on model parameters. Hence, the effective potential during inflation has the power-law form

$$V(\phi) = \mu^{4-p} \phi^p. \quad (111)$$

For the simplest scenario described above one has  $p = 2/3$ . However, formulas (110) and (111) still hold for the variant considered in [77] with  $p = 2/5$  as long as one replaces  $u_1$  by a more complicated linear combination of coordinates. These relations also hold for axion monodromy models [78] with  $p = 1$  and  $u_1$  replaced by the axion,  $c$ , arising from a 2-form RR potential  $C^{(2)}$  integrated over a 2-cycle  $\Sigma_2$ . For all models of the form of (111) the number of  $e$ -foldings  $N$  from  $\phi = \phi(N)$  to the end of inflation (which occurs at  $\phi = p/\sqrt{2}$  when the slow-roll parameter  $\epsilon(\phi) = 1$ ) is given by

$$\begin{aligned} N &= \frac{1}{2p} \frac{\phi^2(N)}{M_p^2} - \frac{p}{4} \\ &= \frac{1}{2p} \frac{B^2}{M_p^2} u_1^{2/p} - \frac{p}{4}. \end{aligned} \quad (112)$$

Since the spectator branes are located at  $u_1 = i/M$ , the bursts of particle production must occur at times  $N = N_i$  where

$$N_i = \frac{1}{2p} \frac{B^2}{M_p^2} \left(\frac{i}{M}\right)^{2/p} - \frac{p}{4}. \quad (113)$$

The location  $k = k_i$  of the  $i$ th feature is defined, roughly, by the scale leaving the horizon at the moment  $N = N_i$ . Hence, the distribution of features for brane/axion monodromy models is given by

$$\ln\left[\frac{k_i}{H}\right] \cong \tilde{B}^2 i^{2/p} - \frac{p}{4} \quad (114)$$

with  $p = 2/3$  or  $p = 2/5$  for brane monodromy and  $p = 1$  for axion monodromy. In (114) the dimensionless number  $\tilde{B}$  depends on model parameters.

Although the distribution of features (114) is not exactly the same as the evenly space distribution considered in Section 6.4, the two are essentially indistinguishable over the range of scales which are probed observationally (corresponding to roughly 10  $e$ -foldings of inflation). The reason for this is simple: the inflaton is nearly constant during the first 10  $e$ -foldings of inflation and hence  $\delta N \sim \delta\phi \sim \delta u_1$  within the observable region. It follows that  $k_i \cong e^{(i-1)\Delta} k_1$  to very good approximation for a huge class of models. To see this more concretely in the case at hand, let us compute  $dN/du_1$  from (112). It is straightforward to show that

$$\frac{dN}{du_1} = \frac{1}{p^p} \frac{1}{[2\epsilon(\phi)]^{1-p/2}} \left(\frac{B}{M_p}\right)^p, \quad (115)$$

where

$$\epsilon(\phi) \equiv \frac{M_p^2}{2} \left(\frac{V'}{V}\right)^2 = \frac{p^2}{2} \left(\frac{M_p}{\phi}\right)^2 \quad (116)$$

is the usual slow-roll parameter. Observational constraints on the running of the spectral index imply that  $\epsilon(\phi)$  cannot change much over the observable 10  $e$ -foldings of inflation. Since  $dN/du_1 \cong \text{const}$  to very high accuracy it follows trivially that  $N = N(u_1)$  is very close to linear and  $k_i \cong e^{(i-1)\Delta} k_1$  as desired.

In the context of axion monodromy inflation models [78] the multiple bump features discussed here will be complimentary to the oscillatory features described in [79] which result from the sinusoidal modulation of the inflaton potential by instanton effects. If the bursts of particle production are sufficiently densely spaced, then signal from IR cascading may appear oscillatory; however, it differs from the effect discussed in [79] in both physical origin and also functional form.

Let us now estimate the effective value of the couplings  $g_i^2$  appearing in the prototype interaction (105) that are predicted from the simplest brane monodromy model. A complete calculation would involve dimensionally reducing the DBI action describing the brane motion and requires knowledge of the full 10-dimensional geometry with the various embedded branes. For our purposes, however, a simple heuristic estimate for the collision of two D4-branes will suffice. When  $N$  D-branes become coincident the symmetry is enhanced from  $U(1)^N$  to a  $U(N)$  Yang Mills gauge theory. The gauge coupling for this Yang Mills theory is given by

$$g_{\text{YM}}^2 = \frac{g_s (2\pi)^2}{L}, \quad (117)$$

where  $L$  is the volume of the 1 cycle that the D4 branes wrap and is given by (5). If the inflationary brane is at position  $u_1$  and the  $i$ th spectator brane is at  $u_{1,i}$ , then the distance between the two branes is given by

$$d^2 = \alpha' L_{u_1}^2 (u_1 - u_{1,i})^2. \quad (118)$$

The mass of the gauge bosons corresponding to the enhanced symmetry is

$$M_i^2 = g_{\text{YM}}^2 \frac{d^2}{(2\pi)^2 (\alpha')^2} = \frac{g_s L_{u_1}^2 (u_1 - u_{1,i})^2}{\alpha' \sqrt{L_{u_2}^2 + L_x^2 M^2 u_1^2}}. \quad (119)$$

To put this in the prototype form  $M_i^2 = g_i^2 (\phi - \phi_i)^2$  we must first convert to the canonical variable  $\phi$  using formula (110) with  $p = 2/3$  and

$$B = \frac{M^{1/2} L_{u_1} L_x^{1/2}}{6\pi^2 \sqrt{g_s \alpha'}}. \quad (120)$$

Next, we must Taylor expand the resulting equation about the minimum  $\phi = \phi_i$ . We find

$$M_i^2 \cong g_i^2 (\phi - \phi_i)^2 + \dots, \quad (121)$$

$$\begin{aligned} g_i^2 &= \frac{16g_s^2 \pi^4}{M L_x u_{1,i}} \frac{1}{\sqrt{L_{u_2}^2 + L_x^2 M^2 u_{1,i}^2}} \\ &= \frac{16g_s^2 \pi^4}{L_x i} \frac{1}{\sqrt{L_{u_2}^2 + L_x^2 i^2}}, \end{aligned} \quad (122)$$

where on the second line of (122) we have used the fact that  $u_{1,i} = i/M$  (with  $i$  integer) in the simplest models. We see that the effective couplings  $g_i^2$  become larger as the D4 unwinds during inflation. (The apparent divergence for  $u_{1,i} = 0$  in formula (122) is an artifact of the fact that relation (110) is not valid at small values of  $u_1$ . This will not concern us here since inflation has already terminated at the point that our formulas break down.)

To compute the amplitude of the bump-like feature produced by brane monodromy inflation we should take into account also combinatorial factors. When two branes become coincident the symmetry is enhanced from  $U(1)^2$  to  $U(2)$  so there are  $2^2 - 2 = 2$  additional massless spin-1 fields appearing at the brane collision. Thus, using (107), the amplitude of the feature that will be imprinted in the CMB is

$$A_{i,\text{eff}} = 2 \times (2^2 - 2) \times \left[ 10^{-6} \cdot g_i^{15/4} \right], \quad (123)$$

where the extra factor of 2 counts the polarizations of the massless spin-1 fields. This combinatorial enhancement can be much larger if the inflationary brane collides with a *stack* of spectators.

The above discussion is predicated on the assumption that the original brane monodromy setup 1 is supplemented by additional spectator branes. This may seem like an unnecessary contrivance; however, in order for this model to reheat successfully it may be *necessary* to include spectator branes. For example, with the reheating mechanism proposed in

[177], semirealistic particle phenomenology can be obtained by confining the standard model (SM) to a D6 brane which wraps the compact space. In order to reheat into SM degrees of freedom, we orient this brane so that its world-volume is parallel to the mobile (inflationary) D4. In this case the end of inflation involves multiple oscillations of the D4 about the minimum of its potential. At each oscillation the D4 collides with the D6 and SM particles are produced by parametric resonance preheating [92]. However, due to the periodic structure of the compactification, D4/D6 collisions will necessarily occur also *during* inflation, leading to IR cascading features in the CMB.

The timing of these D4/D6 collisions was computed in [177] for the minimal  $p = 2/3$  brane monodromy model, assuming the same choices of parameters used in [77]. For this particular case there is only one collision (and hence one feature) during the first 10  $e$ -foldings of inflation, and the phenomenology is essentially the same as that considered in Section 6.3. What is the amplitude of this feature? Assuming, again, the parameters employed in [77] and noting that the first collision takes place at  $i = 13$  [177], (122) gives  $g_1^2 \cong 0.001$ . From (123) we find the effective amplitude of the feature to be  $A_{1,\text{eff}}/A_s \cong 0.01$ . This value is well within the observational bounds derived in Section 6.3.

We stress that the conclusions in the previous paragraph apply *only* for the particular choice of model parameters employed in [77]. There exist other consistent parameter choices for which the simplest brane monodromy model predicts a much higher density of features with much larger amplitude.

Note that both brane and axion monodromy models may be used to realize trapped inflation [54]. Here we are restricting ourselves to the large-field regime where the potential  $V = \mu^{4-p} \phi^p$  is flat enough to drive inflation without the need for trapping effects. For a given choice of parameters one should verify that this classical potential dominates over the quantum corrections from particle production.

**7.3. A Supersymmetric Model.** Another microscopic realization of multiple bursts of particle production and IR cascading during inflation which does not rely on string theory can be obtained from the so-called ‘‘distributed mass’’ model derived in [76] with warm inflation [82] in mind; however, the theory works equally well for our scenario. This model is based on  $\mathcal{N} = 1$  global SUSY and allows for the inclusion of multiple points along the inflationary trajectory where both scalar degrees of freedom and also their associated fermion super-partners become massless. The distribution of features in this set-up is essentially arbitrary.

## 8. Nongaussianity from Particle Production and IR Cascading

We have seen that the dynamics of particle production, rescattering, and IR cascading during inflation in model (5) leads to a bump-like contribution to the primordial power spectrum of the observable cosmological fluctuations.

However, this dynamics must also have a nontrivial impact on nongaussian statistics, such as the bispectrum. Indeed, it is evident from the analytical formalism presented in Sections 4 and 5 that the inflaton fluctuations  $\delta\phi$  generated via the rescattering diagram in Figure 1 may be significantly nongaussian. (This is evident since the particular solution of (49) is bi-linear in the Gaussian field  $\chi$ .) In this section we characterize this nongaussianity. The results of this section appeared for the first time in [12, 14, 15].

**8.1. Probability Density Function and Cummulants.** As discussed in the introduction, nongaussianity is often characterized by the size, shape, and running of the bispectrum  $B(k_i)$ . A popular measure of the size of the nongaussianity is the magnitude of  $B(k_i)$  on equilateral triangles. This approach has the advantage of being straightforward; however, it may give misleading results when one wishes to compare bispectra with different shapes [178] or scaling behaviour. This is particularly true in models, such as (5), where the bispectrum is large only for triangles with a characteristic size and where the nongaussian part of  $\zeta$  is uncorrelated with the gaussian part. A more meaningful single number to compare between models is the skewness (defined below), which contains information about the bispectrum integrated over all scales (up to some UV smoothing scale) and all shape configurations. (See also [179] for a related discussion and alternative methodology.)

In what follows it will be useful to define a normalized curvature perturbation

$$\xi \equiv \frac{\zeta}{\sigma_\zeta}, \quad (124)$$

where  $\zeta$  is the usual gauge invariant curvature perturbation, and  $\sigma_\zeta$  is the variance

$$\sigma_\zeta^2 = \langle \zeta^2 \rangle = \int \frac{d^3 k_1}{(2\pi)^{3/2}} \frac{d^3 k_2}{(2\pi)^{3/2}} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \rangle. \quad (125)$$

If  $\zeta$  is generated by the quantum vacuum fluctuations of the inflaton field, then we have  $\sigma_\zeta \cong 10^{-9/2}$ .

Let us define the probability density function (PDF),  $P(\xi)$ , as the probability that the (normalized) curvature perturbation has fluctuations of size  $\xi$ . The  $n$ th central moment of the PDF is simply

$$\langle \xi^n \rangle = \int_{-\infty}^{+\infty} \xi^n P(\xi) d\xi. \quad (126)$$

In order to quantify the nongaussianity of the PDF it is useful to define the cummulants

$$\hat{S}_n \equiv \langle \xi^n \rangle_c = \frac{\langle \zeta^n \rangle_c}{\sigma_\zeta^n}, \quad (127)$$

where the subscript c is a reminder that only the connected part of the correlation function should be included. For a gaussian PDF we have  $\hat{S}_n = 0$  for  $n = 3, 4, \dots$ , hence the cummulants provide a measure of nongaussianity. (Notice that, since  $\xi$  is constructed to have unit variance, we do not

need to distinguish between the reduced and dimensionless cummulants.)

In the case that  $\zeta$  (and hence  $\xi$ ) is a free field then the central limit theorem implies that the PDF is gaussian:

$$P(\xi) d\xi = \frac{d\xi}{\sqrt{2\pi}} e^{-\xi^2/2}. \quad (128)$$

This expression admits a simple generalization to the case where  $P(\xi)$  is close to (but not exactly) gaussian. This generalization is the Edgeworth expansion

$$P(\xi) d\xi \cong \frac{d\xi}{\sqrt{2\pi}} e^{-\xi^2/2} \times \left[ 1 + \frac{\hat{S}_3}{6} H_3(\xi) + \left( \frac{\hat{S}_4}{24} H_4(\xi) + \frac{\hat{S}_3^2}{72} H_3(\xi) \right) + \dots \right] \quad (129)$$

(See [27] for a review and [180] for an alternative derivation.) Where  $H_n(\xi)$  denotes the Hermite polynomials of order  $n$  and  $\hat{S}_n$  are the dimensionless cummulants, defined by (127). Note that result (129) is an expansion in cummulants. This expression provides an accurate approximation of the true PDF provided the cummulants are well ordered in the sense that

$$1 \gg |\hat{S}_3| \gg |\hat{S}_4| \gg \dots \quad (130)$$

See [181, 182] for more discussion on the structure of the correlation functions and the validity of cosmological perturbation theory.

Finally, it remains to relate the cummulants  $\hat{S}_n$  ( $n \geq 3$ ) to the correlation functions (such as the bispectrum) of the primordial curvature perturbation. We can write the dimensionless skewness,  $\hat{S}_3$ , as an integral over the 3-point correlation function [27, 180]:

$$\hat{S}_3 = \frac{1}{\sigma_\zeta^3} \int \frac{d^3 k_1}{(2\pi)^{3/2}} \frac{d^3 k_2}{(2\pi)^{3/2}} \frac{d^3 k_3}{(2\pi)^{3/2}} \langle \zeta_{\mathbf{k}_1} \zeta_{\mathbf{k}_2} \zeta_{\mathbf{k}_3} \rangle_c. \quad (131)$$

A similar expression may be derived for the dimensionless kurtosis,  $\hat{S}_4$ . From (131) it is clear that the skewness provides a measure of the magnitude of the bispectrum integrated over all triangles. Hence, this provides a reasonable single number to compare the size of nongaussianity between different models. Moreover, the skewnesses (and higher cummulants) are of observational interest since these determine the probability of rare fluctuations [27, 166–168, 183–185]. Observables such as the abundance of collapsed objects may therefore be used to constrain  $\hat{S}_3$ ,  $\hat{S}_4$ , and so forth.

We have reviewed the derivation of the Edgeworth expansion (129) only to illustrate that the dimensionless skewness (131) encodes information about the bispectrum integrated over all momenta. However, our actual calculation of the PDF in the next subsection will be fully nonperturbative and hence does *not* rely on the validity of (129) or (130).

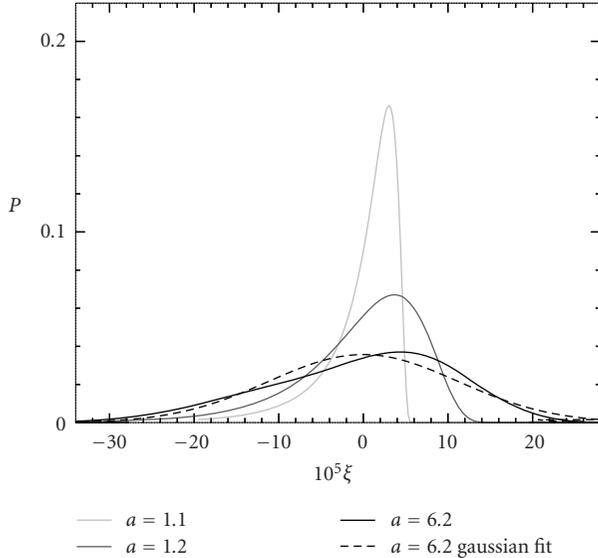


FIGURE 13: Probability density function of  $\zeta$  for  $g^2 = 0.1$  at a series of different values of the scale factor,  $a$ . The dashed curve shows a Gaussian fit at late time  $a = 6.2$ , and we have normalized the scale factor so that  $a = 1$  at the moment when particle production occurs. For illustration, we have chosen standard chaotic inflation  $V(\phi) = (m^2/2)\phi^2$ ; however, our qualitative results are independent of the background inflation model.

### 8.2. Magnitude of the Nongaussianity from IR Cascading.

We have argued in the last subsection that the moments of the PDF provide a useful tool for quantifying the size of nongaussianity and drawing comparisons between microscopic models whose bispectra may have very different shapes and scaling properties. Our goal now is to compute the skewness and kurtosis generated by particle production, rescattering and, IR cascading in model (5). We will then compare this to the skewness that would be generated by more familiar constructions, such as the local model  $\zeta = \zeta_g + (3/5)f_{NL}\zeta_g^2$ .

It is straightforward to extract the PDF numerically from our HLattice simulations by computing the fraction of the simulation box which contains the field  $\delta\phi \propto \zeta$  at a particular value. We again stress that this numerical evaluation of the PDF is non-perturbative; our results do not rely on any prior assumptions about the size or ordering of the cumulants.

We present our results for the numerical evaluation of the PDF from IR cascading in Figure 13. For illustration we have assumed a chaotic potential  $V(\phi) = m^2\phi^2/2$ ; however, our results are largely insensitive to the choice of background inflation model.

We can understand physically the behaviour of the PDF plotted in Figure 13. Shortly after the initial burst of particle production, the curvature perturbation is extremely nongaussian, due to the sudden appearance of the source term  $J \propto \chi^2$  in the equation of motion for the inflaton fluctuations  $\delta\phi$ . Quickly, in less than an  $e$ -folding, nonlinear interactions begin to drive the system towards gaussianity. A similar

TABLE 4: Dimensionless cumulants for IR cascading.

$g^2$	Skewness ( $\hat{S}_3$ )	Kurtosis ( $\hat{S}_4$ )	Equivalent $f_{NL}^{\text{local}}$
1	-0.51	0.16	-2800
0.1	-0.49	-0.08	-2700
0.01	-0.033	$\mathcal{O}(10^{-3})$	-183

behaviour has been observed in lattice simulations of out-of-equilibrium interacting scalar fields during preheating [186, 187]. In the case of rescattering during preheating, the system will eventually become gaussian when the fields thermalize. However, in our case the universe is still inflating during the process of rescattering and IR cascading. As a result, nongaussian inflaton fluctuations generated by rescattering are stretched out by the quasi-de Sitter expansion and must freeze once their wavelength crosses the Hubble scale. Hence, at late times the PDF does not become completely gaussian but rather freezes-in with some non-trivial skewness. Within a few  $e$ -foldings from the moment of particle production the time evolution of the PDF has become completely negligible.

Given our numerical results for the PDF, it is straightforward to compute dimensionless cumulants such as the skewness ( $\hat{S}_3$ ) and kurtosis ( $\hat{S}_4$ ) for various values of the coupling  $g^2$ . We have summarized our results in Table 4. At small coupling,  $g^2 < 1$ , we always have  $|\hat{S}_3| > |\hat{S}_4|$  which suggests that the cumulants are well ordered. In particular, note that for  $g^2 = 0.01$ , the kurtosis  $\hat{S}_4$  is too small to be measured accurately from our HLattice simulations.

In order to give a sense of the magnitude of the nongaussianity from particle production we have also computed an “equivalent  $f_{NL}^{\text{local}}$ ” defined by  $5\hat{S}_3/(18\sigma_\zeta)$ . For a given  $g^2$  this effective  $f_{NL}^{\text{local}}$  is the magnitude of  $f_{NL}$  which would be necessary to reproduce the skewness  $\hat{S}_3$  of the IR cascading PDF using a local ansatz  $\zeta = \zeta_g + (3/5)f_{NL}\zeta_g^2$ . (Our sign conventions for  $f_{NL}$  are consistent with WMAP [6]. See [27] for a discussion of various conventions employed in the literature.)

From Table 4 we see that IR cascading during inflation may generate significant nongaussianity. Even taking  $g^2 = 0.01$  (which is compatible with cosmological data for any choice of  $\phi_0$ ), we still obtain a skewness  $\hat{S}_3 = -0.033$ , which is the same value that would be produced by a local model with  $f_{NL} = -183$ . This “equivalent” local nongaussianity is larger than current observational bounds, which suggests that nongaussian features from particle production during inflation might be observable for reasonable values of the coupling  $g^2$ . Note, however, that observational bounds on  $f_{NL}^{\text{local}}$  cannot be directly applied to our model since the bispectrum in our case is uncorrelated with the vacuum fluctuations and is far from scale invariant.

Depending on the value of  $\phi_0$  (and hence the location of the feature in the power spectrum) the nongaussianity from IR cascading may lead to a variety of observational signatures. If the feature is localized on scales relevant

for CMB experiments, then the key observable is the primordial bispectrum,  $B(k_i)$ . Efficient computation of the observational constraints on nongaussianity from inflationary particle production will require a separable template for the bispectrum, an issue which we briefly comment on in the next subsection and which will be considered in more detail in [15]. On the other hand, the feature might show up on smaller scales and leave an imprint on large scale structures. In this case there are at least two observables which could be used to constrain our model: higher order correlations of LSS probes (such as the galaxy bispectrum) and the abundance of collapsed objects. In the first case, we require the evolved bispectrum while in the latter case we require the PDF of the evolved density field  $\delta\rho/\rho$ . Our model robustly predicts negative skewness for both the curvature perturbation  $\zeta$  and the density field. Hence we should expect *less* rare objects (as compared to the gaussian case) on some characteristic scale. (The recent weak lensing measurement of the dark matter mass of the high-redshift galaxy cluster XMMUJ2235.3-2557 [188] has been construed as a possible hint of nongaussian initial conditions [184]. Unfortunately, our model does not produce the correct sign of skewness to explain such observations.)

When computing the statistics of density fluctuations on some given length scale  $R$  one should smooth the perturbation field over wave-numbers larger than  $k_{\max} \sim R^{-1}$ . This may be implemented in PDF (129) by the formal substitution  $\zeta_{\mathbf{k}} \rightarrow \mathcal{W}(k)\zeta_{\mathbf{k}}$  in expressions (125) and (131) for the variance and skewness. Here  $\mathcal{W}(k)$  is some window function which tends to zero for  $k \gg R^{-1}$  (e.g., a standard choice of window is a top-hat in real space). This substitution makes any predictions for the statistics of the density fluctuations implicitly dependent on the smoothing scale,  $R$ . (It also renders the momentum integrals UV finite.) In our analysis we have computed the PDF numerically by measuring the fraction of the simulation box in which the inflaton field takes a given value. Thus, our cummulants are automatically smoothed on the UV cutoff scale  $k_{\max} \sim \Delta x^{-1}$  defined by the lattice spacing  $\Delta x$ .

One may wonder how sensitively our results for the dimensionless cummulants depend on this particular choice of smoothing. Recall that the bispectrum from IR cascading (73) is strongly peaked on triangles with a particular size, corresponding to the location of the bump-like distortion in the power spectrum. If we denote this characteristic scale by  $k_{\text{bump}}$ , then it should be clear that the cummulants in our model are insensitive to the smoothing scale, as long as  $R \lesssim k_{\text{bump}}^{-1}$ .

**8.3. Shape of the Bispectrum.** In Section 4 we presented a schematic discussion of how to compute analytically the primordial bispectrum from IR cascading, (73). In [14], this expression is evaluated analytically. We find that the bispectrum (73) from particle production has a unique shape that has not been considered in the previous literature. In order to describe this novel shape we would like to first factor out that strong overall dependence of  $B_\phi(k_i)$  on the size of

the triangle. To this end we define a ‘‘shape function’’  $S(k_i)$  in terms of the bispectrum as follows:

$$S(k_i) = N^{-1}(k_1 k_2 k_3)^2 B(k_i), \quad (132)$$

where  $B(k_i)$  is related to  $B_\phi(k_i)$  by (104), and  $N$  is a normalization factor to be discussed shortly. The function  $S(k_i)$  has the advantage that the strong  $k^6$  running of the bispectrum is extracted. Hence, any residual scaling behaviour displayed by  $S(k_i)$  must be a result of nonlinear interactions.

Since we expect the 3-point correlation function to be of order  $P_\zeta^2(k)$ , a natural choice of normalization is  $N = (2\pi)^4 \mathcal{P}_\zeta^2$  where  $\mathcal{P}_\zeta^{1/2} = 5 \times 10^{-5}$  is the usual amplitude of the scale invariant fluctuations from inflation. With this choice of normalization our function  $S(k_i)$  coincides with the quantity  $\mathcal{G}(k_i)/(k_1 k_2 k_3)$  which was used to study localized nongaussian features from models with steps in the inflaton potential in [59, 60].

Symmetry of the bispectrum under permutations of momenta implies that we can focus only on the region  $k_1 \geq k_2 \geq k_3$ , to avoid counting the same configuration twice. Moreover, the triangle inequality implies that  $1 - k_2/k_1 \leq k_3/k_1$ . Therefore we can completely specify the shape of the bispectrum for a given size of triangle  $k$  by plotting  $S(k, kx_2, kx_3)$  in the region  $x_3 \leq x_2 \leq 1$  and  $1 - x_2 \leq x_3$ . Because our bispectrum is very far from being scale invariant, it follows that this shape function is sensitive to the choice of  $k$ . Therefore, in Figure 14 we choose several representative choices:  $\ln(k/k_{\text{bump}}) = -1, 0, 1, 2$ .

We see that a rich array of shapes is possible: for  $k \lesssim k_{\text{bump}}$ , the bispectrum is qualitatively similar to the equilateral model; however, at slightly larger  $k$  there is also considerable support on flattened triangles. Note that for  $k \gtrsim 7.4 k_{\text{bump}}$  the shape of the bispectrum is extremely unusual and is not easily comparable to any shape that has been proposed in the previous literature.

We find that the bispectrum from IR cascading is, to a first approximation, factorisable in the sense that  $B(k_1, k_2, k_3) \cong \prod_{i=1}^3 F(k_i)$  where  $F(k) \sim C_1 k^3 e^{-C_2 k^2}$  for some constants  $C_1$  and  $C_2$ . This separable form is important since it permits fast algorithms for forecasting, data analysis, and simulations [189].

In this section we have attempted to characterize the nongaussian signature associated with inflationary particle production. This signature is rather unique in the literature. Our model predicts uncorrelated localized nongaussian features with a unique shape of bispectrum. We have quantified the size of this effect by studying the dimensionless cummulants (such as the skewness) and have argued that this nongaussianity can be significant, depending on  $g^2$ . We leave a detailed study of the observational constraints on such nongaussianities to future studies [15].

**8.4. Multiple Bursts of Particle Production.** Note that our discussion of nongaussianity generalizes easily to the case

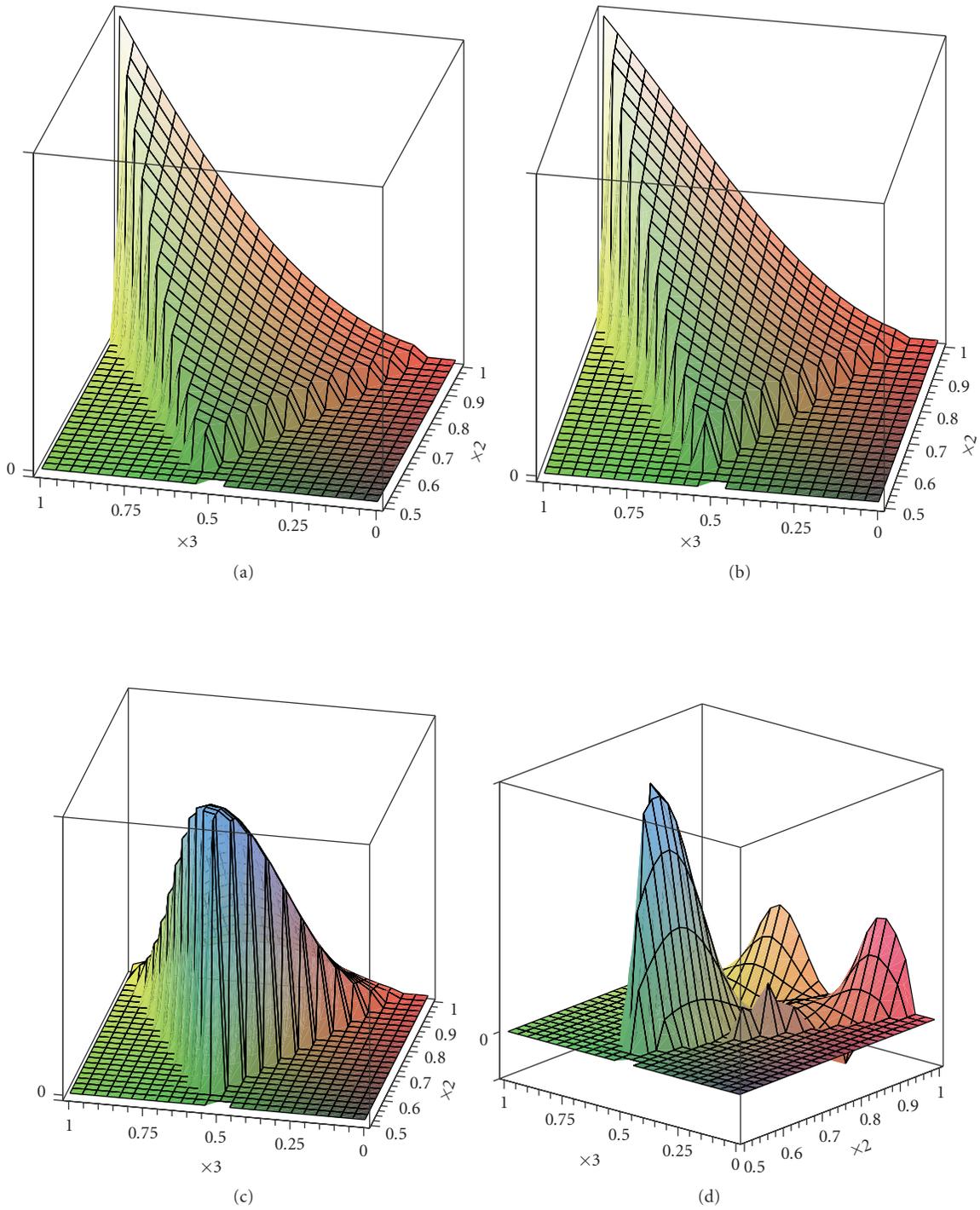


FIGURE 14: The shape function  $S(k, kx_2, kx_3)$ , defined by (132), as a function of the dimensionless quantities  $x_2, x_3$  which parametrize the shape of the triangle. (a) corresponds to  $k = e^{-1}k_{\text{bump}}$ , (b) is  $k = k_{\text{bump}}$ , (c) is  $k = e^{+1}k_{\text{bump}}$ , and (d) is  $k = e^{+2}k_{\text{bump}}$ . In the IR ( $k \leq k_{\text{bump}}$ ) the shape of the bispectrum is similar to the equilateral shape; however, there is also some support on flattened triangles near  $k \sim e^{+1}k_{\text{bump}}$ . At larger values of  $k$  the shape is unlike any other template proposed in the literature.

where there are multiple points  $\phi_i$  ( $i = 0, \dots, n$ ) along the inflaton trajectory where new degrees of freedom  $\chi_i$  become massless, such as model (105). Such a construction may be quite natural in the context of brane/axion monodromy inflation [54, 77–79]. In this case the nongaussian features in  $B(k_i)$  from each burst of particle production may superpose to generate a broad-band signal. Such a construction may (but need not) be associated with trapped inflation [54]. Depending on the spacing of the points  $\phi_i$  and the couplings  $g_i^2$ , a rich variety of nongaussianities may be possible. In this case that the points  $\phi_i$  are sufficiently densely spaced, we expect a nongaussian signal which is close to equilateral but also has some support on flattened triangles, consistent with the analysis of [54, 81]. We leave a detailed discussion to future studies.

## 9. Conclusions

In this paper we have considered the possibility that some non-inflation (iso-curvature) particles were produced during the observable range of  $e$ -foldings of inflation. Inflationary particle production might occur as a result of a phase transition, parametric resonance, or other non-adiabatic processes. In order to illustrate the basic physics of inflationary particle production we restricted our analysis to a simple prototype model with coupling  $(g^2/2)(\phi - \phi_0)^2\chi^2$  between the inflaton  $\phi$  and iso-inflaton  $\chi$ . However, we expect that our qualitative results will apply also to SUSY models, gauged interactions, higher spin iso-inflaton fields, and phase transitions.

Models of the type we study have attracted considerable interest recently in connections with trapped inflation, trans-Planckian effects, and observable features/nongaussianity in the primordial curvature fluctuations from inflation. Moreover, such models are quite natural from the microscopic perspective and may be obtained in popular models of open string inflation, such as brane/axion monodromy.

We have shown that inflationary particle production in the model  $(g^2/2)(\phi - \phi_0)^2\chi^2$  leads to a new mechanism for generating cosmological fluctuations. This mechanism is qualitatively different from previous proposals in that this approach does not rely on the quantum vacuum fluctuations of light fields during inflation. Rather, the scenario involves the production of massive  $\chi$  particles during inflation, which subsequently rescatter off the slow-roll condensate  $\phi(t)$  to emit bremsstrahlung radiation of light inflaton fluctuations  $\delta\phi$ . We have studied this dynamics using classical lattice field theory simulations, analytical QFT computations, and also second-order cosmological perturbation theory. All of these approaches yield consistent results. We have found that rescattering proceeds with a time scale short compared to the expansion time. Moreover, the emission of long-wavelength inflaton fluctuations is very energetically inexpensive. The combination of these two effects leads to a rapid build-up of power in IR inflation fluctuations shortly after the moment of particle production. This dynamical process is called IR cascading.

Our numerical and analytical studies of rescattering and IR cascading during inflation may have relevance for trapped inflation, preheating, moduli trapping, and also non-equilibrium QFT more generally. For instance, we have seen that, even with a small number of out-of-equilibrium  $\chi$  particles, multiple rescatterings can nevertheless generate long-wavelength  $\delta\phi$  fluctuations with huge occupation numbers. We have also observed, for the first time, the dynamical approach to the scaling regime discovered in [96, 97].

IR cascading during inflation leads to observable features in the primordial cosmological fluctuations. In particular, this process generates a bump-like contribution to the primordial scalar power spectrum. This signature is very different from what would be obtained in a model with transient violation of slow roll during inflation (such as a step-like feature in  $V(\phi)$ ), contrary to some claims in the literature.

We have studied the observational constraints on bump-like features from inflationary particle production during inflation. We found that relatively large distortions, of order 10% of the usual scale invariant vacuum fluctuations, are compatible with current data. We have derived observational bounds on the coupling  $g^2$  for a given  $\phi_0$ , which play a crucial role in determining the detectability of nongaussianity from particle production. Our observational bounds on particle production during inflation have implications for brane/axion monodromy inflation models and other microscopic constructions.

IR cascading also has a nontrivial impact on nongaussian statistics, such as the bispectrum. The model  $(g^2/2)(\phi - \phi_0)^2\chi^2$  leads to a very novel nongaussian signature: uncorrelated, localized nongaussian features with a unique shape of bispectrum. For reasonable values of the coupling,  $g^2 \lesssim 0.01$ , this new kind of nongaussianity may be detectable in future missions.

The nongaussian signature predicted by inflationary particle production is rather unusual, as compared to other models of inflation which are frequently studied in the literature. However, the underlying field theory description is extremely simple and rather generic from the low-energy perspective. In order to obtain observable nongaussianity it was not necessary to fine-tune the inflaton trajectory or appeal to re-summation of an infinite series of high-dimension operators. Indeed, the only “tuning” which is required for our signal to be observable is the requirement that  $\phi = \phi_0$  during the observable range of  $e$ -foldings.

There are a variety of directions for future studies. From the theoretical perspective, it would be interesting to explicitly generalize our results to more complicated models with particle production during inflation (such as SUSY models, fermion  $\chi$  fields, and phase transitions). There are also a wide range of interesting phenomenological possibilities. Varying the location of the feature we can have a variety of possible signatures for the CMB and LSS. We expect that IR cascading will also have implications for the spectrum of gravity waves from inflation and also primordial black holes. We can also imagine superposing multiple bursts of particle production to obtain an even

richer variety of signatures. We leave these possibilities for future investigation.

## Appendices

### A. Backreaction Effects

Once  $\phi$  rolls past the massless point  $\phi = \phi_0$ , a gas of  $\chi$  particles is produced with occupation number given by (13). This gas costs energy, which must be drained from the condensate  $\phi(t)$ . Hence, in order for the system to conserve energy, the inflaton must slow down slightly. As discussed in [12, 61, 64, 80, 92], this slowing down can be studied using the mean field equation (14). Using solution (45) to compute the vacuum average  $\langle \chi^2 \rangle$ , we find

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} + g \frac{k_*^3}{(2\pi)^3} \frac{\Theta(t)}{a^3(t)} \frac{(\phi - \phi_0)}{|\phi - \phi_0|} = 0. \quad (\text{A.1})$$

The step function  $\Theta(t)$  reflects the fact that the impact of particle production is felt only *after*  $\phi$  passes through  $\phi_0$  and  $a^{-3}$  comes from the volume dilution of non-relativistic particles. The final term in (A.1) may be interpreted as a quantum correction to the effective force due to particle production. (Note that we have implicitly subtracted the Coleman-Weinberg potential, which may be justified by the assumption of softly broken SUSY.)

The solutions of (A.1) display the expected slowing down behaviour and have been studied analytically in [61, 64, 80]. In [12] this slowing down was studied using inhomogeneous lattice field theory simulations, and the results were found to be compatible with the mean field treatment (see Figure 2). Here, we consider simple energetic arguments in order to clear up some common misconceptions. The effective inflaton potential, including the effects of particle production, is

$$V_{\text{eff}} = V(\phi) + g\Theta(t) \frac{k_*^3}{(2\pi)^3} \frac{|\phi - \phi_0|}{a^3}. \quad (\text{A.2})$$

For  $t \gtrsim k_*^{-1}$ , the  $\chi$  particles are non-relativistic, and their energy density is dominated by potential (rather than kinetic) energy. Hence, we have

$$\rho_\chi \cong g \frac{k_*^3}{(2\pi)^3} \frac{|\phi - \phi_0|}{a^3} \cong \frac{k_*^5}{(2\pi)^3 H} N e^{-3N} \quad (\text{A.3})$$

with  $N = Ht$  is the number of  $e$ -foldings measured from the moment when  $\phi = \phi_0$ . Shortly after particle production this energy density grows, corresponding to the fact that as  $\phi$  moves away from  $\phi_0$  the  $\chi$  particles become even more massive. However, this growth in the energy density cannot continue forever. At  $N = 1/3$  the energy density  $\rho_\chi$  peaks, and at later times it decays exponentially as  $e^{-3N}$ , corresponding to the volume dilution of the massive  $\chi$ . Thus, the energy density in the produced  $\chi$  is always bounded as

$$\rho_\chi < \frac{1}{24\pi^3 e} \frac{k_*^5}{H}. \quad (\text{A.4})$$

On the short time scales relevant for particle production,  $|\Delta N| \lesssim 1$ , it is still sensible to talk about energy conservation. The energy in produced  $\chi$  particles must therefore be balanced by a dip in the kinetic energy of the inflaton. (Intuitively this is to be expected since the ‘‘extra’’ term in (A.1) represents a force which tends to pull the inflaton back towards the point  $\phi = \phi_0$ .) To get a sense of the magnitude of this effect, let us compare  $\rho_\chi$  to the initial kinetic energy of the inflaton

$$K_{\text{in}} = \frac{\dot{\phi}_{\text{in}}^2}{2} = \frac{k_*^4}{2g^2}. \quad (\text{A.5})$$

Using (11) it is easy to see that

$$\frac{\rho_\chi}{K_{\text{in}}} < 0.35 g^{5/2} \ll 1. \quad (\text{A.6})$$

Hence, the total energy density that goes into  $\chi$  particles is small compared to the inflaton kinetic energy, and hence the velocity dip must also be a small effect, roughly  $\Delta\dot{\phi}/\dot{\phi} < 0.18 g^{5/2}$ . This simple estimate is consistent with a more quantitative treatment [12].

Note that since inflation is driven by potential energy, we have  $V \gg K_{\text{in}} \gg \rho_\chi$ . Hence, even during particle production the expansion rate  $H$  will still be dominated by the classical inflaton potential:  $H \cong \sqrt{V/(3M_p^2)}$ . Therefore, a single burst of particle production will *not* terminate inflation. Nor will this burst decohere the condensate since the occupation number of produced particles (13) is always less than unity. (We again remind the reader that, unlike the case of broadband resonant preheating after inflation, we have only a single burst of particle production.)

Note that throughout this appendix we are assuming standard slow roll inflation. In particular, our argument does not apply for trapped inflation [54] where the effect of backreaction on  $H$  and  $\phi(t)$  is much more significant.

### B. Detailed Computation of $P(k)$

In this appendix we discuss some of the technical details associated with the computation of the renormalized power spectrum (71). First, notice that using (41) and (45) we can write the quantity appearing in each renormalized Wick contraction as

$$\begin{aligned} & \chi^k(\tau) \chi_k^*(\tau') - f_k(\tau) f_k^*(\tau') \\ & \cong \frac{1}{k_*^2} \frac{1}{\sqrt{a(\tau)a(\tau')}} \frac{1}{\sqrt{t(\tau)t(\tau')}} \\ & \times \left[ n_k \cos\left(\frac{k_*^2 t^2(\tau)}{2} - \frac{k_*^2 t^2(\tau')}{2}\right) \right. \\ & \quad \left. + \sqrt{n_k} \sqrt{1+n_k} \sin\left(\frac{k_*^2 t^2(\tau)}{2} - \frac{k_*^2 t^2(\tau')}{2}\right) \right], \end{aligned} \quad (\text{B.1})$$

where the occupation number  $n_k$  is defined by (13). Plugging (B.1) into (71) we find

$$\begin{aligned}
P_\phi(k) = & \frac{g^2 k^3}{8\pi^5} \left[ \int d^3 k' n_{k-k'} n_{k'} \times \int d\tau' d\tau'' \frac{G_k(\tau - \tau')}{a(\tau)} \frac{G_k(\tau - \tau'')}{a(\tau)} \cos^2 \left[ \frac{k_\star^2 t^2(\tau')}{2} - \frac{k_\star^2 t^2(\tau'')}{2} \right] \right. \\
& + \int d^3 k' \sqrt{n_{k-k'} n_{k'}} \sqrt{1 + n_{k-k'}} \sqrt{1 + n_{k'}} \\
& \times \int d\tau' d\tau'' \frac{G_k(\tau - \tau')}{a(\tau)} \frac{G_k(\tau - \tau'')}{a(\tau)} \sin^2 \left[ \frac{k_\star^2 t^2(\tau')}{2} + \frac{k_\star^2 t^2(\tau'')}{2} \right] \\
& + \int d^3 k' \left( n_{k-k'} \sqrt{n_{k'}} \sqrt{1 + n_{k'}} + n_{k'} \sqrt{n_{k-k'}} \sqrt{1 + n_{k-k'}} \right) \\
& \left. \times \int d\tau' d\tau'' \frac{G_k(\tau - \tau')}{a(\tau)} \frac{G_k(\tau - \tau'')}{a(\tau)} \cos \left[ \frac{k_\star^2 t^2(\tau')}{2} - \frac{k_\star^2 t^2(\tau'')}{2} \right] \sin \left[ \frac{k_\star^2 t^2(\tau')}{2} + \frac{k_\star^2 t^2(\tau'')}{2} \right] \right]. \tag{B.2}
\end{aligned}$$

Notice that the time and phase space integrations in (B.2) decouple. This is the key simplification which makes an analytical evaluation of this expression tractable. Let us consider these integrations separately.

*B.1. Time Integrals.* All of the integrals over conformal time that appear in (B.2) can be expressed in terms of two characteristic integrals which we call  $I_1$  and  $I_2$ . Explicitly, these are defined as

$$\begin{aligned}
I_1(k, \tau) &= \frac{1}{a(\tau)} \int d\tau' G_k(\tau - \tau') e^{ik_\star^2 t^2(\tau')}, \\
I_2(k, \tau) &= \frac{1}{a(\tau)} \int d\tau' G_k(\tau - \tau'). \tag{B.3}
\end{aligned}$$

The second characteristic integral,  $I_2$ , can be evaluated analytically. However, the resulting expression is not particularly enlightening. Evaluation of  $I_1$ , on the other hand, requires numerical methods.

Let us now show how the various integrals appearing in (B.2) may be rewritten in terms of  $I_1$ ,  $I_2$ . First, consider the first line of (B.2) where the following integral appears:

$$\begin{aligned}
& \int d\tau' d\tau'' \frac{G_k(\tau - \tau')}{a(\tau)} \frac{G_k(\tau - \tau'')}{a(\tau)} \\
& \times \cos^2 \left[ \frac{k_\star^2 t^2(\tau')}{2} - \frac{k_\star^2 t^2(\tau'')}{2} \right] \\
& = \frac{|I_1(k, \tau)|^2}{2} + \frac{I_2(k, \tau)^2}{2}. \tag{B.4}
\end{aligned}$$

Next, consider the second line of (B.2) where the following integral appears:

$$\begin{aligned}
& \int d\tau' d\tau'' \frac{G_k(\tau - \tau')}{a(\tau)} \frac{G_k(\tau - \tau'')}{a(\tau)} \\
& \times \sin^2 \left[ \frac{k_\star^2 t^2(\tau')}{2} + \frac{k_\star^2 t^2(\tau'')}{2} \right] \\
& = -\frac{\text{Re}[I_1(k, \tau)^2]}{2} + \frac{I_2(k, \tau)^2}{2}. \tag{B.5}
\end{aligned}$$

Finally, consider the fourth line of (B.2) where the following integral appears:

$$\begin{aligned}
& \int d\tau' d\tau'' \frac{G_k(\tau - \tau')}{a(\tau)} \frac{G_k(\tau - \tau'')}{a(\tau)} \\
& \times \cos \left[ \frac{k_\star^2 t^2(\tau')}{2} - \frac{k_\star^2 t^2(\tau'')}{2} \right] \sin \left[ \frac{k_\star^2 t^2(\tau')}{2} + \frac{k_\star^2 t^2(\tau'')}{2} \right] \\
& = \text{Im}[I_1(k, \tau) I_2(k, \tau)]. \tag{B.6}
\end{aligned}$$

In expressions (B.5) and (B.6) the notations Re and Im denote the real and imaginary parts, respectively.

*B.2. Phase Space Integrals.* As a warm-up to the subsequent calculation consider the following integral:

$$\begin{aligned}
& \int d^3 k' n_{k-k'}^a n_{k'}^b \\
& = \int d^3 k' \exp[-a\pi |\mathbf{k} - \mathbf{k}'|^2 / k_\star^2] \exp[-b\pi |\mathbf{k}'|^2 / k_\star^2] \\
& = \frac{k_\star^3}{(a+b)^{3/2}} \exp \left[ -\frac{ab}{a+b} \frac{\pi k^2}{k_\star^2} \right]. \tag{B.7}
\end{aligned}$$

This formula is valid when  $a, b$  are positive real numbers. Notice that this expression is symmetric under interchange of  $a$  and  $b$ .

The phase space integral in the first line of (B.2) is computed by a trivial application of identity (B.7):

$$\int d^3 k' n_{k-k'} n_{k'} = \frac{k_\star^3}{2\sqrt{2}} e^{-\pi k^2 / (2k_\star^2)}. \tag{B.8}$$

However, the remaining phase space integrals appearing in (B.2) cannot be obtained exactly in closed form because they contain terms like  $\sqrt{1 + n_{k'}}$  where the gaussian factors appear under the square root. In order to deal with such expressions, we note because  $n_k \ll 1$  over most of the domain of integration, it is reasonable to replace  $\sqrt{1 + n_{k'}} \cong 1 + n_{k'}/2$ .

Let us now proceed in this manner. The phase space integral on the second line of (B.2) is

$$\begin{aligned} & \int d^3 k' \sqrt{n_{k-k'}} \sqrt{n_{k'}} \sqrt{1+n_{k-k'}} \sqrt{1+n_{k'}} \\ & \cong \int d^3 k' \left[ n_{k-k'}^{1/2} n_{k'}^{1/2} + \frac{1}{2} n_{k-k'}^{3/2} n_{k'}^{1/2} + \frac{1}{2} n_{k-k'}^{1/2} n_{k'}^{3/2} \right] \quad (\text{B.9}) \\ & = k_*^3 \left[ \exp\left(-\frac{\pi k^2}{4k_*^2}\right) + \frac{1}{2\sqrt{2}} \exp\left(-\frac{3\pi k^2}{8k_*^2}\right) \right]. \end{aligned}$$

Finally, consider the phase space integral on the third line of (B.2):

$$\begin{aligned} & \int d^3 k' \left[ n_{k-k'} \sqrt{n_{k'}} \sqrt{1+n_{k'}} + n_{k'} \sqrt{n_{k-k'}} \sqrt{1+n_{k-k'}} \right] \\ & \cong \int d^3 k' \\ & \quad \times \left[ n_{k-k'} n_{k'}^{1/2} + n_{k'} n_{k-k'}^{1/2} + \frac{1}{2} n_{k-k'} n_{k'}^{3/2} + \frac{1}{2} n_{k'} n_{k-k'}^{3/2} \right] \\ & = k_*^3 \left[ \frac{4\sqrt{2}}{3\sqrt{3}} \exp\left(-\frac{\pi k^2}{3k_*^2}\right) + \frac{2\sqrt{2}}{5\sqrt{5}} \exp\left(-\frac{3\pi k^2}{5k_*^2}\right) \right]. \quad (\text{B.10}) \end{aligned}$$

We have verified formulae (B.9) and (B.10) numerically. In both cases that the numerical results agree with these semianalytical expressions up to the percent level.

We can now, finally, insert results (B.4), (B.5), (B.6) and (B.8), (B.9), (B.10) into expression (B.2). Doing so, we arrive at our main analytical result, which is (71).

## Acknowledgments

This work has benefited considerably from the interactions with a number of people. Thanks are due to L. Kofman and D. Pogosian for collaboration on the original paper [12] which instigated this work. The author is especially grateful to Z. Huang for collaboration on the works of [12, 13] and also for numerous discussions and help with numerical simulations and several figures. Finally, thanks are also to T. Battefeld, C. Burgess, J. Cline, N. Dalal, H. Firouzjahi, L. Hoi, I. Huston, K. Malik, P. McDonald, A. E. Romano, M. Sasaki, D. Seery, S. Shandera, and A. Tolley for helpful comments, discussions, and input at various stages during the completion of this project. This work is dedicated to the memory of Lev Kofman.

## References

- [1] V. F. Mukhanov and G. V. Chibisov, "Quantum fluctuation and 'nonsingular' universe," *JETP Letters*, vol. 33, p. 532, 1981, *Pisma Zh. Eksp. Teor. Fiz.* vol. 33, p. 549, 1981.
- [2] S. W. Hawking, "The development of irregularities in a single bubble inflationary universe," *Physics Letters B*, vol. 115, no. 4, pp. 295–297, 1982.
- [3] A. A. Starobinsky, "Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations," *Physics Letters B*, vol. 117, no. 3–4, pp. 175–178, 1982.

- [4] A. H. Guth and S.-Y. Pi, "Fluctuations in the new inflationary universe," *Physical Review Letters*, vol. 49, no. 15, pp. 1110–1113, 1982.
- [5] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, "Spontaneous creation of almost scale-free density perturbations in an inflationary universe," *Physical Review D*, vol. 28, no. 4, pp. 679–693, 1983.
- [6] E. Komatsu, K. M. Smith, J. Dunkley, et al., "Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation," submitted to *Astrophysical Journal* <http://arxiv.org/abs/1001.4538>.
- [7] L. Kofman, "Probing string theory with modulated cosmological fluctuations," <http://arxiv.org/abs/astro-ph/0303614>.
- [8] F. Bernardeau, L. Kofman, and J.-P. Uzan, "Modulated fluctuations from hybrid inflation," *Physical Review D*, vol. 70, no. 8, Article ID 083004, 2004.
- [9] G. Dvali, A. Gruzinov, and M. Zaldarriaga, "New mechanism for generating density perturbations from inflation," *Physical Review D*, vol. 69, no. 2, Article ID 023505, 6 pages, 2004.
- [10] G. Dvali, A. Gruzinov, and M. Zaldarriaga, "Cosmological perturbations from inhomogeneous reheating, freezeout, and mass domination," *Physical Review D*, vol. 69, no. 9, 5 pages, 2004.
- [11] D. H. Lyth and D. Wands, "Generating the curvature perturbation without an inflaton," *Physics Letters B*, vol. 524, no. 1–2, pp. 5–14, 2002.
- [12] N. Barnaby, Z. Huang, L. Kofman, and D. Pogosyan, "Cosmological fluctuations from infrared cascading during inflation," *Physical Review D*, vol. 80, no. 4, Article ID 043501, 13 pages, 2009.
- [13] N. Barnaby and Z. Huang, "Particle production during inflation: observational constraints and signatures," *Physical Review D*, vol. 80, no. 12, Article ID 126018, 12 pages, 2009.
- [14] N. Barnaby, "On features and nongaussianity from inflationary particle production," <http://arxiv4.library.cornell.edu/abs/1006.4615>.
- [15] N. Barnaby, "Nongaussian signatures of inflationary particle production," *work in progress*.
- [16] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, "Gauge-invariant second-order perturbations and non-Gaussianity from inflation," *Nuclear Physics B*, vol. 667, no. 1–2, pp. 119–148, 2003.
- [17] J. Maldacena, "Non-gaussian features of primordial fluctuations in single field inflationary models," *Journal of High Energy Physics*, vol. 7, no. 5, pp. 233–264, 2003.
- [18] D. Seery and J. E. Lidsey, "Primordial non-Gaussianities in single-field inflation," *Journal of Cosmology and Astroparticle Physics*, no. 6, article no. 003, pp. 45–70, 2005.
- [19] N. Barnaby and J. M. Cline, "Non-Gaussian and nonscale-invariant perturbations from tachyonic preheating in hybrid inflation," *Physical Review D*, vol. 73, no. 10, Article ID 106012, 29 pages, 2006.
- [20] N. Barnaby and J. M. Cline, "Nongaussianity from tachyonic preheating in hybrid inflation," *Physical Review D*, vol. 75, no. 8, Article ID 086004, 14 pages, 2007.
- [21] J. R. Bond, A. V. Frolov, Z. Huang, and L. Kofman, "Non-gaussian curvature spikes from chaotic billiards in inflation preheating," *Physical Review Letters*, vol. 103, no. 7, Article ID 071301, 4 pages, 2009.
- [22] C. T. Byrnes, "Constraints on generating the primordial curvature perturbation and non-Gaussianity from instant preheating," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 1, article no. 011, 2009.

- [23] D. Mulryne, D. Seery, and D. Wesley, “Non-Gaussianity constrains hybrid inflation,” <http://arxiv.org/abs/0911.3550>.
- [24] M. Sasaki, “Multi-brid inflation and non-gaussianity,” *Progress of Theoretical Physics*, vol. 120, no. 1, pp. 159–174, 2008.
- [25] A. Naruko and M. Sasaki, “Large non-gaussianity from multi-brid inflation,” *Progress of Theoretical Physics*, vol. 121, no. 1, pp. 193–210, 2009.
- [26] X. Chen, M.-X. Huang, S. Kachru, and G. Shiu, “Observational signatures and non-Gaussianities of general single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 1, article no. 002, 2007.
- [27] M. LoVerde, A. Miller, S. Shandera, and L. Verde, “Effects of scale-dependent non-Gaussianity on cosmological structures,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article no. 014, 2008.
- [28] J. Kumar, L. Leblond, and A. Rajaraman, “Scale dependent local non-gaussianity from loops,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 4, article no. 024, 2010.
- [29] C. T. Byrnes, S. Nurmi, G. Tasinato, and D. Wands, “Scale dependence of local  $f_{NL}$ ,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 2, article no. 034, 2010.
- [30] E. Sefusatti, M. Liguori, A. P. S. Yadav, M. G. Jackson, and E. Pajer, “Constraining running non-gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article no. 022, 2009.
- [31] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity in the curvaton scenario,” *Physical Review D*, vol. 69, no. 4, Article ID 043503, 7 pages, 2004.
- [32] K. Enqvist and S. Nurmi, “Non-Gaussianity in curvaton models with nearly quadratic potentials,” *Journal of Cosmology and Astroparticle Physics*, no. 10, article no. 013, pp. 163–173, 2005.
- [33] K. A. Malik and D. H. Lyth, “A numerical study of non-Gaussianity in the curvaton scenario,” *Journal of Cosmology and Astroparticle Physics*, no. 9, article no. 008, 2006.
- [34] M. Sasaki, J. Valiviita, and D. Wands, “Nongaussianity of the primordial perturbation in the curvaton model,” *Physical Review D*, vol. 74, no. 10, Article ID 103003, 16 pages, 2006.
- [35] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. Van Tent, “Large non-Gaussianity in multiple-field inflation,” *Physical Review D*, vol. 73, no. 8, Article ID 083522, pp. 1–19, 2006.
- [36] F. Vernizzi and D. Wands, “Non-Gaussianities in twofield inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article no. 019, 2006.
- [37] C. T. Byrnes, K.-Y. Choi, and L. M. H. Hall, “Conditions for large non-Gaussianity in two-field slow-roll inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 10, article no. 008, 2008.
- [38] C. T. Byrnes and G. Tasinato, “Non-gaussianity beyond slow roll in multi-field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article no. 016, 2009.
- [39] X. Chen and Y. Wang, “Quasi-single field inflation and non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 4, article no. 027, 2010.
- [40] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $f_{NL}^{local}$  from WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article no. 006, 2009.
- [41] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, “Constraints on local primordial non-Gaussianity from large scale structure,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article no. 031, 2008.
- [42] N. Barnaby, T. Biswas, and J. M. Cline, “ $p$ -adic inflation,” *Journal of High Energy Physics*, vol. 2007, no. 4, article no. 056, 2007.
- [43] N. Barnaby and J. M. Cline, “Large non-Gaussianity from non-local inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 7, article no. 017, 2007.
- [44] N. Barnaby and J. M. Cline, “Predictions for nongaussianity from nonlocal inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 6, article no. 030, 2008.
- [45] C. Cheung, A. L. Fitzpatrick, J. Kaplan, and L. Senatore, “On the consistency relation of the 3-point function in single field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 2, article no. 021, 2008.
- [46] C. Cheung, A. L. Fitzpatrick, J. Kaplan, L. Senatore, and P. Creminelli, “The effective field theory of inflation,” *Journal of High Energy Physics*, vol. 2008, no. 3, article no. 014, 2008.
- [47] N. Barnaby and N. Kamran, “Dynamics with infinitely many derivatives: the initial value problem,” *Journal of High Energy Physics*, vol. 2008, no. 2, article no. 008, 2008.
- [48] N. Barnaby and N. Kamran, “Dynamics with infinitely many derivatives: variable coefficient equations,” *Journal of High Energy Physics*, vol. 2008, no. 12, article no. 022, 2008.
- [49] N. Barnaby, J. D. Mulryne, J. N. Nunes, and P. Robinson, “Dynamics and stability of light-like tachyon condensation,” *Journal of High Energy Physics*, vol. 2009, no. 3, article no. 018, 2009.
- [50] N. Barnaby, “Nonlocal inflation,” *Canadian Journal of Physics*, vol. 87, no. 3, pp. 189–194, 2009.
- [51] E. Silverstein and D. Tong, “Scalar speed limits and cosmology: acceleration from D-cceleration,” *Physical Review D*, vol. 70, no. 10, Article ID 103505, 18 pages, 2004.
- [52] M. Alishahiha, E. Silverstein, and D. Tong, “DBI in the sky: non-Gaussianity from inflation with a speed limit,” *Physical Review D*, vol. 70, no. 12, Article ID 123505, 15 pages, 2004.
- [53] A. J. Tolley and M. Wyman, “Equilateral non-Gaussianity from multifield dynamics,” *Physical Review D*, vol. 81, no. 4, article no. 043502, 2010.
- [54] D. Green, B. Horn, L. Senatore, and E. Silverstein, “Trapped inflation,” *Physical Review D*, vol. 80, no. 6, Article ID 063533, 2009.
- [55] L. Senatore, K. M. Smith, and M. Zaldarriaga, “Non-Gaussianities in single field inflation and their optimal limits from the WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 1, article no. 028, 2010.
- [56] P. D. Meerburg, J. P. Van Der Schaar, and P. S. Corasaniti, “Signatures of initial state modifications on bispectrum statistics,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article no. 018, 2009.
- [57] P. D. Meerburg, J. P. Van Der Schaar, and M. G. Jackson, “Bispectrum signatures of a modified vacuum in single field inflation with a small speed of sound,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 2, article no. 001, 2010.
- [58] R. Holman and A. J. Tolley, “Enhanced non-Gaussianity from excited initial states,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 5, article no. 001, 2008.
- [59] X. Chen, R. Easther, and E. A. Lim, “Large non-Gaussianities in single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 6, article no. 023, 2007.
- [60] X. Chen, R. Easther, and E. A. Lim, “Generation and characterization of large non-Gaussianities in single field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article no. 010, 2008.

- [61] D. J. H. Chung, E. W. Kolb, A. Riotto, and I. I. Tkachev, "Probing Planckian physics: resonant production of particles during inflation and features in the primordial power spectrum," *Physical Review D*, vol. 62, no. 4, Article ID 043508, pp. 1–13, 2000.
- [62] G. J. Mathews, D. J.H. Chung, K. Ichiki, T. Kajino, and M. Orito, "Constraints on resonant particle production during inflation from the matter and CMB power spectra," *Physical Review D*, vol. 70, no. 8, Article ID 083505, 2004.
- [63] Ø. Elgarøy, S. Hannestad, and T. Haugbølle, "Observational constraints on particle production during inflation," *Journal of Cosmology and Astroparticle Physics*, no. 9, article no. 008, pp. 87–111, 2003.
- [64] A. E. Romano and M. Sasaki, "Effects of particle production during inflation," *Physical Review D*, vol. 78, no. 10, Article ID 103522, 6 pages, 2008.
- [65] D. Langlois and L. Sorbo, "Primordial perturbations and non-Gaussianities from modulated trapping," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article no. 014, 2009.
- [66] P. Brax and E. Cluzel, "Brane bremsstrahlung in DBI inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 3, article no. 016, 2010.
- [67] L. A. Kofman and A. D. Linde, "Generation of density perturbations in inflationary cosmology," *Nuclear Physics, Section B*, vol. 282, pp. 555–588, 1987.
- [68] L. A. Kofman and D. YU. Pogosyan, "Nonflat perturbations in inflationary cosmology," *Physics Letters B*, vol. 214, no. 4, pp. 508–514, 1988.
- [69] D. S. Salopek, J. R. Bond, and J. M. Bardeen, "Designing density fluctuation spectra in inflation," *Physical Review D*, vol. 40, no. 6, pp. 1753–1788, 1989.
- [70] J. Adams, B. Cresswell, and R. Easther, "Inflationary perturbations from a potential with a step," *Physical Review D*, vol. 64, no. 12, Article ID 123514, 2001.
- [71] P. Hunt and S. Sarkar, "Multiple inflation and the WMAP "glitches";" *Physical Review D*, vol. 70, no. 10, Article ID 103518, 2004.
- [72] P. Hunt and S. Sarkar, "Multiple inflation and the WMAP "glitches". II. Data analysis and cosmological parameter extraction," *Physical Review D*, vol. 76, no. 12, Article ID 123504, 21 pages, 2007.
- [73] M. J. Mortonson, C. Dvorkin, H. V. Peiris, and W. Hu, "CMB polarization features from inflation versus reionization," *Physical Review D*, vol. 79, no. 10, Article ID 103519, 15 pages, 2009.
- [74] D. Battefeld, T. Battefeld, H. Firouzjahi, and N. Khosravi, "Brane annihilations during inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 07, article no. 009, 2010.
- [75] M. M. Anber and L. Sorbo, "Naturally inflating on steep potentials through electromagnetic dissipation," *Physical Review D*, vol. 81, no. 4, Article ID 043534, 7 pages, 2010.
- [76] A. Berera and T. W. Kephart, "Ubiquitous inflaton in string-inspired models," *Physical Review Letters*, vol. 83, no. 6, pp. 1084–1087, 1999.
- [77] E. Silverstein and A. Westphal, "Monodromy in the CMB: gravity waves and string inflation," *Physical Review D*, vol. 78, no. 10, Article ID 106003, 21 pages, 2008.
- [78] L. McAllister, E. Silverstein, and A. Westphal, "Gravity waves and linear inflation from axion monodromy," *Physical Review D*, vol. 82, Article ID 046003, 19 pages, 2010.
- [79] R. Flauger, L. McAllister, E. Pajer, A. Westphal, and G. Xu, "Oscillations in the CMB from axion monodromy inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 06, article no. 009, 2010.
- [80] L. Kofman, A. Linde, X. Liu, A. Maloney, L. McAllister, and E. Silverstein, "Beauty is attractive: moduli trapping at enhanced symmetry points," *Journal of High Energy Physics*, vol. 8, no. 5, pp. 671–711, 2004.
- [81] D. Battefeld and T. Battefeld, "A terminal velocity on the landscape: particle production near extra species loci in higher dimensions," *Journal of High Energy Physics*, vol. 2010, no. 7, pp. 1–40, 2010.
- [82] A. Berera, "Warm inflation," *Physical Review Letters*, vol. 75, no. 18, pp. 3218–3221, 1995.
- [83] K. Freese, J. A. Frieman, and A. V. Olinto, "Natural inflation with pseudo Nambu-Goldstone bosons," *Physical Review Letters*, vol. 65, no. 26, pp. 3233–3236, 1990.
- [84] F. C. Adams, J. R. Bond, K. Freese, J. A. Frieman, and A. V. Olinto, "Natural inflation: particle physics models, power-law spectra for large-scale structure, and constraints from the Cosmic Background Explorer," *Physical Review D*, vol. 47, no. 2, pp. 426–455, 1993.
- [85] C. P. Burgess, J. M. Cline, F. Lemieux, and R. Holman, "Decoupling, trans-Planckia and inflation," <http://arxiv.org/abs/astro-ph/0306236>.
- [86] C. P. Burgess, J. M. Cline, and R. Holman, "Effective field theories and inflation," *Journal of Cosmology and Astroparticle Physics*, no. 10, article no. 004, pp. 63–80, 2003.
- [87] C. P. Burgess, J. M. Cline, F. Lemieux, and R. Holman, "Are inflationary predictions sensitive to very high energy physics?" *Journal of High Energy Physics*, vol. 7, no. 2, pp. 1091–1115, 2003.
- [88] J. Berges and J. Serreau, "Progress in nonequilibrium quantum field theory," <http://arxiv.org/abs/hep-ph/0302210>.
- [89] J. Berges and J. Serreau, "Progress in nonequilibrium quantum field theory. II," <http://arxiv.org/abs/hep-ph/0410330>.
- [90] J. Berges and S. Borsányi, "Progress in nonequilibrium quantum field theory III," *Nuclear Physics A*, vol. 785, no. 1–2, pp. 58–67, 2007.
- [91] L. Kofman, A. Linde, and A. A. Starobinsky, "Reheating after inflation," *Physical Review Letters*, vol. 73, no. 24, pp. 3195–3198, 1994.
- [92] L. Kofman, A. Linde, and A. A. Starobinsky, "Towards the theory of reheating after inflation," *Physical Review D*, vol. 56, no. 6, pp. 3258–3295, 1997.
- [93] G. Felder, "Development of equilibrium after preheating," *Physical Review D*, vol. 63, no. 10, Article ID 103503, 14 pages, 2001.
- [94] R. Micha and I. I. Tkachev, "Relativistic turbulence: a long way from preheating to equilibrium," *Physical Review Letters*, vol. 90, no. 12, Article ID 121301, pp. 1–4, 2003.
- [95] R. Micha and I. I. Tkachev, "Turbulent thermalization," *Physical Review D*, vol. 70, no. 4 A, Article ID 043538, 25 pages, 2004.
- [96] J. Berges, A. Rothkopf, and J. Schmidt, "Non-thermal fixed points: effective weak-coupling for strongly correlated systems far from equilibrium," *Physical Review Letters*, vol. 101, no. 4, Article ID 041603, 4 pages, 2008.
- [97] J. Berges and G. Hoffmeister, "Nonthermal fixed points and the functional renormalization group," *Nuclear Physics B*, vol. 813, no. 3, pp. 383–407, 2009.
- [98] S. Hannestad, "Reconstructing the primordial power spectrum—a new algorithm," *Journal of Cosmology and Astroparticle Physics*, no. 4, article no. 002, pp. 19–37, 2004.

- [99] A. Shafieloo and T. Souradeep, "Primordial power spectrum from WMAP," *Physical Review D*, vol. 70, no. 4, Article ID 043523, 14 pages, 2004.
- [100] A. Shafieloo, T. Souradeep, P. Manimaran, P. K. Panigrahi, and R. Rangarajan, "Features in the primordial spectrum from WMAP: a wavelet analysis," *Physical Review D*, vol. 75, no. 12, Article ID 123502, 7 pages, 2007.
- [101] L. Covi, J. Hamann, A. Melchiorri, A. Slosar, and I. Sorbera, "Inflation and WMAP three year data: features have a future!," *Physical Review D*, vol. 74, no. 8, 8 pages, 2006.
- [102] J. Hamann, L. Covi, A. Melchiorri, and A. Slosar, "New constraints on oscillations in the primordial spectrum of inflationary perturbations," *Physical Review D*, vol. 76, no. 2, Article ID 023503, 14 pages, 2007.
- [103] P. Mukherjee and Y. Wang, "Model-independent reconstruction of the primordial power spectrum from Wilkinson microwave anisotropy probe data," *Astrophysical Journal*, vol. 599, no. 1, pp. 1–6, 2003.
- [104] N. Kogo, M. Matsumiya, M. Sasaki, and J. Yokoyama, "Reconstructing the primordial spectrum from WMAP data by the cosmic inversion method," *Astrophysical Journal*, vol. 607, no. 1, pp. 32–39, 2004.
- [105] R. Nagata and J. Yokoyama, "Reconstruction of the primordial fluctuation spectrum from the five-year WMAP data by the cosmic inversion method with band-power decorrelation analysis," *Physical Review D*, vol. 78, no. 12, Article ID 123002, 9 pages, 2008.
- [106] R. Nagata and J. Yokoyama, "Band-power reconstruction of the primordial fluctuation spectrum by the maximum likelihood reconstruction method," *Physical Review D*, vol. 79, no. 4, Article ID 043010, 9 pages, 2009.
- [107] J. M. Cline and L. Hoi, "Inflationary potential reconstruction for a WMAP running power spectrum," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 6, article no. 007, 2006.
- [108] L. Hoi and J. M. Cline, "Testing for features in the primordial power spectrum," *International Journal of Modern Physics D*, vol. 18, no. 2, pp. 1863–1888, 2009.
- [109] G. Nicholson and C. R. Contaldi, "Reconstruction of the primordial power spectrum using temperature and polarisation data from multiple experiments," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 7, article no. 011, 2009.
- [110] K. Ichiki, R. Nagata, and J. Yokoyama, "Cosmic discordance: detection of a modulation in the primordial fluctuation spectrum," *Physical Review D*, vol. 81, no. 8, Article ID 083010, 6 pages, 2010.
- [111] J. Hamann, A. Shafieloo, and T. Souradeep, "Features in the primordial power spectrum? A frequentist analysis," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 4, article no. 010, 2010.
- [112] Z. Huang, "HLattice: a new code for lattice cosmology," *work in progress*.
- [113] N. Barnaby, J. R. Bond, Z. Huang, and L. Kofman, "Preheating after modular inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article no. 021, 2009.
- [114] A. V. Frolov, "DEFROST: a new code for simulating preheating after inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 11, article no. 009, 2008.
- [115] G. Felder and I. Tkachev, "LATTICEASY: a program for lattice simulations of scalar fields in an expanding universe," *Computer Physics Communications*, vol. 178, no. 12, pp. 929–932, 2008.
- [116] E. J. Copeland, A. R. Liddle, D. H. Lyth, E. D. Stewart, and D. Wands, "False vacuum inflation with Einstein gravity," *Physical Review D*, vol. 49, no. 12, pp. 6410–6433, 1994.
- [117] M. Dine, L. Randall, and S. Thomas, "Supersymmetry breaking in the early Universe," *Physical Review Letters*, vol. 75, no. 3, pp. 398–401, 1995.
- [118] L. Kofman and S. Mukohyama, "Rapid roll inflation with conformal coupling," *Physical Review D*, vol. 77, no. 4, Article ID 043519, 9 pages, 2008.
- [119] D. Seery, K. A. Malik, and D. H. Lyth, "Non-Gaussianity of inflationary field perturbations from the field equation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 3, article no. 014, 2008.
- [120] K. A. Malik, "A not so short note on the Klein-Gordon equation at second order," *Journal of Cosmology and Astroparticle Physics*, no. 3, article no. 004, 2007.
- [121] K. A. Malik, "Gauge-invariant perturbations at second order: multiple scalar fields on large scales," *Journal of Cosmology and Astroparticle Physics*, no. 11, pp. 87–109, 2005.
- [122] M. Sasaki, "Large scale quantum fluctuations in the inflationary universe," *Progress of Theoretical Physics*, vol. 76, no. 5, pp. 1036–1046, 1986.
- [123] A. Riotto, "Inflation and the theory of cosmological perturbations," <http://arxiv.org/abs/hep-ph/0210162>.
- [124] D. Langlois and F. Vernizzi, "Non-linear perturbations of cosmological scalar fields," *Journal of Cosmology and Astroparticle Physics*, no. 2, article no. 017, 2007.
- [125] D. Seery, "One-loop corrections to a scalar field during inflation," *Journal of Cosmology and Astroparticle Physics*, no. 11, article no. 025, pp. 1–32, 2007.
- [126] D. Seery, "One-loop corrections to the curvature perturbation from inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 2, article no. 006, 2008.
- [127] K. A. Malik and D. Wands, "Evolution of second-order cosmological perturbations," *Classical and Quantum Gravity*, vol. 21, no. 11, pp. L65–L71, 2004.
- [128] D. H. Lyth and Y. Rodríguez, "Non-Gaussianity from the second-order cosmological perturbation," *Physical Review D*, vol. 71, no. 12, Article ID 123508, 14 pages, 2005.
- [129] D. H. Lyth, K. A. Malik, and M. Sasaki, "A general proof of the conservation of the curvature perturbation," *Journal of Cosmology and Astroparticle Physics*, no. 5, pp. 50–69, 2005.
- [130] S. M. Leach and A. R. Liddle, "Inflationary perturbations near horizon crossing," *Physical Review D*, vol. 63, no. 4, Article ID 043508, 7 pages, 2001.
- [131] S. M. Leach, M. Sasaki, D. Wands, and A. R. Liddle, "Enhancement of superhorizon scale inflationary curvature perturbations," *Physical Review D*, vol. 64, no. 2, Article ID 023512, 5 pages, 2001.
- [132] G. Hinshaw, "Five-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: data processing, sky maps, & basic results," *The Astrophysical Journal*, vol. 180, p. 225, 2009.
- [133] A. Lewis and S. Bridle, "Cosmological parameters from CMB and other data: a Monte Carlo approach," *Physical Review D*, vol. 66, no. 10, Article ID 103511, 16 pages, 2002.
- [134] E. Komatsu, J. Dunkley, M. R. Nolte et al., "Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation," *Astrophysical Journal*, vol. 180, no. 2, pp. 330–376, 2009.
- [135] W. C. Jones, P. A. R. Ade, J. J. Bock et al., "Observations of the temperature and polarization anisotropies with Boomerang 2003," *New Astronomy Reviews*, vol. 50, no. 11–12, pp. 945–950, 2006.

- [136] F. Piacentini, "A measurement of the polarization temperature angular cross power spectrum of the Cosmic Microwave Background from the 2003 flight of BOOMERANG," *The Astrophysical Journal*, vol. 647, p. 833, 2006.
- [137] T. E. Montroy, et al., "A Measurement of the CMB Spectrum from the 2003 Flight of BOOMERANG," *The Astrophysical Journal*, vol. 647, p. 813, 2006.
- [138] M. C. Runyan, P. A. R. Ade, R. S. Bhatia et al., "ACBAR: the arcminute cosmology bolometer array receiver," *Astrophysical Journal*, vol. 149, no. 2, supplement, pp. 265–287, 2003.
- [139] J. H. Goldstein, P. A. R. Ade, J. J. Bock et al., "Estimates of cosmological parameters using the cosmic microwave background angular power spectrum of ACBAR," *Astrophysical Journal*, vol. 599, no. 2, pp. 773–785, 2003.
- [140] C. L. Kuo, P. A. R. Ade, J. J. Bock et al., "Improved measurements of the CMB power spectrum with ACBAR," *Astrophysical Journal*, vol. 664, no. 2 I, pp. 687–701, 2007.
- [141] C. L. Reichardt, P. A. R. Ade, J. J. Bock et al., "High-resolution CMB power spectrum from the complete ACBAR data set," *Astrophysical Journal*, vol. 694, no. 2, pp. 1200–1219, 2009.
- [142] T. J. Pearson, B. S. Mason, A. C. S. Readhead et al., "The anisotropy of the microwave background to  $l = 3500$ : mosaic observations with the cosmic background imager," *Astrophysical Journal*, vol. 591, no. 2, pp. 556–574, 2003.
- [143] A. C. S. Readhead, B. S. Mason, C. R. Contaldi et al., "Extended mosaic observations with the cosmic background imager," *Astrophysical Journal*, vol. 609, no. 2, pp. 498–512, 2004.
- [144] A. C. S. Readhead, S. T. Myers, T. J. Pearson et al., "Polarization observations with the cosmic background imager," *Science*, vol. 306, no. 5697, pp. 836–844, 2004.
- [145] J. L. Sievers, C. Achermann, J. R. Bond et al., "Implications of the cosmic background imager polarization data," *Astrophysical Journal*, vol. 660, no. 2, pp. 976–987, 2007.
- [146] C. Dickinson, R. A. Battye, P. Carreira et al., "High-sensitivity measurements of the cosmic microwave background power spectrum with the extended Very Small Array," *Monthly Notices of the Royal Astronomical Society*, vol. 353, no. 3, pp. 732–746, 2004.
- [147] N. W. Halverson, E. M. Leitch, C. Pryke et al., "Degree angular scale interferometer first results: a measurement of the cosmic microwave background angular power spectrum," *Astrophysical Journal*, vol. 568, no. 1, pp. 38–45, 2002.
- [148] E. M. Leitch, J. M. Kovac, N. W. Halverson, J. E. Carlstrom, C. Pryke, and M. W. E. Smith, "Degree angular scale interferometer 3 year cosmic microwave background polarization results," *Astrophysical Journal*, vol. 624, no. 1, pp. 10–20, 2005.
- [149] S. Hanany, et al., "MAXIMA-1: a measurement of the Cosmic Microwave Background Anisotropy on angular scales of 10 arcminutes to 5 degrees," *The Astrophysical Journal*, vol. 545, no. 1, p. L5, 2000.
- [150] R. A. Sunyaev and Y. B. Zeldovich, "The observations of relic radiation as a test of the nature of X-ray radiation from the clusters of galaxies," *Comments on Astrophysics and Space Physics*, vol. 4, p. 173, 1972.
- [151] R. A. Sunyaev and Y. B. Zeldovich, "Microwave background radiation as a probe of the contemporary structure and history of the universe," *Annual Review of Astronomy and Astrophysics*, vol. 18, pp. 537–560, 1980.
- [152] J. R. Bond, et al., "The Sunyaev-Zeldovich effect in CMB-calibrated theories applied to the Cosmic Background Imager anisotropy power at  $l > 2000$ ," *The Astrophysical Journal*, vol. 626, no. 1, article no. 12, 2005.
- [153] M. Kowalski, et al., "Improved cosmological constraints from new, old and combined supernova datasets," *The Astrophysical Journal*, vol. 686, no. 2, p. 749, 2008.
- [154] S. Cole, W. J. Percival, J. A. Peacock, et al., "The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final dataset and cosmological implications," *Monthly Notices of the Royal Astronomical Society*, vol. 362, no. 2, pp. 505–534, 2005.
- [155] M. Tegmark, D. J. Eisenstein, M. A. Strauss et al., "Cosmological constraints from the SDSS luminous red galaxies," *Physical Review D*, vol. 74, no. 12, Article ID 123507, 34 pages, 2006.
- [156] R. Massey, J. Rhodes, A. Leauthaud et al., "COSMOS: three-dimensional weak lensing and the growth of structure," *Astrophysical Journal*, vol. 172, no. 1, supplement, pp. 239–253, 2007.
- [157] J. Lesgourgues, M. Viel, M. G. Haehnelt, and R. Massey, "A combined analysis of Lyman-alpha forest, 3D Weak Lensing and WMAP year three data," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, article no. 008, 2007.
- [158] H. Hoekstra, et al., "First cosmic shear results from the Canada-France-Hawaii telescope wide synoptic legacy survey," *The Astrophysical Journal*, vol. 647, no. 1, p. 116, 2006.
- [159] C. Schimd, I. Tereno, J.-P. Uzan et al., "Tracking quintessence by cosmic shear Constraints from VIRMOS-Descart and CFHTLS and future prospects," *Astronomy and Astrophysics*, vol. 463, no. 2, pp. 405–421, 2007.
- [160] H. Hoekstra, H. K. C. Yee, M. D. Gladders, L. Felipe Barriento, P. B. Hall, and L. Infante, "A measurement of weak lensing by large-scale structure in red-sequence cluster survey fields," *Astrophysical Journal*, vol. 572, no. 1, pp. 55–65, 2002.
- [161] H. Hoekstra, H. K. C. Yee, and M. D. Gladders, "Constraints on omega m and sigma 8 from weak lensing in RCS fields," *The Astrophysical Journal*, vol. 577, p. 595, 2002.
- [162] L. Van Waerbeke, Y. Mellier, and H. Hoekstra, "Dealing with systematics in cosmic shear studies: new results from the VIRMOS-Descart Survey," *Astronomy and Astrophysics*, vol. 429, p. 75, 2005.
- [163] J. Benjamin, C. Heymans, and E. Semboloni, "Cosmological constraints from the 100-deg<sup>2</sup> weak-lensing survey," *Monthly Notices of the Royal Astronomical Society*, vol. 381, no. 2, pp. 702–712, 2007.
- [164] S. De Lope Amigo, W. Y. Cheung, Z. Huang, and S. P. Ng, "Cosmological constraints on decaying dark matter," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 6, article no. 005, 2009.
- [165] V. Springel, et al., "Simulating the joint evolution of quasars, galaxies and their large-scale distribution," *Nature*, vol. 435, p. 629, 2005.
- [166] N. Dalal, O. Dore, D. Huterer, and A. Shirokov, "The imprints of primordial non-gaussianities on large-scale structure: scale dependent bias and abundance of virialized objects," *Physical Review D*, vol. 77, Article ID 123514, 2008.
- [167] P. McDonald, "Primordial non-Gaussianity: large-scale structure signature in the perturbative bias model," *Physical Review D*, vol. 78, Article ID 123519, 2008.
- [168] N. Afshordi and A. J. Tolley, "Primordial nongaussianity, statistics of collapsed objects, and the Integrated Sachs-Wolfe effect," *Physical Review D*, vol. 78, Article ID 123507, 2008.
- [169] J. J. Blanco-Pillado, et al., "Racetrack inflation," *Journal of High Energy Physics*, vol. 2004, no. 11, article no. 063, 2004.
- [170] J. P. Conlon and F. Quevedo, "Kaehler moduli inflation," *Journal of High Energy Physics*, vol. 2006, no. 1, article no. 146, 2006.

- [171] J. R. Bond, L. Kofman, S. Prokushkin, and P. M. Vaudrevange, "Roulette inflation with Kaehler moduli and their axions," *Physical Review D*, vol. 75, Article ID 123511, 2007.
- [172] M. Cicoli, C. P. Burgess, and F. Quevedo, "Fibre inflation: observable gravity waves from IIB string compactifications," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 3, article no. 013, 2009.
- [173] G. R. Dvali and S. H. H. Tye, "Brane inflation," *Physics Letters B*, vol. 450, no. 72, 1999.
- [174] C. P. Burgess, M. Majumdar, D. Nolte, F. Quevedo, G. Rajesh, and R. J. Zhang, "The inflationary brane-antibrane universe," *Journal of High Energy Physics*, vol. 2001, no. 1, article no. 048, 2001.
- [175] S. Kachru, R. Kallosh, A. Linde, J. M. Maldacena, L. P. McAllister, and S. P. Trivedi, "Towards inflation in string theory," *Journal of Cosmology and Astroparticle Physics*, vol. 2003, no. 10, article no. 013, 2003.
- [176] K. Dasgupta, C. Herdeiro, S. Hirano, and R. Kallosh, "D3-D7 inflationary model and M theory," *Physical Review D*, vol. 65, no. 12, Article ID 126002, 2002.
- [177] R. H. Brandenberger, A. Knauf, and L. C. Lorenz, "Reheating in a brane monodromy inflation model," *Journal of High Energy Physics*, vol. 2008, no. 10, article no. 110, 2008.
- [178] J. R. Fergusson and E. P. S. Shellard, "Shape of primordial non-Gaussianity and the CMB bispectrum," *Physical Review D*, vol. 80, no. 4, Article ID 043510, 26 pages, 2009.
- [179] J. R. Fergusson, M. Liguori, and E. P. S. Shellard, "The CMB bispectrum," <http://arxiv.org/abs/1006.1642>.
- [180] D. Seery and J. C. Hidalgo, "Non-Gaussian corrections to the probability distribution of the curvature perturbation from inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 7, article no. 008, 2006.
- [181] S. Shandera, "The structure of correlation functions in single field inflaton," *Physical Review D*, vol. 79, Article ID 123518, 2009.
- [182] L. Leblond and S. Shandera, "Simple bounds from the perturbative regime of inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article no. 007, 2008.
- [183] A. Slosar, C. Hirata, U. Seljak, S. Ho, and N. Padmanabhan, "Constraints on local primordial non-Gaussianity from large scale structure," *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article no. 031, 2008.
- [184] R. Jimenez and L. Verde, "Implications for Primordial Non-Gaussianity  $f_{NL}$  from weak lensing masses of high- $z$  galaxy clusters," *Physical Review D*, vol. 80, Article ID 127302, 2009.
- [185] L. Verde and S. Matarrese, "Detectability of the effect of Inflationary non-Gaussianity on halo bias," *The Astrophysical Journal*, vol. 706, p. L91, 2009.
- [186] G. N. Felder and L. Kofman, "The development of equilibrium after preheating," *Physical Review D*, vol. 63, Article ID 103503, 2001.
- [187] G. N. Felder and O. Navros, "Inflaton fragmentation after  $\lambda\phi^4$  inflation," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 2, article no. 014, 2007.
- [188] M. J. Jee, et al., "Hubble space telescope weak-lensing study of the galaxy cluster XMMU J2235.3-2557 at  $z=1.4$ : a surprisingly massive galaxy cluster when the universe is one-third of its current age," *The Astrophysical Journal*, vol. 704, p. 672, 2009.
- [189] K. M. Smith and M. Zaldarriaga, "Algorithms for bispectra: forecasting, optimal analysis, and simulation," <http://arxiv.org/abs/astro-ph/0612571>.

## Review Article

# Ekpyrotic Nongaussianity: A Review

**Jean-Luc Lehnert**

*Princeton Center for Theoretical Science, Princeton University, Princeton, NJ 08544, USA*

Correspondence should be addressed to Jean-Luc Lehnert, [jlehnert@princeton.edu](mailto:jlehnert@princeton.edu)

Received 15 January 2010; Revised 6 May 2010; Accepted 9 June 2010

Academic Editor: Sarah Shandera

Copyright © 2010 Jean-Luc Lehnert. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ekpyrotic models and their cyclic extensions solve the standard cosmological flatness, horizon, and homogeneity puzzles by postulating a slowly contracting phase of the universe prior to the big bang. This ekpyrotic phase also manages to produce a nearly scale-invariant spectrum of scalar density fluctuations but, crucially, with significant nongaussian corrections. In fact, some versions of ekpyrosis are on the borderline of being ruled out by observations, while, interestingly, the best-motivated models predict levels of nongaussianity that will be measurable by near-future experiments. Here, we review these predictions in detail, and comment on their implications.

## 1. Motivation and Introduction

The standard big bang cosmology is hugely successful in describing the evolution of our universe from the time of nucleosynthesis onwards. However, a central assumption is that the universe started out in a hot big bang and in a special state: extrapolating back from current knowledge, we know that early on the universe must have been very flat, homogeneous, and isotropic, with in addition small density perturbations with a nearly scale-invariant spectrum and a nearly gaussian distribution. Hence, the “initial” state of the universe was far from random, and its specialness prompts us to try and explain it via a dynamical mechanism.

The most studied such mechanism is the model of inflation, which assumes that there was a phase of rapid, accelerated expansion preceding the hot big bang; for a comprehensive review see [1]. Such a phase can be modeled by having a scalar field (the “inflaton”) with a positive and suitably flat potential. Inflation has the property of flattening the universe, so that, if it lasts long enough, the flatness of the “initial” state can be explained. Moreover, inflation possesses the remarkable byproduct that it generates nearly scale-invariant spectra of scalar and tensor perturbations by amplifying quantum fluctuations. The predicted scalar perturbations are in excellent agreement with current observations, but the tensor perturbations have yet to be observed—their discovery would be a strong indication for

the correctness of the inflationary picture. However, inflation also presents a number of conceptual problems: for example, even though the inflationary phase is supposed to erase all memory of initial conditions, this is not really the case. In order for inflation to start in a given patch of space, that patch must be reasonably smooth over several Planck lengths and the inflaton field must have a small initial velocity (the “patch” and “overshoot” problems, respectively; see e.g., [1, 2]). Also, it has been realized not long ago that inflation is geodesically incomplete towards the past, which means that the predictions of the theory depend on the specification of data on a space-like initial hypersurface [3]. In other words, inflation requires its own initial conditions. Hence, if inflation is correct, it will only form a part of the story. More worrying is the problem of unpredictability, which is associated with the quantum nature and the effectiveness of inflation. Inflation ends when the inflaton field oscillates around a minimum of its potential and “reheats” the universe by decaying into standard model particles. However, for generic initial conditions there will always be regions in which rare but large quantum fluctuations kick the inflaton field back up its potential and keep a fraction of the universe in the inflationary phase. In most of the concrete realizations of inflation, the region that keeps inflating expands so fast that it quickly dominates the overall volume of the universe. Hence, inflation never ends, and the global picture of this process of “eternal inflation” is that of an empty de

Sitter universe punctured by an infinite number of small pockets where inflation has ended (at a random time) [4]. Because inflation ends at a random moment in these pocket universes, the pockets might have become sufficiently flattened or not, they might have acquired scale-invariant perturbations or not. Without a measure which would determine the relative likelihood of the various pockets, it becomes difficult to know exactly what eternal inflation predicts. These problems do not mean that the idea of inflation is wrong, but, if inflation continues to be supported by observations, they will have to be addressed. In the meantime, the seriousness of these open problems means that it is worthwhile considering alternative models for the early universe in parallel.

The present paper deals with one such model in particular, namely, the ekpyrotic model and its extension, the cyclic universe; for a comprehensive overview see [5]. In this model, the inflationary phase is replaced by the ekpyrotic phase, which is a slowly contracting phase preceding the big bang. The ekpyrotic phase can be modeled by having a scalar field with a negative and steep potential. As described in detail below, it also manages to flatten a given region of the universe and generates nearly scale-invariant scalar perturbations but no observable tensor fluctuations. At the linear level, the scalar fluctuations are virtually indistinguishable from the perturbations produced by inflation, but at higher orders the predictions differ. Since primordial gravitational waves might turn out to be rather elusive to measure over the coming years, the most promising way of distinguishing between alternative models of the early universe is therefore by studying these higher-order, nongaussian signatures.

There is a simply, intuitive argument for why the predictions regarding higher-order corrections to the linear perturbations should differ for models of inflation and ekpyrosis. For a scalar field fluctuation  $\delta\varphi$ , the semiclassical probability density is roughly given by  $e^{-S_E(\delta\varphi)}$ , where  $S_E(\delta\varphi)$  is the euclidean action [6]. Since inflation requires a very flat potential, the inflaton is an almost free field. For a free field, the action is quadratic in the field, and hence the probability distribution is simply a gaussian distribution. For an exact gaussian distribution the 3-point function  $\langle\delta\varphi^3\rangle$  vanishes, and hence for inflation, where the field is almost free, we would expect the 3-point function to be nonzero, but small. For ekpyrosis, on the other hand, the potential is steep, and hence the scalar field is necessarily significantly self-coupled. This has the consequence that ekpyrotic models generally predict significant levels of nongaussianity. In fact, some versions of ekpyrosis are already on the borderline of being ruled out by observations, while the best-motivated models predict values that are measurable by near-future experiments. Thus, the nongaussian predictions are crucial in assessing the viability of various cosmological models, and promise to significantly enhance our understanding of the physics of the early universe.

The plan of this paper is to start with a brief summary of the main ideas behind ekpyrotic and cyclic models of the universe. We will then discuss in some detail the generation of linear cosmological perturbations (a good understanding

of the linear perturbations greatly facilitates an understanding of the higher-order ones), before turning to the main subject of the paper, namely, the nongaussian corrections to these linear perturbations. We will conclude with a discussion of the nongaussian predictions and in particular their observability and relation to current observational limits, as well as the consequences of a potential detection.

## 2. Ekpyrotic and Cyclic Cosmology

The ekpyrotic phase is the cornerstone of ekpyrotic and cyclic models of the universe: it is a conjectured, slowly contracting phase preceding the big bang, and it resolves the standard cosmological puzzles [7, 8]. The main feature of ekpyrosis is that during this phase the equation of state

$$w \equiv \frac{p}{\rho} \gg 1 \quad (1)$$

is very large (here  $p$  and  $\rho$  denote the average pressure and energy density of the universe). Let us briefly explore the most direct consequences of such an ultrastiff equation of state. Consider a Friedmann-Robertson-Walker (FRW) metric. (I will mostly use natural units  $\hbar = c = 1$  and  $8\pi G = M_{pl}^{-2} = 1$ .)

$$ds^2 = -dt^2 + a(t)^2 \left( \frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega_2^2 \right), \quad (2)$$

where  $a(t)$  denotes the scale factor of the universe and  $\kappa = -1, 0, 1$  for an open, flat, or closed universe, respectively. If the universe is filled with a number of fluids interacting only via gravity and with energy densities  $\rho_i$  and constant equations of state  $w_i$ , then the equations of continuity

$$\dot{\rho}_i + 3 \frac{\dot{a}}{a} (\rho_i + p_i) = 0, \quad (3)$$

(where dots denote derivatives with respect to time  $t$ ) imply that they will evolve according to

$$\rho_i \propto a^{-3(1+w_i)}. \quad (4)$$

The Einstein equations for this system contain a constraint equation, better known as the Friedmann equation, which involves the Hubble parameter  $H \equiv \dot{a}/a$ :

$$H^2 = \frac{1}{3} \left( \frac{-3\kappa}{a^2} + \frac{\rho_{m,0}}{a^3} + \frac{\rho_{r,0}}{a^4} + \frac{\rho_{a,0}}{a^6} + \dots + \frac{\rho_{\phi,0}}{a^{3(1+w_\phi)}} \right). \quad (5)$$

The  $\rho_{i,0}$ 's are constants giving the energy densities at scale factor  $a = 1$  of the various constituents of the universe: we consider the universe to be composed of nonrelativistic matter (subscript  $m$ ), radiation ( $r$ ), and the energy density associated with anisotropies in the curvature of the universe ( $a$ ). In addition, we consider there to be ekpyrotic (scalar) matter, denoted by the subscript  $\phi$ , and, as usual, there is a contribution due to the average curvature of space.

As the universe contracts, components whose energy density scales with a higher negative power of the scale factor  $a$  will successively come to dominate, first matter, then

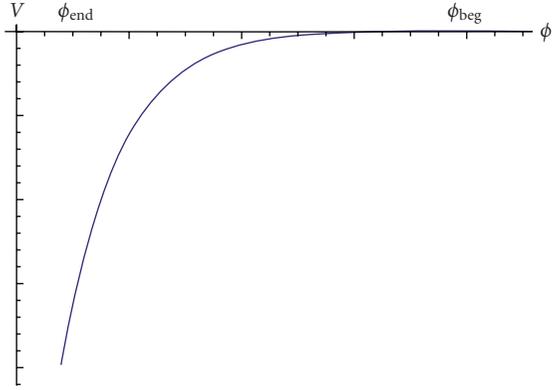


FIGURE 1: The potential during ekpyrosis is negative and steeply falling; it can be modeled by the exponential form  $V(\phi) = -V_0 e^{-c\phi}$ .

radiation, then anisotropies, and eventually, since  $w_\phi \gg 1$  by assumption, the ekpyrotic matter. This means that the relative energy densities in curvature and anisotropies, for example, become smaller and smaller, the longer the ekpyrotic contracting phase lasts. In other words, if ekpyrosis lasts long enough, the flatness problem is solved. We will make this statement quantitative below. Strictly speaking, for the flatness problem to be solved, all we need is a matter component with  $w > 1$ . In the next section, we will see that for realistic ekpyrotic models, typically  $w_\phi \gg 1$ . In passing, we should also point out that there is no horizon problem in ekpyrotic and cyclic models, as there is plenty of time before the big bang for different parts of our currently observable universe to have been in causal contact with each other.

But what form of matter can have the large equation of state that we require? A simple way to model the ekpyrotic matter is to have a scalar field  $\phi$  with a steep and negative potential  $V(\phi)$ . A concrete example is provided by the negative exponential

$$V(\phi) = -V_0 e^{-c\phi}, \quad (6)$$

where  $V_0$  and  $c$  are constants—see Figure 1. In the context of string theory, such scalar fields appear very naturally, and the ekpyrotic potential can then correspond to an attractive force between branes—this picture will be briefly described below.

Given an explicit form of the potential, such as (6), we can solve for the evolution of the universe. In fact it is straightforward to generalize the treatment to having many scalars  $\phi_i$  with potentials  $V_i(\phi_i)$ . Then, in a flat FRW background and neglecting other matter components, the equations of motion become

$$\begin{aligned} \ddot{\phi}_i + 3H\dot{\phi}_i + V_{i,\phi_i} &= 0, \\ H^2 &= \frac{1}{3} \left[ \frac{1}{2} \sum_i \dot{\phi}_i^2 + \sum_i V_i(\phi_i) \right], \end{aligned} \quad (7)$$

where  $V_{i,\phi_i} = \partial V_i / \partial \phi_i$  with no summation implied. If all the fields have negative exponential potentials  $V_i(\phi_i) = -V_i e^{-c_i \phi_i}$  and if  $c_i \gg 1$  for all  $i$ , then the Einstein-scalar equations admit the *scaling solution*

$$a = (-t)^{1/\epsilon}, \quad \phi_i = \frac{2}{c_i} \ln \left( -\sqrt{\frac{c_i^2 V_i}{2t}} \right), \quad \frac{1}{\epsilon} = \sum_i \frac{2}{c_i^2}. \quad (8)$$

Thus, we have a very slowly contracting universe with (constant) equation of state

$$w \equiv \frac{\sum_i (1/2) \dot{\phi}_i^2 - V_i(\phi_i)}{\sum_i (1/2) \dot{\phi}_i^2 + V_j(\phi_j)} = \frac{2\epsilon}{3} - 1 \gg 1. \quad (9)$$

We are using a coordinate system in which the big crunch occurs at  $t = 0$ ; in other words, the time coordinate is negative during the ekpyrotic phase. Here, the parameter  $\epsilon$  corresponds to the *fast-roll* parameter and is typically of  $\mathcal{O}(100)$ ; its definition is identical with that in inflation, where its value is typically of  $\mathcal{O}(1/100)$  and where, correspondingly, it is called the *slow-roll* parameter.

Using this explicit solution, we can get an idea for how long the ekpyrotic phase has to last in order for the flatness problem to be solved. Quantitatively, the problem can be formulated as follows: dividing the Friedmann equation (5) by  $H^2$  we can see that the fractional energy density stored in the average curvature of the universe is given by

$$\frac{\kappa}{(aH)^2}. \quad (10)$$

At the present time, observations imply that this quantity is smaller than  $10^{-2}$  in magnitude [9]. If we assume a radiation-dominated universe, which is a good approximation for this calculation, then  $aH \propto t^{-1/2}$ , and hence, if we extrapolate back to the Planck time, the fractional energy density in curvature must have been smaller than

$$\frac{t_{pl}}{t_0} 10^{-2} \approx 10^{-62}, \quad (11)$$

an incredibly small number. However, from (8), we can see that during the ekpyrotic phase the scale factor  $a$  remains almost constant, while the Hubble parameter  $H \propto t^{-1}$ . Hence  $aH$  grows by a factor of  $10^{30}$  as long as

$$|t_{ek-beg}| \geq e^{60} |t_{ek-end}|, \quad (12)$$

where the subscripts  $ek-beg$  and  $ek-end$  refer to the beginning and the end of the ekpyrotic phase, respectively. As will be discussed in the next section, we need  $t_{ek-end} \approx -10^3 M_{pl}^{-1}$  in order to obtain the observed amplitude of cosmological perturbations, so that we need

$$|t_{ek-beg}| \geq 10^{33} M_{pl}^{-1} = 10^{-10} \text{s}. \quad (13)$$

This is the minimum time the ekpyrotic phase has to last in order to solve the flatness problem. Cosmologically speaking, this is a very short time, attesting to the effectiveness of the ekpyrotic phase.

Before discussing the cosmological perturbations produced during the ekpyrotic phase, it is useful to provide a quick overview of how the ekpyrotic phase might fit into a more complete cosmological model. The crucial ingredient in any such model is the proposed mechanism for how the ekpyrotic contracting phase (with  $H < 0$ ) and the subsequent radiation-dominated expanding phases (with  $H > 0$ ) should link up. The Einstein equations provide the relation

$$\dot{H} = -\frac{1}{2}(\rho + p). \quad (14)$$

All forms of matter that are currently known to exist obey the *null energy condition*

$$\rho + p \geq 0(\text{NEC}), \quad (15)$$

which implies  $\dot{H} \leq 0$  and which thus precludes a smooth transition between a contracting and an expanding universe. This leaves two possibilities for achieving such a transition: either the NEC is violated during the transition, or the transition is classically singular.

In *new ekpyrotic* models [10–12], a smooth reversal from contraction to expansion is achieved by adding a further matter component to the universe which can violate the NEC. The particular example that these models consider is the so-called ghost condensate, which corresponds to the gravitational equivalent of a Higgs phase [13]. It is not clear yet whether or not the ghost condensate can be obtained from a fundamental theory such as string theory [14] (in the more restricted framework of quantum field theory it seems impossible to construct a stable ghost condensate model [15]); however, it is interesting that string theory contains many objects (orientifolds, negative-tension branes) which do violate the NEC. Of course, simply adding such a component is not enough: it must become relevant as the universe contracts, and vanish again as the universe expands. The simplest way to achieve this is by assuming that the ghost condensate itself also plays the role of the ekpyrotic matter, and that after the transition to expansion, it decays into ordinary matter fields. This scenario requires the ghost condensate to possess both a particular form for its kinetic term and a particular potential; for details regarding possible realizations see [10, 12].

The *cyclic* model of the universe [16, 17] is based on the braneworld picture of the universe, in which space time is effectively 5-dimensional, but with one dimension not extending indefinitely but being a line segment; see Figure 2. The endpoints of this line segment (orbifold) are two (3+1)-dimensional boundary branes. In the full string theory setup, there is in addition a 6-dimensional internal manifold at each point in the 5-dimensional space time, for a total of 11 dimensions [18]. This description of the universe stems from string theory and in particular the duality, known as *Hořava-Witten theory* [19], between 11-dimensional supergravity and the  $E_8 \times E_8$  heterotic string theory. All matter and forces, except for gravity, are localized on the branes, while gravity can propagate in the whole space time. Our universe, as we see it, is identified with one of the boundary branes and, as long as the branes are far apart, can interact with the other

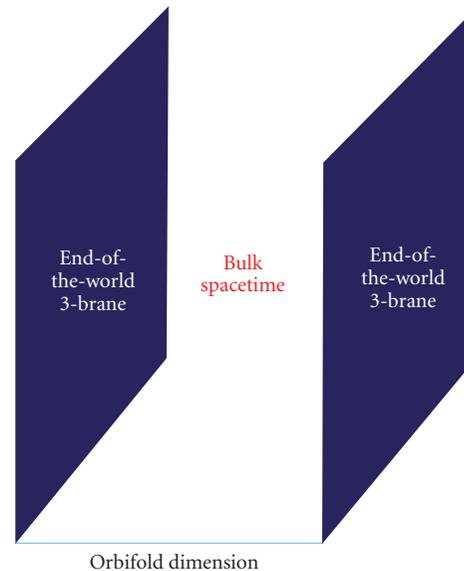


FIGURE 2: The braneworld picture of our universe. Think of a sandwich: the 5-dimensional bulk space time is bounded by two 4-dimensional boundary branes. There is no space “outside” of the sandwich, but the branes can be infinite in all directions perpendicular to the line segment (orbifold). In the M-theory embedding, there are 6 additional internal dimensions at each point of the sandwich.

brane only via gravity. The cyclic model assumes that there is an attractive force between the two branes, which causes the branes to approach each other. This force is modeled by a potential of the form shown in Figure 3. Note that the potential incorporates an ekpyrotic part. From the higher-dimensional point of view, the ekpyrotic phase has the rather nonintuitive property that it flattens the branes to a very high degree. Eventually the two branes collide and bounce off each other. It is this collision that, from the point of view of someone living on one of the branes, looks like the big bang. Classically, the collision is singular, as the orbifold dimension shrinks to zero size. The collision is slightly inelastic and produces matter and radiation on the branes, where the standard cosmological evolution now takes place. However, due to quantum fluctuations, the branes are slightly rippled and do not collide everywhere at exactly the same time. In some places, the branes collide slightly earlier, which means that the universe has a little bit more time to expand and cool. In other places, the collision takes place slightly later, and those regions remain a little hotter. This provides a heuristic picture of the way temperature fluctuations are naturally produced within the model. Shortly after the branes have separated, the distance between the boundary branes gets almost stabilized, but the branes start attracting each other again very slightly. This very slight attraction acts as quintessence and is identified with the dark energy observed in the universe. After a long time, and as the branes become closer again, they start attracting each other more strongly so that we get another ekpyrotic phase and eventually another brane collision with the creation of new matter. In this way,

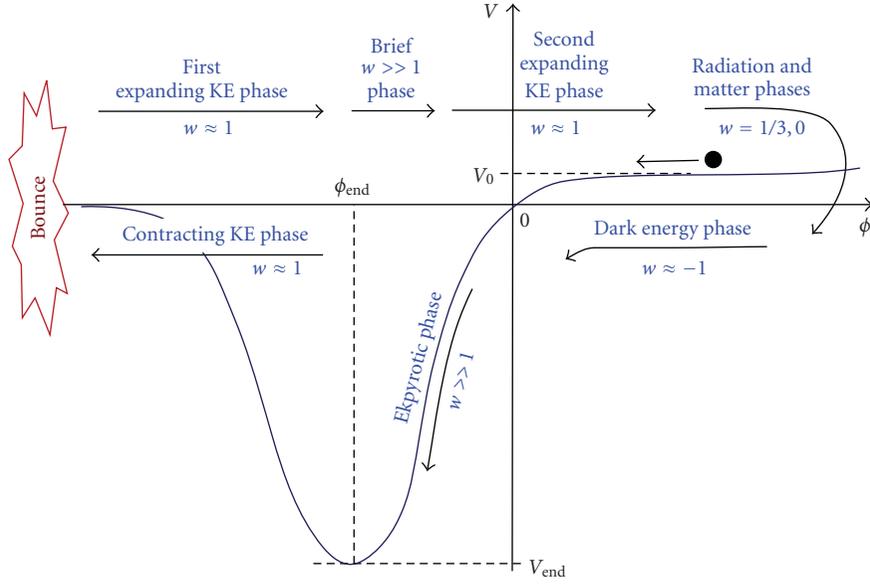


FIGURE 3: The potential for the cyclic universe integrates the ekpyrotic part and a quintessence epoch but is irrelevant at the brane collision. A possible form for the potential is  $V(\phi) = V_0(e^{b\phi} - e^{-c\phi})F(\phi)$ , with  $b \ll 1, c \gg 1$ , and  $F(\phi)$  tends to unity for  $\phi > \phi_{\text{end}}$  and to zero for  $\phi < \phi_{\text{end}}$ . Reproduced with permission from [8].

a cyclic model of the universe emerges. Before continuing, we should mention the main open issues related to the cyclic model: the first one concerns the potential, which at this point is simply conjectured. It will be important to see if a potential of the required shape can be derived from microphysics. And the second is the brane collision, which so far has been extensively studied at the classical and semiclassical level [20], but a full quantum treatment has remained elusive.

In the discussion above, we have mostly focussed on models involving one effective scalar field. However, there are two good reasons to extend the analysis to two or more scalars: first, in embedding the ekpyrotic and cyclic models in M-theory, there are two universal scalars, namely, the radion mode (which determines the distance between the branes) and the volume modulus of the internal 6-dimensional manifold [21]. There can be many more scalar fields (such as the shape moduli of the internal space), but we always must consider these two universal scalars. And secondly, as we will see in the next section, it is much more natural to generate a nearly scale-invariant spectrum of curvature perturbations (in agreement with observations) in models with two scalars than in models with only one. However, multifield ekpyrotic models present some qualitatively new features, which we discuss briefly here.

The 4-dimensional effective action

$$S = \int \sqrt{-g} \left[ R - \frac{1}{2}(\partial\phi_1)^2 - \frac{1}{2}(\partial\phi_2)^2 - V(\phi_1, \phi_2) \right] \quad (16)$$

can be obtained as the low-energy limit of Hořava-Witten theory, where  $\phi_1$  and  $\phi_2$  are related by a field redefinition to the radion and the internal volume modulus [22]. We are

assuming that during the ekpyrotic phase, both fields feel an ekpyrotic-type potential, for example,

$$V(\phi_1, \phi_2) = -V_1 e^{-c_1 \phi_1} - V_2 e^{-c_2 \phi_2}. \quad (17)$$

Then it is much more natural to discuss the dynamics in terms of the new variables  $\sigma$  and  $s$  pointing transverse and perpendicular to the field velocity, respectively [23, 24]; they are defined, up to unimportant additive constants which we will fix below, via

$$\sigma \equiv \frac{\dot{\phi}_1 \phi_1 + \dot{\phi}_2 \phi_2}{\dot{\sigma}}, \quad s \equiv \frac{\dot{\phi}_1 \phi_2 - \dot{\phi}_2 \phi_1}{\dot{\sigma}}, \quad (18)$$

with  $\dot{\sigma} \equiv (\dot{\phi}_1^2 + \dot{\phi}_2^2)^{1/2}$ . It is also useful to define the angle  $\theta$  of the trajectory in field space via [25]

$$\cos \theta = \frac{\dot{\phi}_1}{\dot{\sigma}}, \quad \sin \theta = \frac{\dot{\phi}_2}{\dot{\sigma}}. \quad (19)$$

In terms of these new variables, the potential can be re-expressed as

$$V_{ek} = -V_0 e^{\sqrt{2}\epsilon\sigma} \left[ 1 + \epsilon s^2 + \frac{\kappa_3}{3!} \epsilon^{3/2} s^3 + \frac{\kappa_4}{4!} \epsilon^2 s^4 + \dots \right], \quad (20)$$

where for exact exponentials of the form (17), one has  $\kappa_3 = 2\sqrt{2}(c_1^2 - c_2^2)/|c_1 c_2|$  and  $\kappa_4 = 4(c_1^6 + c_2^6)/(c_1^2 c_2^2 (c_1^2 + c_2^2))$ . However, in the absence of a microphysical derivation of the potential, we will simply take  $\kappa_3, \kappa_4 \sim \mathcal{O}(1)$  and express all results in terms of  $\kappa_3, \kappa_4$ . See also Figure 4 for an illustration of the potential. The ekpyrotic scaling solution becomes

$$a(t) = (-t)^{1/\epsilon}, \quad \sigma = -\sqrt{\frac{2}{\epsilon}} \ln \left( -\sqrt{\epsilon V_0 t} \right), \quad s = 0, \quad (21)$$

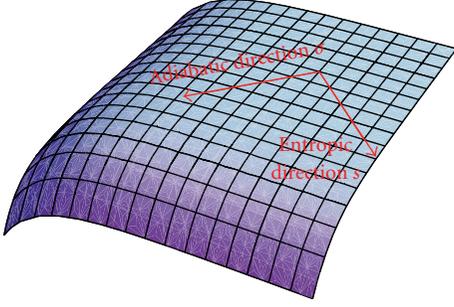


FIGURE 4: After a rotation in field space, the two-field ekpyrotic potential can be viewed as composed of an ekpyrotic direction ( $\sigma$ ) and a transverse tachyonic direction ( $s$ ). The ekpyrotic scaling solution corresponds to motion along the ridge of the potential. Perturbations along the direction of the trajectory are adiabatic/curvature perturbations while perturbations transverse to the trajectory are entropy/isocurvature perturbations.

with the angle  $\theta$  being constant. The solution corresponds to motion along a ridge in the potential, as is evident from the figure. Hence, in contrast to the single field case, the multifield ekpyrotic background evolution is unstable to small perturbations [26, 27]. This implies that the trajectory must be localized near the ridge with extreme precision at the beginning of the ekpyrotic phase, the condition being that the field should stray no more than a value of  $e^{-60}$  (at best) in Planck units from the ridge at the beginning of ekpyrosis [11]. Thus, at first sight, it looks as if the multifield ekpyrotic phase has not managed to solve the problem of initial conditions. However, there currently exist two approaches addressing this issue: the authors of [11] considered the existence of a “pre-ekpyrotic” phase during which the potential is curved upwards and during which the trajectory is localized. Meanwhile, in the context of the cyclic universe, there is a natural resolution of the issue of initial conditions, not involving any new ingredients of the model: indeed, the multifield cyclic universe *selects* those regions that happen to correspond to trajectories sufficiently close to the ridge, in the sense that these regions are vastly amplified over the course of one cycle due to the phases of radiation, matter, and dark energy domination (note that the ekpyrotic phase shrinks the universe by a negligible amount). At the same time, the regions corresponding to trajectories not sufficiently close to the ridge (this would include the vast majority of trajectories) do not undergo a full ekpyrotic phase, and after these regions undergo chaotic mixmaster behavior close to the big crunch, they simply collapse (presumably they will end up forming black holes) and stop both growing and cycling. In this way the global structure of the universe becomes of the *phoenix* type, in which vast habitable regions are interspersed with small collapsed ones. The important point is that the habitable regions, which are the only regions of interest to us here, automatically correspond to the regions that had the right “initial conditions” at the beginning of their preceding ekpyrotic phase. This is discussed in detail in [28]; see also the essay [29].

### 3. Linear Perturbations

**3.1. Single Scalar Field.** In the last section, we have dealt with the classical evolution during the ekpyrotic phase. We will now add quantum fluctuations, and we will see that, just as in inflation, the quantum fluctuations get amplified into classical density perturbations. Hence, on top of resolving the standard cosmological puzzles, the ekpyrotic phase can also be the source of the primordial temperature fluctuations whose imprint is seen in maps of the cosmic microwave background, provided that the amplitude and spectrum of the fluctuations match observations.

From the study of inflationary models, we have developed the intuition that quantum fluctuations that get stretched to superhorizon scales turn into classical perturbations, roughly speaking because the fluctuations go out of causal contact with themselves, do not remember locally that they are in fact fluctuations, and end up as local, classical perturbations to the background evolution. In inflation, this effect occurs because the horizon is approximately constant in size while the wavelengths of the quantum modes get stretched exponentially with time (the scale factor of the universe grows exponentially). For ekpyrosis, the scaling solution (8) shows that the scale factor is almost constant, so that the mode wavelengths remain almost constant too. However, the horizon, which is proportional to  $1/H \sim t$ , shrinks rapidly as  $t \rightarrow 0$ , and hence the modes automatically become of superhorizon size. (Since tensor modes/gravitational waves depend on the evolution of the scale factor alone and since the scale factor shrinks imperceptibly slowly during ekpyrosis, there are no substantial gravity waves produced during the ekpyrotic phase (the background space time is almost Minkowski) [30]. In fact, the dominant gravitational waves that are produced from ekpyrosis are those that arise from the backreaction of the scalar fluctuations onto the metric, at second order in perturbation theory [31].) We will now discuss in some detail what amplitude and spectrum these modes obtain. We will first concentrate on the single field case, before discussing two fields.

Since the scale factor evolves very little during the ekpyrotic phase, one is tempted to simply turn gravity off as a first approximation and to consider the theory consisting only of a scalar field with a steep and negative potential [32]:

$$S = \int d^4x \left[ -\frac{1}{2} (\partial\phi_1)^2 + V_1 e^{-c_1\phi_1} \right]. \quad (22)$$

Then, if we define scalar fluctuations  $\delta\phi$  via  $\phi_1 \equiv \bar{\phi}_1(t) + \delta\phi(t, \underline{x})$ , where  $\bar{\phi}_1 = (2/c_1) \ln(-\sqrt{c_1^2 V_1/2t})$  denotes the background evolution, the equation of motion for the fluctuations is given by

$$\ddot{\delta\phi} - \nabla^2 \delta\phi + V_{,\phi_1\phi_1} \delta\phi = 0, \quad (23)$$

where  $V_{,\phi_1\phi_1} = -2/t^2$ . We then expand the fluctuation field  $\delta\phi$  into Fourier modes

$$\delta\phi = \int \frac{d^3k}{(2\pi)^3} a_k \chi_k e^{ik \cdot x} + h.c. \quad (24)$$

where the  $\chi_k$ s are the positive frequency mode functions (due to the assumed cosmological symmetries, they depend only on the magnitude  $k = |k|$ ). We proceed to quantize the field by imposing the canonical commutation relations

$$[a_k, a_{k'}] = [a_k^\dagger, a_{k'}^\dagger] = 0, \quad [a_k, a_{k'}^\dagger] = (2\pi)^3 \delta(k - k'). \quad (25)$$

In the process, the  $a_k$ 's have been promoted to (annihilation) operators, and the vacuum state  $|0\rangle$  is defined by  $a_k|0\rangle = 0$ . The mode functions obey the equation of motion

$$\ddot{\chi}_k + k^2 \chi_k - \frac{2}{t^2} \chi_k = 0, \quad (26)$$

which admits the two solutions  $\chi_k \propto e^{-ikt}(1 - (i/kt))$ ,  $e^{ikt}(1 + (i/kt))$ . However, as  $t \rightarrow -\infty$  the modes should asymptote to the Minkowski space-free particle state  $\chi_k \rightarrow e^{-ikt}/\sqrt{2k}$  (note that in that limit (26) reduces to the equation of a simple harmonic oscillator), and this fixes the solution to be

$$\chi_k = \frac{1}{\sqrt{2k}} e^{-ikt} \left(1 - \frac{i}{kt}\right). \quad (27)$$

Towards the end of the ekpyrotic phase, we have  $|kt| \ll 1$ , and then the solution can be well approximated by

$$\chi_k \approx \frac{-i}{\sqrt{2k^{3/2}t}}. \quad (28)$$

The quantum fluctuations have a mean, that is, zero,  $\langle 0|\delta\phi|0\rangle = 0$ . However, the *variance*  $\Delta_\phi^2(k)$ , which is defined by  $\langle 0|\delta\phi^2|0\rangle \equiv \int (dk/k) \Delta_\phi^2(k)$ , does not vanish. It is conventional to write the variance as

$$\Delta_\phi^2(k) = \Delta_\phi^2(k_0) \left(\frac{k}{k_0}\right)^{n_s-1}, \quad (29)$$

where  $k_0$  denotes a reference scale and  $n_s$  is the *spectral index*.

A related concept in momentum space is the *power spectrum*  $P(k)$ , defined by

$$P(k) \equiv |\chi_k|^2 = \frac{2\pi^2}{k^3} \Delta_\phi^2(k). \quad (30)$$

It is the Fourier transform of the 2-point correlation function, and we can equivalently define it as

$$\langle \zeta_k \zeta_{k'} \rangle \equiv (2\pi)^3 P(k) \delta^3(k + k'), \quad (31)$$

where isotropy dictates that  $P$  only depends on  $k = |k|$ . We will find this definition useful later on.

In our case, we have that at late times

$$\langle 0|\delta\phi^2|0\rangle = \int \frac{d^3k}{(2\pi)^3} \chi_k^* \chi_k = \int \frac{4\pi k^2 dk}{(2\pi)^3} \frac{1}{2k^3 t^2}, \quad (32)$$

so that the variance is given by

$$\Delta_\phi^2(k) = \frac{1}{4\pi^2 t^2} \rightarrow n_s = 1. \quad (33)$$

The variance is independent of  $k$ , and hence we obtain a scale-invariant spectrum for  $\delta\phi$ . This looks very promising. However, we really must include gravity in our analysis and calculate the spectrum for the curvature perturbation  $\zeta$ , which is the quantity, that is, measured to have a nearly scale-invariant spectrum of perturbations.

Once we add gravity, it is easiest to perform the calculation in so-called  $\zeta$ -gauge, where the perturbations in the scalar field are gauged away and all perturbations are expressed via dilatations of the 3-metric

$$\delta\phi = 0, \quad ds^2 = -dt^2 + a^2(t) e^{2\zeta(t,x)} dx_j dx^j, \quad (34)$$

where  $j = 1, 2, 3$ . Then, using the background scaling solution (8), the action reduces to an action for  $\zeta$  which is given by [33]

$$S = - \int \epsilon g^{\mu\nu} \partial_\mu \zeta \partial_\nu \zeta. \quad (35)$$

During ekpyrosis,  $\epsilon$  is typically nearly constant. In fact, in the scaling solution used above, we have already made the approximation that  $\epsilon$  is constant, and with this approximation, the equation of motion for  $\zeta$  resulting from the action above is particularly simple: in Fourier space it is given by

$$\ddot{\zeta}_k + 3H\dot{\zeta}_k + \frac{k^2}{a^2} \zeta_k = 0. \quad (36)$$

If we use conformal time  $\tau$ , defined via  $dt \equiv a d\tau$  and the notation  $' \equiv (d/d\tau)$ , the above equation becomes

$$\zeta_k'' + 2\frac{a'}{a} \zeta_k' + k^2 \zeta_k = 0. \quad (37)$$

After a further change of variables to  $y \equiv a\zeta/\sqrt{-k\tau}$  and  $x \equiv -k\tau$ , the equation turns into a Bessel equation  $x^2(d^2y/dx^2) + x(dy/dx) + (x^2 - \alpha^2)y = 0$ , with  $\alpha = \sqrt{\tau^2 a''/a + 1/4} \approx 1/2$  since  $a \approx \text{constant}$ . Hence the solutions are given by the Hankel functions  $y \propto H_{1/2}^{(1)}(-k\tau)$ ,  $H_{1/2}^{(2)}(-k\tau)$ , and with the boundary condition that we want  $\zeta \rightarrow e^{-ik\tau}/\sqrt{2k}$  as  $\tau \rightarrow -\infty$ , we obtain the solution (up to a phase). (Useful asymptotic expressions are  $H_\alpha^{(1)}(x) \rightarrow \sqrt{(2/\pi x)} e^{i(x - \alpha\pi/(2 - \pi/4))}$  when  $x \gg \alpha$  and  $H_\alpha^{(1)}(x) \rightarrow -(i/\pi)\Gamma(\alpha)(2/x)^\alpha$  when  $x \ll \alpha$  and for  $\alpha > 0$ .)

$$\zeta = \frac{\sqrt{-\tau}}{a} H_{1/2}^{(1)}(-k\tau). \quad (38)$$

At late times, the variance becomes

$$\langle 0|\zeta^2|0\rangle = \int \frac{d^3k}{(2\pi)^3} \frac{(-\tau)}{a^2} |H_{1/2}^{(1)}(-k\tau)|^2 \sim \int \frac{dk}{k} k^2, \quad (39)$$

and hence we get a spectral index  $n_s = 3$ . This spectrum is *blue* as there is more power on smaller scales, and it is in disagreement with observations [32–34]. Hence, the scale-invariant spectrum of the scalar perturbation in the no-gravity theory did not get transferred to the curvature

perturbation  $\zeta$ . A closer analysis reveals that these two perturbations correspond to two physically distinct modes, the former is a time-delay mode to the big crunch, and the latter a local dilatation in space. In a contracting universe, these two modes are distinct, and they do not mix. It is conceivable that they might mix at the big crunch/big bang transition [35, 36], in which case the scale-invariant contribution would be the dominant one on the large scales of interest, but this possibility is still insufficiently understood to make definite predictions. As we will show next, this is also unnecessary, as there is a very natural *entropic* mechanism which generates scale-invariant curvature perturbations before the big bang, as long as there is more than one scalar field present. (Recently, Khoury and Steinhardt have also pointed out that right at the onset of the single-field ekpyrotic phase, a range of scale-invariant modes can be produced [37]. However, contrary to the cases that we have discussed so far, this *adiabatic* mechanism requires the universe to already be contracting when the equation of state is still near  $w \approx -1$ . If viable, this mechanism would produce an interesting nongaussian signal; but as it is currently not known how to incorporate this mechanism into a more complete cosmological model, we will not discuss this mechanism here. See also [38] for the challenges that this scenario must address.) But before continuing, it might be useful to add a few remarks concerning the validity of our approach: indeed, the reader might be worried about the validity of perturbation theory, since the background quantities, such as the Hubble rate, as well as the perturbations themselves blow up as  $t \rightarrow 0$ . However, as shown in [33], after switching to synchronous gauge, it is straightforward to see that the universe evolves to become closer and closer to the unperturbed background solution, and hence perturbation theory is valid. Also, even though the background quantities blow up as seen from the 4-dimensional viewpoint, in fact in the higher-dimensional colliding branes picture the ekpyrotic phase has the effect of flattening the branes and hence of rendering the curvatures *small* as the big crunch is approached [20].

**3.2. Two Fields: The Entropic Mechanism.** As discussed at the end of the last section, it is rather unnatural to consider only a single scalar field in the effective theory, since there are two universal scalars that are always present in a higher-dimensional context: the radion field, determining the distance between the two end-of-the-world branes, and the volume modulus of the internal manifold. But as soon as there is more than one scalar field present, one can have entropy, or isocurvature, perturbations, which are growing mode perturbations in a collapsing universe [39]. Entropy perturbations can source the curvature perturbation, and hence (provided the entropy perturbations acquire a nearly scale-invariant spectrum), nearly scale-invariant curvature perturbations can be generated just before the bounce [26, 40]. These then turn into growing mode perturbations in the ensuing expanding phase.

For the two scalar fields, we will again assume the potential (20), but with the slight generalization that we allow the fast-roll parameter  $\epsilon$  to be slowly varying. There are

two gauge-invariant scalar perturbation modes: the entropy perturbation  $\delta s = \cos\theta\delta\phi_2 - \sin\theta\delta\phi_1$  corresponds to perturbations transverse to the background trajectory, see Figure 4, while the adiabatic, or curvature, perturbation  $\zeta$  is the gauge-invariant quantity expressing perturbations along the background trajectory; see [25, 41] for a detailed exposition. For a straight trajectory ( $\theta = 0$ ), the linearized equation of motion for  $\delta s$  is

$$\ddot{\delta s} + 3H\dot{\delta s} + \left(\frac{k^2}{a^2} + V_{ss}\right)\delta s = 0, \quad (40)$$

where  $V_{ss}$  denotes the second derivative of the potential with respect to  $s$ . In conformal time, and for the rescaled variable  $\delta S = a(\tau)\delta s$ , we obtain

$$\delta S'' + \left(k^2 - \frac{a''}{a} + a^2 V_{ss}\right)\delta S = 0. \quad (41)$$

To proceed, we must relate  $a''/a$  and  $V_{ss}$  to the fast-roll parameter  $\epsilon$  and its derivative with respect to the number of e-folds of expansion  $N$ , where  $dN \equiv d \ln a$ . By requiring  $\epsilon$  to vary slowly, what is meant is that we will keep terms in  $d\epsilon/dN$  but not higher-order terms such as  $d^2\epsilon/dN^2$ . Then, by differentiating  $\epsilon = \dot{\sigma}^2/(2H^2)$  twice, and using  $\ddot{\sigma} + 3H\dot{\sigma} + V_\sigma = 0$  as well as  $V_{ss} = V_{\sigma\sigma}$ , one can derive the following expressions, valid to subleading order in  $\epsilon$ :

$$\begin{aligned} \frac{a''}{a} &= H^2 a^2 (2 - \epsilon), \\ V_{ss} &= H^2 \left(-2\epsilon^2 + 6\epsilon + \frac{5}{2} \epsilon_{,N}\right). \end{aligned} \quad (42)$$

Using in addition that  $aH = (1 + (1/\epsilon) + (\epsilon_{,N}/\epsilon^2))/\epsilon\tau$ , (41) finally reads

$$\delta S'' + \left(k^2 - \frac{2(1 - (3/2)\epsilon) + (3/4)(\epsilon_{,N}/\epsilon^2)}{\tau^2}\right)\delta S = 0. \quad (43)$$

In analogy with our discussion of the single-field case, this equation can be solved in terms of the Hankel functions, supplemented by the boundary condition of approaching the Minkowski vacuum state in the far past, to yield (up to a phase)

$$\delta S = \frac{\sqrt{-k\tau}}{2} H_\nu^{(1)}(-k\tau), \quad \nu = \frac{3}{2} \left(1 - \frac{2}{3\epsilon} + \frac{\epsilon_{,N}}{3\epsilon^2}\right). \quad (44)$$

At late times  $(-k\tau) \rightarrow 0$ , and we obtain

$$\delta S \approx \frac{1}{\sqrt{2(-\tau)}k^\nu}, \quad (45)$$

implying that at the end of the ekpyrotic phase, the entropy perturbation is given by

$$\delta s(t_{ek\text{-end}}) \approx \frac{|\epsilon V_{ek\text{-end}}|^{1/2}}{\sqrt{2}k^\nu}. \quad (46)$$

Following the same steps as in the single field case above, it is straightforward to see that the spectral index of the entropy perturbation is now given by [26]

$$n_s - 1 = \frac{2}{\epsilon} - \frac{\epsilon, \mathcal{N}}{\epsilon^2}. \quad (47)$$

The first term on the right-hand side is a gravitational contribution, which, being positive, tends to make the spectrum blue. The second term is a nongravitational contribution, which tends to make the spectrum red. A simple way to estimate the natural range of  $n_s$  is to rewrite the above expression in terms of  $\mathcal{N}$ , the number of e-folds before the end of the ekpyrotic phase (where  $d\mathcal{N} = d \ln(aH)$ ):

$$n_s - 1 = \frac{2}{\epsilon} - \frac{d \ln \epsilon}{d\mathcal{N}}. \quad (48)$$

In this expression,  $\epsilon(\mathcal{N})$  measures the equation of state during the ekpyrotic phase, which decreases from a value much greater than unity to a value of order unity in the last  $\mathcal{N}$  e-folds. If we estimate  $\epsilon \approx \mathcal{N}^\alpha$  [42], then the spectral tilt is

$$n_s - 1 \approx \frac{2}{\mathcal{N}^\alpha} - \frac{\alpha}{\mathcal{N}}. \quad (49)$$

Here we see that the sign of the tilt is sensitive to  $\alpha$ . For nearly exponential potentials ( $\alpha \approx 1$ ), the spectral tilt is  $n_s \approx 1 + 1/\mathcal{N} \approx 1.02$ , slightly blue, because the first term dominates. However, in the cyclic model the steepness of the potential must decrease in order for the ekpyrotic phase to come to an end, and  $\alpha$  parameterizes these cases. If  $\alpha > 1.14$ , the spectral tilt is red. For example,  $n_s = 0.97$  for  $\alpha \approx 2$ . These examples represent the range that can be achieved by the entropic mechanism, roughly

$$0.97 < n_s < 1.02. \quad (50)$$

These are in good agreement with the present observational bounds obtained by the WMAP satellite, which are  $n_s = 0.96 \pm 0.03$  at the  $2\sigma$  level [9].

Now that we have shown how an approximately scale-invariant spectrum of entropy perturbations may be generated by scalar fields in a contracting universe, we will discuss how these perturbations may be converted to curvature perturbations. Since the entropy perturbations of interest are all of superhorizon scales, we can now restrict our study to large scales only, where spatial gradients can be neglected. On these scales, the evolution equation for the curvature perturbation is given by [25]

$$\dot{\zeta} = -\frac{2H}{\dot{\sigma}} \dot{\theta} \delta s = \sqrt{\frac{2}{\epsilon}} \dot{\theta} \delta s. \quad (51)$$

Hence, as soon as the background trajectory bends ( $\dot{\theta} \neq 0$ ), the entropy perturbations become a source for the curvature perturbations.

There are at least two ways in which such a bending can occur: the first makes use of the instability of the two-field ekpyrotic potential; compare again Figure 4. If the

background trajectory strays sufficiently far from the ridge of the potential, the trajectory will turn and fall off one of the steep sides of the potential [10, 23, 24]. The turning of the trajectory then immediately results in the conversion of entropy into curvature perturbations. Since this conversion occurs during the ekpyrotic phase, we will term this process *ekpyrotic conversion*. It is straightforward to estimate the amplitude of the resulting curvature perturbation (its spectrum will be identical to the spectrum of the entropy perturbations, as (51) is  $k$ -independent): if we approximate the entropy perturbation as remaining constant during the conversion process and assume a total bending angle of order unity,  $\int \dot{\theta} \sim \mathcal{O}(1)$ , then the resulting curvature perturbation after conversion will be given by

$$\zeta_{\text{conv-end}} \approx \sqrt{\frac{2}{\epsilon_{ek}}} \delta s_{ek\text{-end}}. \quad (52)$$

We should mention straight away that the approximations just made will not be good enough in calculating the nongaussian corrections to the linear calculation, but for the present purposes, they will do. Since the fast-roll parameter  $\epsilon_{ek} \sim \mathcal{O}(10^2)$ , we find that

$$\zeta_{\text{conv-end}} \sim \frac{1}{10} \delta s_{ek\text{-end}}. \quad (53)$$

The second way in which a bending of the trajectory can occur is suggested by the embedding of the cyclic model in M-theory [26] and applies when the background field trajectory remains straight throughout the ekpyrotic phase. At the end of the ekpyrotic phase, the potential turns off, and the evolution becomes dominated by the kinetic energy of the two scalar fields. This kinetic phase corresponds to the final approach of the branes in the higher-dimensional picture. During this approach, there is a generic effect that occurs, but that cannot be seen in the 4-dimensional effective theory. The branes that are bound to collide with each other are of opposite tension. Now, it turns out that just before the collision, it always happens that at the location of the negative tension brane, the internal 6-dimensional manifold tries to shrink to zero size [21]. However, just about any type of matter present on the negative tension brane will smoothly cause the internal manifold to grow again [43]. This effect is due to the special properties of gravity on a negative tension object. When this effect is translated back into the effective theory that we have been using all along, the location in field space where the internal manifold reaches zero size is given by the  $\phi_2 = 0$  line. This line thus constitutes a boundary to field space. And the presence of matter on the negative tension brane gives rise to an effective repulsive potential in the vicinity of the  $\phi_2 = 0$  line. Hence, during the kinetic phase, the background trajectory automatically bends, just before the trajectory shoots off to  $-\infty$  where the brane collision/big bang occurs. What we have just discussed is a concrete example originating from string theory of how a bend in the trajectory can occur during the kinetic phase. However, more generally our results will apply whatever the microphysical cause of the bending and of the effective repulsive potential; see Figure 5 for an illustration

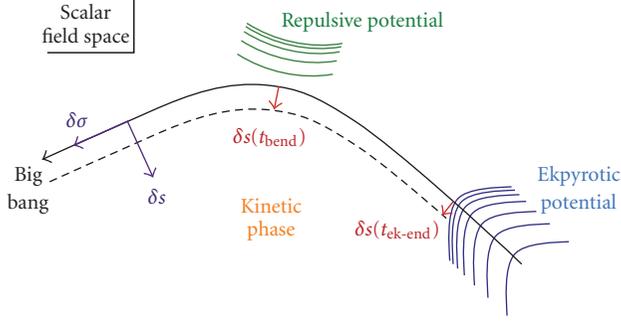


FIGURE 5: After the ekpyrotic phase, the trajectory in scalar field space enters the kinetic phase and bends - this bending is described by the existence of an effective repulsive potential (the potentials are indicated by their contour lines). A trajectory adjacent to the background evolution can be characterized by the entropy perturbation  $\delta s(t_{ek-end})$  at the end of the ekpyrotic phase, leading to a corresponding off-set  $\delta s(t_{bend})$ , or equivalently  $\delta V(t_{bend})$ , at the time of bending.

of the general case that we have in mind. Again, it is quite straightforward to estimate the amplitude of the curvature perturbation resulting from this process of *kinetic conversion*. On large scales, the linearized equation of motion for the entropy perturbation is given by

$$\ddot{\delta s} + 3H\dot{\delta s} + (V_{ss} + 3\dot{\theta}^2)\delta s = 0, \quad (54)$$

where, incidentally, we have the useful relation  $\dot{\theta} = -V_s/\dot{\sigma}$ . Then, during the kinetic phase and away from the repulsive potential, the Einstein equations immediately yield

$$H = \frac{1}{3t}, \quad \dot{\sigma} = \frac{-\sqrt{2}}{\sqrt{3}t} \quad (\text{kinetic phase}). \quad (55)$$

Thus, (54) simplifies to  $\ddot{\delta s} + \dot{\delta s}/t = 0$  which implies that the entropy perturbation grows logarithmically during the potential-free kinetic phase. We can ignore this insignificant growth. However, the evolution of the entropy perturbation during the process of conversion turns out to be important. We can estimate it by assuming that the trajectory bends with a constant  $\dot{\theta} \sim 1/\Delta t$ , where  $\Delta t$  denotes the duration of the conversion process. We can further assume that the repulsive potential depends only on  $\phi_2$ . Then  $\dot{\theta}$ ,  $V_s$ , and  $V_{ss}$  can all be related to  $V_{,\phi_2}$ ,  $V_{,\phi_2\phi_2}$ , evaluated during the conversion, and it is not difficult to show that this leads to  $V_{ss} \approx \dot{\phi}_1\dot{\theta}/(t_{bend}\dot{\phi}_2)$ , where  $t_{bend}$  corresponds to the time halfway through the bending of the trajectory [44]. For the particular example where the cyclic model is embedded in M-theory, we have that  $\dot{\phi}_1 = -\sqrt{3}\dot{\phi}_2$ , and specializing to this example, we have  $V_{ss} \approx (2-3)\dot{\theta}^2$ . Hence, (54) becomes (where we can neglect the term in  $\delta s$ )

$$\ddot{\delta s} + 6\dot{\theta}^2\delta s \approx 0, \quad (56)$$

and thus, during the conversion, the entropy perturbation evolves sinusoidally

$$\delta s \approx \cos\left[\omega(t - t_{conv-beg})\right]\delta s(t_{ek-end}), \quad (57)$$

where  $t_{conv-beg}$  denotes the time at which the trajectory starts to bend, and  $\omega \approx 2.5/\Delta t$ . Now we can immediately evaluate the resulting linear curvature perturbation by integrating (51) to get

$$\begin{aligned} \zeta_L &= \int_{bend} -\frac{2H}{\dot{\sigma}}\dot{\theta}\delta s \\ &\approx \sqrt{\frac{2}{3}}\frac{\dot{\theta}}{\omega}\sin(\omega\Delta t)\delta s(t_{ek-end}) \\ &\approx \frac{1}{5}\delta s(t_{ek-end}). \end{aligned} \quad (58)$$

Thus, the amplitude is very similar in magnitude to the value estimated above for the process of ekpyrotic conversion.

We are now in a position to calculate the variance of the generated curvature perturbation, which, on account of (46), is given by

$$\langle \zeta^2 \rangle \approx \int \frac{d^3k}{(2\pi)^3} \frac{\epsilon_{ek} V_{ek-end}}{50k^{2\nu}} = \int \frac{dk}{k} \frac{\epsilon_{ek} V_{ek-end}}{100\pi^2} k^{n_s-1}. \quad (59)$$

Hence, the amplitude is in agreement with the current WMAP bounds of  $\Delta_\zeta^2(0.002 Mpc^{-1}) = (2.4 \pm 0.2) \times 10^{-9}$  [9], as long as  $|V_{ek-end}| \approx (10^{-2} M_{Pl})^4$ , that is, the minimum of the potential has to be roughly at the grand unified scale for models using kinetic conversion [26]. This scale is also the natural scale of Hořava-Witten theory, and thus it is the scale where one would expect the potential to turn around. Note that for models using ekpyrotic conversion, this result implies that the bending must occur at a specific time, namely, when the potential reaches the grand unified scale. In the latter models, this may or may not also correspond to the bottom of the potential.

Finally, we should state an important assumption that we have been making implicitly up to now, namely, we assumed that the curvature perturbation passes through the big crunch/big bang transition essentially unchanged. The reason for doing so is that the perturbations we are considering are vastly larger than the horizon size around the time of the crunch, and hence, due to causality, it seems reasonable to assume that long-wavelength modes suffer no change—this viewpoint is discussed in much more detail in [33]. In new ekpyrotic models, in which the bounce is smooth and describable entirely within a 4-dimensional effective theory, this assumption certainly holds true. In the case of a classically singular bounce, this remains an assumption subject to possible revision in the future. (In this context, we can also mention the possibility that no conversion of entropy to curvature perturbations might occur before the big crunch, but that this conversion could happen during the phase shortly following the bang through modulated reheating [45]: if massive matter fields are produced copiously at the brane collision and dominate the energy density immediately after the bang and if, furthermore, these fields couple to ordinary matter via a function of  $\delta s$ , then their decay into ordinary matter will occur at slightly different times depending on the value of  $\delta s$ . In this way, the ordinary matter perturbations would also inherit the entropic perturbation spectrum.)

## 4. Higher-Order Perturbations and Predictions for Nongaussianity

*4.1. Definitions and Local Nongaussianity.* Now that we have seen in detail how the ekpyrotic phase generates linear, nearly scale-invariant density perturbations via the entropic mechanism, we can inquire as to whether the higher-order corrections might lead to an observable signal. We will only calculate nongaussian corrections for perturbations generated via the entropic mechanism, because, as discussed in the previous section, this is the only robust and well-understood mechanism to date that generates ekpyrotic perturbations in agreement with observations. As we saw earlier, the linear perturbations are related to observations of the 2-point correlation function. Similarly, quadratic and cubic corrections to these perturbations are related to observations of the 3- and 4-point functions, respectively. For an exactly gaussian probability distribution, all  $n$ -point functions for which  $n$  is odd vanish while for  $n$  even, the  $n$ -point functions are related to the 2-point function. Thus, the simplest way in which we could detect a departure from exact gaussianity would be due to the presence of a nonvanishing 3-point function.

In momentum space, the 3-point function corresponds to a configuration of 3 momenta, which form a closed triangle due to momentum conservation. Hence, the 3-point function is specified not only by its magnitude on different scales but also by its magnitude for different shapes of the triangle. Or, turning this reasoning around, when we make predictions for nongaussianity, we must predict both the amplitude and the shape of the momentum space triangle that we would like to observe. Let us make all of this more precise now. Earlier, we defined the power spectrum as the Fourier transform of the 2-point function

$$\langle \zeta_{k_1} \zeta_{k_2} \rangle = (2\pi)^3 \delta^3(k_1 + k_2) P(k_1). \quad (60)$$

Similarly, the *bispectrum*, which is the Fourier transform of the 3-point function, is given by

$$\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \rangle = (2\pi)^3 \delta^3(k_1 + k_2 + k_3) B(k_1, k_2, k_3), \quad (61)$$

the *trispectrum*, the Fourier transform of the 4-point function, via

$$\langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \zeta_{k_4} \rangle_c = (2\pi)^3 \delta^3(k_1 + k_2 + k_3 + k_4) T(k_1, k_2, k_3, k_4), \quad (62)$$

and so on. The  $\delta$ -functions result from momentum conservation, while  $B$  and  $T$  are shape functions (for a triangle and a quadrangle, resp.). In the last expression, the subscript  $c$  indicates that we only need to consider the connected part of the 4-point function, that is, the part that is not captured by products of 2-point functions.

For nongaussianity of the so-called *local* form, it is useful to define (in real space) the following expansion of the curvature perturbation on uniform energy density surfaces,

$$\zeta = \zeta_L + \frac{3}{5} f_{\text{NL}} \zeta_L^2 + \frac{9}{25} g_{\text{NL}} \zeta_L^3, \quad (63)$$

with  $\zeta_L$  being the linear, gaussian part of  $\zeta$ . The factors of 3/5 are a historical accident; they arose because this type of expansion was first defined for a different variable. In momentum space,  $B$  is then given by

$$B = \frac{6}{5} f_{\text{NL}} [P(k_1)P(k_2) + 2 \text{ permutations}], \quad (64)$$

as can be verified straightforwardly by combining (60), (61), and (63). Similarly, the momentum space 4-point function corresponding to nongaussianity of the local form can be expressed as

$$T = \tau_{\text{NL}} [P(k_{13})P(k_3)P(k_4) + 11 \text{ perms.}] + \frac{54}{25} g_{\text{NL}} [P(k_2)P(k_3)P(k_4) + 3 \text{ perms.}], \quad (65)$$

where  $\tau_{\text{NL}}$  and  $g_{\text{NL}}$  parameterize the two relevant shape functions; see, for example, [46] for more details. For cosmological models in which the perturbations originate from the fluctuations of a single field (in our case the entropy field),  $\tau_{\text{NL}}$  is directly related to the square of  $f_{\text{NL}}$ , explicitly

$$\tau_{\text{NL}} = \frac{36}{25} f_{\text{NL}}^2. \quad (66)$$

Concentrating now on the bispectrum, we can see that, since for a scale-invariant spectrum  $P(k) \sim k^{-3}$ , we have

$$B \sim f_{\text{NL}} \left( \frac{1}{k_1^3 k_2^3} + \frac{1}{k_2^3 k_3^3} + \frac{1}{k_3^3 k_1^3} \right) = f_{\text{NL}} \frac{\sum k_i^3}{\prod k_i^3}. \quad (67)$$

This is the typical momentum dependence for local nongaussianity [47], which is also the relevant one for ekpyrotic models, as we will show shortly. The signal is the largest when one of the momenta is very small—this automatically requires the other two momenta to be almost equal, and hence the local form of nongaussianity corresponds to having the largest signal generated for *squeezed* triangles in momentum space.

It is instructive to calculate explicitly the tree-level 3-point function for the entropy perturbation generated during the ekpyrotic phase, that is, before the conversion to curvature perturbations. (Note that quantum corrections from loop diagrams will be suppressed by factors of  $\hbar$ .) Maldacena described in [48] how the expectation value for the 3-point function is given by

$$\langle (\delta s)^3 \rangle = -i \int dt' \langle [(\delta s)^3(t), H_{\text{int}}(t')] \rangle, \quad (68)$$

where  $H_{\text{int}}(t') = V_{\text{sss}}(\delta s)^3/3! = -\sqrt{\epsilon}\kappa_3/(3!t'^2)$  is the cubic interaction Hamiltonian. In Fourier space, this can be rewritten as [12, 49]

$$\begin{aligned} \langle (\delta s)^3 \rangle &= (2\pi)^3 \delta(\Sigma_i k_i) \\ &\times 2\text{Re} \left\{ -i \delta s_{k_1}(t) \delta s_{k_2}(t) \delta s_{k_3}(t) \right. \\ &\quad \left. \int_{-\infty}^t \frac{(-\sqrt{\epsilon}\kappa_3)}{t'^2} \delta s_{k_1}(t') \delta s_{k_2}(t') \delta s_{k_3}(t') \right\}. \end{aligned} \quad (69)$$

For  $\epsilon$  large, the mode functions are given approximately by (cf. (27))

$$\delta s_k(t) = \frac{1}{\sqrt{2k}} e^{-ikt} \left( 1 - \frac{i}{kt} \right), \quad (70)$$

so that we get

$$\langle (\delta s)^3 \rangle = (2\pi)^3 \delta(\Sigma_i k_i) \frac{\sqrt{\epsilon}\kappa_3}{6t^4} \frac{\Sigma_i k_i^3}{\Pi k_i^3}, \quad (71)$$

where we have used the result that

$$\begin{aligned} &i(1 + ik_1 t)(1 + ik_2 t)(1 + ik_3 t) e^{-iKt} \\ &\times \int (1 - ik_1 t')(1 - ik_2 t')(1 - ik_3 t') e^{iKt'} (t')^{-5} + c.c. \quad (72) \\ &= \frac{1}{2} \Sigma_i k_i^3 + \dots, \end{aligned}$$

with  $K = k_1 + k_2 + k_3$  and where the dots indicate terms suppressed by powers of  $k_i t$ . The calculation shows two things: first, the momentum dependence in (71) is of the local form, and, comparing with (64) and using  $P_{\delta s}(k) = 1/(2t^2 k^3)$  from (30) and (33), it corresponds to having an entropy perturbation of the form

$$\delta s = \delta s_L + \frac{\sqrt{\epsilon}\kappa_3}{8} \delta s_L^2, \quad (73)$$

where  $\delta s_L$  is the linear, gaussian part of the entropy perturbation  $\delta s$ . Secondly, the endresult is dominated by the terms for which  $|k_i t| \ll 1$ —in other words, the dominant contribution to nongaussianity is generated by the (classical) evolution on superhorizon scales, and the same holds true for the 4-point function also. Hence, in determining the predictions for the nongaussian curvature perturbation, we can simply perform the calculation using the higher-order classical equations of motion on large scales, up to the required order in perturbation theory.

The strategy for calculating the nonlinearity parameters defined in (63) is straightforward: first we solve the equation of motion for the entropy perturbation up to third order in perturbation theory. This allows us to integrate the equation of motion for  $\zeta$ , at the first three orders in perturbation theory, and then we obtain the nonlinearity parameters by evaluating

$$f_{\text{NL}} = \frac{5}{3} \frac{\int_{t_{\text{ek-beg}}}^{t_{\text{conv-end}}} \zeta(2)'}{\left( \int_{t_{\text{ek-beg}}}^{t_{\text{conv-end}}} \zeta(1)'' \right)^2}, \quad g_{\text{NL}} = \frac{25}{9} \frac{\int_{t_{\text{ek-beg}}}^{t_{\text{conv-end}}} \zeta(3)'}{\left( \int_{t_{\text{ek-beg}}}^{t_{\text{conv-end}}} \zeta(1)'' \right)^3}, \quad (74)$$

where the integrals are evaluated from the time that the ekpyrotic phase begins until the conversion phase has ended and  $\zeta$  has evolved to a constant value. A note on notation is that we expand the entropy perturbation (and similarly the curvature perturbation) as

$$\delta s = \delta s^{(1)} + \delta s^{(2)} + \delta s^{(3)} \quad (75)$$

without factorial factors and where we use  $\delta s^{(1)}$  and  $\delta s_L$  interchangeably (in Section 3, where we dealt exclusively with linear perturbations, we sometimes wrote  $\delta s$  instead of  $\delta s^{(1)}$ , but hopefully this will not confuse the reader).

The relevant equations of third-order cosmological perturbation theory with multiple scalar fields were developed in [50], and we will use the results of that paper. The derivations of the equations are lengthy and do not provide any further insight into the topic of this paper. Hence, we will simply use the equations as we need them. Interestingly, it turns out that all results regarding the conversion process can be well approximated by various simple techniques that we will present further below, both for ekpyrotic and for kinetic conversion.

**4.2. Generation of Entropy Perturbations.** During the ekpyrotic phase, the equation of motion for the entropy perturbation, up to third order in perturbation theory, is given by [50]

$$\begin{aligned} &\ddot{\delta s} + 3H\dot{\delta s} + V_{,ss}\delta s + \frac{1}{2}V_{,sss}(\delta s)^2 \\ &+ \frac{V_{,\sigma\sigma}}{\dot{\sigma}^3} (\dot{\delta s})^3 + \left( \frac{V_{,\sigma\sigma}}{\dot{\sigma}^2} + 3\frac{V_{,\sigma}^2}{\dot{\sigma}^4} + 3H\frac{V_{,\sigma}}{\dot{\sigma}^3} - 2\frac{V_{,ss}}{\dot{\sigma}^2} \right) (\dot{\delta s})^2 \delta s \\ &+ \left( -\frac{3}{2\dot{\sigma}}V_{,ss\sigma} - 5\frac{V_{,\sigma}V_{,ss}}{\dot{\sigma}^3} - 3H\frac{V_{,ss}}{\dot{\sigma}^2} \right) \dot{\delta s}(\delta s)^2 \\ &+ \left( \frac{1}{6}V_{,ssss} + 2\frac{V_{,ss}^2}{\dot{\sigma}^2} \right) (\delta s)^3 = 0. \end{aligned} \quad (76)$$

Using the following expressions, valid during the ekpyrotic phase:

$$\begin{aligned} \dot{\sigma} &= -\frac{\sqrt{2}}{\sqrt{\epsilon}t}, & V &= -\frac{1}{\epsilon t^2}, \\ V_{,\sigma} &= -\frac{\sqrt{2}}{\sqrt{\epsilon}t^2}, & V_{,\sigma\sigma} &= -\frac{2}{t^2}, & V_{,\sigma\sigma} &= 0, & V_{,ss\sigma} &= -\frac{2\sqrt{2}\epsilon}{t^2}, \\ V_{,s} &= 0, & V_{,ss} &= -\frac{2}{t^2}, & V_{,sss} &= -\frac{\kappa_3\sqrt{\epsilon}}{t^2}, & V_{,ssss} &= -\frac{\kappa_4\epsilon}{t^2}, \end{aligned} \quad (77)$$

it is not difficult to find by iteration that the entropy perturbation, to leading order in  $1/\epsilon$ , is given by

$$\delta s = \delta s_L + \frac{\kappa_3\sqrt{\epsilon}}{8} \delta s_L^2 + \epsilon \left( \frac{\kappa_4}{60} + \frac{\kappa_3^2}{80} - \frac{2}{5} \right) \delta s_L^3, \quad (78)$$

where, just as before,  $\delta s_L \propto 1/t$ . Note that the quadratic term agrees exactly with (73). At each order in perturbation

theory, the nonlinear corrections depend on the parameters of the potential at that order; compare the parametrization of the potential in (20). The above equation specifies the initial conditions for the start of the conversion phase.

Both during the ekpyrotic phase and during the conversion process, when the background trajectory bends, the entropy perturbation sources the curvature perturbation on large scales according to [50]

$$\begin{aligned} \zeta \approx & \frac{2H}{\dot{\sigma}^2} \left[ V_{,s} \delta s - \frac{1}{2\dot{\sigma}} V_{,\sigma} \delta s \dot{\delta s} + \left( \frac{1}{2} V_{,ss} + \frac{2}{\dot{\sigma}^2} V_{,s}^2 \right) (\delta s)^2 \right] \\ & + \frac{2H}{\dot{\sigma}^2} \left[ \left( -\frac{\dot{\theta}}{6\dot{\sigma}^2} V_{,\sigma} - \frac{1}{2\dot{\sigma}} V_{,s\sigma} - \frac{2}{\dot{\sigma}^3} V_{,s} V_{,\sigma} \right) (\delta s)^2 \dot{\delta s} \right. \\ & \left. + \left( \frac{1}{6} V_{,sss} + \frac{2}{\dot{\sigma}^2} V_{,s} V_{,ss} + \frac{4}{\dot{\sigma}^4} V_{,s}^3 \right) (\delta s)^3 \right]. \end{aligned} \quad (79)$$

The inelegant but sure-fire thing to do now is to simply integrate this equation numerically during the two phases of generation and conversion and to deduce the nonlinearity parameters using (74), as was done in [50]. However, this approach does not give much insight into the final result. This is why we will present more physically transparent techniques first, which allow us to estimate the nonlinearity parameters  $f_{\text{NL}}$  and  $g_{\text{NL}}$  pretty well, and subsequently we will compare these estimates with the results of numerical integration and discuss the predictions.

**4.3. Ekpyrotic Conversion.** We start by analyzing the case where the conversion of entropy into curvature modes occurs during the ekpyrotic phase. For this scenario, it was shown by Koyama et al. [49] that the so-called  $\delta N$  formalism [51, 52] is well suited. For ekpyrotic conversion, the calculation is most easily performed, and the result most easily expressed, in terms of the potential (17), which is why we are adopting this restricted form here. In working with a parameterized potential like (20), the bending of the trajectory can be more complicated, in the sense that there can be multiple turns, and one has to decide when to stop the evolution. In this case, the results are strongly cutoff dependent, and without a precisely defined model specifying the subsequent evolution, it is impossible to make any generic predictions.

The  $\delta N$  formalism is particularly well suited to the case of ekpyrotic conversion, because the background evolution is simple. In fact, it turns out that by making the approximation that the bending is instantaneous, it is very easy to find an approximate formula for the nonlinearity parameters at any chosen order in perturbation theory. The following calculation was first presented in [49] for the bispectrum and trivially extended to the trispectrum in [50].

In order to implement the  $\delta N$  formalism, we have to calculate the integrated expansion  $N = \int H dt$  along the background trajectory. Initially, the trajectory is close to the scaling solution (21). Then, we assume that at a fixed field value  $\delta s_B$  away from the ridge, the trajectory instantly

bends and rolls off along the  $\phi_2$  direction. At this point, the trajectory follows the single-field evolution

$$a(t) = (-t)^{2/c_2^2}, \quad \phi_2 = \frac{2}{c_2} \ln(-t) + \text{constant}, \quad \phi_1 = \text{constant}. \quad (80)$$

Approximating the bending as instantaneous, it is easy to evaluate the integrated expansion, with the result that

$$N = -\frac{2}{c_1^2} \ln|H_B| + \text{constant}, \quad (81)$$

where  $H_B$  denotes the Hubble parameter at the instant that the bending occurs. Note that all  $c_2$ -dependence has canceled out of the formula above. At the end of the conversion process, we are interested in evaluating the curvature perturbation on a surface of constant energy density. But, on a surface of constant energy density, the curvature perturbation is equal to a perturbation in the integrated expansion [53]. If we assume that the integrated expansion depends on a single variable  $\alpha$ , we can write

$$\zeta = \delta N = N_{,\alpha} \delta \alpha + \frac{1}{2} N_{,\alpha\alpha} (\delta \alpha)^2 + \frac{1}{6} N_{,\alpha\alpha\alpha} (\delta \alpha)^3. \quad (82)$$

In our example, we indeed expect a change in  $N$  to depend solely on a change in the initial value of the entropy perturbation  $\delta s$ . Now, from (78), we know that  $\delta s_L \propto 1/t \propto H$ , and hence we can parameterize different initial values of the entropy perturbation by writing

$$\delta s_L = \alpha H. \quad (83)$$

Note that since  $\delta s_L$  is gaussian, so is  $\alpha$ . With this identification, we have

$$\delta s = \alpha H + \frac{\kappa_3 \sqrt{\epsilon}}{8} (\alpha H)^2 + \epsilon \left( \frac{\kappa_4}{60} + \frac{\kappa_3^2}{80} - \frac{2}{5} \right) (\alpha H)^3, \quad (84)$$

so that at the fixed value  $\delta s = \delta s_B$ , we have

$$\alpha \propto \frac{1}{H_B}. \quad (85)$$

Now we can immediately evaluate

$$N_{,\alpha} = N_{,H_B} \frac{dH_B}{d\alpha} = \frac{2}{c_1^2} \alpha \quad (86)$$

and similarly

$$N_{,\alpha\alpha} = -\frac{2}{c_1^2 \alpha^2}, \quad N_{,\alpha\alpha\alpha} = \frac{4}{c_1^2 \alpha^3}. \quad (87)$$

In this way, with very little work, we can estimate the nonlinearity parameters

$$\begin{aligned} f_{\text{NL}} &= \frac{5N_{,\alpha\alpha}}{6N_{,\alpha}^2} = -\frac{5}{12} c_1^2, \\ \tau_{\text{NL}} &= \frac{36}{25} f_{\text{NL}}^2 = \frac{1}{4} c_1^4, \\ g_{\text{NL}} &= \frac{25N_{,\alpha\alpha\alpha}}{54N_{,\alpha}^3} = \frac{25}{108} c_1^4. \end{aligned} \quad (88)$$

We are now in a position to compare these estimates to the numerical results obtained in [50] by solving and integrating the equations of motion (76) and (79). (Approximate analytic solutions to the equations of motion were first presented in [54]). There, the initial conditions were chosen such that they are given by the scaling solution (21), except for a 0.1 percent increase in the initial field velocity  $|\dot{\phi}_2|$ . This causes the trajectory to eventually roll off in the  $\phi_2$  direction and to quickly approach the single-field solution (80). The results for several values of  $c_1$  and  $c_2$  are shown in Table 1, alongside the values estimated by the  $\delta N$  formulae.

It is immediately apparent that the general trend is accurately captured by the  $\delta N$  formulae. However, one may notice that the agreement is slightly less good at third order than at second, and also, that the  $\delta N$  formulae tend to slightly overestimate  $\tau_{\text{NL}}$  and slightly underestimate  $g_{\text{NL}}$ . But given the quickness of the  $\delta N$  calculation and the complexity of the third-order equations of motion, the agreement is pretty impressive. Of course, the  $\delta N$  formula was derived subject to the instantaneous bending approximation. Without this approximation, we would expect a numerical scheme that uses the  $\delta N$  formalism to yield results in close agreement with the numerical results.

**4.4. Kinetic Conversion.** As discussed in more detail at the end of Section 3, in models motivated by M-theory a bend in the trajectory happens naturally just before the big crunch, during the final approach of the two end-of-the-world branes. This bend takes place after the ekpyrotic phase has come to an end and while the evolution of the universe is dominated by the kinetic energies of the scalar fields—see also Figure 5. Again, there exists a simple approach to estimate the nongaussianity parameters of the curvature perturbation generated by this process of kinetic conversion, and we will review it here. This simplified approach was first presented in [55], based on previous work in [44, 50, 54, 56].

This estimating procedure is based on the fact that the physics of the kinetic phase is really quite simple, and moreover, except for the fact that its initial conditions involve the entropy perturbation  $\delta s$ , the kinetic phase has no memory of the details of the ekpyrotic phase. In particular, only the total  $\delta s$  in (78) matters, and the way we choose to decompose it into linear, second-, and third-order parts is irrelevant at this point. What is more, since  $\delta s_L \ll 1$ , the second- and third-order terms in (78) are highly subdominant, and we can approximate the evolution of  $\delta s$  by that of the linear term  $\delta s_L$  throughout the kinetic phase. This realization is the first ingredient of the calculation.

The second is a compact and very useful expression for the evolution of the curvature perturbation  $\zeta$  on large scales and on surfaces of constant energy density [53, 54]:

$$\dot{\zeta} = \frac{2\bar{H}\delta V}{\bar{\sigma}^2 - 2\delta V}, \quad (89)$$

where  $\delta V \equiv V(t, x^i) - \bar{V}(t)$  and a bar denotes a background quantity. This equation is exact in the limit where spatial gradients can be neglected and can thus be expanded up

to the desired order in perturbation theory if required. If expanded, it corresponds precisely to (79) [55].

The third and last ingredient of the calculation is the simple relationship between  $\delta V$  and  $\delta s$  during the conversion process. As we saw earlier, during the ekpyrotic phase, the curvature perturbation picks up a blue spectrum and is hence completely negligible on large scales. To be precise, since  $\delta V \neq 0$  during ekpyrosis, there is already some conversion of entropy into curvature perturbations occurring at this stage. However, this contribution is entirely negligible compared to the subsequent conversion (note that since  $V_{,s} = 0$  during ekpyrosis,  $\delta V$  starts out at subleading order), and hence we can take  $\zeta(t_{ek\text{-end}}) \approx 0$ . Moreover, as we saw when we were discussing the linear perturbations, at the end of the conversion process  $\zeta$  is still significantly smaller than  $\delta s$ , being given by

$$\zeta_L \approx \frac{1}{5} \delta s_L. \quad (90)$$

Hence, during the conversion process, we can take the potential to depend only on  $\delta s$ . And since the repulsive potential is monotonic and we are interested in small departures  $\delta s \ll 1$  from the background trajectory, it is intuitively clear that  $\delta V$  is directly proportional to  $\delta s$  during the bending. A numerical calculation readily confirms this simple relationship. (For completeness, we mention that the curvature perturbation can also be sourced by perturbations in the comoving energy density [41]. However, the ekpyrotic phase massively suppresses comoving energy density perturbations on large scales; since the kinetic phase is relatively short, they do not have time to grow and become significant, so that we can neglect them in our calculation—see also [56].)

During the conversion, the effect of the repulsive potential is to cause the entropy perturbation to behave approximately sinusoidally, as shown in (57). As we will confirm below, during the conversion process  $\delta V \ll \bar{\sigma}^2$ , so that (89) simplifies further to

$$\dot{\zeta} \approx \frac{2\bar{H}}{\bar{\sigma}^2} \delta V, \quad (91)$$

which, when integrated and upon use of (55) immediately reproduces (90). But, as argued above,  $\delta s$  as a whole must behave approximately in this way during the conversion phase, and subsequently analogous relationships must hold at higher orders too:

$$\zeta^{(2)} \approx \frac{1}{5} \frac{\kappa_3 \sqrt{\epsilon}}{8} \delta s_L^2, \quad \zeta^{(3)} \approx \frac{1}{5} \left( \frac{\kappa_4}{60} + \frac{\kappa_3^2}{80} - \frac{2}{5} \right) \epsilon \delta s_L^3. \quad (92)$$

These expressions immediately allow us to calculate the nonlinearity parameters

$$f_{\text{NL}} \equiv \frac{5}{3} \frac{\zeta^{(2)}}{\zeta_L^2} \approx 3\kappa_3 \sqrt{\epsilon}, \quad (93)$$

$$g_{\text{NL}} \equiv \frac{25}{9} \frac{\zeta^{(3)}}{\zeta_L^3} \approx 70 \left( \frac{\kappa_4}{60} + \frac{\kappa_3^2}{80} - \frac{2}{5} \right) \epsilon.$$

TABLE 1: Ekpyrotic conversion: the values of the nonlinearity parameters estimated by the  $\delta N$  formalism (using the instantaneous bending approximation) compared to the numerical results obtained by directly integrating the equations of motion, for potentials of the form  $V = -V_1 e^{-c_1 \phi_1} - V_2 e^{-c_2 \phi_2}$ .

$c_1$	$c_2$	$f_{\text{NL},\delta N}$	$\tau_{\text{NL},\delta N}$	$g_{\text{NL},\delta N}$	$f_{\text{NL}}$	$\tau_{\text{NL}}$	$g_{\text{NL}}$
10	10	-41.67	2500	2315	-39.95	2298	2591
10	15	-41.67	2500	2315	-40.45	2356	2813
10	20	-41.67	2500	2315	-40.62	2377	3030
15	10	-93.75	12660	11720	-91.01	11930	13100
15	15	-93.75	12660	11720	-92.11	12220	13830
15	20	-93.75	12660	11720	-92.49	12320	14440
20	10	-166.7	40000	37040	-162.5	38020	41320
20	15	-166.7	40000	37040	-164.4	38930	43170
20	20	-166.7	40000	37040	-165.1	39240	44490

Thus, without much work at all, we find these simple estimates for the nonlinearity parameters.

Before discussing this result, let us briefly pause to verify the approximation made in obtaining (91): during the kinetic phase, we can rewrite (89) as

$$\dot{\zeta} = \frac{t \delta V}{1 - 3t^2 \delta V}. \quad (94)$$

The approximation made above consists in writing  $\dot{\zeta} \approx t \delta V$ , and this leads to  $\zeta \approx (1/2)t_{\text{bend}}^2 \delta V(t_{\text{bend}})$ . But we know that by the end of the conversion process  $\zeta \approx (1/5)\delta s$ , and hence we find that

$$3t_{\text{bend}}^2 \delta V \approx \delta s \ll 1, \quad (95)$$

which shows that the approximation is self-consistent and confirms the validity of (91).

The results above indicate that the nonlinearity that was present in the entropy perturbation gets transferred straightforwardly (i.e., only with an overall numerical coefficient) to the nonlinearity in the curvature perturbation, essentially due to the simplicity of the kinetic phase. This calculation therefore leads us to expect no significant additional constant terms in  $f_{\text{NL}}$  or constants and  $\kappa_3$ -dependent terms in  $g_{\text{NL}}$ ; we will see shortly that this expectation is indeed borne out.

We can now compare these estimates to the results of brute-force numerical integration [50]. As discussed in Section 3, the repulsive potential can in principle be calculated, given a specific matter configuration on the negative-tension brane [43]. Here, we consider four different examples for the repulsive potential

$$V_{\text{rep}} \propto \phi_2^{-2}, \phi_2^{-2} + \phi_2^{-6}, (\sinh \phi_2)^{-2}, \quad (96)$$

$$(\sinh \phi_2)^{-2} + (\sinh \phi_2)^{-4},$$

with the overall magnitude adjusted in order to obtain various values for the duration of the conversion (see below). These potential forms should give an indication of the range of values that one can expect the nonlinearity parameters to take. Note that, without loss of generality, we take the

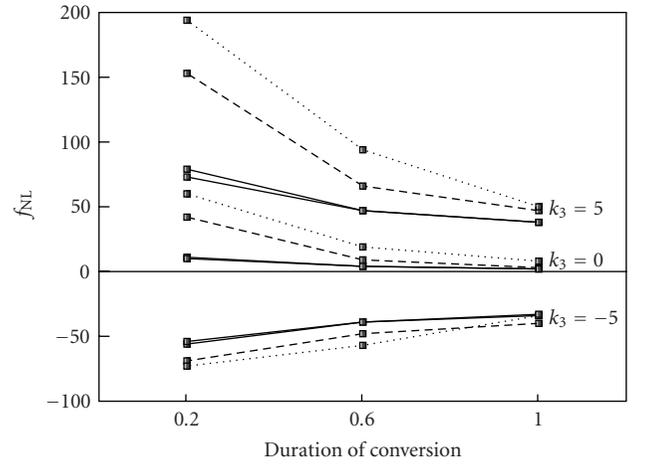


FIGURE 6:  $f_{\text{NL}}$  as a function of the duration of conversion and for the values  $\kappa_3 = -5, 0, 5$  and  $\epsilon = 36$ . In each case, we have plotted the results for four different reflection potentials, with the simplest potentials ( $\phi_2^{-2}$ ,  $(\sinh \phi_2)^{-2}$ ) indicated by solid lines, while the dashed ( $(\sinh \phi_2)^{-2} + (\sinh \phi_2)^{-4}$ ) and dotted ( $\phi_2^{-2} + \phi_2^{-6}$ ) lines give an indication of the range of values that can be expected. As the conversions become smoother, the predicted range of values narrows, and smooth conversions lead to a natural range of about  $-50 \lesssim f_{\text{NL}} \lesssim +60$  or so.

boundary to be located at  $\phi_2 = 0$ , and we only consider conversions during which  $\dot{\theta} > 0$ —other cases can be related to these by an appropriate change of coordinates.

An important parameter turns out to be the duration of the conversion [50, 56], measured by the number of Hubble times during which most, say 90 percent, of the conversion takes place, that is, the duration is measured by the number of e-folds by which the scale factor shrinks during conversion. Conversions lasting less than about 0.2 Hubble times correspond to what we call sharp conversions while those that last on the order of 1 Hubble time correspond to smooth conversions. It turns out that the estimating procedure presented above works best for the case of smooth conversions. For  $f_{\text{NL}}$ , the range of predicted values is considerably narrower as the conversion becomes

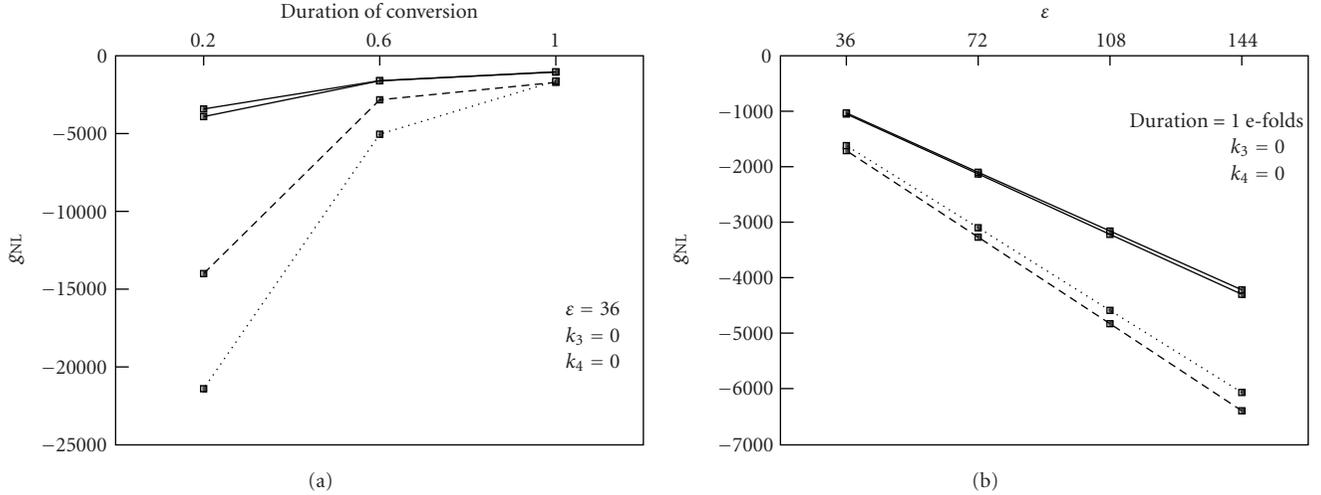


FIGURE 7: (a)  $g_{\text{NL}}$  is shown as a function of the duration of conversion, with  $\kappa_3 = \kappa_4 = 0$  and for four different reflection potentials, and with the same line style assignments as in Figure 6. As the conversions become smoother, the predicted range of values narrows considerably, allowing rather definite predictions. (b)  $g_{\text{NL}}$  can be seen to be proportional to  $\epsilon$ , that is, proportional to the equation of state  $w_{\text{ek}}$ .

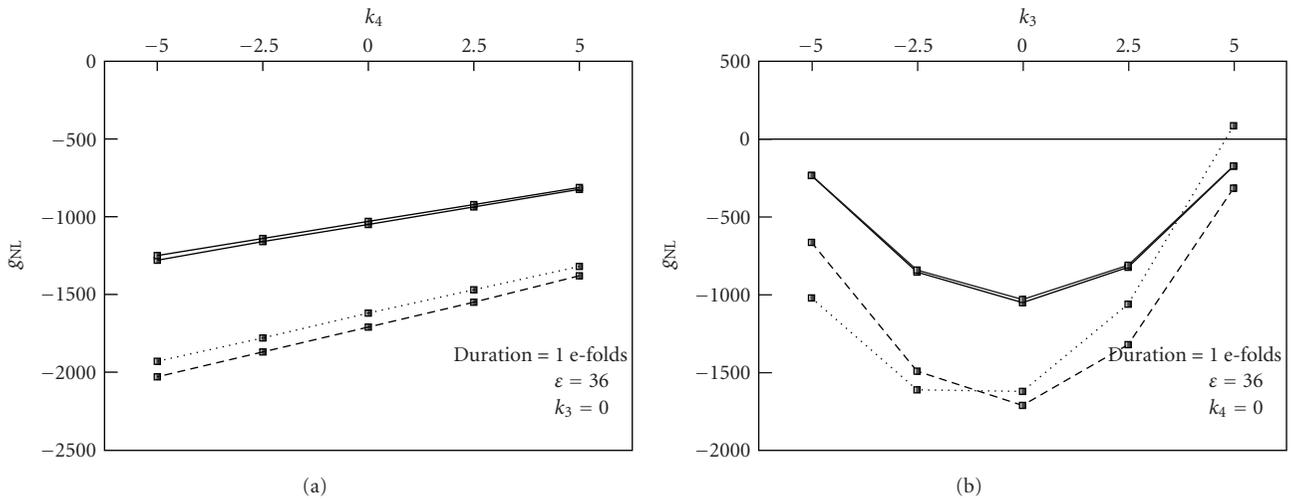


FIGURE 8: (a) indicates that  $g_{\text{NL}}$  depends linearly on  $\kappa_4$ , the parameter we are using to specify the fourth derivative of the ekpyrotic potential with respect to the entropic direction. Similarly, (b)  $g_{\text{NL}}$  can be seen to depend approximately quadratically on  $\kappa_3$ , the parameter determining the third derivative of the ekpyrotic potential with respect to the entropic direction.

smoother [44, 56], as illustrated in Figure 6. Moreover,  $f_{\text{NL}}$  can be well fitted by

$$f_{\text{NL}} = \frac{3}{2}\kappa_3\sqrt{\epsilon} + 5, \quad (97)$$

in good agreement with the simple estimate (93).

We will show the results for  $g_{\text{NL}}$  slightly more explicitly; they are illustrated in Figures 7 and 8. In each case, we have plotted the results obtained for the four repulsive potentials (96). The left panel of Figure 7 shows that, even more so than for  $f_{\text{NL}}$ , the range of predicted values for  $g_{\text{NL}}$  narrows drastically as the conversion process becomes smoother. In fact, for sharp conversions, typical values are very large in magnitude, and we expect these to be observationally disfavored shortly, if they are not ruled out

already. Thus, phenomenologically speaking, it is much more interesting to focus on smooth conversions. The right panel in Figure 7 shows that  $g_{\text{NL}}$  is proportional to  $\epsilon$  while Figure 8 indicates that  $g_{\text{NL}}$  depends linearly on  $\kappa_4$  and approximately quadratically on  $\kappa_3$ . In fact, all numerical results indicate that  $g_{\text{NL}}$  scales with  $\epsilon$ ,  $\kappa_3$ , and  $\kappa_4$  exactly as the third-order coefficient in the expression (78) for the entropy perturbation during the ekpyrotic phase, and the data can be fitted by the approximate formula

$$g_{\text{NL}} \approx 100 \left( \frac{\kappa_4}{60} + \frac{\kappa_3^2}{80} - \frac{2}{5} \right) \epsilon, \quad (98)$$

again in good agreement with the estimate (101).

## 5. Discussion of the Results

In this paper, we have focussed on the predictions for nongaussianity generated via the entropic mechanism in ekpyrotic models and their cyclic extensions. The reason for concentrating on the entropic mechanism is that this is currently the most robust, best-motivated, and best-understood mechanism for generating nearly scale-invariant curvature perturbations during a contracting phase of the universe. As indicated by its name, the entropic mechanism achieves this by first generating nearly scale-invariant entropy perturbations, which are subsequently converted into curvature perturbations. There are essentially two distinct possibilities for when this conversion can happen: either directly during the ekpyrotic phase (*ekpyrotic conversion*) or during the subsequent phase of scalar field kinetic energy domination right before the big crunch/big bang transition (*kinetic conversion*). In both cases, the curvature perturbations pick up nongaussian corrections of the local form, although the magnitude of the second- and third-order nonlinearity parameters  $f_{\text{NL}}$  and  $g_{\text{NL}}$  differ substantially for the two modes of conversion. For convenience, we will repeat the predictions here. (We should caution the reader that what we are presenting here is the predictions from the ekpyrotic phase alone, that is, we have assumed that the primordial density perturbations generated by the ekpyrotic phase have not been modified by the dynamics of the big crunch/big bang transition, nor that there are additional effects of relevance during the first stages of the subsequent expanding phase of the universe (it is conceivable, for example, that models might be constructed which utilize the ekpyrotic phase, but where there are additional fields that are relevant during the expanding phase and which would contribute to the curvature perturbation—in this case, the final predictions will evidently be model dependent).)

$$f_{\text{NL}} = -\frac{5}{12}c_1^2 \quad \text{Ekpyrotic Conversion,} \quad (99)$$

$$g_{\text{NL}} = \frac{25}{108}c_1^4,$$

$$f_{\text{NL}} = \frac{3}{2}\kappa_3\sqrt{\epsilon} + 5 \quad \text{Kinetic Conversion,} \quad (100)$$

$$g_{\text{NL}} = \left(\frac{5}{3}\kappa_4 + \frac{5}{4}\kappa_3^2 - 40\right)\epsilon.$$

For ekpyrotic conversion, the results are presented in terms of the parameters of the ekpyrotic potential as expressed in (17). The first thing to note is that the signs are unambiguously fixed:  $f_{\text{NL}}$  is predicted to be always negative while  $g_{\text{NL}}$  is always positive. In order for the power spectrum of the perturbations to be in agreement with observations, one needs  $c_1 \gtrsim 10$ , which implies  $f_{\text{NL}} \lesssim -40$  and  $g_{\text{NL}} \gtrsim 2500$ . The current observational bounds are that  $f_{\text{NL}} = 38 \pm 21$ , where the errors are quoted at  $1\sigma$  [57], while currently no strong constraints exist yet on  $g_{\text{NL}}$ . These values put the predicted values for  $f_{\text{NL}}$  for ekpyrotic conversion at  $4\sigma$  or more from the central value, and thus this type of conversion is on the verge of being ruled out by observations. What this means for model building is that models that have made

use of the ekpyrotic conversion mechanism, such as new ekpyrotic models [10, 12], might have to be modified in a way such that the conversion occurs only once the equation of state becomes small. This is not unnatural, as the ekpyrotic phase must come to an end anyhow before the bounce phase begins, and if the conversion occurs after the end of the ekpyrotic phase, the predictions will be closer to those predicted by the kinetic conversion mechanism, which we turn our attention to now.

For kinetic conversion, the results (100) are presented in terms of the parameters of the ekpyrotic potential as written in (20). The parameters  $\kappa_3, \kappa_4$  are expected to be of  $\mathcal{O}(1)$ . The fast-roll parameter  $\epsilon$  is typically of  $\mathcal{O}(10^2)$ , with a lower bound of about 50 in order for the power spectrum to be in agreement with observations. Thus, the fitting formulae predict the bispectrum parameter  $f_{\text{NL}}$  to be of order a few tens, with the sign being typically determined by the sign of  $\kappa_3$ . This is in good agreement with current bounds for very natural values of the parameters, such as  $\epsilon = 100$  and  $-1 \lesssim \kappa_3 \lesssim 5$ , for example. The associated prediction for the trispectrum  $g_{\text{NL}}$  is rather interesting, owing to the fact that the  $\kappa_3, \kappa_4$ -independent contribution is quite large. It sets the “central” value of the prediction at  $-40\epsilon$  and thus tends to make  $g_{\text{NL}}$  of order a few thousand and *negative* in sign. Even with largish values of  $\kappa_3 \sim 5$  and  $\kappa_4 \sim 5$ , the prediction is still negative  $g_{\text{NL}} \sim -\epsilon$ , and hence the negative sign of  $g_{\text{NL}}$  is a rather robust prediction of the kinetic conversion mechanism. For completeness, we note that the second trispectrum parameter  $\tau_{\text{NL}}$  is always given in terms of  $f_{\text{NL}}$  according to (66), and thus it is predicted to be positive and of order a few hundred. Observational limits on  $\tau_{\text{NL}}$  will thus provide a consistency check of the predictions and will give an indication whether or not the fluctuations of a single field (in this case the entropy field) were at the origin of the final spectrum of curvature perturbations.

What are the implications of these predictions? Assuming that the observational error bars quoted above will shrink in the near future and that the ekpyrotic conversion mechanism will be ruled out, we will focus here on the predictions of the kinetic conversion mechanism. The most interesting feature is that the natural ranges of the nonlinearity parameters are at a level that will be observable by near-future experiments, perhaps already with the PLANCK satellite. Hence, in a few years, we will know whether or not ekpyrotic models, in combination with the entropic mechanism, will be viable or even preferred by the data. In this vein, it is useful to briefly contrast the predictions discussed here with those of inflationary models: for simple, single-field inflationary models the predicted values for all nonlinearity parameters are very small, of order 1 or smaller [48, 58]. (Here, we only contrast the predictions for nongaussianity of the local form. Noncanonical kinetic terms additionally generate nongaussianities with different shapes in momentum space, such as equilateral triangles in the case of the bispectrum. This is well understood for inflation but has not been much explored yet for ekpyrotic models—though see [59] for a related study.) Even though these values lie within the predicted ranges of the kinetic conversion mechanism, it is clear that, since these values arise naturally for single-field

inflationary models whereas obtaining the same small values in ekpyrosis would require a very finely tuned potential, in case of a nondetection of nongaussianity the simple inflationary models will be strongly preferred. Conversely, since multifield inflationary models can give rise to just about any values of the nonlinearity parameters [60], in case of a detection of nongaussianity compatible with the values predicted by the kinetic conversion mechanism, those latter models will be preferred by the data. At that point it will be necessary to include also the observational evidence for or against primordial scale-invariant gravitational waves. Such gravity waves would strongly favor inflationary models whereas their absence is compatible with ekpyrotic models. The absence of primordial scale-invariant gravitational waves, combined with local nongaussianity parameters in the ranges predicted by (100), would be nothing short of spectacular, as they would point towards the existence of a contracting phase prior to the big bang and open up the possibility of a multiverse in time, with correspondingly vast timescales unlike anything we are used to in cosmology right now.

## Acknowledgments

I would like to thank Sébastien Renaux-Petel and Paul Steinhardt for very enjoyable collaborations on the subject of this paper. I would also like to thank Justin Khoury, Kazuya Koyama, Burt Ovrut, Neil Turok, and David Wands for stimulating and informative discussions.

## References

- [1] D. Baumann, “TASI lectures on inflation,” <http://arxiv.org/abs/0907.5424>.
- [2] R. Brustein and P. J. Steinhardt, “Challenges for superstring cosmology,” *Physics Letters B*, vol. 302, no. 2-3, pp. 196–201, 1993.
- [3] A. Borde, A. H. Guth, and A. Vilenkin, “Inflationary spacetimes are incomplete in past directions,” *Physical Review Letters*, vol. 90, no. 15, Article ID 151301, 4 pages, 2003.
- [4] A. H. Guth, “Eternal inflation and its implications,” *Journal of Physics A*, vol. 40, no. 25, pp. 6811–6826, 2007.
- [5] J.-L. Lehners, “Ekpyrotic and cyclic cosmology,” *Physics Reports*, vol. 465, no. 6, pp. 223–263, 2008.
- [6] D. Seery and J. E. Lidsey, “Primordial non-Gaussianities in single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 9, article 011, 2005.
- [7] J. Khoury, B. A. Ovrut, P. J. Steinhardt, and N. Turok, “Ekpyrotic universe: colliding branes and the origin of the hot big bang,” *Physical Review D*, vol. 64, no. 12, Article ID 123522, 2001.
- [8] J. K. Erickson, S. Gratton, P. J. Steinhardt, and N. Turok, “Cosmic perturbations through the cyclic ages,” *Physical Review D*, vol. 75, no. 12, Article ID 123507, 2007.
- [9] E. Komatsu, J. Dunkley, M. R.olta et al., “Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [10] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, “New ekpyrotic cosmology,” *Physical Review D*, vol. 76, no. 12, Article ID 123503, 2007.
- [11] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, “On the initial conditions in new ekpyrotic cosmology,” *Journal of High Energy Physics*, vol. 2007, no. 11, article 076, 2007.
- [12] P. Creminelli and L. Senatore, “A smooth bouncing cosmology with scale invariant spectrum,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, article 010, 2007.
- [13] N. Arkani-Hamed, H.-C. Cheng, M. A. Luty, and S. Mukohyama, “Ghost condensation and a consistent infrared modification of gravity,” *Journal of High Energy Physics*, vol. 8, no. 5, article 074, pp. 1729–1764, 2004.
- [14] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, “Causality, analyticity and an IR obstruction to UV completion,” *Journal of High Energy Physics*, vol. 2006, no. 10, article 014, 2006.
- [15] R. Kallosh, J. U. Kang, A. D. Linde, and V. Mukhanov, “The new ekpyrotic ghost,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 4, article 018, 2008.
- [16] P. J. Steinhardt and N. Turok, “A cyclic model of the universe,” *Science*, vol. 296, no. 5572, pp. 1436–1439, 2002.
- [17] P. J. Steinhardt and N. Turok, “Cosmic evolution in a cyclic universe,” *Physical Review D*, vol. 65, no. 12, Article ID 126003, 2002.
- [18] A. Lukas, B. A. Ovrut, K. S. Stelle, and D. Waldram, “Universe as a domain wall,” *Physical Review D*, vol. 59, no. 8, Article ID 086001, 1999.
- [19] P. Hořava and E. Witten, “Heterotic and type I string dynamics from eleven dimensions,” *Nuclear Physics B*, vol. 460, no. 3, pp. 506–524, 1996.
- [20] N. Turok, M. Perry, and P. J. Steinhardt, “M theory model of a big crunch/big bang transition,” *Physical Review D*, vol. 70, no. 10, Article ID 106004, 2004.
- [21] J.-L. Lehners, P. McFadden, and N. Turok, “Colliding branes in heterotic M theory,” *Physical Review D*, vol. 75, no. 10, Article ID 103510, 2007.
- [22] J.-L. Lehners, P. McFadden, and N. Turok, “Effective actions for heterotic M theory,” *Physical Review D*, vol. 76, no. 2, Article ID 023501, 2007.
- [23] K. Koyama and D. Wands, “Ekpyrotic collapse with multiple fields,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 4, article 008, 2007.
- [24] K. Koyama, S. Mizuno, and D. Wands, “Curvature perturbations from ekpyrotic collapse with multiple fields,” *Classical and Quantum Gravity*, vol. 24, no. 15, pp. 3919–3931, 2007.
- [25] C. Gordon, D. Wands, B. A. Bassett, and R. Maartens, “Adiabatic and entropy perturbations from inflation,” *Physical Review D*, vol. 63, no. 2, Article ID 023506, 2001.
- [26] J.-L. Lehners, P. McFadden, N. Turok, and P. J. Steinhardt, “Generating ekpyrotic curvature perturbations before the big bang,” *Physical Review D*, vol. 76, no. 10, Article ID 103501, 2007.
- [27] A. J. Tolley and D. H. Wesley, “Scale invariance in expanding and contracting universes from two-field models,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 5, article 006, 2007.
- [28] J.-L. Lehners and P. J. Steinhardt, “Dark energy and the return of the phoenix universe,” *Physical Review D*, vol. 79, no. 6, Article ID 063503, 2009.
- [29] J.-L. Lehners, P. J. Steinhardt, and N. Turok, “The return of the phoenix universe,” *International Journal of Modern Physics D*, vol. 18, no. 14, pp. 2231–2235, 2009.
- [30] L. A. Boyle, P. J. Steinhardt, and N. Turok, “Cosmic gravitational-wave background in a cyclic universe,” *Physical Review D*, vol. 69, no. 12, Article ID 127302, 2004.

- [31] D. Baumann, P. Steinhardt, K. Takahashi, and K. Ichiki, "Gravitational wave spectrum induced by primordial scalar perturbations," *Physical Review D*, vol. 76, no. 8, Article ID 084019, 2007.
- [32] J. Khoury, B. A. Ovrut, P. J. Steinhardt, and N. Turok, "Density perturbations in the ekpyrotic scenario," *Physical Review D*, vol. 66, no. 4, Article ID 046005, 2002.
- [33] P. Creminelli, A. Nicolis, and M. Zaldarriaga, "Perturbations in bouncing cosmologies: dynamical attractor versus scale invariance," *Physical Review D*, vol. 71, no. 6, Article ID 063505, 12 pages, 2005.
- [34] D. H. Lyth, "The primordial curvature perturbation in the ekpyrotic universe," *Physics Letters B*, vol. 524, no. 1-2, pp. 1-4, 2002.
- [35] A. J. Tolley, N. Turok, and P. J. Steinhardt, "Cosmological perturbations in a big-crunch—big-bang space-time," *Physical Review D*, vol. 69, no. 10, Article ID 106005, 2004.
- [36] P. L. McFadden, N. Turok, and P. J. Steinhardt, "Solution of a braneworld big crunch/big bang cosmology," *Physical Review D*, vol. 76, no. 10, Article ID 104038, 2007.
- [37] J. Khoury and P. J. Steinhardt, "Adiabatic ekpyrosis: scale-invariant curvature perturbations from a single scalar field in a contracting universe," *Physical Review Letters*, vol. 104, no. 9, 2010.
- [38] A. Linde, V. Mukhanov, and A. Vikman, "On adiabatic perturbations in the ekpyrotic scenario," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 2, article 006, 2010.
- [39] F. Finelli, "A contracting universe driven by two scalar fields," *Physics Letters B*, vol. 545, no. 1-2, pp. 1-7, 2002.
- [40] A. Notari and A. Riotto, "Isocurvature perturbations in the Ekpyrotic Universe," *Nuclear Physics B*, vol. 644, no. 1-2, pp. 371-382, 2002.
- [41] D. Langlois and F. Vernizzi, "Non-linear perturbations of cosmological scalar fields," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 2, article 017, 2007.
- [42] J. Khoury, P. J. Steinhardt, and N. Turok, "Inflation versus cyclic predictions for spectral tilt," *Physical Review Letters*, vol. 91, no. 16, Article ID 161301, 4 pages, 2003.
- [43] J.-L. Lehners and N. Turok, "Bouncing negative-tension branes," *Physical Review D*, vol. 77, no. 2, Article ID 023516, 2008.
- [44] J.-L. Lehners and P. J. Steinhardt, "Intuitive understanding of non-Gaussianity in ekpyrotic and cyclic models," *Physical Review D*, vol. 78, no. 2, Article ID 023506, 2008.
- [45] T. Battefeld, "Modulated perturbations from instant preheating after new ekpyrosis," *Physical Review D*, vol. 77, no. 6, Article ID 063503, 2008.
- [46] C. T. Byrnes, M. Sasaki, and D. Wands, "Primordial trispectrum from inflation," *Physical Review D*, vol. 74, no. 12, Article ID 123519, 2006.
- [47] D. Babich, P. Creminelli, and M. Zaldarriaga, "The shape of non-Gaussianities," *Journal of Cosmology and Astroparticle Physics*, vol. 2004, no. 8, article 009, pp. 199-217, 2004.
- [48] J. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," *Journal of High Energy Physics*, vol. 7, no. 5, article 013, pp. 233-264, 2003.
- [49] K. Koyama, S. Mizuno, F. Vernizzi, and D. Wands, "Non-Gaussianities from ekpyrotic collapse with multiple fields," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 11, article 024, pp. 1-21, 2007.
- [50] J.-L. Lehners and S. Renaux-Petel, "Multifield cosmological perturbations at third order and the ekpyrotic trispectrum," *Physical Review D*, vol. 80, no. 6, Article ID 063503, 2009.
- [51] A. A. Starobinsky, "Multicomponent de Sitter (inflationary) stages and the generation of perturbations," *Journal of Experimental and Theoretical Physics Letters*, vol. 42, no. 3, p. 152, 1985.
- [52] M. Sasaki and E. D. Stewart, "A general analytic formula for the spectral index of the density perturbations produced during inflation," *Progress of Theoretical Physics*, vol. 95, no. 1, pp. 71-78, 1996.
- [53] D. H. Lyth, K. A. Malik, and M. Sasaki, "A general proof of the conservation of the curvature perturbation," *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 5, article 004, pp. 50-69, 2005.
- [54] E. I. Buchbinder, J. Khoury, and B. A. Ovrut, "Non-Gaussianities in new ekpyrotic cosmology," *Physical Review Letters*, vol. 100, no. 17, Article ID 171302, 2008.
- [55] J.-L. Lehners and P. J. Steinhardt, "Non-Gaussianity generated by the entropic mechanism in bouncing cosmologies made simple," *Physical Review D*, vol. 80, no. 10, Article ID 103520, 2009.
- [56] J.-L. Lehners and P. J. Steinhardt, "Non-Gaussian density fluctuations from entropically generated curvature perturbations in ekpyrotic models," *Physical Review D*, vol. 77, no. 6, Article ID 063533, 2008.
- [57] K. M. Smith, L. Senatore, and M. Zaldarriaga, "Optimal limits on fNLlocal from WMAP 5-year data," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article 006, 2009.
- [58] D. Seery and J. E. Lidsey, "Non-Gaussianity from the inflationary trispectrum," *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 1, article 008, 2007.
- [59] J. Khoury and F. Piazza, "Rapidly-varying speed of sound, scale invariance and non-Gaussian signatures," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 7, article 026, 2009.
- [60] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, "Non-Gaussianity from inflation: theory and observations," *Physics Reports*, vol. 402, no. 3-4, pp. 103-266, 2004.

## Review Article

# Second-Order Gauge-Invariant Cosmological Perturbation Theory: Current Status

**Kouji Nakamura**

*Optical and Infrared Astronomy Division, National Astronomical Observatory of Japan, Osawa, Mitaka, Tokyo 181-8588, Japan*

Correspondence should be addressed to Kouji Nakamura, [kouji.nakamura@nao.ac.jp](mailto:kouji.nakamura@nao.ac.jp)

Received 19 April 2010; Accepted 12 July 2010

Academic Editor: Eiichiro Komatsu

Copyright © 2010 Kouji Nakamura. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The current status of the recent developments of the second-order gauge-invariant cosmological perturbation theory is reviewed. To show the essence of this perturbation theory, we concentrate only on the universe filled with a single scalar field. Through this paper, we point out the problems which should be clarified for the further theoretical sophistication of this perturbation theory. We also expect that this theoretical sophistication will be also useful to discuss the theoretical predictions of non-Gaussianity in CMB and comparison with observations.

## 1. Introduction

The general relativistic cosmological *linear* perturbation theory has been developed to a high degree of sophistication during the last 30 years [1–3]. One of the motivations of this development was to clarify the relation between the scenarios of the early universe and cosmological data, such as the cosmic microwave background (CMB) anisotropies. Recently, the first-order approximation of our universe from a homogeneous isotropic one was revealed through the observation of the CMB by the Wilkinson Microwave Anisotropy Probe (WMAP) [4, 5], the cosmological parameters are accurately measured, we have obtained the standard cosmological model, and the so-called “precision cosmology” has begun. These developments in observations were also supported by the theoretical sophistication of the linear-order cosmological perturbation theory.

The observational results of CMB also suggest that the fluctuations of our universe are adiabatic and Gaussian at least in the first-order approximation. We are now on the stage to discuss the deviation from this first-order approximation from the observational [5] and theoretical sides [6–21] through the non-Gaussianity, the nonadiabaticity, and so on. These will be goals of future satellite missions. With the increase of precision of the CMB data, the study of relativistic cosmological perturbations beyond linear order is a topical

subject. The *second-order* cosmological perturbation theory is one of such perturbation theories beyond linear order.

Although the second-order perturbation theory in general relativity is an old topic, a general framework of the gauge-invariant formulation of the general relativistic second-order perturbation has been proposed [22, 23]. This general formulation is an extension of the works of Bruni et al. [24] and has also been applied to cosmological perturbations: the derivation of the second-order Einstein equation in a gauge-invariant manner without any gauge fixing [25, 26]; applicability in more generic situations [27]; confirmation of the consistency between all components of the second-order Einstein equations and equations of motions [28]. We also note that the radiation case has recently been discussed by treating the Boltzmann equation up to second order [29, 30] along the gauge-invariant manner of the above series of papers by the present author.

In this paper, we summarize the current status of this development of the second-order gauge-invariant cosmological perturbation theory through the simple system of a scalar field. Through this paper, we point out the problems which should be clarified and directions of the further development of the theoretical sophistication of the general relativistic higher-order perturbation theory, especially in cosmological perturbations. We expect that this sophistication will be also useful to discuss the theoretical predictions

of non-Gaussianity in CMB and comparison with observations.

The organization of this paper is as follows. In Section 2, we review the general framework of the second-order gauge-invariant perturbation theory developed in [22, 23, 25, 26, 31]. This review also includes additional explanation not given in those papers. In Section 3, we present also the derivations of the second-order perturbation of the Einstein equation and the energy-momentum tensor from general point of view. For simplicity, in this paper, we only consider a single scalar field as a matter content. The ingredients of Sections 2 and 3 will be applicable to perturbation theory in any theory with general covariance, if the decomposition formula (23) for the linear-order metric perturbation is correct. In Section 4, we summarize the Einstein equations in the case of a background homogeneous isotropic universe, which are used in the derivation of the first- and second-order Einstein equations. In Section 5, the first-order perturbation of the Einstein equations and the Klein-Gordon equations are summarized. The derivation of the second-order perturbations of the Einstein equations and the Klein-Gordon equations, and their consistency are reviewed in Section 6. The final section, Section 7, is devoted to a summary and discussions.

## 2. General Framework of the General Relativistic Gauge-Invariant Perturbation Theory

In this section, we review the general framework of the gauge-invariant perturbation theory developed in [22–26, 31–39]. To develop the general relativistic gauge-invariant perturbation theory, we first explain the general arguments of the Taylor expansion on a manifold without introducing an explicit coordinate system in Section 2.1. Further, we also have to clarify the notion of “gauge” in general relativity to develop the gauge-invariant perturbation theory from general point of view, which is explained in Section 2.2. After clarifying the notion of “gauge” in general relativistic perturbations, in Section 2.3, we explain the formulation of the general relativistic gauge-invariant perturbation theory from general point of view. Although our understanding of “gauge” in general relativistic perturbations essentially is different from “degree of freedom of coordinates” as in many literature, “a coordinate transformation” is induced by our understanding of “gauge.” This situation is explained in Section 2.4. To exclude “gauge degree of freedom” which is unphysical degree of freedom in perturbations, we construct “gauge-invariant variables” of perturbations as reviewed in Section 2.5. These “gauge-invariant variables” are regarded as physical quantities.

*2.1. Taylor Expansion of Tensors on a Manifold.* First, we briefly review the issues on the general form of the Taylor expansion of tensors on a manifold  $\mathcal{M}$ . The gauge issue of general relativistic perturbation theories which we will discuss is related to the coordinate transformation. Therefore, we have to discuss the general form of the Taylor

expansion without the explicit introduction of coordinate systems. Although we only consider the Taylor expansion of a scalar function  $f : \mathcal{M} \mapsto \mathbb{R}$ , here, the resulting formula is extended to that for any tensor field on a manifold as in Appendix A. We have to emphasize that the general formula of the Taylor expansion shown here is the starting point of our gauge-invariant formulation of the second-order general relativistic perturbation theory.

The Taylor expansion of a function  $f$  is an approximated form of  $f(q)$  at  $q \in \mathcal{M}$  in terms of the variables at  $p \in \mathcal{M}$ , where  $q$  is in the neighborhood of  $p$ . To derive the formula for the Taylor expansion of  $f$ , we have to compare the values of  $f$  at the different points on the manifold. To accomplish this, we introduce a one-parameter family of diffeomorphisms  $\Phi_\lambda : \mathcal{M} \mapsto \mathcal{M}$ , where  $\Phi_\lambda(p) = q$  and  $\Phi_{\lambda=0}(p) = p$ . One example of a diffeomorphisms  $\Phi_\lambda$  is an exponential map with a generator. However, we consider a more general class of diffeomorphisms.

The diffeomorphism  $\Phi_\lambda$  induces the pull-back  $\Phi_\lambda^*$  of the function  $f$  and this pull-back enables us to compare the values of the function  $f$  at different points. Further, the Taylor expansion of the function  $f(q)$  is given by

$$\begin{aligned} f(q) &= f(\Phi_\lambda(p)) =: (\Phi_\lambda^* f)(p) \\ &= f(p) + \frac{\partial}{\partial \lambda} (\Phi_\lambda^* f) \Big|_p \lambda + \frac{1}{2} \frac{\partial^2}{\partial \lambda^2} (\Phi_\lambda^* f) \Big|_p \lambda^2 + O(\lambda^3). \end{aligned} \quad (1)$$

Since this expression hold for an arbitrary smooth function  $f$ , the function  $f$  in (1) can be regarded as a dummy. Therefore, we may regard the Taylor expansion (1) to be the expansion of the pull-back  $\Phi_\lambda^*$  of the diffeomorphism  $\Phi_\lambda$ , rather than the expansion of the function  $f$ .

According to this point of view, Sonego and Bruni [36] showed that there exist vector fields  $\xi_1^a$  and  $\xi_2^a$  such that the expansion (1) is given by

$$\begin{aligned} f(q) &= (\Phi_\lambda^* f)(p) \\ &= f(p) + (\xi_{\xi_1} f) \Big|_p \lambda + \frac{1}{2} (\xi_{\xi_2} + \xi_{\xi_1}^2) f \Big|_p \lambda^2 + O(\lambda^3), \end{aligned} \quad (2)$$

without loss of generality (see Appendix A). Equation (2) is not only the representation of the Taylor expansion of the function  $f$ , but also the definitions of the generators  $\xi_1^a$  and  $\xi_2^a$ . These generators of the one-parameter family of diffeomorphisms  $\Phi_\lambda$  represent the direction along which the Taylor expansion is carried out. The generator  $\xi_1^a$  is the first-order approximation of the flow of the diffeomorphism  $\Phi_\lambda$ , and the generator  $\xi_2^a$  is the second-order correction to this flow. We should regard the generators  $\xi_1^a$  and  $\xi_2^a$  to be independent. Further, as shown in Appendix A, the representation of the Taylor expansion of an arbitrary scalar function  $f$  is extended to that for an arbitrary tensor field  $Q$  just through the replacement  $f \rightarrow Q$ .

We must note that, in general, the representation (2) of the Taylor expansion is different from an usual exponential map which is generated by a vector field. In general,

$$\Phi_\sigma \circ \Phi_\lambda \neq \Phi_{\sigma+\lambda}, \quad \Phi_\lambda^{-1} \neq \Phi_{-\lambda}. \quad (3)$$

As noted in [24], if the second-order generator  $\xi_2$  in (2) is proportional to the first-order generator  $\xi_1$  in (2), the diffeomorphism  $\Phi_\lambda$  is reduced to an exponential map. Therefore, one may reasonably doubt that  $\Phi_\lambda$  forms a group except under very special conditions. However, we have to note that the properties (3) do not directly mean that  $\Phi_\lambda$  does not form a group. There will be possibilities that  $\Phi_\lambda$  form a group in a different sense from exponential maps, in which the properties (3) will be maintained.

Now, we give an intuitive explanation of the representation (2) of the Taylor expansion through the case where the scalar function  $f$  in (2) is a coordinate function. When two points  $p, q \in \mathcal{M}$  in (2) are in the neighborhood of each other, we can apply a coordinate system  $\mathcal{M} \mapsto \mathbb{R}^n$  ( $n = \dim \mathcal{M}$ ), which is denoted by  $\{x^\mu\}$ , to an open set which includes these two points. Then, we can measure the relative position of these two points  $p$  and  $q$  in  $\mathcal{M}$  in terms of this coordinate system in  $\mathbb{R}^n$  through the Taylor expansion (2). In this case, we may regard that the scalar function  $f$  in (2) is a coordinate function  $x^\mu$  and (2) yields

$$\begin{aligned} x^\mu(q) &= (\Phi_\lambda^* x^\mu)(p) \\ &= x^\mu(p) + \lambda \xi_1(p) + \frac{1}{2} \lambda^2 (\xi_2 + \xi_1^\nu \partial_\nu \xi_1^\mu) \Big|_p + O(\lambda^3). \end{aligned} \quad (4)$$

The second term  $\lambda \xi_1(p)$  in the right-hand side of (4) is familiar. This is regarded as the vector which point from the point  $x^\mu(p)$  to the points  $x^\mu(q)$  in the sense of the first-order correction as shown in Figure 1(a). However, in the sense of the second order, this vector  $\lambda \xi_1(p)$  may fail to point to  $x^\mu(q)$ . Therefore, it is necessary to add the second-order correction as shown in Figure 1(b). As a correction of the second order, we may add the term  $(1/2)\lambda^2 \xi_1^\nu(p) \partial_\nu \xi_1^\mu(p)$ . This second-order correction corresponds to that coming from the exponential map which is generated by the vector field  $\xi_1^\mu$ . However, this correction completely determined by the vector field  $\xi_1^\mu$ . Even if we add this correction that comes from the exponential map, there is no guarantee that the corrected vector  $\lambda \xi_1(p) + (1/2)\lambda^2 \xi_1^\nu(p) \partial_\nu \xi_1^\mu(p)$  does point to  $x^\mu(q)$  in the sense of the second order. Thus, we have to add the new correction  $(1/2)\lambda^2 \xi_2^\nu(p)$  of the second order, in general.

Of course, without this correction  $(1/2)\lambda^2 \xi_2^\nu(p)$ , the vector which comes only from the exponential map generated by the vector field  $\xi_1$  might point to the point  $x^\mu(q)$ . Actually, this is possible if we carefully choose the vector field  $\xi_1^\mu$  taking into account the deviations at the second order. However, this means that we have to take care of the second-order correction when we determine the first-order correction. This contradicts to the philosophy of the Taylor expansion as a perturbative expansion, in which we can determine everything order-by-order. Therefore, we should regard that the correction  $(1/2)\lambda^2 \xi_2^\nu(p)$  is necessary in general situations.

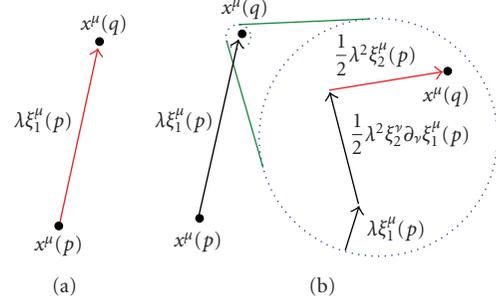


FIGURE 1: (a) The second term  $\lambda \xi_1(p)$  in (4) is the vector which points from the point  $x^\mu(p)$  to the point  $x^\mu(q)$  in the sense of the first-order correction. (b) If we look at the neighborhood of the point  $x^\mu(q)$  in detail, the vector  $\lambda \xi_1(p)$  may fail to point to  $x^\mu(q)$  in the sense of the second order. Therefore, it is necessary to add the second-order correction  $(1/2)\lambda^2(\xi_2^\mu + \xi_1^\nu \partial_\nu \xi_1^\mu)(p)$ .

**2.2. Gauge Degree of Freedom in General Relativity.** Since we want to explain the gauge-invariant perturbation theory in general relativity, first of all, we have to explain the notion of “gauge” in general relativity [31]. General relativity is a theory with general covariance, which intuitively states that there is no preferred coordinate system in nature. This general covariance also introduces the notion of “gauge” in the theory. In the theory with general covariance, these “gauges” give rise to the unphysical degree of freedom and we have to fix the “gauges” or to extract some invariant quantities to obtain physical result. Therefore, treatments of “gauges” are crucial in general relativity and this situation becomes more delicate in general relativistic perturbation theory as explained below.

In 1964, Sachs [32] pointed out that there are two kinds of “gauges” in general relativity. Sachs called these two “gauges” as the first- and the second-kind of gauges, respectively. Here, we review these concepts of “gauge.”

**2.2.1. First-Kind Gauge.** The first-kind gauge is a coordinate system on a single manifold  $\mathcal{M}$ . Although this first-kind gauge is not important in this paper, we explain this to emphasize the “gauge” discuss in this paper is different from this first-kind gauge.

In the standard text book of manifolds (e.g., see [40]), the following property of a manifold is written: on a manifold, we can always introduce a coordinate system as a diffeomorphism  $\psi_\alpha$  from an open set  $O_\alpha \subset \mathcal{M}$  to an open set  $\psi_\alpha(O_\alpha) \subset \mathbb{R}^n$  ( $n = \dim \mathcal{M}$ ). This diffeomorphism  $\psi_\alpha$ , that is, coordinate system of the open set  $O_\alpha$ , is called *gauge choice* (of the first-kind). If we consider another open set in  $O_\beta \subset \mathcal{M}$ , we have another gauge choice  $\psi_\beta : O_\beta \mapsto \psi_\beta(O_\beta) \subset \mathbb{R}^n$  for  $O_\beta$ . If these two open sets  $O_\alpha$  and  $O_\beta$  have the intersection  $O_\alpha \cap O_\beta \neq \emptyset$ , we can consider the diffeomorphism  $\psi_\beta \circ \psi_\alpha^{-1}$ . This diffeomorphism  $\psi_\beta \circ \psi_\alpha^{-1}$  is just a coordinate transformation:  $\psi_\alpha(O_\alpha \cap O_\beta) \subset \mathbb{R}^n \mapsto \psi_\beta(O_\alpha \cap O_\beta) \subset \mathbb{R}^n$ , which is called *gauge transformation* (of the first-kind) in general relativity.

According to the theory of a manifold, coordinate system are not on a manifold itself but we can always introduce a

coordinate system through a map from an open set in the manifold  $\mathcal{M}$  to an open set of  $\mathbb{R}^n$ . For this reason, general covariance in general relativity is automatically included in the premise that our spacetime is regarded as a single manifold. The first-kind gauge does arise due to this general covariance. The gauge issue of the first-kind is represented by the question, which coordinate system is convenient? The answer to this question depends on the problem which we are addressing, that is, what we want to clarify. In some case, this gauge issue of the first-kind is an important. However, in many case, it becomes harmless if we apply a covariant theory on the manifold.

*2.2.2. Second-Kind Gauge.* The *second-kind gauge* appears in perturbation theories in a theory with general covariance. This notion of the second-kind “gauge” is the main issue of this paper. To explain this, we have to remind our selves what we are doing in perturbation theories.

First, in any perturbation theories, we always treat two spacetime manifolds. One is the physical spacetime  $\mathcal{M}$ . This physical spacetime  $\mathcal{M}$  is our nature itself and we want to describe the properties of this physical spacetime  $\mathcal{M}$  through perturbations. The other is the background spacetime  $\mathcal{M}_0$ . This background spacetime have nothing to do with our nature and this is a fictitious manifold which is prepared by us. This background spacetime is just a reference to carry out perturbative analyses. We emphasize that these two spacetime manifolds  $\mathcal{M}$  and  $\mathcal{M}_0$  are distinct. Let us denote the physical spacetime by  $(\mathcal{M}, \bar{g}_{ab})$  and the background spacetime by  $(\mathcal{M}_0, g_{ab})$ , where  $\bar{g}_{ab}$  is the metric on the physical spacetime manifold,  $\mathcal{M}$ , and  $g_{ab}$  is the metric on the background spacetime manifold,  $\mathcal{M}_0$ . Further, we formally denote the spacetime metric and the other physical tensor fields on  $\mathcal{M}$  by  $Q$  and its background value on  $\mathcal{M}_0$  by  $Q_0$ .

Second, in any perturbation theories, we always write equations for the perturbation of the physical variable  $Q$  in the form

$$Q(\text{“}p\text{”}) = Q_0(p) + \delta Q(p). \quad (5)$$

Usually, this equation is simply regarded as a relation between the physical variable  $Q$  and its background value  $Q_0$ , or as the definition of the deviation  $\delta Q$  of the physical variable  $Q$  from its background value  $Q_0$ . However, (5) has deeper implications. Keeping in our mind that we always treat two different spacetimes,  $\mathcal{M}$  and  $\mathcal{M}_0$ , in perturbation theory, (5) is a rather curious equation in the following sense: the variable on the left-hand side of (5) is a variable on  $\mathcal{M}$ , while the variables on the right-hand side of (5) are variables on  $\mathcal{M}_0$ . Hence, (5) gives a relation between variables on two different manifolds.

Further, through (5), we have implicitly identified points in these two different manifolds. More specifically,  $Q(\text{“}p\text{”})$  on the left-hand side of (5) is a field on  $\mathcal{M}$ , and  $\text{“}p\text{”} \in \mathcal{M}$ . Similarly, we should regard the background value  $Q_0(p)$  of  $Q(\text{“}p\text{”})$  and its deviation  $\delta Q(p)$  of  $Q(\text{“}p\text{”})$  from  $Q_0(p)$ , which are on the right-hand side of (5), as fields on  $\mathcal{M}_0$ , and  $p \in \mathcal{M}_0$ . Because (5) is regarded as an equation for field variables, it implicitly states that the points  $\text{“}p\text{”} \in \mathcal{M}$  and

$p \in \mathcal{M}_0$  are same. This represents the implicit assumption of the existence of a map  $\mathcal{M}_0 \rightarrow \mathcal{M} : p \in \mathcal{M}_0 \mapsto \text{“}p\text{”} \in \mathcal{M}$ , which is usually called a *gauge choice* (of the second-kind) in perturbation theory [33–35].

It is important to note that the second-kind gauge choice between points on  $\mathcal{M}_0$  and  $\mathcal{M}$ , which is established by such a relation as (5), is not unique to the theory with general covariance. Rather, (5) involves the degree of freedom corresponding to the choice of the map  $\mathcal{X} : \mathcal{M}_0 \mapsto \mathcal{M}$ . This is called the *gauge degree of freedom* (of the second-kind). Such a degree of freedom always exists in perturbations of a theory with general covariance. General covariance intuitively means that there is no preferred coordinate system in the theory as mentioned above. If general covariance is not imposed on the theory, there is a preferred coordinate system in the theory, and we naturally introduce this preferred coordinate system onto both  $\mathcal{M}_0$  and  $\mathcal{M}$ . Then, we can choose the identification map  $\mathcal{X}$  using this preferred coordinate system. However, there is no such coordinate system in general relativity due to the general covariance, and we have no guiding principle to choose the identification map  $\mathcal{X}$ . Indeed, we may identify  $\text{“}p\text{”} \in \mathcal{M}$  with  $q \in \mathcal{M}_0$  ( $q \neq p$ ) instead of  $p \in \mathcal{M}_0$ . In the above understanding of the concept of “gauge” (of the second-kind) in general relativistic perturbation theory, a gauge transformation is simply a change of the map  $\mathcal{X}$ .

These are the basic ideas of gauge degree of freedom (of the second-kind) in the general relativistic perturbation theory which are pointed out by Sacks [32] and mathematically clarified by Stewart and Walker [33–35]. Based on these ideas, higher-order perturbation theory has been developed in [22–28, 31, 38, 39, 41].

*2.3. Formulation of Perturbation Theory.* To formulate the above understanding in more detail, we introduce an infinitesimal parameter  $\lambda$  for the perturbation. Further, we consider the  $4 + 1$ -dimensional manifold  $\mathcal{N} = \mathcal{M} \times \mathbb{R}$ , where  $4 = \dim \mathcal{M}$  and  $\lambda \in \mathbb{R}$ . The background spacetime  $\mathcal{M}_0 = \mathcal{N}|_{\lambda=0}$  and the physical spacetime  $\mathcal{M} = \mathcal{M}_\lambda = \mathcal{N}|_{\mathbb{R}=\lambda}$  are also submanifolds embedded in the extended manifold  $\mathcal{N}$ . Each point on  $\mathcal{N}$  is identified by a pair  $(p, \lambda)$ , where  $p \in \mathcal{M}_\lambda$ , and each point in  $\mathcal{M}_0 \subset \mathcal{N}$  is identified by  $\lambda = 0$ .

Through this construction, the manifold  $\mathcal{N}$  is foliated by four-dimensional submanifolds  $\mathcal{M}_\lambda$  of each  $\lambda$ , and these are diffeomorphic to  $\mathcal{M}$  and  $\mathcal{M}_0$ . The manifold  $\mathcal{N}$  has a natural differentiable structure consisting of the direct product of  $\mathcal{M}$  and  $\mathbb{R}$ . Further, the perturbed spacetimes  $\mathcal{M}_\lambda$  for each  $\lambda$  must have the same differential structure with this construction. In other words, we require that perturbations be continuous in the sense that  $\mathcal{M}$  and  $\mathcal{M}_0$  are connected by a continuous curve within the extended manifold  $\mathcal{N}$ . Hence, the changes of the differential structure resulting from the perturbation, for example, the formation of singularities and singular perturbations in the sense of fluid mechanics, are excluded from consideration.

Let us consider the set of field equations

$$\mathcal{E}[Q_\lambda] = 0 \quad (6)$$

on the physical spacetime  $\mathcal{M}_\lambda$  for the physical variables  $Q_\lambda$  on  $\mathcal{M}_\lambda$ . The field equation (6) formally represents the Einstein equation for the metric on  $\mathcal{M}_\lambda$  and the equations for matter fields on  $\mathcal{M}_\lambda$ . If a tensor field  $Q_\lambda$  is given on each  $\mathcal{M}_\lambda$ ,  $Q_\lambda$  is automatically extended to a tensor field on  $\mathcal{N}$  by  $Q(p, \lambda) := Q_\lambda(p)$ , where  $p \in \mathcal{M}_\lambda$ . In this extension, the field equation (6) is regarded as an equation on the extended manifold  $\mathcal{N}$ . Thus, we have extended an arbitrary tensor field and the field equations (6) on each  $\mathcal{M}_\lambda$  to those on the extended manifold  $\mathcal{N}$ .

Tensor fields on  $\mathcal{N}$  obtained through the above construction are necessarily ‘‘tangent’’ to each  $\mathcal{M}_\lambda$ . To consider the basis of the tangent space of  $\mathcal{N}$ , we introduce the normal form and its dual, which are normal to each  $\mathcal{M}_\lambda$  in  $\mathcal{N}$ . These are denoted by  $(d\lambda)_a$  and  $(\partial/\partial\lambda)^a$ , respectively, and they satisfy  $(d\lambda)_a(\partial/\partial\lambda)^a = 1$ . The form  $(d\lambda)_a$  and its dual,  $(\partial/\partial\lambda)^a$ , are normal to any tensor field extended from the tangent space on each  $\mathcal{M}_\lambda$  through the above construction. The set consisting of  $(d\lambda)_a$ ,  $(\partial/\partial\lambda)^a$ , and the basis of the tangent space on each  $\mathcal{M}_\lambda$  is regarded as the basis of the tangent space of  $\mathcal{N}$ .

Now, we define the perturbation of an arbitrary tensor field  $Q$ . We compare  $Q$  on  $\mathcal{M}_\lambda$  with  $Q_0$  on  $\mathcal{M}_0$ , and it is necessary to identify the points of  $\mathcal{M}_\lambda$  with those of  $\mathcal{M}_0$  as mentioned above. This point identification map is the gauge choice of the second-kind as mentioned above. The gauge choice is made by assigning a diffeomorphism  $\mathcal{X}_\lambda : \mathcal{N} \rightarrow \mathcal{N}$  such that  $\mathcal{X}_\lambda : \mathcal{M}_0 \rightarrow \mathcal{M}_\lambda$ . Following the paper of Bruni et al. [24], we introduce a gauge choice  $\mathcal{X}_\lambda$  as an one-parameter groups of diffeomorphisms, that is, an exponential map, for simplicity. We denote the generator of this exponential map by  ${}_x\eta^a$ . This generator  ${}_x\eta^a$  is decomposed by the basis on  $\mathcal{N}$  which are constructed above. Although the generator  ${}_x\eta^a$  should satisfy some appropriate properties [22], the arbitrariness of the gauge choice  $\mathcal{X}_\lambda$  is represented by the tangential component of the generator  ${}_x\eta^a$  to  $\mathcal{M}_\lambda$ .

The pull-back  $\mathcal{X}_\lambda^*Q$ , which is induced by the exponential map  $\mathcal{X}_\lambda$ , maps a tensor field  $Q$  on the physical manifold  $\mathcal{M}_\lambda$  to a tensor field  $\mathcal{X}_\lambda^*Q$  on the background spacetime. In terms of this generator  ${}_x\eta^a$ , the pull-back  $\mathcal{X}_\lambda^*Q$  is represented by the Taylor expansion

$$\begin{aligned} Q(r) &= Q(\mathcal{X}_\lambda(p)) = \mathcal{X}_\lambda^*Q(p) \\ &= Q(p) + \lambda \mathcal{E}_{{}_x\eta} Q \Big|_p + \frac{1}{2} \lambda^2 \mathcal{E}_{{}_x\eta}^2 Q \Big|_p + O(\lambda^3), \end{aligned} \quad (7)$$

where  $r = \mathcal{X}_\lambda(p) \in \mathcal{M}_\lambda$ . Because  $p \in \mathcal{M}_0$ , we may regard the equation

$$\begin{aligned} \mathcal{X}_\lambda^*Q(p) &= Q_0(p) + \lambda \mathcal{E}_{{}_x\eta} Q \Big|_{\mathcal{M}_0}(p) \\ &\quad + \frac{1}{2} \lambda^2 \mathcal{E}_{{}_x\eta}^2 Q \Big|_{\mathcal{M}_0}(p) + O(\lambda^3) \end{aligned} \quad (8)$$

as an equation on the background spacetime  $\mathcal{M}_0$ , where  $Q_0 = Q|_{\mathcal{M}_0}$  is the background value of the physical variable of  $Q$ . Once the definition of the pull-back of the gauge choice  $\mathcal{X}_\lambda$  is given, the first- and the second-order perturbations  ${}_x^{(1)}Q$

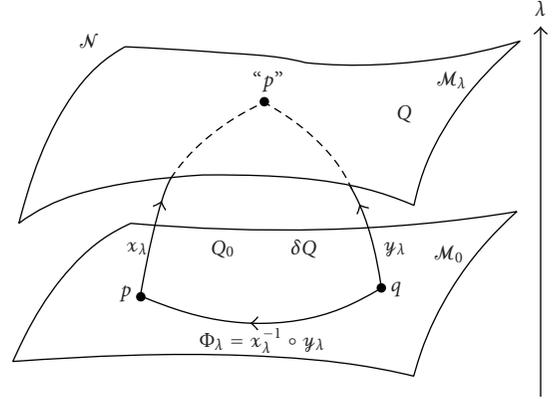


FIGURE 2: The second-kind gauge is a point-identification between the physical spacetime  $\mathcal{M}_\lambda$  and the background spacetime  $\mathcal{M}_0$  on the extended manifold  $\mathcal{N}$ . Through (5), we implicitly assume the existence of a point-identification map between  $\mathcal{M}_\lambda$  and  $\mathcal{M}_0$ . However, this point-identification is not unique by virtue of the general covariance in the theory. We may choose the gauge of the second-kind so that  $p \in \mathcal{M}_0$  and ‘‘ $p$ ’’  $\in \mathcal{M}_\lambda$  is same ( $\mathcal{X}_\lambda$ ). We may also choose the gauge so that  $q \in \mathcal{M}_0$  and ‘‘ $p$ ’’  $\in \mathcal{M}_\lambda$  is same ( $\mathcal{Y}_\lambda$ ). These are different gauge choices. The gauge transformation  $\mathcal{X}_\lambda \rightarrow \mathcal{Y}_\lambda$  is given by the diffeomorphism  $\Phi = \mathcal{X}_\lambda^{-1} \circ \mathcal{Y}_\lambda$ .

and  ${}_x^{(2)}Q$  of a tensor field  $Q$  under the gauge choice  $\mathcal{X}_\lambda$  are simply given by the expansion

$$\mathcal{X}_\lambda^*Q \Big|_{\mathcal{M}_0} = Q_0 + \lambda {}_x^{(1)}Q + \frac{1}{2} \lambda^2 {}_x^{(2)}Q + O(\lambda^3) \quad (9)$$

with respect to the infinitesimal parameter  $\lambda$ . Comparing (8) and (9), we define the first- and the second-order perturbations of a physical variable  $Q_\lambda$  under the gauge choice  $\mathcal{X}_\lambda$  by

$${}_x^{(1)}Q := \mathcal{E}_{{}_x\eta} Q \Big|_{\mathcal{M}_0}, \quad {}_x^{(2)}Q := \mathcal{E}_{{}_x\eta}^2 Q \Big|_{\mathcal{M}_0}. \quad (10)$$

We note that all variables in (9) are defined on  $\mathcal{M}_0$ .

Now, we consider two *different gauge choices* based on the above understanding of the second-kind gauge choice. Suppose that  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$  are two exponential maps with the generators  ${}_x\eta^a$  and  ${}_y\eta^a$  on  $\mathcal{N}$ , respectively. In other words,  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$  are two gauge choices (see Figure 2). Then, the integral curves of each  ${}_x\eta^a$  and  ${}_y\eta^a$  in  $\mathcal{N}$  are the orbits of the actions of the gauge choices  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$ , respectively. Since we choose the generators  ${}_x\eta^a$  and  ${}_y\eta^a$  so that these are transverse to each  $\mathcal{M}_\lambda$  everywhere on  $\mathcal{N}$ , the integral curves of these vector fields intersect with each  $\mathcal{M}_\lambda$ . Therefore, points lying on the same integral curve of either of the two are to be regarded as *the same point* within the respective gauges. When these curves are not identical, that is, the tangential components to each  $\mathcal{M}_\lambda$  of  ${}_x\eta^a$  and  ${}_y\eta^a$  are different, these point identification maps  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$  are regarded as *two different gauge choices*.

We next introduce the concept of *gauge invariance*. In particular, in this paper, we consider the concept of *order-by-order gauge invariance* [27]. Suppose that  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$  are two different gauge choices which are generated by the vector

fields  ${}_x\eta^a$  and  ${}_y\eta^a$ , respectively. These gauge choices also pull back a generic tensor field  $Q$  on  $\mathcal{M}$  to two other tensor fields,  $\mathcal{X}_\lambda^* Q$  and  $\mathcal{Y}_\lambda^* Q$ , for any given value of  $\lambda$ . In particular, on  $\mathcal{M}_0$ , we now have three tensor fields associated with a tensor field  $Q$ ; one is the background value  $Q_0$  of  $Q$ , and the other two are the pulled-back variables of  $Q$  from  $\mathcal{M}_\lambda$  to  $\mathcal{M}_0$  by the two different gauge choices

$$\begin{aligned} {}_x Q_\lambda &:= \mathcal{X}_\lambda^* Q \Big|_{\mathcal{M}_0} \\ &= Q_0 + \lambda {}_x^{(1)} Q + \frac{1}{2} \lambda^2 {}_x^{(2)} Q + O(\lambda^3), \end{aligned} \quad (11)$$

$$\begin{aligned} {}_y Q_\lambda &:= \mathcal{Y}_\lambda^* Q \Big|_{\mathcal{M}_0} \\ &= Q_0 + \lambda {}_y^{(1)} Q + \frac{1}{2} \lambda^2 {}_y^{(2)} Q + O(\lambda^3). \end{aligned} \quad (12)$$

Here, we have used (9). Because  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$  are gauge choices which map from  $\mathcal{M}_0$  to  $\mathcal{M}_\lambda$ ,  ${}_x Q_\lambda$  and  ${}_y Q_\lambda$  are the different representations on  $\mathcal{M}_0$  in the two different gauges of the same perturbed tensor field  $Q$  on  $\mathcal{M}_\lambda$ . The quantities  ${}_x^{(k)} Q$  and  ${}_y^{(k)} Q$  in (11) and (12) are the perturbations of  $O(k)$  in the gauges  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$ , respectively. We say that the  $k$ th-order perturbation  ${}_x^{(k)} Q$  of  $Q$  is *order-by-order gauge-invariant* if and only if for any two gauges  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$  the following holds:

$${}_x^{(k)} Q = {}_y^{(k)} Q. \quad (13)$$

Now, we consider the *gauge transformation rules* between different gauge choices. In general, the representation  ${}_x Q_\lambda$  on  $\mathcal{M}_0$  of the perturbed variable  $Q$  on  $\mathcal{M}_\lambda$  depends on the gauge choice  $\mathcal{X}_\lambda$ . If we employ a different gauge choice, the representation of  $Q_\lambda$  on  $\mathcal{M}_0$  may change. Suppose that  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$  are different gauge choices, which are the point identification maps from  $\mathcal{M}_0$  to  $\mathcal{M}_\lambda$ , and the generators of these gauge choices are given by  ${}_x\eta^a$  and  ${}_y\eta^a$ , respectively. Then, the change of the gauge choice from  $\mathcal{X}_\lambda$  to  $\mathcal{Y}_\lambda$  is represented by the diffeomorphism

$$\Phi_\lambda := (\mathcal{X}_\lambda)^{-1} \circ \mathcal{Y}_\lambda. \quad (14)$$

This diffeomorphism  $\Phi_\lambda$  is the map  $\Phi_\lambda : \mathcal{M}_0 \rightarrow \mathcal{M}_0$  for each value of  $\lambda \in \mathbb{R}$ . The diffeomorphism  $\Phi_\lambda$  does change the point identification, as expected from the understanding of the gauge choice discussed above. Therefore, the diffeomorphism  $\Phi_\lambda$  is regarded as the gauge transformation  $\Phi_\lambda : \mathcal{X}_\lambda \rightarrow \mathcal{Y}_\lambda$ .

The gauge transformation  $\Phi_\lambda$  induces a pull-back from the representation  ${}_x Q_\lambda$  of the perturbed tensor field  $Q$  in the gauge choice  $\mathcal{X}_\lambda$  to the representation  ${}_y Q_\lambda$  in the gauge choice  $\mathcal{Y}_\lambda$ . Actually, the tensor fields  ${}_x Q_\lambda$  and  ${}_y Q_\lambda$ , which are defined on  $\mathcal{M}_0$ , are connected by the linear map  $\Phi_\lambda^*$  as

$$\begin{aligned} {}_y Q_\lambda &= \mathcal{Y}_\lambda^* Q \Big|_{\mathcal{M}_0} = \left( \mathcal{Y}_\lambda^* (\mathcal{X}_\lambda \mathcal{X}_\lambda^{-1})^* Q \right) \Big|_{\mathcal{M}_0} \\ &= (\mathcal{X}_\lambda^{-1} \mathcal{Y}_\lambda)^* (\mathcal{X}_\lambda^* Q) \Big|_{\mathcal{M}_0} = \Phi_\lambda^* {}_x Q_\lambda. \end{aligned} \quad (15)$$

According to generic arguments concerning the Taylor expansion of the pull-back of a tensor field on the same manifold, given in Section 2.1, it should be possible to express the gauge transformation  $\Phi_\lambda^* {}_x Q_\lambda$  in the form

$$\Phi_\lambda^* {}_x Q = {}_x Q + \lambda \xi_{\xi_1} {}_x Q + \frac{\lambda^2}{2} \{ \xi_{\xi_2} + \xi_{\xi_1}^2 \} {}_x Q + O(\lambda^3), \quad (16)$$

where the vector fields  $\xi_1^a$  and  $\xi_2^a$  are the generators of the gauge transformation  $\Phi_\lambda$  (see (2)).

Comparing the representation (16) of the Taylor expansion in terms of the generators  $\xi_1^a$  and  $\xi_2^a$  of the pull-back  $\Phi_\lambda^* {}_x Q$  and that in terms of the generators  ${}_x\eta^a$  and  ${}_y\eta^a$  of the pull-back  $\mathcal{Y}_\lambda^* \circ (\mathcal{X}_\lambda^{-1})^* {}_x Q (= \Phi_\lambda^* {}_x Q)$ , we readily obtain explicit expressions for the generators  $\xi_1^a$  and  $\xi_2^a$  of the gauge transformation  $\Phi = \mathcal{X}_\lambda^{-1} \circ \mathcal{Y}_\lambda$  in terms of the generators  ${}_x\eta^a$  and  ${}_y\eta^a$  of each gauge choices as follows:

$$\xi_1^a = {}_y\eta^a - {}_x\eta^a, \quad \xi_2^a = [{}_y\eta, {}_x\eta]^a. \quad (17)$$

Further, because the gauge transformation  $\Phi_\lambda$  is a map within the background spacetime  $\mathcal{M}_0$ , the generator should consist of vector fields on  $\mathcal{M}_0$ . This can be satisfied by imposing some appropriate conditions on the generators  ${}_y\eta^a$  and  ${}_x\eta^a$ .

We can now derive the relation between the perturbations in the two different gauges. Up to second order, these relations are derived by substituting (11) and (12) into (16):

$${}_y^{(1)} Q - {}_x^{(1)} Q = \xi_{\xi_1} Q_0, \quad (18)$$

$${}_y^{(2)} Q - {}_x^{(2)} Q = 2\xi_{\xi_1} {}_x^{(1)} Q + \{ \xi_{\xi_2} + \xi_{\xi_1}^2 \} Q_0. \quad (19)$$

Here, we should comment on the gauge choice in the above explanation. We have introduced an exponential map  $\mathcal{X}_\lambda$  (or  $\mathcal{Y}_\lambda$ ) as the gauge choice, for simplicity. However, this simplified introduction of  $\mathcal{X}_\lambda$  as an exponential map is not essential to the gauge transformation rules (18) and (19). Actually, we can generalize the diffeomorphism  $\mathcal{X}_\lambda$  from an exponential map. For example, the diffeomorphism whose pull-back is represented by the Taylor expansion (2) is a candidate of the generalization. If we generalize the diffeomorphism  $\mathcal{X}_\lambda$ , the representation (8) of the pulled-back variable  $\mathcal{X}_\lambda^* Q(p)$ , the representations of the perturbations (10), and the relations (17) between generators of  $\Phi_\lambda$ ,  $\mathcal{X}_\lambda$ , and  $\mathcal{Y}_\lambda$  will be changed. However, the gauge transformation rules (18) and (19) are direct consequences of the generic Taylor expansion (16) of  $\Phi_\lambda$ . Generality of the representation of the Taylor expansion (16) of  $\Phi_\lambda$  implies that the gauge transformation rules (18) and (19) will not be changed, even if we generalize the each gauge choice  $\mathcal{X}_\lambda$ . Further, the relations (17) between generators also imply that, even if we employ simple exponential maps as gauge choices, both of the generators  $\xi_1^a$  and  $\xi_2^a$  are naturally induced by the generators of the original gauge choices. Hence, we conclude that the gauge transformation rules (18) and (19) are quite general and irreducible. In this paper, we review the development of a second-order gauge-invariant cosmological perturbation theory based on

the above understanding of the gauge degree of freedom only through the gauge transformation rules (18) and (19). Hence, the developments of the cosmological perturbation theory presented below will not be changed even if we generalize the gauge choice  $\mathcal{X}_\lambda$  from a simple exponential map.

We also have to emphasize the physical implication of the gauge transformation rules (18) and (19). According to the above construction of the perturbation theory, gauge degree of freedom, which induces the transformation rules (18) and (19), is unphysical degree of freedom. As emphasized above, the physical spacetime  $\mathcal{M}_\lambda$  is our nature itself, while there is no background spacetime  $\mathcal{M}_0$  in our nature. The background spacetime  $\mathcal{M}_0$  is a fictitious spacetime and it has nothing to do with our nature. Since the gauge choice  $\mathcal{X}_\lambda$  just gives a relation between  $\mathcal{M}_\lambda$  and  $\mathcal{M}_0$ , the gauge choice  $\mathcal{X}_\lambda$  also has nothing to do with our nature. On the other hand, any observations and experiments are carried out only on the physical spacetime  $\mathcal{M}_\lambda$  through the physical processes on the physical spacetime  $\mathcal{M}_\lambda$ . Therefore, any direct observables in any observations and experiments should be independent of the gauge choice  $\mathcal{X}_\lambda$ , that is, should be gauge-invariant. Keeping this fact in our mind, the gauge transformation rules (18) and (19) imply that the perturbations  ${}^{(1)}_x Q$  and  ${}^{(2)}_x Q$  include unphysical degree of freedom, that is, gauge degree of freedom, if these perturbations are transformed as (18) or (19) under the gauge transformation  $\mathcal{X}_\lambda \rightarrow \mathcal{Y}_\lambda$ . If the perturbations  ${}^{(1)}_x Q$  and  ${}^{(2)}_x Q$  are independent of the gauge choice, these variables are order-by-order gauge-invariant. Therefore, order-by-order gauge-invariant variables does not include unphysical degree of freedom and should be related to the physics on the physical spacetime  $\mathcal{M}_\lambda$ .

**2.4. Coordinate Transformations Induced by the Second Kind Gauge Transformation.** In many literature, gauge degree of freedom is regarded as the degree of freedom of the coordinate transformation. In the linear-order perturbation theory, these two degree of freedom are equivalent with each other. However, in the higher-order perturbations, we should regard that these two degree of freedom are different. Although the essential understanding of the gauge degree of freedom (of the second-kind) is as that explained above, the gauge transformation (of the second-kind) also induces the infinitesimal coordinate transformation on the physical spacetime  $\mathcal{M}_\lambda$  as a result. In many cases, the understanding of “gauges” in perturbations based on coordinate transformations leads mistakes. Therefore, we did not use any ingredient of this subsection in our series of papers [22, 23, 25–28] concerning about higher-order general relativistic gauge-invariant perturbation theory. However, we comment on the relations between the coordinate transformation, briefly. Details can be seen in [22, 37, 38].

To see that the gauge transformation of the second-kind induces the coordinate transformation, we introduce the coordinate system  $\{O_\alpha, \psi_\alpha\}$  on the “background spacetime”  $\mathcal{M}_0$ , where  $O_\alpha$  are open sets on the background spacetime and  $\psi_\alpha$  are diffeomorphisms from  $O_\alpha$  to  $\mathbb{R}^4$  ( $4 = \dim \mathcal{M}_0$ ). The coordinate system  $\{O_\alpha, \psi_\alpha\}$  is the set of the collection of

the pair of open sets  $O_\alpha$  and diffeomorphism  $O_\alpha \mapsto \mathbb{R}^4$ . If we employ a gauge choice  $\mathcal{X}_\lambda$ , we have the correspondence of  $\mathcal{M}_\lambda$  and  $\mathcal{M}_0$ . Together with the coordinate system  $\psi_\alpha$  on  $\mathcal{M}_0$ , this correspondence between  $\mathcal{M}_\lambda$  and  $\mathcal{M}_0$  induces the coordinate system on  $\mathcal{M}_\lambda$ . Actually,  $X_\lambda(O_\alpha)$  for each  $\alpha$  is an open set of  $\mathcal{M}_\lambda$ . Then,  $\psi_\alpha \circ \mathcal{X}_\lambda^{-1}$  becomes a diffeomorphism from an open set  $X_\lambda(O_\alpha) \subset \mathcal{M}_\lambda$  to  $\mathbb{R}^4$ . This diffeomorphism  $\psi_\alpha \circ \mathcal{X}_\lambda^{-1}$  induces a coordinate system of an open set on  $\mathcal{M}_\lambda$ .

When we have two different gauge choices  $\mathcal{X}_\lambda$  and  $\mathcal{Y}_\lambda$ ,  $\psi_\alpha \circ \mathcal{X}_\lambda^{-1}$  and  $\psi_\alpha \circ \mathcal{Y}_\lambda^{-1}$  become different coordinate systems on  $\mathcal{M}_\lambda$ . We can also consider the coordinate transformation from the coordinate system  $\psi_\alpha \circ \mathcal{X}_\lambda^{-1}$  to another coordinate system  $\psi_\alpha \circ \mathcal{Y}_\lambda^{-1}$ . Since the gauge transformation  $\mathcal{X}_\lambda \rightarrow \mathcal{Y}_\lambda$  is induced by the diffeomorphism  $\Phi_\lambda$  defined by (14), the induced coordinate transformation is given by

$$y^\mu(q) := x^\mu(p) = \left( (\Phi^{-1})^* x^\mu \right) (q) \quad (20)$$

in the *passive* point of view [22, 37, 38]. If we represent this coordinate transformation in terms of the Taylor expansion in Section 2.1, up to third order, we have the coordinate transformation

$$y^\mu(q) = x^\mu(q) - \lambda \xi_1^\mu(q) + \frac{\lambda^2}{2} \left\{ -\xi_2^\mu(q) + \xi_1^\nu(q) \partial_\nu \xi_1^\mu(q) \right\} + O(\lambda^3). \quad (21)$$

**2.5. Gauge-Invariant Variables.** Here, inspecting the gauge transformation rules (18) and (19), we define the gauge-invariant variables for a metric perturbation and for arbitrary matter fields (tensor fields). Employing the idea of order-by-order gauge invariance for perturbations [27], we proposed a procedure to construct gauge-invariant variables of higher-order perturbations [22]. This proposal is as follows. First, we decompose a linear-order metric perturbation into its gauge-invariant and variant parts. The procedure for decomposing linear-order metric perturbations is extended to second-order metric perturbations, and we can decompose the second-order metric perturbation into gauge-invariant and variant parts. Then, we can define the gauge-invariant variables for the first- and second-order perturbations of an arbitrary field other than the metric by using the gauge variant parts of the first- and second-order metric perturbations. Although the procedure for finding gauge-invariant variables for linear-order metric perturbations is highly nontrivial, once we know this procedure, we can easily define the gauge-invariant variables of a higher-order perturbation through a simple extension of the procedure for the linear-order perturbations.

Now, we review the above strategy to construct gauge-invariant variables. To consider a metric perturbation, we expand the metric on the physical spacetime  $\mathcal{M}_\lambda$ , which is pulled back to the background spacetime  $\mathcal{M}_0$  using a gauge choice in the form given in (9):

$$\mathcal{X}_\lambda^* \bar{g}_{ab} = g_{ab} + \lambda x_\lambda h_{ab} + \frac{\lambda^2}{2} x_\lambda l_{ab} + O^3(\lambda), \quad (22)$$

where  $g_{ab}$  is the metric on  $\mathcal{M}_0$ . Of course, the expansion (22) of the metric depends entirely on the gauge choice  $\mathcal{X}_\lambda$ .

Nevertheless, henceforth, we do not explicitly express the index of the gauge choice  $\mathcal{X}_\lambda$  in an expression if there is no possibility of confusion.

Our starting point to construct gauge-invariant variables is the assumption that *we already know the procedure for finding gauge-invariant variables for the linear metric perturbations*. Then, a linear metric perturbation  $h_{ab}$  is decomposed as

$$h_{ab} =: \mathcal{H}_{ab} + \varepsilon_X g_{ab}, \quad (23)$$

where  $\mathcal{H}_{ab}$  and  $X^a$  are the gauge-invariant and variant parts of the linear-order metric perturbations, that is, under the gauge transformation (18), these are transformed as

$${}_y\mathcal{H}_{ab} - {}_x\mathcal{H}_{ab} = 0, \quad {}_yX^a - {}_xX^a = \xi_1^a. \quad (24)$$

The first-order metric perturbation (23) together with the gauge transformation rules (24) does satisfy the gauge transformation rule (18) for the first-order metric perturbation, that is,

$${}^{(1)}{}_y h_{ab} - {}^{(1)}{}_x h_{ab} = \varepsilon_{\xi_1} g_{ab}. \quad (25)$$

As emphasized in our series of papers [22, 23, 25–28], the above assumption is quite non-trivial and it is not simple to carry out the systematic decomposition (23) on an arbitrary background spacetime, since this procedure depends completely on the background spacetime  $(\mathcal{M}_0, g_{ab})$ . However, as we will show below, this procedure exists at least in the case of cosmological perturbations of a homogeneous and isotropic universe in Section 5.1.

Once we accept this assumption for linear-order metric perturbations, we can always find gauge-invariant variables for higher-order perturbations [22]. According to the gauge transformation rule (19), the second-order metric perturbation  $l_{ab}$  is transformed as

$${}^{(2)}{}_y l_{ab} - {}^{(2)}{}_x l_{ab} = 2\varepsilon_{\xi_1} {}_x h_{ab} + \{\varepsilon_{\xi_2} + \varepsilon_{\xi_1}^2\} g_{ab} \quad (26)$$

under the gauge transformation  $\Phi_\lambda = (\mathcal{X}_\lambda)^{-1} \circ \mathcal{Y}_\lambda : \mathcal{X}_\lambda \rightarrow \mathcal{Y}_\lambda$ . Although this gauge transformation rule is slightly complicated, inspecting this gauge transformation rule, we first introduce the variable  $\hat{L}_{ab}$  defined by

$$\hat{L}_{ab} := l_{ab} - 2\varepsilon_X h_{ab} + \varepsilon_X^2 g_{ab}. \quad (27)$$

Under the gauge transformation  $\Phi_\lambda = (\mathcal{X}_\lambda)^{-1} \circ \mathcal{Y}_\lambda : \mathcal{X}_\lambda \rightarrow \mathcal{Y}_\lambda$ , the variable  $\hat{L}_{ab}$  is transformed as

$${}_y\hat{L}_{ab} - {}_x\hat{L}_{ab} = \varepsilon_\sigma g_{ab}, \quad (28)$$

$$\sigma^a := \xi_2^a + [\xi_1, X]^a. \quad (29)$$

The gauge transformation rule (28) is identical to that for a linear metric perturbation. Therefore, we may apply the above procedure to decompose  $h_{ab}$  into  $\mathcal{H}_{ab}$  and  $X_a$  when we decompose of the components of the variable  $\hat{L}_{ab}$ . Then,  $\hat{L}_{ab}$  can be decomposed as

$$\hat{L}_{ab} = \mathcal{L}_{ab} + \varepsilon_Y g_{ab}, \quad (30)$$

where  $\mathcal{L}_{ab}$  is the gauge-invariant part of the variable  $\hat{L}_{ab}$ , or equivalently, of the second-order metric perturbation  $l_{ab}$ , and  $Y^a$  is the gauge variant part of  $\hat{L}_{ab}$ , that is, the gauge variant part of  $l_{ab}$ . Under the gauge transformation  $\Phi_\lambda = (\mathcal{X}_\lambda)^{-1} \circ \mathcal{Y}_\lambda$ , the variables  $\mathcal{L}_{ab}$  and  $Y^a$  are transformed as

$${}_y\mathcal{L}_{ab} - {}_x\mathcal{L}_{ab} = 0, \quad {}_yY^a - {}_xY^a = \sigma^a, \quad (31)$$

respectively. Thus, once we accept the assumption (23), the second-order metric perturbations are decomposed as

$$l_{ab} =: \mathcal{L}_{ab} + 2\varepsilon_X h_{ab} + (\varepsilon_Y - \varepsilon_X^2) g_{ab}, \quad (32)$$

where  $\mathcal{L}_{ab}$  and  $Y^a$  are the gauge-invariant and variant parts of the second order metric perturbations, that is,

$${}_y\mathcal{L}_{ab} - {}_x\mathcal{L}_{ab} = 0, \quad {}_yY^a - {}_xY^a = \xi_2^a + [\xi_1, X]^a. \quad (33)$$

Furthermore, as shown in [22], using the first- and second-order gauge variant parts,  $X^a$  and  $Y^a$ , of the metric perturbations, the gauge-invariant variables for an arbitrary field  $Q$  other than the metric are given by

$${}^{(1)}\mathcal{Q} := {}^{(1)}Q - \varepsilon_X Q_0, \quad (34)$$

$${}^{(2)}\mathcal{Q} := {}^{(2)}Q - 2\varepsilon_X {}^{(1)}Q - \{\varepsilon_Y - \varepsilon_X^2\} Q_0. \quad (35)$$

It is straightforward to confirm that the variables  ${}^{(p)}\mathcal{Q}$  defined by (34) and (35) are gauge-invariant under the gauge transformation rules (18) and (19), respectively.

Equations (34) and (35) have very important implications. To see this, we represent these equations as

$${}^{(1)}Q = {}^{(1)}\mathcal{Q} + \varepsilon_X Q_0, \quad (36)$$

$${}^{(2)}Q = {}^{(2)}\mathcal{Q} + 2\varepsilon_X {}^{(1)}Q + \{\varepsilon_Y - \varepsilon_X^2\} Q_0. \quad (37)$$

These equations imply that any perturbation of first and second order can always be decomposed into gauge-invariant and gauge-variant parts as (36) and (37), respectively. These decomposition formulae (36) and (37) are important ingredients in the general framework of the second-order general relativistic gauge-invariant perturbation theory.

### 3. Perturbations of the Field Equations

In terms of the gauge-invariant variables defined last section, we derive the field equations, that is, Einstein equations and the equation for a matter field. To derive the perturbation of the Einstein equations and the equation for a matter field (Klein-Gordon equation), first of all, we have to derive the perturbative expressions of the Einstein tensor [23]. This is reviewed in Section 3.1. We also derive the first and the second order perturbations of the energy momentum tensor for a scalar field and the Klein-Gordon equation [27] in Section 3.2. Finally, we consider the first- and the second-order the Einstein equations in Section 3.3.

*3.1. Perturbations of the Einstein Curvature.* The relation between the curvatures associated with the metrics on the physical spacetime  $\mathcal{M}_\lambda$  and the background spacetime  $\mathcal{M}_0$  is given by the relation between the pulled-back operator  $\mathcal{X}_\lambda^* \bar{\nabla}_a (\mathcal{X}_\lambda^{-1})^*$  of the covariant derivative  $\bar{\nabla}_a$  associated with the metric  $\bar{g}_{ab}$  on  $\mathcal{M}_\lambda$  and the covariant derivative  $\nabla_a$  associated with the metric  $g_{ab}$  on  $\mathcal{M}_0$ . The pulled-back covariant derivative  $\mathcal{X}_\lambda^* \bar{\nabla}_a (\mathcal{X}_\lambda^{-1})^*$  depends on the gauge choice  $\mathcal{X}_\lambda$ . The property of the derivative operator  $\mathcal{X}_\lambda^* \bar{\nabla}_a (\mathcal{X}_\lambda^{-1})^*$  as the covariant derivative on  $\mathcal{M}_\lambda$  is given by

$$\mathcal{X}_\lambda^* \bar{\nabla}_a \left( (\mathcal{X}_\lambda^{-1})^* \mathcal{X}_\lambda^* \bar{g}_{ab} \right) = 0, \quad (38)$$

where  $\mathcal{X}_\lambda^* \bar{g}_{ab}$  is the pull-back of the metric on  $\mathcal{M}_\lambda$ , which is expanded as (22). In spite of the gauge dependence of the operator  $\mathcal{X}_\lambda^* \bar{\nabla}_a (\mathcal{X}_\lambda^{-1})^*$ , we simply denote this operator by  $\bar{\nabla}_a$ , because our calculations are carried out only on  $\mathcal{M}_0$  in the same gauge choice  $\mathcal{X}_\lambda$ . Further, we denote the pulled-back metric  $\mathcal{X}_\lambda^* \bar{g}_{ab}$  on  $\mathcal{M}_\lambda$  by  $\bar{g}_{ab}$ , as mentioned above.

Since the derivative operator  $\bar{\nabla}_a (= \mathcal{X}_\lambda^* \bar{\nabla}_a (\mathcal{X}_\lambda^{-1})^*)$  may be regarded as a derivative operator on  $\mathcal{M}_0$  that satisfies the property (38), there exists a tensor field  $C^c{}_{ab}$  on  $\mathcal{M}_0$  such that

$$\bar{\nabla}_a \omega_b = \nabla_a \omega_b - C^c{}_{ab} \omega_c, \quad (39)$$

where  $\omega_a$  is an arbitrary one-form on  $\mathcal{M}_0$ . From the property (38) of the covariant derivative operator  $\bar{\nabla}_a$  on  $\mathcal{M}_\lambda$ , the tensor field  $C^c{}_{ab}$  on  $\mathcal{M}_0$  is given by

$$C^c{}_{ab} = \frac{1}{2} \bar{g}^{cd} \left( \nabla_a \bar{g}_{db} + \nabla_b \bar{g}_{da} - \nabla_d \bar{g}_{ab} \right), \quad (40)$$

where  $\bar{g}^{ab}$  is the inverse of  $\bar{g}_{ab}$  (see Appendix B). We note that the gauge dependence of the covariant derivative  $\bar{\nabla}_a$  appears only through  $C^c{}_{ab}$ . The Riemann curvature  $\bar{R}_{abc}{}^d$  on  $\mathcal{M}_\lambda$ , which is also pulled back to  $\mathcal{M}_0$ , is given by [42]

$$\bar{R}_{abc}{}^d = R_{abc}{}^d - 2 \nabla_{[a} C^d{}_{b]c} + 2 C^e{}_{c[a} C^d{}_{b]e}, \quad (41)$$

where  $R_{abc}{}^d$  is the Riemann curvature on  $\mathcal{M}_0$ . The perturbative expression for the curvatures are obtained from the expansion of (41) through the expansion of  $C^c{}_{ab}$ .

The first- and the second order perturbations of the Riemann, the Ricci, the scalar, the Weyl curvatures, and the Einstein tensors on the general background spacetime are summarized in [23]. We also derived the perturbative form of the divergence of an arbitrary tensor field of second rank to check the perturbative Bianchi identities in [23]. In this paper, we only present the perturbative expression for the Einstein tensor, and its derivations in Appendix B.

We expand the Einstein tensor  $\bar{G}_a{}^b := \bar{R}_a{}^b - (1/2) \delta_a{}^b \bar{R}$  on  $\mathcal{M}_\lambda$  as

$$\bar{G}_a{}^b = G_a{}^b + \lambda^{(1)} \bar{G}_a{}^b + \frac{1}{2} \lambda^{(2)} G_a{}^b + O(\lambda^3). \quad (42)$$

As shown in Appendix B, each order perturbation of the Einstein tensor is given by

$${}^{(1)}G_a{}^b = {}^{(1)}\mathcal{G}_a{}^b[\mathcal{H}] + \xi_X G_a{}^b, \quad (43)$$

$${}^{(2)}G_a{}^b = {}^{(1)}\mathcal{G}_a{}^b[\mathcal{L}] + {}^{(2)}\mathcal{G}_a{}^b[\mathcal{H}, \mathcal{H}] + 2\xi_X^{(1)} \bar{G}_a{}^b + \{\xi_Y - \xi_X^2\} G_a{}^b, \quad (44)$$

where

$${}^{(1)}\mathcal{G}_a{}^b[A] := {}^{(1)}\Sigma_a{}^b[A] - \frac{1}{2} \delta_a{}^b {}^{(1)}\Sigma_c{}^c[A], \quad (45)$$

$${}^{(1)}\Sigma_a{}^b[A] := -2 \nabla_{[a} H_{d]}{}^{bd}[A] - A^{cb} R_{ac},$$

$${}^{(2)}\mathcal{G}_a{}^b[A, B] := {}^{(2)}\Sigma_a{}^b[A, B] - \frac{1}{2} \delta_a{}^b {}^{(2)}\Sigma_c{}^c[A, B], \quad (46)$$

$$\begin{aligned} {}^{(2)}\Sigma_a{}^b[A, B] := & 2R_{ad} B_c{}^{(b} A^{d)c} + 2H_{[a}{}^{de}[A] H_{d]}{}^{b}{}_e[B] \\ & + 2H_{[a}{}^{de}[B] H_{d]}{}^{b}{}_e[A] + 2A_e{}^d \nabla_{[a} H_{d]}{}^{be}[B] \\ & + 2B_e{}^d \nabla_{[a} H_{d]}{}^{be}[A] + 2A_c{}^b \nabla_{[a} H_{d]}{}^{cd}[B] \\ & + 2B_c{}^b \nabla_{[a} H_{d]}{}^{cd}[A], \end{aligned} \quad (47)$$

$$H_{ab}{}^c[A] := \nabla_{(a} A_{b)}{}^c - \frac{1}{2} \nabla^c A_{ab}, \quad (48)$$

$$H_{abc}[A] := g_{cd} H_{ab}{}^d[A], \quad H_a{}^{bc}[A] := g^{bd} H_{ad}{}^c[A], \quad (49)$$

$$H_a{}^b{}_c[A] := g_{cd} H_a{}^{bd}[A].$$

We note that  ${}^{(1)}\mathcal{G}_a{}^b[*]$  and  ${}^{(2)}\mathcal{G}_a{}^b[*]$  in (43) and (44) are the gauge-invariant parts of the perturbative Einstein tensors, and (43) and (44) have the same forms as (34) and (37), respectively. The expression of  ${}^{(2)}\mathcal{G}_a{}^b[A, B]$  in (46) with (47) is derived by the consideration of the general relativistic gauge-invariant perturbation theory with two infinitesimal parameters in [22, 23].

We also note that  ${}^{(1)}\mathcal{G}_a{}^b[*]$  and  ${}^{(2)}\mathcal{G}_a{}^b[*]$  defined by (45)–(47) satisfy the identities

$$\nabla_a {}^{(1)}\mathcal{G}_a{}^b[A] = -H_{ca}{}^a[A] G_b{}^c + H_{ba}{}^c[A] G_c{}^a, \quad (50)$$

$$\begin{aligned} \nabla_a {}^{(2)}\mathcal{G}_b{}^a[A, B] = & -H_{ca}{}^a[A] {}^{(1)}\mathcal{G}_b{}^c[B] - H_{ca}{}^a[B] {}^{(1)}\mathcal{G}_b{}^c[A] \\ & + H_{ba}{}^e[A] {}^{(1)}\mathcal{G}_e{}^a[B] + H_{ba}{}^e[B] {}^{(1)}\mathcal{G}_e{}^a[A] \\ & - \left( H_{bad}[B] A^{dc} + H_{bad}[A] B^{dc} \right) G_c{}^a \\ & + \left( H_{cad}[B] A^{ad} + H_{cad}[A] B^{ad} \right) G_b{}^c, \end{aligned} \quad (51)$$

for arbitrary tensor fields  $A_{ab}$  and  $B_{ab}$ , respectively. We can directly confirm these identities without specifying arbitrary tensors  $A_{ab}$  and  $B_{ab}$  of the second rank, respectively. This implies that our general framework of the second-order gauge-invariant perturbation theory discussed here gives a self-consistent formulation of the second-order perturbation theory. These identities (50) and (51) guarantee the first- and

second-order perturbations of the Bianchi identity  $\bar{\nabla}_b \bar{G}_a^b = 0$  and are also useful when we check whether the derived components of (45) and (46) are correct.

**3.2. Perturbations of the Energy Momentum Tensor and Klein-Gordon Equation.** Here, we consider the perturbations of the energy momentum tensor of the equation of motion. As a model of the matter field, we only consider the scalar field, for simplicity. Then, equation of motion for a scalar field is the Klein-Gordon equation.

The energy momentum tensor for a scalar field  $\bar{\varphi}$  is given by

$$\bar{T}_a^b = \bar{\nabla}_a \bar{\varphi} \bar{\nabla}^b \bar{\varphi} - \frac{1}{2} \delta_a^b (\bar{\nabla}_c \bar{\varphi} \bar{\nabla}^c \bar{\varphi} + 2V(\bar{\varphi})), \quad (52)$$

where  $V(\bar{\varphi})$  is the potential of the scalar field  $\bar{\varphi}$ . We expand the scalar field  $\bar{\varphi}$  as

$$\bar{\varphi} = \varphi + \lambda \hat{\varphi}_1 + \frac{1}{2} \lambda^2 \hat{\varphi}_2 + O(\lambda^3), \quad (53)$$

where  $\varphi$  is the background value of the scalar field  $\bar{\varphi}$ . Further, following to the decomposition formulae (34) and (35), each order perturbation of the scalar field  $\bar{\varphi}$  is decomposed as

$$\begin{aligned} \hat{\varphi}_1 &=: \varphi_1 + \xi_X \varphi, \\ \hat{\varphi}_2 &=: \varphi_2 + 2\xi_X \hat{\varphi}_1 + (\xi_Y - \xi_X^2) \varphi, \end{aligned} \quad (54)$$

where  $\varphi_1$  and  $\varphi_2$  are the first- and the second-order gauge-invariant perturbations of the scalar field, respectively.

Through the perturbative expansions (53) and (B.2) of the scalar field  $\bar{\varphi}$  and the inverse metric, the energy momentum tensor (52) is also expanded as

$$\bar{T}_a^b = T_a^b + \lambda^{(1)} (T_a^b) + \frac{1}{2} \lambda^2 (T_a^b) + O(\lambda^3). \quad (55)$$

The background energy momentum tensor  $T_a^b$  is given by the replacement  $\bar{\varphi} \rightarrow \varphi$  in (52). Further, through the decompositions (23), (32), (54), the perturbations of the energy momentum tensor  $^{(1)}(T_a^b)$  and  $^{(2)}(T_a^b)$  are also decomposed as

$$^{(1)}(T_a^b) = :^{(1)}\mathcal{T}_a^b + \xi_X T_a^b, \quad (56)$$

$$^{(2)}(T_a^b) = :^{(2)}\mathcal{T}_a^b + 2\xi_X ^{(1)}(T_a^b) + (\xi_Y - \xi_X^2) T_a^b, \quad (57)$$

where the gauge-invariant parts  $^{(1)}\mathcal{T}_a^b$  and  $^{(2)}\mathcal{T}_a^b$  of the first and the second order are given by

$$\begin{aligned} ^{(1)}\mathcal{T}_a^b &:= \nabla_a \varphi \nabla^b \varphi_1 - \nabla_a \varphi \mathcal{H}^{bc} \nabla_c \varphi + \nabla_a \varphi_1 \nabla^b \varphi \\ &\quad - \delta_a^b \left( \nabla_c \varphi \nabla^c \varphi_1 - \frac{1}{2} \nabla_c \varphi \mathcal{H}^{dc} \nabla_d \varphi + \varphi_1 \frac{\partial V}{\partial \varphi} \right), \end{aligned} \quad (58)$$

$$\begin{aligned} ^{(2)}\mathcal{T}_a^b &:= \nabla_a \varphi \nabla^b \varphi_2 + \nabla_a \varphi_2 \nabla^b \varphi - \nabla_a \varphi g^{bd} \mathcal{L}_{dc} \nabla^c \varphi \\ &\quad - 2 \nabla_a \varphi \mathcal{H}^{bc} \nabla_c \varphi_1 + 2 \nabla_a \varphi \mathcal{H}^{bd} \mathcal{H}_{dc} \nabla^c \varphi \\ &\quad + 2 \nabla_a \varphi_1 \nabla^b \varphi_1 - 2 \nabla_a \varphi_1 \mathcal{H}^{bc} \nabla_c \varphi \\ &\quad - \delta_a^b \left( \nabla_c \varphi \nabla^c \varphi_2 - \frac{1}{2} \nabla^c \varphi \mathcal{L}_{dc} \nabla^d \varphi \right. \\ &\quad \left. + \nabla^c \varphi \mathcal{H}^{de} \mathcal{H}_{ec} \nabla_d \varphi - 2 \nabla_c \varphi \mathcal{H}^{dc} \nabla_d \varphi_1 \right. \\ &\quad \left. + \nabla_c \varphi_1 \nabla^c \varphi_1 + \varphi_2 \frac{\partial V}{\partial \varphi} + \varphi_1^2 \frac{\partial^2 V}{\partial \varphi^2} \right). \end{aligned} \quad (59)$$

We note that (56) and (57) have the same form as (36) and (37), respectively.

Next, we consider the perturbation of the Klein-Gordon equation

$$\bar{C}_{(K)} := \bar{\nabla}^a \bar{\nabla}_a \bar{\varphi} - \frac{\partial V}{\partial \bar{\varphi}}(\bar{\varphi}) = 0. \quad (60)$$

Through the perturbative expansions (53) and (22), the Klein-Gordon equation (60) is expanded as

$$\bar{C}_{(K)} =: C_{(K)} + \lambda C_{(K)}^{(1)} + \frac{1}{2} \lambda^2 C_{(K)}^{(2)} + O(\lambda^3). \quad (61)$$

$C_{(K)}$  is the background Klein-Gordon equation

$$C_{(K)} := \nabla_a \nabla^a \varphi - \frac{\partial V}{\partial \varphi}(\varphi) = 0. \quad (62)$$

The first- and the second-order perturbations  $C_{(K)}^{(1)}$  and  $C_{(K)}^{(2)}$  are also decomposed into the gauge-invariant and the gauge-variant parts as

$$C_{(K)}^{(1)} =: \mathcal{C}_{(K)}^{(1)} + \xi_X C_{(K)}, \quad (63)$$

$$C_{(K)}^{(2)} =: \mathcal{C}_{(K)}^{(2)} + 2\xi_X C_{(K)}^{(1)} + (\xi_Y - \xi_X^2) C_{(K)},$$

where

$$\mathcal{C}_{(K)}^{(1)} := \nabla^a \nabla_a \varphi_1 - H_a^{ac} [\mathcal{H}] \nabla_c \varphi - \mathcal{H}^{ab} \nabla_a \nabla_b \varphi - \varphi_1 \frac{\partial^2 V}{\partial \varphi^2}(\varphi), \quad (64)$$

$$\begin{aligned} \mathcal{C}_{(K)}^{(2)} &:= \nabla^a \nabla_a \varphi_2 - H_a^{ac} [\mathcal{L}] \nabla_c \varphi + 2H_a^{ad} [\mathcal{H}] \mathcal{H}_{cd} \nabla^c \varphi \\ &\quad - 2H_a^{ac} [\mathcal{H}] \nabla_c \varphi_1 + 2\mathcal{H}^{ab} H_{ab}^c [\mathcal{H}] \nabla_c \varphi \\ &\quad - \mathcal{L}^{ab} \nabla_a \nabla_b \varphi + 2\mathcal{H}^a_d \mathcal{H}^{db} \nabla_a \nabla_b \varphi - 2\mathcal{H}^{ab} \nabla_a \nabla_b \varphi_1 \\ &\quad - \varphi_2 \frac{\partial^2 V}{\partial \varphi^2}(\varphi) - (\varphi_1)^2 \frac{\partial^3 V}{\partial \varphi^3}(\varphi). \end{aligned} \quad (65)$$

Here, we note that (63) have the same form as (36) and (37).

By virtue of the order-by-order evaluations of the Klein-Gordon equation, the first- and the second-order perturbation of the Klein-Gordon equation are necessarily given in gauge-invariant form as

$$\mathcal{C}_{(K)}^{(1)} = 0, \quad \mathcal{C}_{(K)}^{(2)} = 0. \quad (66)$$

We should note that, in [27], we summarized the formulae of the energy momentum tensors for an perfect fluid, an imperfect fluid, and a scalar field. Further, we also summarized the equations of motion of these three matter fields, that is, the energy continuity equation and the Euler equation for a perfect fluid; the energy continuity equation and the Navier-Stokes equation for an imperfect fluid; the Klein-Gordon equation for a scalar field. All these formulae also have the same form as the decomposition formulae (36) and (37). In this sense, we may say that the decomposition formulae (36) and (37) are universal.

*3.3. Perturbations of the Einstein Equation.* Finally, we impose the perturbed Einstein equation of each order,

$${}^{(1)}G_a{}^b = 8\pi G {}^{(1)}T_a{}^b, \quad {}^{(2)}G_a{}^b = 8\pi G {}^{(2)}T_a{}^b. \quad (67)$$

Then, the perturbative Einstein equation is given by

$${}^{(1)}\mathcal{G}_a{}^b[\mathcal{H}] = 8\pi G {}^{(1)}\mathcal{T}_a{}^b \quad (68)$$

at linear order and

$${}^{(1)}\mathcal{G}_a{}^b[\mathcal{L}] + {}^{(2)}\mathcal{G}_a{}^b[\mathcal{H}, \mathcal{H}] = 8\pi G {}^{(2)}\mathcal{T}_a{}^b \quad (69)$$

at second order. These explicitly show that, order-by-order, the Einstein equations are necessarily given in terms of gauge-invariant variables only.

Together with (66), we have seen that the first- and the second-order perturbations of the Einstein equations and the Klein-Gordon equation are necessarily given in gauge-invariant form. This implies that we do not have to consider the gauge degree of freedom, at least in the level where we concentrate only on the equations of the system.

We have reviewed the general outline of the second-order gauge-invariant perturbation theory. We also note that the ingredients of this section are independent of the explicit form of the background metric  $g_{ab}$ , except for the decomposition assumption (23) for the linear-order metric perturbations and are valid not only in cosmological perturbation case but also the other generic situations if (23) is correct. Within this general framework, we develop a second-order cosmological perturbation theory in terms of the gauge-invariant variables.

#### 4. Cosmological Background Spacetime and Equations

The background spacetime  $\mathcal{M}_0$  considered in cosmological perturbation theory is a homogeneous, isotropic universe that is foliated by the three-dimensional hypersurface  $\Sigma(\eta)$ , which is parametrized by  $\eta$ . Each hypersurface of  $\Sigma(\eta)$  is a

maximally symmetric three-space [43], and the spacetime metric of this universe is given by

$$g_{ab} = a^2(\eta) \left( -(d\eta)_a (d\eta)_b + \gamma_{ij} (dx^i)_a (dx^j)_b \right), \quad (70)$$

where  $a = a(\eta)$  is the scale factor,  $\gamma_{ij}$  is the metric on the maximally symmetric 3-space with curvature constant  $K$ , and the indices  $i, j, k, \dots$  for the spatial components run from 1 to 3.

To study the Einstein equation for this background spacetime, we introduce the energy-momentum tensor for a scalar field, which is given by

$$\begin{aligned} T_a{}^b &= \nabla_a \varphi \nabla^b \varphi - \frac{1}{2} \delta_a{}^b (\nabla_c \varphi \nabla^c \varphi + 2V(\varphi)) \\ &= - \left( \frac{1}{2a^2} (\partial_\eta \varphi)^2 + V(\varphi) \right) (d\eta)_a \left( \frac{\partial}{\partial \eta} \right)^b \\ &\quad + \left( \frac{1}{2a^2} (\partial_\eta \varphi)^2 - V(\varphi) \right) \gamma_a{}^b, \end{aligned} \quad (71)$$

where we assumed that the scalar field  $\varphi$  is homogeneous

$$\varphi = \varphi(\eta) \quad (72)$$

and  $\gamma_a{}^b$  are defined as

$$\gamma_{ab} := \gamma_{ij} (dx^i)_a (dx^j)_b, \quad \gamma_a{}^b := \gamma_i{}^j (dx^i)_a \left( \frac{\partial}{\partial x^j} \right)^b. \quad (73)$$

The background Einstein equations  $G_a{}^b = 8\pi G T_a{}^b$  for this background spacetime filled with the single scalar field are given by

$$\mathcal{H}^2 + K = \frac{8\pi G}{3} a^2 \left( \frac{1}{2a^2} (\partial_\eta \varphi)^2 + V(\varphi) \right), \quad (74)$$

$$2\partial_\eta \mathcal{H} + \mathcal{H}^2 + K = 8\pi G \left( -\frac{1}{2} (\partial_\eta \varphi)^2 + a^2 V(\varphi) \right).$$

We also note that (74) lead to

$$\mathcal{H}^2 + K - \partial_\eta \mathcal{H} = 4\pi G (\partial_\eta \varphi)^2. \quad (75)$$

Equation (75) is also useful when we derive the perturbative Einstein equations.

Next, we consider the background Klein-Gordon equation which is the equation of motion  $\nabla_a T_b{}^a = 0$  for the scalar field

$$\partial_\eta^2 \varphi + 2\mathcal{H} \partial_\eta \varphi + a^2 \frac{\partial V}{\partial \varphi} = 0. \quad (76)$$

The Klein-Gordon equation (76) is also derived from the Einstein equations (74). This is a well known fact and is just due to the Bianchi identity of the background spacetime. However, these types of relation are useful to check whether the derived system of equations is consistent.

## 5. Equations for the First-Order Cosmological Perturbations

On the cosmological background spacetime in the last section, we develop the perturbation theory in the gauge-invariant manner. In this section, we summarize the first-order perturbation of the Einstein equation and the Klein-Gordon equations. In Section 5.1, we show that the assumption on the decomposition (23) of the linear-order metric perturbation is correct. In Section 5.2, we summarize the first-order perturbation of the Einstein equation. Finally, in Section 5.3, we show the first-order perturbation of the Klein-Gordon equation.

*5.1. Gauge-Invariant Metric Perturbations.* Here, we consider the first-order metric perturbation  $h_{ab}$  and show the assumption on the decomposition (23) is correct in the background metric (70). To accomplish the decomposition (23), first, we assume the existence of the Green functions  $\Delta^{-1} := (D^i D_i)^{-1}$ ,  $(\Delta + 2K)^{-1}$ , and  $(\Delta + 3K)^{-1}$ , where  $D_i$  is the covariant derivative associated with the metric  $\gamma_{ij}$  and  $K$  is the curvature constant of the maximally symmetric three space. Next, we consider the decomposition of the linear-order metric perturbation  $h_{ab}$  as

$$\begin{aligned} h_{ab} &= h_{\eta\eta}(d\eta)_a(d\eta)_b \\ &+ 2(D_i h_{(VL)} + h_{(V)i})(d\eta)_{(a}(dx^i)_{b)} \\ &+ a^2 \left\{ h_{(L)}\gamma_{ij} + \left( D_i D_j - \frac{1}{3}\gamma_{ij}\Delta \right) h_{(TL)} \right. \\ &\quad \left. + 2D_i h_{(TV)j} + h_{(TT)ij} \right\} (dx^i)_a (dx^j)_b, \end{aligned} \quad (77)$$

where  $h_{(V)i}$ ,  $h_{(TV)j}$ , and  $h_{(TT)ij}$  satisfy the properties

$$\begin{aligned} D^i h_{(V)i} &= 0, & D^i h_{(TV)i} &= 0, \\ h_{(TT)ij} &= h_{(TT)ji}, & h_{(T)}^i{}_{i} &:= \gamma^{ij} h_{(T)ij} = 0, \\ D^i h_{(TT)ij} &= 0. \end{aligned} \quad (78)$$

The gauge-transformation rules for the variables  $h_{\eta\eta}$ ,  $h_{(VL)}$ ,  $h_{(V)i}$ ,  $h_{(L)}$ ,  $h_{(TL)}$ ,  $h_{(TV)j}$  and  $h_{(TT)ij}$  are derived from (25). Inspecting these gauge-transformation rules, we define the gauge-variant part  $X_a$  in (23):

$$\begin{aligned} X_a &:= \left( h_{(VL)} - \frac{1}{2}a^2 \partial_\eta h_{(TL)} \right) (d\eta)_a \\ &+ a^2 \left( h_{(TV)i} + \frac{1}{2}D_i h_{(TL)} \right) (dx^i)_a. \end{aligned} \quad (79)$$

We can easily check this vector field  $X_a$  satisfies (24). Subtracting gauge variant-part  $\xi_X g_{ab}$  from  $h_{ab}$ , we have the gauge-invariant part  $\mathcal{H}_{ab}$  in (23):

$$\begin{aligned} \mathcal{H}_{ab} &= a^2 \left\{ -2 \overset{(1)}{\Phi} (d\eta)_a (d\eta)_b + 2 \overset{(1)}{\nu}_i (d\eta)_{(a} (dx^i)_{b)} \right. \\ &\quad \left. + \left( -2 \overset{(1)}{\Psi} \gamma_{ij} + \overset{(1)}{\chi}_{ij} \right) (dx^i)_a (dx^j)_b \right\}, \end{aligned} \quad (80)$$

where the properties  $D^i \overset{(1)}{\nu}_i := \gamma^{ij} D_i \overset{(1)}{\nu}_j = \overset{(1)}{\chi}_i{}^i := \gamma^{ij} \overset{(1)}{\chi}_{ij} = D^i \overset{(1)}{\chi}_{ij} = 0$  are satisfied as consequences of (78).

Thus, we may say that our assumption for the decomposition (23) in linear-order metric perturbation is correct in the case of cosmological perturbations. However, we have to note that to accomplish (23), we assumed the existence of the Green functions  $\Delta^{-1}$ ,  $(\Delta + 2K)^{-1}$ , and  $(\Delta + 3K)^{-1}$ . As shown in [25, 26], this assumption is necessary to guarantee the one to one correspondence between the variables  $\{h_{\eta\eta}, h_{i\eta}, h_{ij}\}$  and  $\{h_{\eta\eta}, h_{(VL)}, h_{(V)i}, h_{(L)}, h_{(TL)}, h_{(TV)j}, h_{(TT)ij}\}$ , but excludes some perturbative modes of the metric perturbations which belong to the kernel of the operator  $\Delta$ ,  $(\Delta + 2K)$ , and  $(\Delta + 3K)$  from our consideration. For example, homogeneous modes belong to the kernel of the operator  $\Delta$  and are excluded from our consideration. If we have to treat these modes, the separate treatments are necessary. In this paper, we ignore these modes, for simplicity.

We also note the fact that the definition (23) of the gauge-invariant variables is not unique. This comes from the fact that we can always construct new gauge-invariant quantities by the combination of the gauge-invariant variables. For example, using the gauge-invariant variables  $\overset{(1)}{\Phi}$  and  $\overset{(1)}{\nu}_i$  of the first-order metric perturbation, we can define a vector field  $Z_a$  by  $Z_a := -a \overset{(1)}{\Phi} (d\eta)_a + a \overset{(1)}{\nu}_i (dx^i)_a$  which is gauge invariant. Then, we can rewrite the decomposition formula (23) for the linear-order metric perturbation as

$$\begin{aligned} h_{ab} &= \mathcal{H}_{ab} - \xi_Z g_{ab} + \xi_Z g_{ab} + \xi_X g_{ab}, \\ &=: \mathcal{K}_{ab} + \xi_{X+Z} g_{ab}, \end{aligned} \quad (81)$$

where we have defined new gauge invariant variable  $K_{ab}$  by  $\mathcal{K}_{ab} := \mathcal{H}_{ab} - \xi_Z g_{ab}$ . Clearly,  $\mathcal{K}_{ab}$  is gauge-invariant and the vector field  $X^a + Z^a$  satisfies (24). In spite of this nonuniqueness, we specify the components of the tensor  $\mathcal{H}_{ab}$  as (80), which is the gauge-invariant part of the linear-order metric perturbation associated with the longitudinal gauge.

The non-uniqueness of the definitions of gauge-invariant variables is related to the ‘‘gauge-fixing’’ for the linear-order metric perturbations. Due to this non-uniqueness, we can consider the gauge-fixing in the first-order metric perturbation from two different points of view. The first point of view is that the gauge-fixing is to specify the gauge-variant part  $X^a$ . For example, the longitudinal gauge is realized by the gauge fixing  $X^a = 0$ . Due to this gauge fixing  $X^a = 0$ , we can regard the fact that perturbative variables in the longitudinal gauge are the completely gauge fixed variables. On the other hand, we may also regard that the gauge fixing is the specification of the gauge-invariant vector field  $Z^a$  in (81). In this point of view, we do not specify the vector field  $X^a$ . Instead, we have to specify the gauge-invariant vector  $Z^a$  or equivalently to specify the gauge-invariant metric perturbation  $\mathcal{K}_{ab}$  without specifying  $X^a$  so that the first-order metric perturbation  $h_{ab}$  coincides with the gauge-invariant variables  $\mathcal{K}_{ab}$  when we fix the gauge  $X^a$  so that  $X^a + Z^a = 0$ . These two different point of view of ‘‘gauge fixing’’ is equivalent with each other due to the non-uniqueness of the definition (81) of the gauge-invariant variables.

5.2. *First-Order Einstein Equations.* Here, we derive the linear-order Einstein equation (68). To derive the components of the gauge-invariant part of the linearized Einstein tensor  ${}^{(1)}\mathcal{G}_a{}^b[\mathcal{H}]$ , which is defined by (45), we first derive the components of the tensor  $\mathcal{H}_{ab}{}^c[\mathcal{H}]$ , which is defined in (48) with  $A_{ab} = \mathcal{H}_{ab}$  and its component (80). These components are summarized in [25, 26].

From (45), the component of  ${}^{(1)}\mathcal{G}_a{}^b[\mathcal{H}]$  are summarized as

$$\begin{aligned}
{}^{(1)}\mathcal{G}_\eta{}^\eta[\mathcal{H}] &= -\frac{1}{a^2} \left\{ (-6\mathcal{H}\partial_\eta + 2\Delta + 6K) {}^{(1)}\Psi - 6\mathcal{H}^2 {}^{(1)}\Phi \right\}, \\
{}^{(1)}\mathcal{G}_i{}^\eta[\mathcal{H}] &= -\frac{1}{a^2} \left( 2\partial_\eta D_i {}^{(1)}\Psi + 2\mathcal{H}D_i {}^{(1)}\Phi - \frac{1}{2}(\Delta + 2K) {}^{(1)}\nu_i \right), \\
{}^{(1)}\mathcal{G}_\eta{}^i[\mathcal{H}] &= \frac{1}{a^2} \left\{ 2\partial_\eta D^i {}^{(1)}\Psi + 2\mathcal{H}D^i {}^{(1)}\Phi \right. \\
&\quad \left. + \frac{1}{2}(-\Delta + 2K + 4\mathcal{H}^2 - 4\partial_\eta \mathcal{H}) \nu^i \right\}, \\
{}^{(1)}\mathcal{G}_i{}^j[\mathcal{H}] &= \frac{1}{a^2} \left[ D_i D^j \left( {}^{(1)}\Psi - {}^{(1)}\Phi \right) \right. \\
&\quad \left. + \left\{ (-\Delta + 2\partial_\eta^2 + 4\mathcal{H}\partial_\eta - 2K) {}^{(1)}\Psi \right. \right. \\
&\quad \left. \left. + (2\mathcal{H}\partial_\eta + 4\partial_\eta \mathcal{H} + 2\mathcal{H}^2 + \Delta) {}^{(1)}\Phi \right\} \nu_i{}^j \right. \\
&\quad \left. - \frac{1}{2a^2} \partial_\eta \left\{ a^2 \left( D_i \nu^j + D^j \nu_i \right) \right\} \right. \\
&\quad \left. + \frac{1}{2} \left( \partial_\eta^2 + 2\mathcal{H}\partial_\eta + 2K - \Delta \right) \chi_i{}^j \right]. \tag{82}
\end{aligned}$$

Straightforward calculations show that these components of the first-order gauge-invariant perturbation  ${}^{(1)}\mathcal{G}_a{}^b[\mathcal{H}]$  of the Einstein tensor satisfies the identity (50). Although this confirmation is also possible without specification of the tensor  $\mathcal{H}_{ab}$ , the confirmation of (50) through the explicit components (82) implies that we have derived the components of  ${}^{(1)}\mathcal{G}_a{}^b[\mathcal{H}]$  consistently.

Next, we summarize the first-order perturbation of the energy momentum tensor for a scalar field. Since, at the background level, we assume that the scalar field  $\varphi$  is homogeneous as (72), the components of the gauge-invariant part of the first-order energy-momentum tensor  ${}^{(1)}\mathcal{T}_a{}^b$  are given by

$$\begin{aligned}
{}^{(1)}\mathcal{T}_\eta{}^\eta &= -\frac{1}{a^2} \left( \partial_\eta \varphi_1 \partial_\eta \varphi - {}^{(1)}\Phi \left( \partial_\eta \varphi \right)^2 + a^2 \frac{dV}{d\varphi} \varphi_1 \right), \\
{}^{(1)}\mathcal{T}_i{}^\eta &= -\frac{1}{a^2} D_i \varphi_1 \partial_\eta \varphi, \tag{83}
\end{aligned}$$

$${}^{(1)}\mathcal{T}_\eta{}^i = \frac{1}{a^2} \partial_\eta \varphi \left( D^i \varphi_1 + \left( \partial_\eta \varphi \right) \nu^i \right), \tag{84}$$

$${}^{(1)}\mathcal{T}_i{}^j = \frac{1}{a^2} \gamma_i{}^j \left( \partial_\eta \varphi_1 \partial_\eta \varphi - {}^{(1)}\Phi \left( \partial_\eta \varphi \right)^2 - a^2 \frac{dV}{d\varphi} \varphi_1 \right).$$

The second equation in (84) shows that there is no anisotropic stress in the energy-momentum tensor of the single scalar field. Then, we obtain that

$${}^{(1)}\Phi = {}^{(1)}\Psi. \tag{85}$$

From (82)–(84) and (85), the components of scalar parts of the linearized Einstein equation (68) are given as [3]

$$\begin{aligned}
&(\Delta - 3\mathcal{H}\partial_\eta + 4K - \partial_\eta \mathcal{H} - 2\mathcal{H}^2) {}^{(1)}\Phi \\
&= 4\pi G \left( \partial_\eta \varphi_1 \partial_\eta \varphi + a^2 \frac{dV}{d\varphi} \varphi_1 \right), \tag{86}
\end{aligned}$$

$$\partial_\eta {}^{(1)}\Phi + \mathcal{H} {}^{(1)}\Phi = 4\pi G \varphi_1 \partial_\eta \varphi, \tag{87}$$

$$\begin{aligned}
&(\partial_\eta^2 + 3\mathcal{H}\partial_\eta + \partial_\eta \mathcal{H} + 2\mathcal{H}^2) {}^{(1)}\Phi \\
&= 4\pi G \left( \partial_\eta \varphi_1 \partial_\eta \varphi - a^2 \frac{dV}{d\varphi} \varphi_1 \right). \tag{88}
\end{aligned}$$

In the derivation of (86)–(88), we have used (75). We also note that only two of these equations are independent. Further, the vector part of the component  ${}^{(1)}\mathcal{G}_i{}^\eta[\mathcal{H}] = 8\pi G {}^{(1)}\mathcal{T}_i{}^\eta$  shows that

$${}^{(1)}\nu_i = 0. \tag{89}$$

The equation for the tensor mode  $\chi_{ij}$  is given by

$$(\partial_\eta^2 + 2\mathcal{H}\partial_\eta + 2K - \Delta) \chi_i{}^j = 0. \tag{90}$$

Combining (86) and (88), we eliminate the potential term of the scalar field and thereby obtain that

$$(\partial_\eta^2 + \Delta + 4K) {}^{(1)}\Phi = 8\pi G \partial_\eta \varphi_1 \partial_\eta \varphi. \tag{91}$$

Further, using (87) to express  $\partial_\eta \varphi_1$  in terms of  $\partial_\eta {}^{(1)}\Phi$  and  $\Phi$ , we also eliminate  $\partial_\eta \varphi_1$  in (91). Hence, we have

$$\begin{aligned}
&\left\{ \partial_\eta^2 + 2 \left( \mathcal{H} - \frac{2\partial_\eta^2 \varphi}{\partial_\eta \varphi} \right) \partial_\eta - \Delta - 4K + 2 \left( \partial_\eta \mathcal{H} - \frac{\mathcal{H}\partial_\eta^2 \varphi}{\partial_\eta \varphi} \right) \right\} {}^{(1)}\Phi \\
&= 0. \tag{92}
\end{aligned}$$

This is the master equation for the scalar mode perturbation of the cosmological perturbation in universe filled with a single scalar field. It is also known that (92) reduces to a simple equation through a change of variables [3].

5.3. *First-Order Klein-Gordon Equations.* Next, we consider the first-order perturbation of the Klein-Gordon equation (64). By the straightforward calculations using (70), (80), (72), (76), and the components  $H_a{}^{ac}$  summarized in [25, 26], the gauge-invariant part  $\mathcal{C}_{(K)}^{(1)}$  of the first-order Klein-Gordon equation defined by (64) is given by

$$\begin{aligned} -a^2 \mathcal{C}_{(K)}^{(1)} &= \partial_\eta^2 \varphi_1 + 2\mathcal{H}\partial_\eta \varphi_1 - \Delta\varphi_1 - \left( \partial_\eta \Phi^{(1)} + 3\partial_\eta \Psi^{(1)} \right) \partial_\eta \varphi \\ &\quad + 2a^2 \Phi^{(1)} \frac{\partial V}{\partial \varphi}(\varphi) + a^2 \varphi_1 \frac{\partial^2 V}{\partial \varphi^2}(\varphi) \\ &= 0. \end{aligned} \quad (93)$$

Through the background Einstein equations (74) and the first-order perturbations (87) and (92) of the Einstein equation, we can easily derive the first-order perturbation of the Klein-Gordon equation (93) [28]. Hence, the first-order perturbation of the Klein-Gordon equation is not independent of the background and the first-order perturbation of the Einstein equation. Therefore, from the viewpoint of the Cauchy problem, any information obtained from the first-order perturbation of the Klein-Gordon equation should also be obtained from the set of the background and the first-order the Einstein equation, in principle.

## 6. Equations for the Second-Order Cosmological Perturbations

Now, we develop the second-order perturbation theory on the cosmological background spacetime in Section 4 within the general framework of the gauge-invariant perturbation theory reviewed in Section 2. Since we have already confirmed the important step of our general framework, that is, the assumption for the decomposition (23) of the linear-order metric perturbation is correct. Hence, the general framework reviewed in Section 2 is applicable. Applying this framework, we define the second-order gauge-invariant variables of the metric perturbation in Section 6.1. In Section 6.2, we summarize the explicit components of the gauge-invariant parts of the second-order perturbation of the Einstein tensor. In Section 6.3, we summarize the explicit components of the second-order perturbation of the energy-momentum tensor and the Klein-Gordon equations. Then, in Section 6.4, we derive the second-order Einstein equations in terms of gauge-invariant variables. The resulting equations have the source terms which constitute of the quadratic terms of the linear-order perturbations. Although these source terms have complicated forms, we give identities which comes from the consistency of all the second-order perturbations of the Einstein equation and the Klein-Gordon equation in Section 6.5.

6.1. *Gauge-Invariant Metric Perturbations.* First, we consider the components of the gauge-invariant variables for the metric perturbation of second order. The variable  $\hat{L}_{ab}$

defined by (27) is transformed as (28) under the gauge transformation and we may regard the generator  $\sigma_a$  defined by (29) as an arbitrary vector field on  $\mathcal{M}_0$  from the fact that the generator  $\xi_a^g$  in (29) is arbitrary. We can apply the procedure to find gauge-invariant variables for the first-order metric perturbations (80) in Section 5.1. Then, we can accomplish the decomposition (30). Following to the same argument as in the linear case, we may choose the components of the gauge-invariant variables  $\mathcal{L}_{ab}$  in (32) as

$$\begin{aligned} \mathcal{L}_{ab} &= -2a^2 \Phi^{(2)} (\partial_\eta)_a (\partial_\eta)_b + 2a^2 \mathcal{V}_i^{(2)} (\partial_\eta)_{(a} (dx^i)_{b)} \\ &\quad + a^2 \left( -2 \Psi^{(2)} \gamma_{ij} + \chi_{ij}^{(2)} \right) (dx^i)_a (dx^j)_b, \end{aligned} \quad (94)$$

where  $\mathcal{V}_i^{(2)}$  and  $\chi_{ij}^{(2)}$  satisfy the equations

$$D^i \mathcal{V}_i^{(2)} = 0, \quad \chi^i{}_i = 0, \quad D^i \chi_{ij}^{(2)} = 0. \quad (95)$$

The gauge-invariant variables  $\Phi^{(2)}$  and  $\Psi^{(2)}$  are the scalar mode perturbations of second order, and  $\mathcal{V}_i^{(2)}$  and  $\chi_{ij}^{(2)}$  are the second-order vector and tensor modes of the metric perturbations, respectively.

Here, we also note the fact that the decomposition (32) is not unique. This situation is similar to the case of the linear order, but more complicated. In the definition of the gauge-invariant variables of the second-order metric perturbation, we may replace

$$X^a = X'^a - Z'^a, \quad (96)$$

where  $Z'^a$  is gauge invariant and  $X'^a$  is transformed as

$$y X'^a - x X'^a = \xi_1^a \quad (97)$$

under the gauge transformation  $\mathcal{X}_\lambda \rightarrow \mathcal{Y}_\lambda$ . This  $Z'^a$  may be different from the vector  $Z^a$  in (81). By the replacement (96), the second-order metric perturbation (32) is given in the form

$$l_{ab} =: \mathcal{F}_{ab} + 2\xi_{X'} h_{ab} + (\xi_{Y'} - \xi_{X'}^2) g_{ab}, \quad (98)$$

where we defined

$$\begin{aligned} \mathcal{F}_{ab} &:= \mathcal{L}_{ab} - \xi_W g_{ab} - 2\xi_{Z'} \mathcal{K}_{ab} \\ &\quad - 2\xi_{Z'} \xi_Z g_{ab} + \xi_{Z'}^2 g_{ab}, \end{aligned} \quad (99)$$

$$Y'^a := Y^a + W^a + [X', Z']^a. \quad (100)$$

Here, the vector field  $W^a$  in (100) constitute of some components of gauge-invariant second-order metric perturbation  $\mathcal{L}_{ab}$  like  $Z^a$  in (81). The tensor field  $\mathcal{F}_{ab}$  is manifestly gauge invariant. The gauge transformation rule of the new gauge-invariant part  $Y'^a$  of the second-order metric perturbation is given by

$$y Y'^a - x Y'^a = \xi_{(2)}^a + [\xi_{(1)}, X']^a. \quad (101)$$

Although (98) is similar to (32), the tensor fields  $\mathcal{L}_{ab}$  and  $\mathcal{J}_{ab}$  are different from each other. Thus, the definition of the gauge-invariant variables for the second-order metric perturbation is not unique in a more complicated way than the linear order. This nonuniqueness of gauge-invariant variables for the metric perturbations propagates to the definition (34) and (35) of the gauge-invariant variables for matter fields.

In spite of the existence of infinitely many definitions of the gauge-invariant variables, in this paper, we consider the components of  $\mathcal{L}_{ab}$  given by (94). Equation (94) corresponds to the second-order extension of the longitudinal gauge, which is called Poisson gauge  $X^a = Y^a = 0$ .

**6.2. Einstein Tensor.** Here, we evaluate the second-order perturbation of the Einstein tensor (44) with the cosmological background (70). We evaluate the term  ${}^{(1)}\mathcal{G}_a{}^b[\mathcal{L}]$  and  ${}^{(2)}\mathcal{G}_a{}^b[\mathcal{H}, \mathcal{H}]$  in the Einstein equation (69).

First, we evaluate the term  ${}^{(1)}\mathcal{G}_a{}^b[\mathcal{L}]$  in the Einstein equation (69). Because the components (94) of  $\mathcal{L}_{ab}$  are obtained through the replacements

$$\begin{aligned} \Phi &\longrightarrow \Phi, & \nu_i &\longrightarrow \nu_i, & \Psi &\longrightarrow \Psi, & \chi_{ij} &\longrightarrow \chi_{ij} \end{aligned} \quad (102)$$

in the components (80) of  $\mathcal{H}_{ab}$ , we easily obtain the components of  ${}^{(1)}\mathcal{G}_a{}^b[\mathcal{L}]$  through the replacements (102) in (82).

From (80), we can derive the components of  ${}^{(2)}\mathcal{G}_a{}^b = {}^{(2)}\mathcal{G}_a{}^b[\mathcal{H}, \mathcal{H}]$  defined by (46)–(49) in a straightforward manner. Here, we use the results (85) and (89) of the first-order Einstein equations, for simplicity. Then the explicit components  ${}^{(2)}\mathcal{G}_a{}^b = {}^{(2)}\mathcal{G}_a{}^b[\mathcal{H}, \mathcal{H}]$  are summarized as

$$\begin{aligned} {}^{(2)}\mathcal{G}_\eta{}^\eta &= \frac{2}{a^2} \left[ -3D_k \Phi D^k \Phi - 8\Phi \Delta \Phi - 3\left(\partial_\eta \Phi\right)^2 \right. \\ &\quad - 12(\mathcal{H}^2 + K) \left(\Phi\right)^2 + D_l D_k \Phi \chi^{lk} \\ &\quad + \frac{1}{8} \partial_\eta \chi^{kl} \left(\partial_\eta + 8\mathcal{H}\right) \chi_{kl} + \frac{1}{2} D_k \chi_{lm} D^{[l} \chi^{k]m} \\ &\quad \left. - \frac{1}{8} D_k \chi_{lm} D^k \chi^{ml} - \frac{1}{2} \chi^{lm} (\Delta - K) \chi_{lm} \right], \\ {}^{(2)}\mathcal{G}_\eta{}^i &= \frac{2}{a^2} \left[ 8\Phi \partial_\eta D^i \Phi - D_j \Phi \partial_\eta \chi^{ij} \right. \\ &\quad - \left( \partial_\eta D_j \Phi + 2\mathcal{H} D_j \Phi \right) \chi^{ij} \\ &\quad \left. + \frac{1}{4} \partial_\eta \chi_{jk} D^i \chi^{kj} + \chi_{kl} \partial_\eta D^{[i} \chi^{k]l} \right], \end{aligned}$$

$$\begin{aligned} {}^{(2)}\mathcal{G}_i{}^\eta &= \frac{2}{a^2} \left[ 8\mathcal{H} \Phi D_i \Phi - 2D_i \Phi \partial_\eta \Phi + D^j \Phi \partial_\eta \chi_{ij} \right. \\ &\quad \left. - \partial_\eta D^j \Phi \chi_{ij} - \frac{1}{4} \partial_\eta \chi^{kj} D_i \chi_{kj} + \chi^{kj} \partial_\eta D_{[j} \chi_{i]k} \right], \\ {}^{(2)}\mathcal{G}_i{}^j &= \frac{2}{a^2} \left[ \left\{ -3D_k \Phi D^k \Phi - 4\Phi (\Delta + K) \Phi \right. \right. \\ &\quad - \partial_\eta \Phi \partial_\eta \Phi - 8\mathcal{H} \Phi \partial_\eta \Phi \\ &\quad \left. \left. - 4\left(2\partial_\eta \mathcal{H} + \mathcal{H}^2\right) \left(\Phi\right)^2 \right\} \gamma_i{}^j \right. \\ &\quad + 2D_i \Phi D^j \Phi + 4\Phi D_i D^j \Phi \\ &\quad + \chi_i{}^j \left( \partial_\eta^2 + 2\mathcal{H} \partial_\eta \right) \Phi \\ &\quad + D_k \Phi \left( D_i \chi^{jk} + D^j \chi_{ik} \right) - 2D^k \Phi D_k \chi_i{}^j \\ &\quad - 2\Phi (\Delta - 2K) \chi_i{}^j - \Delta \Phi \chi_i{}^j \\ &\quad + D_k D_i \Phi \chi^{jk} + D^m D^j \Phi \chi_{im} - D_l D_k \Phi \chi^{lk} \gamma_i{}^j \\ &\quad - \frac{1}{2} \partial_\eta \chi_{ik} \partial_\eta \chi^{kj} + D_k \chi_{il} D^{[k} \chi^{l]j} \\ &\quad + \frac{1}{4} D^j \chi_{lk} D_i \chi^{lk} \\ &\quad + \frac{1}{2} \chi_{lm} D_i D^j \chi^{ml} - \frac{1}{2} \chi_{lm} D^l D_i \chi^{mj} \\ &\quad - \frac{1}{2} \chi^{lm} D_l D^j \chi_{mi} + \frac{1}{2} \chi^{lm} D_m D_l \chi_i{}^j \\ &\quad - \frac{1}{2} \chi^{jk} \left( \partial_\eta^2 + 2\mathcal{H} \partial_\eta - \Delta + 2K \right) \chi_{ik} \\ &\quad + \frac{1}{2} \left[ \frac{3}{4} \partial_\eta \chi_{lk} \partial_\eta \chi^{kl} \right. \\ &\quad \left. + \chi^{kl} \left( \partial_\eta^2 + 2\mathcal{H} \partial_\eta - \Delta + K \right) \chi^{lk} \right. \\ &\quad \left. - \frac{1}{4} D_k \chi_{lm} D^k \chi^{ml} + D_k \chi_{lm} D^{[l} \chi^{k]m} \right\} \gamma_i{}^j \right]. \end{aligned} \quad (103)$$

We have checked the identity (51) through (103). Then, we may say that the expressions (103) are self-consistent.

6.3. *Energy-Momentum Tensor and Klein-Gordon Equation.* Here, we summarize the explicit components of the gauge-invariant parts (59) of the second-order perturbation of energy momentum tensor for a single scalar field in terms of gauge-invariant variables. Through (72), (80), and (94), the components of (59) are derived by the straightforward calculations. In this paper, we just summarize the components of  ${}^{(2)}\mathcal{T}_a^b$  in the situation where the first-order Einstein equations (85) and (89) are satisfied:

$$\begin{aligned}
a^{2(2)}\mathcal{T}_\eta^\eta &= -\partial_\eta\varphi\partial_\eta\varphi_2 + (\partial_\eta\varphi)^2\overset{(2)}{\Phi} - a^2\varphi_2\frac{\partial V}{\partial\varphi} \\
&\quad + 4\partial_\eta\varphi\overset{(1)}{\Phi}\partial_\eta\varphi_1 - 4(\partial_\eta\varphi)^2\left(\overset{(1)}{\Phi}\right)^2 - (\partial_\eta\varphi_1)^2 \\
&\quad - D_i\varphi_1D^i\varphi_1 - a^2(\varphi_1)^2\frac{\partial^2 V}{\partial\varphi^2}, \\
a^{2(2)}\mathcal{T}_i^\eta &= -\partial_\eta\varphi D_i\varphi_2 + 4\partial_\eta\varphi D_i\varphi_1\overset{(1)}{\Phi} - 2D_i\varphi_1\partial_\eta\varphi_1, \\
a^{2(2)}\mathcal{T}_\eta^i &= \partial_\eta\varphi D^i\varphi_2 + 2\partial_\eta\varphi_1D^i\varphi_1 + 4\partial_\eta\varphi\overset{(1)}{\Phi}D^i\varphi_1 \\
&\quad - 2\partial_\eta\varphi\overset{(1)}{\chi}{}^{ij}D_j\varphi_1, \\
a^{2(2)}\mathcal{T}_i^j &= D_i\varphi_1D^j\varphi_1 \\
&\quad + \frac{1}{2}\gamma_i{}^j\left\{\partial_\eta\varphi\partial_\eta\varphi_2 - 4\partial_\eta\varphi\overset{(1)}{\Phi}\partial_\eta\varphi_1\right. \\
&\quad \left.+ 4(\partial_\eta\varphi)^2\left(\overset{(1)}{\Phi}\right)^2 - (\partial_\eta\varphi)^2\overset{(2)}{\Phi} + (\partial_\eta\varphi_1)^2\right. \\
&\quad \left.- D_i\varphi_1D^i\varphi_1 - a^2\varphi_2\frac{\partial V}{\partial\varphi} - a^2(\varphi_1)^2\frac{\partial^2 V}{\partial\varphi^2}\right\}. \tag{104}
\end{aligned}$$

More generic formulae for the components of  ${}^{(2)}\mathcal{T}_a^b$  are given in [27].

Next, we show the gauge-invariant second-order the Klein-Gordon equation. We only consider the simple situation where (85) and (89) are satisfied. The formulae for more generic situation is given in [27]. Through (80), (94), and (72), the second-order perturbation of the Klein-Gordon equation (65) is given by

$$\begin{aligned}
-a^2\overset{(2)}{\mathcal{C}}_{(K)} &= \partial_\eta^2\varphi_2 + 2\mathcal{H}\partial_\eta\varphi_2 - \Delta\varphi_2 \\
&\quad - \left(\partial_\eta\overset{(2)}{\Phi} + 3\partial_\eta\overset{(2)}{\Psi}\right)\partial_\eta\varphi \\
&\quad + 2a^2\overset{(2)}{\Phi}\frac{\partial V}{\partial\varphi}(\varphi) + a^2\varphi_2\frac{\partial^2 V}{\partial\varphi^2}(\varphi) - \Xi_{(K)} \\
&= 0, \tag{105}
\end{aligned}$$

where we defined

$$\begin{aligned}
\Xi_{(K)} &:= 8\partial_\eta\overset{(1)}{\Phi}\partial_\eta\varphi_1 + 8\overset{(1)}{\Phi}\Delta\varphi_1 - 4a^2\overset{(1)}{\Phi}\varphi_1\frac{\partial^2 V}{\partial\varphi^2}(\varphi) \\
&\quad - a^2(\varphi_1)^2\frac{\partial^3 V}{\partial\varphi^3}(\varphi) + 8\overset{(1)}{\Phi}\partial_\eta\overset{(1)}{\Phi}\partial_\eta\varphi \\
&\quad - 2\overset{(1)}{\chi}{}^{ij}D_jD_i\varphi_1 + \partial_\eta\varphi\overset{(1)}{\chi}{}^{ij}\partial_\eta\overset{(1)}{\chi}{}_{ij}. \tag{106}
\end{aligned}$$

In (105),  $\Xi_{(K)}$  is the source term which is the collection of the quadratic terms of the linear-order perturbations in the second-order perturbation of the Klein-Gordon equation. If we ignore this source term, (105) coincides with the first-order perturbation of the Klein-Gordon equation. From this source term (106) of the Klein-Gordon equation, we can see that the mode-mode coupling due to the nonlinear effects appear in the second-order Klein-Gordon equation.

We cannot discuss solutions to (105) only through this equation, since this includes metric perturbations. To determine the behavior of the metric perturbations, we have to treat the Einstein equations simultaneously. The second-order Einstein equation is shown in Section 6.4.

6.4. *Einstein Equations.* Here, we show all the components of the second-order Einstein equation (69). All components of (69) are summarized as

$$\begin{aligned}
(-3\mathcal{H}\partial_\eta + \Delta + 3K)\overset{(2)}{\Psi} + (-\partial_\eta\mathcal{H} - 2\mathcal{H}^2 + K)\overset{(2)}{\Phi} \\
- 4\pi G\left(\partial_\eta\varphi\partial_\eta\varphi_2 + a^2\varphi_2\frac{\partial V}{\partial\varphi}\right) = \Gamma_0, \tag{107}
\end{aligned}$$

$$\begin{aligned}
2\partial_\eta D_i\overset{(2)}{\Psi} + 2\mathcal{H}D_i\overset{(2)}{\Phi} - \frac{1}{2}(\Delta + 2K)\overset{(2)}{\nu}_i \\
- 8\pi GD_i\varphi_2\partial_\eta\varphi = \Gamma_i, \tag{108}
\end{aligned}$$

$$\begin{aligned}
D_iD_j\left(\overset{(2)}{\Psi} - \overset{(2)}{\Phi}\right) \\
+ \left\{(-\Delta + 2\partial_\eta^2 + 4\mathcal{H}\partial_\eta - 2K)\overset{(2)}{\Psi}\right. \\
\left.+ (2\mathcal{H}\partial_\eta + 2\partial_\eta\mathcal{H} + 4\mathcal{H}^2 + \Delta + 2K)\overset{(2)}{\Phi}\right\}\gamma_{ij} \\
- \frac{1}{a^2}\partial_\eta\left(a^2D_{(i}\overset{(2)}{\nu}_{j)}\right) + \frac{1}{2}(\partial_\eta^2 + 2\mathcal{H}\partial_\eta + 2K - \Delta)\overset{(2)}{\chi}{}_{ij} \\
- 8\pi G\left(\partial_\eta\varphi\partial_\eta\varphi_2 - a^2\varphi_2\frac{\partial V}{\partial\varphi}(\varphi)\right)\gamma_{ij} = \Gamma_{ij}, \tag{109}
\end{aligned}$$

where  $\Gamma_0$ ,  $\Gamma_i$  and  $\Gamma_{ij}$  are the collection of the quadratic term of the first-order perturbations as follows:

$$\begin{aligned}
\Gamma_0 := & 4\pi G \left( (\partial_\eta \varphi_1)^2 + D_i \varphi_1 D^i \varphi_1 + a^2 (\varphi_1)^2 \frac{\partial^2 V}{\partial \varphi^2} \right) \\
& - 4 \partial_\eta \mathcal{H} \left( \Phi^{(1)} \right)^2 - 2 \Phi^{(1)} \partial_\eta^2 \Phi^{(1)} \\
& - 3 D_k \Phi^{(1)} D^k \Phi^{(1)} - 10 \Phi^{(1)} \Delta \Phi^{(1)} \\
& - 3 \left( \partial_\eta \Phi^{(1)} \right)^2 - 16 K \left( \Phi^{(1)} \right)^2 - 8 \mathcal{H}^2 \left( \Phi^{(1)} \right)^2 \quad (110) \\
& + D_l D_k \Phi^{(1)} \chi^{lk} + \frac{1}{8} \partial_\eta \chi^{lk} \partial_\eta \chi^{kl} + \mathcal{H} \chi^{kl} \partial_\eta \chi^{lk} \\
& - \frac{3}{8} D_k \chi^{lm} D^k \chi^{ml} + \frac{1}{4} D_k \chi^{lm} D^l \chi^{mk} \\
& - \frac{1}{2} \chi^{lm} \Delta \chi_{lm} + \frac{1}{2} K \chi^{lm} \chi^{lm},
\end{aligned}$$

$$\begin{aligned}
\Gamma_i := & 16\pi G \partial_\eta \varphi_1 D_i \varphi_1 - 4 \partial_\eta \Phi^{(1)} D_i \Phi^{(1)} + 8 \mathcal{H} \Phi^{(1)} D_i \Phi^{(1)} \\
& - 8 \Phi^{(1)} \partial_\eta D_i \Phi^{(1)} + 2 D_j \Phi^{(1)} \partial_\eta \chi_{ji} - 2 \partial_\eta D^j \Phi^{(1)} \chi_{ij} \quad (111) \\
& - \frac{1}{2} \partial_\eta \chi_{jk} D_i \chi^{kj} - \chi_{kl} \partial_\eta D_i \chi^{lk} + \chi^{kl} \partial_\eta D_k \chi_{il},
\end{aligned}$$

$$\begin{aligned}
\Gamma_{ij} := & 16\pi G D_i \varphi_1 D_j \varphi_1 \\
& + 8\pi G \left\{ (\partial_\eta \varphi_1)^2 - D_l \varphi_1 D^l \varphi_1 - a^2 (\varphi_1)^2 \frac{\partial^2 V}{\partial \varphi^2} \right\} \gamma_{ij} \\
& - 4 D_i \Phi^{(1)} D_j \Phi^{(1)} - 8 \Phi^{(1)} D_i D_j \Phi^{(1)} \\
& + \left( 6 D_k \Phi^{(1)} D^k \Phi^{(1)} + 4 \Phi^{(1)} \Delta \Phi^{(1)} + 2 \left( \partial_\eta \Phi^{(1)} \right)^2 + 8 \partial_\eta \mathcal{H} \left( \Phi^{(1)} \right)^2 \right. \\
& \quad \left. + 16 \mathcal{H}^2 \left( \Phi^{(1)} \right)^2 + 16 \mathcal{H} \Phi^{(1)} \partial_\eta \Phi^{(1)} - 4 \Phi^{(1)} \partial_\eta^2 \Phi^{(1)} \right) \gamma_{ij} \\
& - 4 \mathcal{H} \partial_\eta \Phi^{(1)} \chi_{ij} - 2 \partial_\eta^2 \Phi^{(1)} \chi_{ij} - 4 D^k \Phi^{(1)} D_{(i} \chi_{j)k} \\
& + 4 D^k \Phi^{(1)} D_k \chi_{ij} - 8 K \Phi^{(1)} \chi_{ij} \\
& + 4 \Phi^{(1)} \Delta \chi_{ij} - 4 D^k D_{(i} \Phi^{(1)} \chi_{j)k} \\
& + 2 \Delta \Phi^{(1)} \chi_{ij} + 2 D_l D_k \Phi^{(1)} \chi^{lk} \gamma_{ij} + \partial_\eta \chi_{ik} \partial_\eta \chi_j^k \\
& - D^k \chi_{il} D_k \chi_j^l + D^k \chi_{il} D^l \chi_{jk} - \frac{1}{2} D_i \chi^{lk} D_j \chi_{lk} \\
& - \chi_{lm}^{(1)} D_i D_j \chi^{ml} + 2 \chi^{lm} D_l D_{(i} \chi_{j)m} - \chi^{lm} D_m D_l \chi_{ij} \\
& - \frac{1}{4} \left( 3 \partial_\eta \chi_{lk} \partial_\eta \chi^{kl} - 3 D_k \chi_{lm} D^k \chi^{ml} \right. \\
& \quad \left. + 2 D_k \chi_{lm} D^l \chi^{mk} - 4 K \chi_{lm} \chi^{lm} \right) \gamma_{ij}. \quad (112)
\end{aligned}$$

Here, we used (75), (85), (87), (89), and (91).

The tensor part of (109) is given by

$$\begin{aligned}
& (\partial_\eta^2 + 2\mathcal{H}\partial_\eta + 2K - \Delta) \chi_{ij}^{(2)} \\
& = 2\Gamma_{ij} - \frac{2}{3} \gamma_{ij} \Gamma_k^k - 3 \left( D_i D_j - \frac{1}{3} \gamma_{ij} \Delta \right) (\Delta + 3K)^{-1} \\
& \quad \times \left( \Delta^{-1} D^k D_l \Gamma_k^l - \frac{1}{3} \Gamma_k^k \right) \\
& \quad + 4 \left\{ D_{(i} (\Delta + 2K)^{-1} D_{j)} \Delta^{-1} D^l D_k \Gamma_l^k \right. \\
& \quad \quad \left. - D_{(i} (\Delta + 2K)^{-1} D^k \Gamma_{j)k} \right\}. \quad (113)
\end{aligned}$$

This tensor mode is also called the second-order gravitational waves.

Further, the vector part of (108) yields the initial value constraint and the evolution equation of the vector mode  $\mathcal{V}_i^{(2)}$ :

$$\mathcal{V}_i^{(2)} = \frac{2}{\Delta + 2K} \left\{ D_i \Delta^{-1} D^k \Gamma_k - \Gamma_i \right\}, \quad (114)$$

$$\partial_\eta \left( a^2 \mathcal{V}_i^{(2)} \right) = \frac{2a^2}{\Delta + 2K} \left\{ D_i \Delta^{-1} D^k D_l \Gamma_k^l - D_k \Gamma_i^k \right\}.$$

Finally, scalar part of (108) are summarized as

$$2\partial_\eta \Psi^{(2)} + 2\mathcal{H} \Phi^{(2)} - 8\pi G \varphi_2 \partial_\eta \varphi = \Delta^{-1} D^k \Gamma_k, \quad (115)$$

$$\Psi^{(2)} - \Phi^{(2)} = \frac{3}{2} (\Delta + 3K)^{-1} \left\{ \Delta^{-1} D^i D_j \Gamma_i^j - \frac{1}{3} \Gamma_k^k \right\}, \quad (116)$$

$$\begin{aligned}
& \left( -\partial_\eta^2 - 5\mathcal{H}\partial_\eta + \frac{4}{3} \Delta + 4K \right) \Psi^{(2)} \\
& - \left( 2\partial_\eta \mathcal{H} + \mathcal{H}\partial_\eta + 4\mathcal{H}^2 + \frac{1}{3} \Delta \right) \Phi^{(2)} - 8\pi G a^2 \varphi_2 \frac{\partial V}{\partial \varphi} \\
& = \Gamma_0 - \frac{1}{6} \Gamma_k^k, \quad (117)
\end{aligned}$$

$$\begin{aligned}
& \left\{ \partial_\eta^2 + 2 \left( \mathcal{H} - \frac{\partial_\eta^2 \varphi}{\partial_\eta \varphi} \right) \partial_\eta - \Delta - 4K + 2 \left( \partial_\eta \mathcal{H} - \frac{\partial_\eta^2 \varphi}{\partial_\eta \varphi} \mathcal{H} \right) \right\} \Phi^{(2)} \\
& = -\Gamma_0 - \frac{1}{2} \Gamma_k^k + \Delta^{-1} D^i D_j \Gamma_i^j + \left( \partial_\eta - \frac{\partial_\eta^2 \varphi}{\partial_\eta \varphi} \right) \Delta^{-1} D^k \Gamma_k \\
& - \frac{3}{2} \left\{ \partial_\eta^2 - \left( \frac{2\partial_\eta^2 \varphi}{\partial_\eta \varphi} - \mathcal{H} \right) \partial_\eta \right\} (\Delta + 3K)^{-1} \\
& \quad \times \left\{ \Delta^{-1} D^i D_j \Gamma_i^j - \frac{1}{3} \Gamma_k^k \right\}, \quad (118)
\end{aligned}$$

where  $\Gamma_i^j := \gamma^{kj} \Gamma_{ik}$  and  $\Gamma_k^k = \gamma^{ij} \Gamma_{ij}$ . Equation (118) is the second-order extension of (92), which is the master equation of scalar mode of the second-order cosmological perturbation in a universe filled with a single scalar field.

Thus, we have a set of ten equations for the second-order perturbations of a universe filled with a single scalar field, (113)–(118). To solve this system of equations of the second-order Einstein equation, first of all, we have to solve

the linear-order system. This is accomplished by solving (92) to obtain the potential  $\Phi^{(1)}$ ,  $\varphi_1$  is given through (87), and the tensor mode  $\chi_{ij}^{(1)}$  is given by solving (90). Next, we evaluate the quadratic terms  $\Gamma_0$ ,  $\Gamma_i$ , and  $\Gamma_{ij}$  of the linear-order perturbations, which are defined by (110)–(112). Then, using the information of (110)–(112), we estimate the source term in (118). If we know the two independent solutions to the linear-order master equation (92), we can solve (118) through the method using the Green functions. After constructing the solution  $\Phi^{(2)}$  to (118), we can obtain the second-order metric perturbation  $\Psi^{(2)}$  through (116). Thus, we have obtained the second-order gauge-invariant perturbation  $\varphi_2$  of the scalar field through (115). Thus, the all scalar modes  $\Phi^{(2)}$ ,  $\Psi^{(2)}$ , and  $\varphi_2$  are obtained. Equation (117) is then used to check the consistency of the second-order perturbation of the Klein Gordon equation (105) as in Section 6.5.

For the vector-mode,  $\mathcal{V}_i^{(1)}$  of the first-order identically vanishes due to the momentum constraint (89) for the linear-order metric perturbations. On the other hand, in the second-order, we have evolution equation (114) of the vector mode  $\mathcal{V}_i^{(2)}$  with the initial value constraint. This evolution equation of the second-order vector mode should be consistent with the initial value constraint, which is confirmed in Section 6.5. Equations (114) also imply that the second-order vector-mode perturbation may be generated by the mode couplings of the linear-order perturbations. As the simple situations, the generation of the second-order vector mode due to the scalar-scalar mode coupling is discussed in [44–47].

The second-order tensor mode is also generated by the mode-coupling of the linear-order perturbations through the source term in (113). Note that (113) is almost same as (90) for the linear-order tensor mode, except for the existence of the source term in (113). If we know the solution to the linear-order Einstein equations (90) and (92), we can evaluate the source term in (113). Further, we can solve (113) through the Green function method. This leads the generation of the gravitational wave of the second order. Actually, in the simple situation where the first-order tensor mode neglected, the generation of the second-order gravitational waves discussed in some literature [48–54].

### 6.5. Consistency of Equations for Second-Order Perturbations.

Now, we consider the consistency of the second-order perturbations of the Einstein equations (115)–(118) for the scalar modes, (114) for vector mode, and the Klein-Gordon equation (105). The consistency check of these equations are necessary to guarantee that the derived equations are correct, since the second-order Einstein equations have complicated forms owing to the quadratic terms of the linear-order perturbations that arise from the nonlinear effects of the Einstein equations.

Since the first equation in (114) is the initial value constraint for the vector mode  $\mathcal{V}_i^{(2)}$  and it should be consistent

with the evolution equation, that is, the second equation of (114). these equations should be consistent with each other from the general arguments of the Einstein equation. Explicitly, these equations are consistent with each other if the equation

$$\partial_\eta \Gamma_k + 2\mathcal{H}\Gamma_k - D^i \Gamma_{ik} = 0 \quad (119)$$

is satisfied. Actually, through the first-order perturbative Einstein equations (87), (92), and (90), we can confirm (119). This is a trivial result from a general viewpoint, because the Einstein equation is the first class constrained system. However, this trivial result implies that we have derived the source terms  $\Gamma_i$  and  $\Gamma_{ij}$  of the second-order Einstein equations consistently.

Next, we consider (117). Through the second-order Einstein equations (115), (116), (118), and the background Klein-Gordon equation (76), we can confirm that (117) is consistent with the set of the background, first-order and other second-order Einstein equation if the equation

$$(\partial_\eta + 2\mathcal{H})D^k \Gamma_k - D^j D^i \Gamma_{ij} = 0 \quad (120)$$

is satisfied under the background and first-order Einstein equations. Actually, we have already seen that (119) is satisfied under the background and first-order Einstein equations. Taking the divergence of (119), we can immediately confirm (120). Then, (117) gives no information.

Thus, we have seen that the derived Einstein equations of the second-order (114)–(118) are consistent with each other through (119). This fact implies that the derived source terms  $\Gamma_i$  and  $\Gamma_{ij}$  of the second-order perturbations of the Einstein equations, which are defined by (111) and (112), are correct source terms of the second-order Einstein equations. On the other hand, for  $\Gamma_0$ , we have to consider the consistency between the perturbative Einstein equations and the perturbative Klein-Gordon equation as seen below.

Now, we consider the consistency of the second-order perturbation of the Klein-Gordon equation and the Einstein equations. The second-order perturbation of the Klein-Gordon equation is given by (105) with the source term (106). Since the vector mode  $\mathcal{V}_i^{(2)}$  and tensor mode  $\chi_{ij}^{(2)}$  of the second-order do not appear in the expressions (105) of the second-order perturbation of the Klein-Gordon equation, we may concentrate on the Einstein equations for scalar mode of the second order, that is, (115), (116), and (118) with the definitions (110)–(112) of the source terms. As in the linear case, the second-order perturbation of the Klein-Gordon equation should also be derived from the set of equations consisting of the second-order perturbations of the Einstein equations (115), (116), (118), the first-order perturbations of the Einstein equations (85), (87), (92), and the background Einstein equations (74). Actually, from these equations, we can show that the second-order perturbation of the Klein-Gordon equation is consistent with the background and the second-order Einstein equations if the equation

$$2(\partial_\eta + H)\Gamma_0 - D^k \Gamma_k + \mathcal{H}\Gamma_k^k + 8\pi G\partial_\eta \varphi \Xi_{(K)} = 0 \quad (121)$$

is satisfied under the background and the first-order Einstein equations. Further, we can also confirm (121) through the background Einstein equations (74), the scalar part of the first-order perturbation of the momentum constraint (87), and the evolution equations (92) and (90) for the first-order scalar and tensor modes in the Einstein equation.

As shown in [28], the first-order perturbation of the Klein-Gordon equation is derived from the background and the first-order perturbations of the Einstein equation. In the case of the second-order perturbation, the Klein-Gordon equation (105) can be also derived from the background, the first-order, and the second-order Einstein equations. The second-order perturbations of the Einstein equation and the Klein-Gordon equation include the source terms  $\Gamma_0$ ,  $\Gamma_i$ ,  $\Gamma_{ij}$ , and  $\Xi_{(K)}$  due to the mode-coupling of the linear-order perturbations. The second-order perturbation of the Klein-Gordon equation gives the relation (121) between the source terms  $\Gamma_0$ ,  $\Gamma_i$ ,  $\Gamma_{ij}$ ,  $\Xi_{(K)}$ , and we have also confirmed that (121) is satisfied due to the background, the first-order perturbation of the Einstein equations, and the Klein-Gordon equation. Thus, the second-order perturbation of the Klein-Gordon equation is not independent of the set of the background, the first-order, and the second-order Einstein equations if we impose on the Einstein equation at any conformal time  $\eta$ . This also implies that the derived formulae of the source terms  $\Gamma_0$ ,  $\Gamma_i$ ,  $\Gamma_{ij}$ , and  $\Xi_{(K)}$  are consistent with each other. In this sense, we may say that the formulae (110)–(112) and (106) for these source terms are correct.

## 7. Summary and Discussions

In this paper, we summarized the current status of the formulation of the gauge-invariant second-order cosmological perturbations. Although the presentation in this paper is restricted to the case of the universe filled by a single scalar field, the essence of the general framework of the gauge-invariant perturbation theory is transparent through this simple case. The general framework of the general relativistic higher-order gauge-invariant perturbation theory can be separated into three parts. First one is the general formulation to derive the gauge-transformation rules (18) and (19). Second one is the construction of the gauge-invariant variables for the perturbations on the generic background spacetime inspecting gauge-transformation rules (18) and (19) and the decomposition formula (36) and (37) for perturbations of any tensor field. Third one is the application of the above general framework of the gauge-invariant perturbation theory to the cosmological situations.

To derive the gauge-transformation rules (18) and (19), we considered the general arguments on the Taylor expansion of an arbitrary tensor field on a manifold, the general class of the diffeomorphism which is wider than the usual exponential map, and the general formulation of the perturbation theory. This general class of diffeomorphism is represented in terms of the Taylor expansion (2) of its pull-back. As commented in Section 2.1, this general class of diffeomorphism does not form a one-parameter group

of diffeomorphism as shown through (3). However, the properties (3) do not directly mean that this general class of diffeomorphism does not form a group. One of the key points of the properties of this diffeomorphism is the noncommutativity of generators  $\xi_1^a$  and  $\xi_2^a$  of each order. Although the expression of the  $n$ th-order Taylor expansion of the pull-back of this general class is discussed in [41], when we consider the situation of the  $n$ th-order perturbation, this noncommutativity becomes important [22]. Therefore, to clarify the properties of this general class of diffeomorphism, we have to take care of this noncommutativity of generators. Thus, there is a room to clarify the properties of this general class of diffeomorphism.

Further, in Section 2.3, we introduced a gauge choice  $\mathcal{X}_\lambda$  as an exponential map, for simplicity. On the other hand, we have the concept of the general class of diffeomorphism which is wider than the class of the exponential map. Therefore, we may introduce a gauge choice as one of the element of this general class of diffeomorphism. However, the gauge-transformation rules (18) and (19) will not be changed even if we generalize the definition of a each gauge choice as emphasized in Section 2.3. Although there is a room to sophisticate in logical arguments to derive the gauge-transformation rules (18) and (19), these are harmless to the development of the general framework of the gauge-invariant perturbation theory shown in Sections 2.3, 2.5, 3, and their application to cosmological perturbations in Section 4.

As emphasized in Section 2.5, our starting point to construct gauge-invariant variables is the assumption that *we already know the procedure for finding gauge-invariant variables for the linear metric perturbations as (23)*. This is highly nontrivial assumption on a generic background spacetime. The procedure to accomplish the decomposition (23) completely depends on the details of the background spacetime. In spite of this nontriviality, this assumption is almost correct in some background spacetime [55–59]. Further, once we accept this assumption, we can develop the higher-order perturbation theory in an independent manner of the details of the background spacetime. We also expect that this general framework of the gauge-invariant perturbation theory is extensible to  $n$ th-order perturbation theory, since our procedure to construct gauge-invariant variables can be extended to the third-order perturbation theory with two-parameter [22]. Due to this situation, in [27], we propose the conjecture which states that the above assumption for the decomposition of the linear-order metric perturbation is correct for any background spacetime. We may also say that the most nontrivial part of our general framework of higher-order gauge-invariant perturbation theory is in the above assumption. Further, as emphasized in Section 5.1, we assumed the existence of some Green functions to accomplish the decomposition (23) and this assumption exclude some perturbative modes of the metric perturbations from our consideration, even in the case of cosmological perturbations. For example, homogeneous modes of perturbations are excluded in our current arguments of the cosmological perturbation theory. These homogeneous modes will be necessary to discuss

the comparison with the arguments based on the long-wavelength approximation. Therefore, we have to say that there is a room to clarify even in the cosmological perturbation theory.

Even if the assumption is correct on any background spacetime, the other problem is in the interpretations of the gauge-invariant variables. We have commented on the nonuniqueness in the definitions of the gauge-invariant variables through (81) and (98). This nonuniqueness in the definition of gauge-invariant variables also leads some ambiguities in the interpretations of gauge-invariant variables. On the other hand, as emphasized in Section 2.3, any observations and experiments are carried out only on the physical spacetime through the physical processes on the physical spacetime. For this reason, any direct observables in any observations and experiments should be independent of the gauge choice. Further, the nonuniqueness in the definitions the gauge-invariant variables expressed by (81) and (98) have the same form as the decomposition formulae (36) and (37). Therefore, if the statement that *any direct observables in any observations and experiments is independent of the gauge choice* is really true, this also confirms that the nonuniqueness of the definition of the gauge-invariant variables also have nothing to do with the direct observables in observations and experiments. These will be confirmed by the clarification of the relations between gauge-invariant variables and observables in experiments and observations. To accomplish this, we have to specify the concrete process of experiments and observations and clarify the problem of what are the direct observables in the experiments and observations and derive the relations between the gauge-invariant variables and observables in concrete observations and experiments. If these arguments are completed, we will be able to show that the gauge degree of freedom is just an unphysical degree of freedom and the nonuniqueness of the gauge-invariant variables has nothing to do with the direct observables in the concrete observation or experiment, and then, we will be able to clarify the precise physical interpretation of the gauge-invariant variables.

For example, in the case of the CMB physics, we can easily see that the first-order perturbation of the CMB temperature is automatically gauge invariant from (36), because the background temperature of CMB is homogeneous. On the other hand, the decomposition formula (37) of the second order yields that the theoretical prediction of the second-order perturbation of the CMB temperature may depend on gauge choice, since we do know the existence of the first-order fluctuations as the temperature anisotropy in CMB. However, as emphasized above, the direct observables in observations should be gauge invariant and the gauge-variant term in (37) should be disappear in the direct observables. Therefore, we have to clarify how the gauge-invariant variables are related to the observed temperature fluctuations and the gauge-variant term disappears in the observable.

Although there are some rooms to accomplish the complete formulation of the second-order cosmological perturbation theory, we derived all the components of the second-order perturbation of the Einstein equation without

ignoring any types modes (scalar-, vector-, tensor-types) of perturbations in the case of a scalar field system. In our formulation, any gauge fixing is not necessary and we can obtain all equations in the gauge-invariant form, which are equivalent to the complete gauge fixing. In other words, our formulation gives complete gauge-fixed equations without any gauge fixing. Therefore, equations obtained in a gauge-invariant manner cannot be reduced without physical restrictions any more. In this sense, the equations shown here are irreducible. This is one of the advantages of the gauge-invariant perturbation theory.

The resulting Einstein equations of the second order show that any type of mode-coupling appears as the quadratic terms of the linear-order perturbations owing to the nonlinear effect of the Einstein equations, in principle. Perturbations in cosmological situations are classified into three types: scalar, vector, and tensor. In the second-order perturbations, we also have these three types of perturbations as in the case of the first-order perturbations. Furthermore, in the equations for the second-order perturbations, there are many quadratic terms of linear-order perturbations owing to the nonlinear effects of the system. Owing to these nonlinear effects, the above three types of perturbations couple with each other. In the scalar field system shown in this paper, the first-order vector mode does not appear due to the momentum constraint of the first-order perturbation of the Einstein equation. Therefore, we have seen that three types of mode-coupling appear in the second-order Einstein equations, that is, scalar-scalar, scalar-tensor, and tensor-tensor type of mode coupling. In general, all types of mode-coupling may appear in the second-order Einstein equations. Actually, in [28], we also derived all the components of the Einstein equations for a perfect fluid system and we can see all types of mode-coupling, that is, scalar-scalar, scalar-vector, scalar-tensor, vector-vector, vector-tensor, and tensor-tensor types mode-coupling, appear in the second-order Einstein equation, in general. Of course, in the some realistic situations of cosmology, we may neglect some modes. In this case, we may neglect some mode-coupling. However, even in this case, we should keep in mind the fact that all types of mode-couplings may appear in principle when we discuss the realistic situations of cosmology. We cannot deny the possibility that the mode-couplings of any type produces observable effects when the quite high accuracy of observations is accomplished.

Even in the case of the single scalar field discussed in this paper, the source terms of the second-order Einstein equation show the mode-coupling of scalar-scalar, scalar-tensor, and the tensor-tensor types as mentioned above. Since the tensor mode of the linear order is also generated due to quantum fluctuations during the inflationary phase, the mode-couplings of the scalar-tensor and tensor-tensor types may appear in the inflation. If these mode-couplings occur during the inflationary phase, these effects will depend on the scalar-tensor ratio  $r$ . If so, there is a possibility that the accurate observations of the second-order effects in the fluctuations of the scalar type in our universe also restrict the scalar-tensor ratio  $r$  or give some consistency relations between the other observations such as the measurements

of the B-mode of the polarization of CMB. This will be a new effect that gives some information on the scalar-tensor ratio  $r$ .

Furthermore, we have also checked the consistency between the second-order perturbations of the equations of motion of matter field and the Einstein equations. In the case of a scalar field, we checked the consistency between the second-order perturbations of the Klein-Gordon equation and the Einstein equations. Due to this consistency check, we have obtained the consistency relations between the source terms in these equations  $\Gamma_0$ ,  $\Gamma_i$ ,  $\Gamma_{ij}$ , and  $\Xi_{(K)}$ , which are given by (119) and (121). We note that the relation (119) comes from the consistency in the Einstein equations of the second order by itself, while the relation (121) comes from the consistency between the second-order perturbation of the Klein-Gordon equation and the Einstein equation. We also showed that these relations between the source terms are satisfied through the background and the first-order perturbation of the Einstein equations in [28]. This implies that the set of all equations are self-consistent and the derived source terms  $\Gamma_0$ ,  $\Gamma_i$ ,  $\Gamma_{ij}$ , and  $\Xi_{(K)}$  are correct. We also note that these relations are independent of the details of the potential of the scalar field.

Thus, we have derived the self-consistent set of equations of the second-order perturbation of the Einstein equations and the evolution equations of matter fields in terms of gauge-invariant variables. As the current status of the second-order gauge-invariant cosmological perturbation theory, we may say that the curvature terms in the second-order Einstein tensor (69), that is, the second-order perturbations of the Einstein tensor, are almost completely derived although there remains the problem of homogeneous modes as mentioned above. After complete the problem of homogeneous modes, we have to clarify the physical behaviors of the second-order cosmological perturbation in the single scalar field system in the context of the inflationary scenario. This is the preliminary step to clarify the quantum behaviors of second-order perturbations in the inflationary universe. Further, we also have to carry out the comparison with the result by long-wavelength approximations. If these issues are completed, we may say that we have completely understood the properties of the second-order perturbation of the Einstein tensor. The next task is to clarify the nature of the second-order perturbation of the energy-momentum tensor through the extension to multi-fluid or multi-field systems. Further, we also have to extend our arguments to the Einstein Boltzmann system to discuss CMB physics, since we have to treat photon and neutrinos through the Boltzmann distribution functions. This issue is also discussed in some literature [13–21, 29, 30]. If we accomplish these extension, we will be able to clarify the Non-linear effects in CMB physics.

Finally, readers might think that the ingredients of this paper is too mathematical as Astronomy. However, we have to emphasize that a high degree of the theoretical sophistication leads unambiguous theoretical predictions in many cases. As in the case of the linear-order cosmological perturbation theory, the developments in observations are also supported by the theoretical sophistication and the

theoretical sophistication are accomplished motivated by observations. In this sense, now, we have an opportunity to develop the general relativistic second-order perturbation theory to a high degree of sophistication which is motivated by observations. We also expect that this theoretical sophistication will be also useful to discuss the theoretical predictions of non-Gaussianity in CMB and comparison with observations. Therefore, I think that this opportunity is opened not only for observational cosmologists but also for theoretical and mathematical physicists.

## Appendices

### A. Derivation of the Generic Representation of the Taylor Expansion of Tensors on a Manifold

In this section, we derive the representation of the coefficients of the formal Taylor expansion (2) of the pull-back of a diffeomorphism in terms of the suitable derivative operators. The guide principle of our arguments is the following theorem [38, 40].

**Theorem A.** *Let  $\mathcal{D}$  be a derivative operator acting on the set of all the tensor fields defined on a differentiable manifold  $\mathcal{M}$  and satisfying the following conditions: (i) it is linear and satisfies the Leibniz rule; (ii) it is tensor-type preserving; (iii) it commutes with every contraction of a tensor field; and (iv) it commutes with the exterior differentiation  $d$ . Then,  $\mathcal{D}$  is equivalent to the Lie derivative operator with respect to some vector field  $\xi$ , that is,  $\mathcal{D} = \mathcal{L}_\xi$ .*

The proof of the assertion of Theorem A is given in [38] as follows. When acting on functions, the derivative operator  $\mathcal{D}$  defines a vector field  $\xi$  through the relation

$$\mathcal{D}f =: \xi(f) = \xi_\xi f, \quad \forall f \in \mathcal{F}(M). \quad (\text{A.1})$$

The assertion of the theorem for an arbitrary tensor field holds if and only if the assertions for an arbitrary scalar function and for an arbitrary vector field  $V$  hold. To do this, we consider the scalar function  $V(f)$  and we obtain that

$$\mathcal{D}(V(f)) = \xi(V(f)) \quad (\text{A.2})$$

through (A.1). Through the conditions (i)–(iv) of  $\mathcal{D}$ ,  $\mathcal{D}(V(f))$  is also given by

$$\begin{aligned} \mathcal{D}(V(f)) &= \mathcal{D}(df(V)) = \mathcal{D}\{\mathcal{C}(df \otimes V)\} \\ &= \mathcal{C}\{\mathcal{D}(df \otimes V)\} \\ &= \mathcal{C}\{\mathcal{D}(df) \otimes V + df \otimes \mathcal{D}V\} \\ &= \mathcal{C}\{d(\mathcal{D}f) \otimes V + df \otimes \mathcal{D}V\} \\ &= d(\mathcal{D}f)(V) + df(\mathcal{D}V) \\ &= V(\mathcal{D}f) + (\mathcal{D}V)(f). \end{aligned} \quad (\text{A.3})$$

Then we obtain that

$$\begin{aligned} (\mathcal{D}V)(f) &= \xi(V(f)) - V(\xi(f)) = [\xi, V](f) \\ &= (\xi_\xi V)(f) \end{aligned} \quad (\text{A.4})$$

for an arbitrary  $f$ , that is,

$$\mathcal{D}V = \xi_\xi V. \quad (\text{A.5})$$

Through (A.1) and (A.5), we can recursively show that

$$\mathcal{D}Q = \xi_\xi Q \quad (\text{A.6})$$

for an arbitrary tensor field  $Q$  [40].

Now, we consider the derivation of the Taylor expansion (1). As in the main text, we first consider the representation of the Taylor expansion of  $\Phi_\lambda^* f$  for an arbitrary scalar function  $f \in \mathcal{F}(M)$ :

$$\begin{aligned} (\Phi_\lambda^* f)(p) &= f(p) + \lambda \left\{ \frac{\partial}{\partial \lambda} (\Phi_\lambda^* f) \right\}_{\lambda=0} \\ &\quad + \frac{1}{2} \lambda^2 \left\{ \frac{\partial^2}{\partial \lambda^2} (\Phi_\lambda^* f) \right\}_{\lambda=0} + O(\lambda^3), \end{aligned} \quad (\text{A.7})$$

where  $\mathcal{F}(M)$  denotes the algebra of  $C^\infty$  functions on  $\mathcal{M}$ . Although the operator  $\partial/\partial\lambda$  in the bracket  $\{*\}_{\lambda=0}$  of (A.7) are simply symbolic notation, we stipulate the properties

$$\left\{ \frac{\partial^2}{\partial \lambda^2} (\Phi_\lambda^* f) \right\}_{\lambda=0} = \left\{ \frac{\partial}{\partial \lambda} \left( \frac{\partial}{\partial \lambda} (\Phi_\lambda^* f) \right) \right\}_{\lambda=0}, \quad (\text{A.8})$$

$$\left\{ \frac{\partial}{\partial \lambda} (\Phi_\lambda^* f)^2 \right\}_{\lambda=0} = \left\{ 2\Phi_\lambda^* f \frac{\partial}{\partial \lambda} (\Phi_\lambda^* f) \right\}_{\lambda=0}, \quad (\text{A.9})$$

for all  $f \in \mathcal{F}(M)$ , where  $n$  is an arbitrary finite integer. These properties imply that the operator  $\partial/\partial\lambda$  is in fact not simply symbolic notation but indeed the usual partial differential operator on  $\mathbb{R}$ . We note that the property (A.9) is the Leibniz rule, which plays important roles when we derive the representation of the Taylor expansion (A.7) in terms of suitable Lie derivatives.

Together with the property (A.9), Theorem A yields that there exists a vector field  $\xi_1$ , so that

$$\left\{ \frac{\partial}{\partial \lambda} (\Phi_\lambda^* f) \right\}_{\lambda=0} =: \xi_{\xi_1} f. \quad (\text{A.10})$$

Actually, the conditions (ii)–(iv) in Theorem A are satisfied from the fact that  $\Phi_\lambda^*$  is the pull-back of a diffeomorphism  $\Phi_\lambda$  and (i) is satisfied due to the property (A.9).

Next, we consider the second-order term in (A.7). Since we easily expect that the second-order term in (A.7) may include  $\mathcal{L}_{\xi_1}^2$ , we define the derivative operator  $\mathcal{L}_2$  by

$$\left\{ \frac{\partial^2}{\partial \lambda^2} (\Phi_\lambda^* f) \right\}_{\lambda=0} =: (\mathcal{L}_2 + a\xi_{\xi_1}^2) f, \quad (\text{A.11})$$

where  $a$  is determined so that  $\mathcal{L}_2$  satisfies the conditions of Theorem A. The conditions (ii)–(iv) in Theorem A for

$\mathcal{L}_2$  are satisfied from the fact that  $\Phi_\lambda^*$  is the pull-back of a diffeomorphism  $\Phi_\lambda$ . Further,  $\mathcal{L}_2$  is obviously linear but we have to check  $\mathcal{L}_2$  satisfy the Leibniz rule, that is,

$$\mathcal{L}_2(f^2) = 2f \mathcal{L}_2 f, \quad (\text{A.12})$$

for all  $f \in \mathcal{F}(M)$ . To do this, we use the properties (A.8) and (A.9), then we can easily see that the Leibniz rule (A.12) is satisfied if and only if  $a = 1$  and we may regard  $\mathcal{L}_2$  as the Lie derivative with respect to some vector field. Then, when and only when  $a = 1$ , there exists a vector field  $\xi_2$  such that

$$\mathcal{L}_2 f = \xi_{\xi_2} f,$$

$$\left\{ \frac{\partial^2}{\partial \lambda^2} (\Phi_\lambda^* f) \right\}_{\lambda=0} =: (\xi_{\xi_2} + \xi_{\xi_1}^2) f. \quad (\text{A.13})$$

Thus, we have seen that the Taylor expansion (A.7) for an arbitrary scalar function  $f$  is given by (2).

Although the formula (2) of the Taylor expansion is for an arbitrary scalar function, we can easily extend this formula to that for an arbitrary tensor field  $Q$  as the assertion of Theorem A. The proof of the extension of the formula (2) to an arbitrary tensor field  $Q$  is completely parallel to the proof of the formula (2) for an arbitrary scalar function if we stipulate the properties

$$\begin{aligned} \left\{ \frac{\partial^2}{\partial \lambda^2} (\Phi_\lambda^* Q) \right\}_{\lambda=0} &= \left\{ \frac{\partial}{\partial \lambda} \left( \frac{\partial}{\partial \lambda} (\Phi_\lambda^* Q) \right) \right\}_{\lambda=0}, \\ \left\{ \frac{\partial}{\partial \lambda} (\Phi_\lambda^* Q)^2 \right\}_{\lambda=0} &= \left\{ 2\Phi_\lambda^* Q \frac{\partial}{\partial \lambda} (\Phi_\lambda^* Q) \right\}_{\lambda=0} \end{aligned} \quad (\text{A.14})$$

instead of (A.8) and (A.9). As the result, we obtain the representation of the Taylor expansion for an arbitrary tensor field  $Q$ .

## B. Derivation of the Perturbative Einstein Tensors

Following the outline of the calculations explained in Section 3.1, we first calculate the perturbative expansion of the inverse metric. The perturbative expansion of the inverse metric can be easily derived from (22) and the definition of the inverse metric

$$\bar{g}^{ab} \bar{g}_{bc} = \delta_c^a. \quad (\text{B.1})$$

We also expand the inverse metric  $\bar{g}^{ab}$  in the form

$$\bar{g}^{ab} = g^{ab} + \lambda^{(1)} \bar{g}^{ab} + \frac{1}{2} \lambda^{2(2)} \bar{g}^{ab}. \quad (\text{B.2})$$

Then, each term of the expansion of the inverse metric is given by

$${}^{(1)}\bar{g}^{ab} = -h^{ab}, \quad {}^{(2)}\bar{g}^{ab} = 2h^{ac} h_c^b - l^{ab}. \quad (\text{B.3})$$

To derive the formulae for the perturbative expansion of the Riemann curvature, we have to derive the formulae for the perturbative expansion of the tensor  $C_{ab}^c$  given by (40).

The tensor  $C^c{}_{ab}$  is also expanded in the same form as (11). The first-order perturbations of  $C^c{}_{ab}$  have the well-known form [42]

$${}^{(1)}C^c{}_{ab} = \nabla_{(a}h_{b)}{}^c - \frac{1}{2}\nabla^c h_{ab} =: H_{ab}{}^c[h], \quad (\text{B.4})$$

where  $H_{ab}{}^c[A]$  is defined by (48) for an arbitrary tensor field  $A_{ab}$  defined on the background spacetime  $\mathcal{M}_0$ . In terms of the tensor field  $H_{ab}{}^c$  defined by (48) the second-order perturbation  ${}^{(2)}C^c{}_{ab}$  of the tensor field  $C^c{}_{ab}$  is given by

$${}^{(2)}C^c{}_{ab} = H_{ab}{}^c[l] - 2h^{cd}H_{abd}[h]. \quad (\text{B.5})$$

The Riemann curvature (41) on the physical spacetime  $\mathcal{M}_\lambda$  is also expanded in the form (11):

$$\bar{R}_{abc}{}^d =: R_{abc}{}^d + \lambda^{(1)}R_{abc}{}^d + \frac{1}{2}\lambda^{2(2)}R_{abc}{}^d + O(\lambda^3). \quad (\text{B.6})$$

The first- and the second-order perturbation of the Riemann curvature are given by

$$\begin{aligned} {}^{(1)}R_{abc}{}^d &= -2\nabla_{[a}{}^{(1)}C^d{}_{b]c}, \\ {}^{(2)}R_{abc}{}^d &= -2\nabla_{[a}{}^{(2)}C^d{}_{b]c} + 4{}^{(1)}C^e{}_{c[a}{}^{(1)}C^d{}_{b]e}. \end{aligned} \quad (\text{B.7})$$

Substituting (B.4) and (B.5) into (B.7), we obtain the perturbative form of the Riemann curvature in terms of the variables defined by (48) and (49):

$${}^{(1)}R_{abc}{}^d = -2\nabla_{[a}H_{b]c}{}^d[h], \quad (\text{B.8})$$

$$\begin{aligned} {}^{(2)}R_{abc}{}^d &= -2\nabla_{[a}H_{b]c}{}^d[l] + 4H_{[a}{}^{de}[h]H_{b]ce}[h] \\ &\quad + 4h^{de}\nabla_{[a}H_{b]ce}[h]. \end{aligned} \quad (\text{B.9})$$

To write down the perturbative curvatures (B.8) and (B.9) in terms of the gauge invariant and variant variables defined by (23) and (32), we first derive an expression for the tensor field  $H_{abc}[h]$  in terms of the gauge-invariant variables, and then, we derive a perturbative expression for the Riemann curvature.

First, we consider the linear-order perturbation (B.8) of the Riemann curvature. Using the decomposition (23) and the identity  $R_{[abc]}{}^d = 0$ , we can easily derive the relation

$$H_{abc}[h] = H_{abc}[\mathcal{H}] + \nabla_a\nabla_b X_c + R_{bca}{}^d X_d, \quad (\text{B.10})$$

where the variable  $H_{abc}[\mathcal{H}]$  is defined by (48) and (49) with  $A_{ab} = \mathcal{H}_{ab}$ . Clearly, the variable  $H_{ab}{}^c[\mathcal{H}]$  is gauge invariant. Taking the derivative and using the Bianchi identity  $\nabla_{[a}R_{bc]de} = 0$ , we obtain that

$${}^{(1)}R_{abc}{}^d = -2\nabla_{[a}H_{b]c}{}^d[\mathcal{H}] + \xi_X R_{abc}{}^d. \quad (\text{B.11})$$

Similar but some cumbersome calculations yield

$$\begin{aligned} {}^{(2)}R_{abc}{}^d &= -2\nabla_{[a}H_{b]c}{}^d[\mathcal{L}] + 4H_{[a}{}^{de}[\mathcal{H}]H_{b]ce}[\mathcal{H}] \\ &\quad + 4\mathcal{H}_c{}^d\nabla_{[a}H_{b]c}{}^e[\mathcal{H}] \\ &\quad + 2\xi_X{}^{(1)}R_{abc}{}^d + (\xi_Y - \xi_X^2)R_{abc}{}^d. \end{aligned} \quad (\text{B.12})$$

Equations (B.11) and (B.12) have the same for as the decomposition formulae (36) and (37), respectively.

Contracting the indices  $b$  and  $d$  in (B.11) and (B.12) of the perturbative Riemann curvature, we can directly derive the formulae for the perturbative expansion of the Ricci curvature: expanding the Ricci curvature

$$\bar{R}_{ab} =: R_{ab} + \lambda^{(1)}R_{ab} + \frac{1}{2}\lambda^{2(2)}R_{ab} + O(\lambda^3), \quad (\text{B.13})$$

we obtain the first-order Ricci curvature as

$${}^{(1)}R_{ab} = -2\nabla_{[a}H_{c]b}{}^c[\mathcal{H}] + \xi_X R_{ab}. \quad (\text{B.14})$$

and we also obtain the second-order Ricci curvature as

$$\begin{aligned} {}^{(2)}R_{ab} &= -2\nabla_{[a}H_{c]b}{}^c[\mathcal{L}] + 4H_{[a}{}^{cd}[\mathcal{H}]H_{c]bd}[\mathcal{H}] \\ &\quad + 4\mathcal{H}_d{}^c\nabla_{[a}H_{b]c}{}^d[\mathcal{H}] \\ &\quad + 2\xi_X{}^{(1)}R_{ab} + (\xi_Y - \xi_X^2)R_{ab}. \end{aligned} \quad (\text{B.15})$$

The scalar curvature on the physical spacetime  $\mathcal{M}$  is given by  $\bar{R} = \bar{g}^{ab}\bar{R}_{ab}$ . To obtain the perturbative form of the scalar curvature, we expand the  $\bar{R}$  in the form (11), that is,

$$\bar{R} =: R + \lambda^{(1)}R + \frac{1}{2}\lambda^{2(2)}R + O(\lambda^3) \quad (\text{B.16})$$

and  $\bar{g}^{ab}\bar{R}_{ab}$  is expanded through the Leibniz rule. Then, the perturbative formula for the scalar curvature at each order is derived from perturbative form of the inverse metric (B.3) and the Ricci curvature (B.14) and (B.15). Straightforward calculations lead to the expansion of the scalar curvature as

$${}^{(1)}R = -2\nabla_{[a}H_{b]}{}^{ab}[\mathcal{H}] - R_{ab}\mathcal{H}^{ab} + \xi_X R, \quad (\text{B.17})$$

$$\begin{aligned} {}^{(2)}R &= -2\nabla_{[a}H_{b]}{}^{ab}[\mathcal{L}] + R^{ab}(2\mathcal{H}_{ca}\mathcal{H}_b{}^c - \mathcal{L}_{ab}) \\ &\quad + 4H_{[a}{}^{cd}[\mathcal{H}]H_{c]}{}^a{}_d[\mathcal{H}] + 4\mathcal{H}_c{}^b\nabla_{[a}H_{b]}{}^{ac}[\mathcal{H}] \\ &\quad + 4\mathcal{H}^{ab}\nabla_{[a}H_{d]b}{}^d[\mathcal{H}] \\ &\quad + 2\xi_X{}^{(1)}R + (\xi_Y - \xi_X^2)R. \end{aligned} \quad (\text{B.18})$$

We also note that the expansion formulae (B.17) and (B.18) have the same for as the decomposition formulae (36) and (37), respectively.

Next, we consider the perturbative form of the Einstein tensor  $\bar{G}_{ab} := \bar{R}_{ab} - (1/2)\bar{g}_{ab}\bar{R}$  and we expand  $\bar{G}_{ab}$  as in the form (11):

$$\bar{G}_{ab} =: G_{ab} + \lambda^{(1)}(G_{ab}) + \frac{1}{2}\lambda^{(2)}(G_{ab}) + O(\lambda^3). \quad (\text{B.19})$$

As in the case of the scalar curvature, straightforward calculations lead to

$$\begin{aligned} {}^{(1)}(G_{ab}) &= -2\nabla_{[a}H_{d]b}{}^d[\mathcal{H}] + g_{ab}\nabla_{[c}H_b]{}^{cd}[\mathcal{H}] \\ &\quad - \frac{1}{2}R\mathcal{H}_{ab} + \frac{1}{2}g_{ab}R_{cd}\mathcal{H}^{cd} + \xi_X G_{ab}, \end{aligned} \quad (\text{B.20})$$

$$\begin{aligned} {}^{(2)}(G_{ab}) &= -2\nabla_{[a}H_{c]b}{}^c[\mathcal{L}] + 4H_{[a}{}^{cd}[\mathcal{H}]H_{c]bd}[\mathcal{H}] \\ &\quad + 4\mathcal{H}_c{}^d\nabla_{[a}H_{d]b}{}^c[\mathcal{H}] \\ &\quad - \frac{1}{2}g_{ab}\left(-2\nabla_{[c}H_d]{}^{cd}[\mathcal{L}] + 2R_{de}\mathcal{H}_c{}^d\mathcal{H}^{ec} \right. \\ &\quad \left. - R_{de}\mathcal{L}^{de} + 4H_{[c}{}^{de}[\mathcal{H}]H_{d]}{}^c{}_e[\mathcal{H}] \right. \\ &\quad \left. + 4\mathcal{H}_e{}^d\nabla_{[c}H_d]{}^{ce}[\mathcal{H}] \right. \\ &\quad \left. + 4\mathcal{H}^{ce}\nabla_{[c}H_{d]}{}^d{}_e[\mathcal{H}]\right) \\ &\quad + 2\mathcal{H}_{ab}\nabla_{[c}H_d]{}^{cd}[\mathcal{H}] + \mathcal{H}_{ab}\mathcal{H}^{cd}R_{cd} - \frac{1}{2}R\mathcal{L}_{ab} \\ &\quad + 2\xi_X{}^{(1)}(G_{ab}) + (\xi_Y - \xi_X^2)G_{ab}. \end{aligned} \quad (\text{B.21})$$

We note again that (B.20) and (B.21) have the same form as the decomposition formulae (36) and (37), respectively.

The perturbative formulae for the perturbation of the Einstein tensor

$$\bar{G}_a{}^b = \bar{g}^{bc}\bar{G}_{ac} \quad (\text{B.22})$$

is derived by the similar manner to the case of the perturbations of the scalar curvature. Through these formulae summarized previously, straightforward calculations leads (43)–(47). We have to note that to derive the formulae (46) with (47), we have to consider the general relativistic gauge-invariant perturbation theory with two infinitesimal parameters which is developed in [22, 23], as commented in the main text.

## Acknowledgments

The author thanks participants in the GCOE/YITP workshop YITP-W-0901 on “Non-linear cosmological perturbations” which was held at YITP in Kyoto, Japan in April, 2009, for valuable discussions, in particular, professor M. Bruni, professor R. Maartens, professor M. Sasaki, professor T. Tanaka, and professor K. Tomita. This paper is an extension of the contribution to this workshop by the author.

## References

- [1] J. M. Bardeen, “Gauge-invariant cosmological perturbations,” *Physical Review D*, vol. 22, no. 8, pp. 1882–1905, 1980.
- [2] H. Kodama and M. Sasaki, “Cosmological perturbation theory,” *Progress of Theoretical Physics Supplement*, no. 78, p. 1, 1984.
- [3] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, “Theory of cosmological perturbations,” *Physics Report*, vol. 215, no. 5-6, pp. 203–333, 1992.
- [4] C. L. Bennett, M. Halpern, G. Hinshaw et al., “First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: preliminary maps and basic results,” *Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 1–27, 2003.
- [5] E. Komatsu, J. Dunkley, M. R. Nolte et al., “Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [6] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant second-order perturbations and non-Gaussianity from inflation,” *Nuclear Physics B*, vol. 667, no. 1-2, pp. 119–148, 2003.
- [7] J. Maldacena, “Non-gaussian features of primordial fluctuations in single field inflationary models,” *Journal of High Energy Physics*, vol. 2003, no. 5, article 013, 2003.
- [8] K. A. Malik and D. Wands, “Evolution of second-order cosmological perturbations,” *Classical and Quantum Gravity*, vol. 21, no. 11, pp. L65–L71, 2004.
- [9] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity in the curvaton scenario,” *Physical Review D*, vol. 69, no. 4, Article ID 043503, 2004.
- [10] N. Bartolo, S. Matarrese, and A. Riotto, “Enhancement of non-gaussianity after inflation,” *Journal of High Energy Physics*, vol. 8, no. 4, article 006, pp. 147–159, 2004.
- [11] D. H. Lyth and Y. Rodríguez, “Non-Gaussianity from the second-order cosmological perturbation,” *Physical Review D*, vol. 71, no. 12, Article ID 123508, 2005.
- [12] F. Vernizzi, “Conservation of second-order cosmological perturbations in a scalar field dominated universe,” *Physical Review D*, vol. 71, no. 6, Article ID 061301, 5 pages, 2005.
- [13] N. Bartolo, S. Matarrese, and A. Riotto, “Evolution of second-order cosmological perturbations and non-Gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, no. 1, article 003, pp. 47–70, 2004.
- [14] N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant temperature anisotropies and primordial non-Gaussianity,” *Physical Review Letters*, vol. 93, no. 23, Article ID 231301, 2004.
- [15] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, “Non-Gaussianity from inflation: theory and observations,” *Physics Reports*, vol. 402, no. 3-4, pp. 103–266, 2004.
- [16] N. Bartolo, S. Matarrese, and A. Riotto, “The full second-order radiation transfer function for large-scale CMB anisotropies,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 010, 2006.
- [17] N. Bartolo, S. Matarrese, and A. Riotto, “Cosmic microwave background anisotropies at second order: I,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 6, article 024, 2006.
- [18] N. Bartolo, S. Matarrese, and A. Riotto, “CMB anisotropies at second-order II: analytical approach,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 1, article 019, 2007.

- [19] D. Nitta, E. Komatsu, N. Bartolo, S. Matarrese, and A. Riotto, “CMB anisotropies at second order III: bispectrum from products of the first-order perturbations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 014, 2009.
- [20] C. Pitrou, J.-P. Uzan, and F. Bernardeau, “Cosmic microwave background bispectrum on small angular scales,” *Physical Review D*, vol. 78, no. 6, Article ID 063526, 2008.
- [21] L. Senatore, S. Tassev, and M. Zaldarriaga, “Cosmological perturbations at second order and recombination perturbed,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 031, 2009.
- [22] K. Nakamura, “Gauge invariant variables in two-parameter nonlinear perturbations,” *Progress of Theoretical Physics*, vol. 110, no. 4, pp. 723–756, 2003.
- [23] K. Nakamura, “Second-order gauge invariant perturbation theory—perturbative curvatures in the two-parameter case,” *Progress of Theoretical Physics*, vol. 113, no. 3, pp. 481–511, 2005.
- [24] M. Bruni, S. Matarrese, S. Mollerach, and S. Sonogo, “Perturbations of spacetime: gauge transformations and gauge invariance at second order and beyond,” *Classical and Quantum Gravity*, vol. 14, no. 9, pp. 2585–2606, 1997.
- [25] K. Nakamura, “Gauge-invariant formulation of second-order cosmological perturbations,” *Physical Review D*, vol. 74, no. 10, Article ID 101301, 2006.
- [26] K. Nakamura, “Second-order gauge invariant cosmological perturbation theory—Einstein equations in terms of gauge invariant variables,” *Progress of Theoretical Physics*, vol. 117, no. 1, pp. 17–74, 2005.
- [27] K. Nakamura, “Perturbations of matter fields in the second-order gauge-invariant cosmological perturbation theory,” *Physical Review D*, vol. 80, no. 12, Article ID 124021, 2009.
- [28] K. Nakamura, “Consistency of equations in the second-order gauge-invariant cosmological perturbation theory,” *Progress of Theoretical Physics*, vol. 121, no. 6, pp. 1321–1360, 2009.
- [29] C. Pitrou, “Gauge-invariant Boltzmann equation and the fluid limit,” *Classical and Quantum Gravity*, vol. 24, no. 24, pp. 6127–6158, 2007.
- [30] C. Pitrou, “The radiative transfer at second order: a full treatment of the Boltzmann equation with polarization,” *Classical and Quantum Gravity*, vol. 26, no. 6, Article ID 065006, 2009.
- [31] K. Nakamura, ““Gauge” in general relativity: second-order general relativistic gauge-invariant perturbation theory,” in *Lie Theory and Its Applications in Physics VII*, V. K. Dobrev et al., Ed., Heron Press, Sofia, Bulgaria, 2008.
- [32] R. K. Sachs, “Gravitational radiation,” in *Relativity, Groups and Topology*, C. DeWitt and B. DeWitt, Eds., Gordon and Breach, New York, NY, USA, 1964.
- [33] J. M. Stewart and M. Walker, “Perturbations of space-times in general relativity,” *Proceedings of the Royal Society of London A*, vol. 341, pp. 49–74, 1974.
- [34] J. M. Stewart, “Perturbations of Friedmann-Robertson-Walker cosmological models,” *Classical and Quantum Gravity*, vol. 7, no. 7, pp. 1169–1180, 1990.
- [35] J. M. Stewart, *Advanced General Relativity*, Cambridge University Press, Cambridge, UK, 1991.
- [36] S. Sonogo and M. Bruni, “Gauge dependence in the theory of non-linear spacetime perturbations,” *Communications in Mathematical Physics*, vol. 193, no. 1, pp. 209–218, 1998.
- [37] S. Matarrese, S. Mollerach, and M. Bruni, “Relativistic second-order perturbations of the Einstein-de Sitter universe,” *Physical Review D*, vol. 58, no. 4, Article ID 043504, 22 pages, 1998.
- [38] M. Bruni, L. Gualtieri, and C. F. Sopuerta, “Two-parameter nonlinear spacetime perturbations: Gauge transformations and gauge invariance,” *Classical and Quantum Gravity*, vol. 20, no. 3, pp. 535–556, 2003.
- [39] C. F. Sopuerta, M. Bruni, and L. Gualtieri, “Nonlinear N-parameter spacetime perturbations: Gauge transformations,” *Physical Review D*, vol. 70, no. 6, Article ID 064002, 2004.
- [40] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry, Volume I*, John Wiley & Sons, New York, NY, USA, 1996.
- [41] M. Bruni and S. Sonogo, “Observables and gauge invariance in the theory of nonlinear spacetime perturbations,” *Classical and Quantum Gravity*, vol. 16, no. 7, pp. L29–L36, 1999.
- [42] R. M. Wald, *General Relativity*, University of Chicago Press, Chicago, Ill, USA, 1984.
- [43] S. Weinberg, *Gravitation and Cosmology*, John Wiley & Sons, New York, NY, USA, 1972.
- [44] F. C. Mena, D. J. Mulryne, and R. Tavakol, “Nonlinear vector perturbations in a contracting universe,” *Classical and Quantum Gravity*, vol. 24, no. 10, p. 2721, 2007.
- [45] T. H.-C. Lu, K. Ananda, and C. Clarkson, “Vector modes generated by primordial density fluctuations,” *Physical Review D*, vol. 77, no. 4, Article ID 043523, 2008.
- [46] T. H.-C. Lu, K. Ananda, C. Clarkson, and R. Maartens, “The cosmological background of vector modes,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 2, article 023, 2009.
- [47] A. J. Christopherson, K. A. Malik, and D. R. Matravers, “Vorticity generation at second order in cosmological perturbation theory,” *Physical Review D*, vol. 79, no. 12, Article ID 123523, 2009.
- [48] K. N. Ananda, C. Clarkson, and D. Wands, “Cosmological gravitational wave background from primordial density perturbations,” *Physical Review D*, vol. 75, no. 12, Article ID 123518, 2007.
- [49] B. Osano, C. Pitrou, P. Dunsby, J.-P. Uzan, and C. Clarkson, “Gravitational waves generated by second order effects during inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 4, article 003, 2007.
- [50] D. Baumann, P. Steinhardt, K. Takahashi, and K. Ichiki, “Gravitational wave spectrum induced by primordial scalar perturbations,” *Physical Review D*, vol. 76, no. 8, Article ID 084019, 2007.
- [51] N. Bartolo, S. Matarrese, A. Riotto, and A. Väihkönen, “Maximal amount of gravitational waves in the curvaton scenario,” *Physical Review D*, vol. 76, no. 6, Article ID 061302, 2007.
- [52] P. Martineau and R. Brandenberger, “A back-reaction induced lower bound on the tensor-to-scalar ratio,” *Modern Physics Letters A*, vol. 23, no. 10, pp. 727–735, 2008.
- [53] R. Saito and J. Yokoyama, “Gravitational-wave background as a probe of the primordial black-hole abundance,” *Physical Review Letters*, vol. 102, no. 16, Article ID 161101, 2009.
- [54] F. Arroja, H. Assadullahi, K. Koyama, and D. Wands, “Cosmological matching conditions for gravitational waves at second order,” *Physical Review D*, vol. 80, no. 12, Article ID 123526, 2009.
- [55] K. Nakamura, A. Ishibashi, and H. Ishihara, “Dynamics of a string coupled to gravitational waves: gravitational wave scattering by a Nambu-Goto straight string,” *Physical Review D*, vol. 62, no. 10, Article ID 101502, 5 pages, 2000.

- [56] K. Nakamura and H. Ishihara, “Dynamics of a string coupled to gravitational waves. II. Perturbations propagate along an infinite Nambu-Goto string,” *Physical Review D*, vol. 63, no. 12, Article ID 127501, 2001.
- [57] K. Nakamura, “Initial condition of a gravitating thick loop cosmic string and linear perturbations,” *Classical and Quantum Gravity*, vol. 19, no. 4, pp. 783–797, 2002.
- [58] K. Nakamura, “Does a Nambu-Goto wall emit gravitational waves? Cylindrical Nambu-Goto wall as an example of gravitating nonspherical walls,” *Physical Review D*, vol. 66, no. 8, Article ID 084005, 14 pages, 2002.
- [59] K. Nakamura, “Comparison of the oscillatory behavior of a gravitating Nambu-Goto string and a test string,” *Progress of Theoretical Physics*, vol. 110, no. 2, pp. 201–232, 2003.

## Review Article

# Non-Gaussianity and the Cosmic Microwave Background Anisotropies

**N. Bartolo,<sup>1,2</sup> S. Matarrese,<sup>1,2</sup> and A. Riotto<sup>2,3</sup>**

<sup>1</sup>*Dipartimento di Fisica “G. Galilei”, Università di Padova, via Marzolo 8, 131 Padova, Italy*

<sup>2</sup>*INFN, Sezione di Padova, via Marzolo 8, 35131 Padova, Italy*

<sup>3</sup>*CERN, Theory Division, CH-1211 Geneva 23, Switzerland*

Correspondence should be addressed to N. Bartolo, bartolo@pd.infn.it

Received 19 January 2010; Revised 14 June 2010; Accepted 29 June 2010

Academic Editor: Eiichiro Komatsu

Copyright © 2010 N. Bartolo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We review in a pedagogical way the present status of the impact of non-Gaussianity (NG) on the cosmic microwave background (CMB) anisotropies. We first show how to set the initial conditions at second order for the CMB anisotropies when some primordial NG is present. However, there are many sources of NG in CMB anisotropies, beyond the primordial one, which can contaminate the primordial signal. We mainly focus on the NG generated from the post inflationary evolution of the CMB anisotropies at second order in perturbation theory at large and small angular scales, such as the ones generated at the recombination epoch. We show how to derive the equations to study the second-order CMB anisotropies and provide analytical computations to evaluate their contamination to primordial NG (complemented with numerical examples). We also offer a brief summary of other secondary effects. This paper requires basic knowledge of the theory of cosmological perturbations at the linear level.

## 1. Introduction

Cosmic microwave background (CMB) anisotropies play a special role in cosmology, as they allow an accurate determination of cosmological parameters and may provide a unique probe of the physics of the early universe and in particular of the processes that gave origin to the primordial perturbations.

Cosmological inflation [1] is nowadays considered the dominant paradigm for the generation of the initial seeds for structure formation. In the inflationary picture, the primordial cosmological perturbations are created from quantum fluctuations “redshifted” out of the horizon during an early period of accelerated expansion of the universe, where they remain “frozen”. They are observable through CMB temperature anisotropies (and polarization) and the large-scale clustering properties of the matter distribution in the Universe.

This picture has recently received further spectacular confirmations from the results of the wilkinson microwave anisotropy probe (WMAP) five year set of data [2]. Since

the observed cosmological perturbations are of the order of  $10^{-5}$ , one might think that first-order perturbation theory will be adequate for all comparisons with observations. This might not be the case, though. Present [2, 3] and future experiments [4] may be sensitive to the nonlinearities of the cosmological perturbations at the level of second- or higher-order perturbation theory. The detection of these nonlinearities through the non-Gaussianity (NG) in the CMB [5] has become one of the primary experimental targets.

There is one fundamental reason why a positive detection of NG is so relevant: it might help in discriminating among the various mechanisms for the generation of the cosmological perturbations. Indeed, various models of inflation, firmly rooted in modern particle physics theory, predict a significant amount of primordial NG generated either during or immediately after inflation when the comoving curvature perturbation becomes constant on superhorizon scales [5]. While standard single-field models of slow-roll inflation [6, 7] and—in general—two (multi)field [8, 9] models of inflation predict a tiny level of NG, “curvaton”-type

models, in which a significant contribution to the curvature perturbation is generated after the end of slow-roll inflation by the perturbation in a field which has a negligible effect on inflation, may predict a high level of NG [10]. Alternatives to the curvaton model are models where a curvature perturbation mode is generated by an inhomogeneity in the decay rate [11–13], the mass [14], or the interaction rate [15] of the particles responsible for the reheating after inflation. Other opportunities for generating the curvature perturbations occur at the end of inflation [16–18], during preheating [19–21], and at a phase-transition producing cosmic strings [22]. Also, within single-field models of inflation, a high level of NG can be generated breaking the standard conditions of canonical kinetic terms and initially vacuum states: for example, this is the case of Dirac-born-infeld (DBI) models of inflation [23], and initially excited states, respectively [24]. For every scenario, there exists a well defined prediction for the *strength* of NG and its *shape* [25, 26] as a function of the parameters.

Statistics like the bispectrum and the trispectrum of the CMB can then be used to assess the level of primordial NG (and possibly its shape) on various cosmological scales and to discriminate it from the one induced by secondary anisotropies and systematic effects [5, 27–30]. A positive detection of a primordial NG in the CMB at some level might therefore confirm and/or rule out a whole class of mechanisms by which the cosmological perturbations have been generated.

Despite the importance of evaluating the impact of primordial NG in a crucial observable like the CMB anisotropy, the vast majority of the literature has been devoted to the computation of the bispectrum of either the comoving curvature perturbation or the gravitational potential on large scales within given inflationary models. These, however, are not the physical quantities which are observed. One should instead provide a full prediction for the second-order radiation transfer function. A preliminary step towards this goal has been taken in [31] (see also [32–37]) where the full second-order radiation transfer function for the CMB anisotropies on large angular scales in a flat universe filled with matter and cosmological constant was computed, including the second-order generalization of the Sachs-Wolfe effect, both the early and late integrated Sachs-Wolfe (ISW) effects and the contribution of the second order tensor modes. (A similar computation of the CMB anisotropies up to third-order from gravitational perturbations has been performed in [38], which is particularly relevant to provide a complete theoretical prediction for cubic nonlinearities characterizing the level of NG in the CMB through the connected four-point correlation function (trispectrum) [27, 28].)

There are many sources of NG in CMB anisotropies, beyond the primordial one. The most relevant sources are the so-called secondary anisotropies, which arise after the last scattering epoch. These anisotropies can be divided into two categories: scattering secondaries, when the CMB photons scatter with electrons along the line of sight, and gravitational secondaries when effects are mediated by gravity [39]. Among the scattering secondaries we may list

the thermal Sunyaev-Zeldovich effect, where hot electrons in clusters transfer energy to the CMB photons, the kinetic Sunyaev-Zeldovich effect produced by the bulk motion of the electrons in clusters, the Ostriker-Vishniac effect, produced by bulk motions modulated by linear density perturbations, and effects due to reionization processes. The scattering secondaries are most significant on small angular scales as density inhomogeneities, bulk and thermal motions grow and become sizeable on small length-scales when structure formation proceeds.

Gravitational secondaries arise from the change in energy of photons when the gravitational potential is time-dependent, the ISW effect, and gravitational lensing. At late times, when the Universe becomes dominated by the dark energy, the gravitational potential on linear scales starts to decay, causing the ISW effect mainly on large angular scales. Other secondaries that result from a time-dependent potential are the Rees-Sciama effect, produced during the matter-dominated epoch by the time evolution of the potential on nonlinear scales.

The fact that the potential never grows appreciably means that most second order effects created by gravitational secondaries are generically small compared to those created by scattering ones. However, when a photon propagates from the last scattering to us, its path may be deflected because of the gravitational lensing. This effect does not create anisotropies, but only modifies existing ones. Since photons with large wavenumbers  $k$  are lensed over many regions ( $\sim k/H$ , where  $H$  is the Hubble rate) along the line of sight, the corresponding second-order effect may be sizeable. The three-point function arising from the correlation of the gravitational lensing and ISW effects generated by the matter distribution along the line of sight [40, 41] and the Sunyaev-Zeldovich effect [42] are large and detectable by Planck [43, 44]. A crucial issue is the level of *contamination* to the extraction of the primordial NG the secondary effects can produce. In Section 8, we briefly summarize some recent results about the level of CMB NG generated by some of these secondary effects.

Another relevant source of NG comes from the physics operating at the recombination. A naive estimate would tell that these nonlinearities are tiny being suppressed by an extra power of the gravitational potential. However, the dynamics at recombination is quite involved because all the nonlinearities in the evolution of the baryon-photon fluid at recombination and the ones coming from general relativity should be accounted for. This complicated dynamics might lead to unexpected suppressions or enhancements of the NG at recombination. A step towards the evaluation of the three-point correlation function has been taken in [45] where some effects were taken into account in the so-called squeezed triangle limit, corresponding to the case when one wavenumber is much smaller than the other two and was outside the horizon at recombination. References [46, 47] (see also [48–50]) present the computation of the full system of Boltzmann equations, describing the evolution of the photon, baryon and cold dark matter (CDM) fluids, at second order and neglecting polarization. These equations allow to follow the time evolution of the CMB anisotropies

at second order on all angular scales from the early epochs, when the cosmological perturbations were generated, to the present time, through the recombination era. These calculations set the stage for the computation of the full second-order radiation transfer function at all scales and for a generic set of initial conditions specifying the level of primordial non-Gaussianity. Of course, for specific effects on small angular scales like Sunyaev-Zel'dovich, gravitational lensing, and so forth, fully nonlinear calculations would provide a more accurate estimate of the resulting CMB anisotropy, however, as long as the leading contribution to second-order statistics like the bispectrum is concerned, second-order perturbation theory suffices.

The goal of this paper is to summarize in a pedagogical form the present status of the evaluation of the impact of NG on the CMB anisotropies. This implies first of all determining how to set the initial conditions at second order for the CMB anisotropy when some source of primordial NG is present. The second step will be determining how primordial NG flows on small angular scales. In this paper we will focus on the study of the second-order effects appearing at the recombination era when the CMB anisotropy is left imprinted. We will show how to derive the equations to evaluate CMB anisotropies, by computing the Boltzmann equations describing the evolution of the baryon-photon fluid up to second order. This permits to follow the time evolution of CMB anisotropies (up to second order) on all scales, from the early epoch, when the cosmological perturbations were generated, to the present time, through the recombination era. We will also provide the reader with some simplified analytical computation to evaluate the contamination of the recombination secondary effects onto the detection of primordial NG. The formalism for a more refined numerical analysis is also displayed and results for some worked examples will be also reported. The paper is mainly based on a series of papers written by the authors along the past years on the subject (with various updates) and, as such, follows both a logic and a chronological order. It requires knowledge of the theory of cosmological perturbation at the linear level (which however we summarize in the Appendices). We have tried to write the different sections in a self-contained way. Nevertheless, we alert the reader that the level of complexity increases with the number of the sections.

The paper is organized as follows. In Section 2, we provide a simple, but illuminating example to show why we do expect some NG present in the CMB anisotropy regardless if there is or not some primordial NG. In Section 3, we provide the reader with the necessary tools to study the dynamics at second order in the gravity sector. In Section 4, we show how to set the initial conditions for the primordial NG, while in Section 5, we provide a gauge-invariant way to define the CMB temperature anisotropy at second order on large scales. In Section 6, we go to small scales and present the full procedure to compute the Boltzmann equations necessary to follow the evolution of the nonlinearities from the recombination epoch down to the present epoch. Section 7 presents some analytical solutions of the Boltzmann equations in the tight coupling limit, along

the same lines of what is done at the linear level. The issue of contamination is addressed in Section 8, while in Section 9 we offer the reader with an analytical estimate of such a contamination. A more refined numerical work is presented in Section 10. Finally, in Section 11, some conclusions are given. This paper has also some hopefully useful appendices: in Appendix A the reader can find the energy-momentum tensors at second-order, Appendix B gives the solutions of Einstein equations for the perturbed fluids up to second-order, while Appendix C offers the analytical solutions of the linearized Boltzmann equations in the tight coupling limit.

## 2. Why Do We Expect NG in the Cosmological Perturbations?

Before tackling the problem of interest—the computation of the cosmological perturbations at second order after the inflationary era—we first provide a simple, but insightful computation, derived from [36], which illustrates why we expect that the cosmological perturbations develop some NG even if the latter is not present at some primordial epoch. This example will help the reader to understand why the cosmological perturbations are inevitably affected by nonlinearities, beyond those arising at some primordial epoch. The reason is clear: gravity is nonlinear and it feeds this property into the cosmological perturbations during the postinflationary evolution of the universe. As gravity talks to all fluids, this transmission is inevitable. To be specific, we focus on the CMB anisotropies. We will adopt the Poisson gauge which eliminates one scalar degree of freedom from the  $g_{0i}$  component of the metric and one scalar and two vector degrees of freedom from  $g_{ij}$ . We will use a metric of the form (see Table 1 for the symbols used in this paper)

$$ds^2 = -e^{2\Phi} dt^2 + 2a(t)\omega_i dx^i dt + a^2(t)\left(e^{-2\Psi}\delta_{ij} + \chi_{ij}\right)dx^i dx^j, \quad (1)$$

where  $a(t)$  is the scale factor as a function of the cosmic time  $t$ , and  $\omega_i$  and  $\chi_{ij}$  are the vector and tensor perturbation modes respectively. Each metric perturbation can be expanded into a linear (first-order) and a second-order part, as for example, the gravitational potential  $\Phi = \Phi^{(1)} + \Phi^{(2)}/2$ . However, in the metric (1) the choice of the exponentials greatly helps in computing the relevant expressions, and thus we will always keep them where it is convenient. From (1), one recovers at linear order the well-known longitudinal gauge while at second order, one finds  $\Phi^{(2)} = \phi^{(2)} - 2(\phi^{(1)})^2$  and  $\Psi^{(2)} = \psi^{(2)} + 2(\psi^{(1)})^2$  where  $\phi^{(1)}$ ,  $\psi^{(1)}$  and  $\phi^{(2)}$ ,  $\psi^{(2)}$  (with  $\phi^{(1)} = \Phi^{(1)}$  and  $\psi^{(1)} = \Psi^{(1)}$ ) are the first and second-order gravitational potentials in the longitudinal (Poisson) gauge adopted in [5, 51] as far as scalar perturbations are concerned.

We now consider the long wavelength modes of the CMB anisotropies, that is, we focus on scales larger than the horizon at last-scattering. We can therefore neglect vector and tensor perturbation modes in the metric. For the vector perturbations, the reason is that they contain gradient terms being produced as nonlinear combination of scalar-modes

and thus they will be more important on small scales (linear vector modes are not generated in standard mechanisms for cosmological perturbations, as inflation). The tensor contribution can be neglected for two reasons. First, the tensor perturbations produced from inflation on large scales give a negligible contribution to the higher-order statistics of the Sachs-Wolfe effect being of the order of (powers of) the slow-roll parameters during inflation (this holds for linear tensor modes as well as for tensor modes generated by the nonlinear evolution of scalar perturbations during inflation). Second, while on large scales the tensor modes have been proven to remain constant in time [52], when they approach the horizon they have a wavelike contribution which oscillates in time with decreasing amplitude [51].

Since we are interested in the cosmological perturbations on large scales, that is in perturbations whose wavelength is larger than the Hubble radius at last scattering, a local observer would see them in the form of a classical—possibly time-dependent—(nearly zero-momentum) homogeneous and isotropic background. Therefore, it should be possible to perform a change of coordinates in such a way as to absorb the super-Hubble modes and work with a metric of an homogeneous and isotropic Universe (plus, of course, cosmological perturbations on scale smaller than the horizon). We split the gravitational potential  $\Phi$  as

$$\Phi = \Phi_\ell + \Phi_s, \quad (2)$$

where  $\Phi_\ell$  stands for the part of the gravitational potential receiving contributions only from the super-Hubble modes;  $\Phi_s$  receives contributions only from the subhorizon modes

$$\begin{aligned} \Phi_\ell &= \int \frac{d^3k}{(2\pi)^3} \theta(aH - k) \Phi_{\vec{k}} e^{i\vec{k}\cdot\vec{x}}, \\ \Phi_s &= \int \frac{d^3k}{(2\pi)^3} \theta(k - aH) \Phi_{\vec{k}} e^{i\vec{k}\cdot\vec{x}}, \end{aligned} \quad (3)$$

where  $H$  is the Hubble rate computed with respect to the cosmic time,  $H = \dot{a}/a$ , and  $\theta(x)$  is the step function. Analogous definitions hold for the other gravitational potential  $\Psi$ .

By construction,  $\Phi_\ell$  and  $\Psi_\ell$  are a collection of Fourier modes whose wavelengths are larger than the horizon length and we may safely neglect their spatial gradients. Therefore,  $\Phi_\ell$  and  $\Psi_\ell$  are only functions of time. This amounts to saying that we can absorb the large-scale perturbations in the metric (1) by the following redefinitions:

$$\begin{aligned} d\bar{t} &= e^{\Phi_\ell} dt, \\ \bar{a} &= ae^{-\Psi_\ell}. \end{aligned} \quad (4)$$

The new metric describes a homogeneous and isotropic Universe

$$ds^2 = -d\bar{t}^2 + \bar{a}^2 \delta_{ij} dx^i dx^j, \quad (5)$$

where for simplicity we have not included the subhorizon modes. On superhorizon scales one can regard the Universe as a collection of regions of size of the Hubble radius evolving like unperturbed patches with metric (5) [52].

Let us now go back to the quantity we are interested in, namely the anisotropies of the CMB as measured today by an observer  $\mathcal{O}$ . If she/he is interested in the CMB anisotropies at large scales, the effect of super-Hubble modes is encoded in the metric (5). During their travel from the last scattering surface—to be considered as the emitter point  $\mathcal{E}$ —to the observer, the CMB photons suffer a redshift determined by the ratio of the emitted frequency  $\bar{\omega}_\mathcal{E}$  to the observed one  $\bar{\omega}_\mathcal{O}$

$$\bar{T}_\mathcal{O} = \bar{T}_\mathcal{E} \frac{\bar{\omega}_\mathcal{O}}{\bar{\omega}_\mathcal{E}}, \quad (6)$$

where  $\bar{T}_\mathcal{O}$  and  $\bar{T}_\mathcal{E}$  are the temperatures at the observer point and at the last scattering surface, respectively.

What is then the temperature anisotropy measured by the observer? (6) shows that the measured large-scale anisotropies are made of two contributions: the intrinsic inhomogeneities in the temperature at the last scattering surface and the inhomogeneities in the scaling factor provided by the ratio of the frequencies of the photons at the departure and arrival points. Let us first consider the second contribution. As the frequency of the photon is the inverse of a time period, we get immediately the fully nonlinear relation

$$\frac{\bar{\omega}_\mathcal{E}}{\bar{\omega}_\mathcal{O}} = \frac{\omega_\mathcal{E}}{\omega_\mathcal{O}} e^{-\Phi_\ell + \Phi_\mathcal{O}}. \quad (7)$$

As for the temperature anisotropies coming from the intrinsic temperature fluctuation at the emission point, it may be worth to recall how to obtain this quantity in the longitudinal gauge at first order. By expanding the photon energy density  $\rho_\gamma \propto T_\gamma^4$ , the intrinsic temperature anisotropies at last scattering are given by  $\delta_1 T_\mathcal{E}/T_\mathcal{E} = (1/4)\delta_1 \rho_\gamma/\rho_\gamma$ . One relates the photon energy density fluctuation to the gravitational perturbation first by implementing the adiabaticity condition  $\delta_1 \rho_\gamma/\rho_\gamma = (4/3)\delta_1 \rho_m/\rho_m$ , where  $\delta_1 \rho_m/\rho_m$  is the relative fluctuation in the matter component, and then using the energy constraint of Einstein equations  $\phi^{(1)} = -(1/2)\delta_1 \rho_m/\rho_m$ . The result is  $\delta_1 T_\mathcal{E}/T_\mathcal{E} = -2\Phi_{1\mathcal{E}}/3$ . Summing this contribution to the anisotropies coming from the redshift factor (7) expanded at first order provides the standard (linear) Sachs-Wolfe effect  $\delta_1 T_\mathcal{O}/T_\mathcal{O} = \Phi_{1\mathcal{E}}/3$ . Following the same steps, we may easily obtain its full nonlinear generalization.

Let us first relate the photon energy density  $\bar{\rho}_\gamma$  to the energy density of the nonrelativistic matter  $\bar{\rho}_m$  by using the adiabaticity condition. Again, here a bar indicates that we are considering quantities in the locally homogeneous Universe described by the metric (5). Using the energy continuity equation on large scales  $\partial \bar{\rho}/\partial \bar{t} = -3\bar{H}(\bar{\rho} + \bar{P})$ , where  $\bar{H} = d \ln \bar{a}/d\bar{t}$  and  $\bar{P}$  is the pressure of the fluid, one can easily show that there exists a conserved quantity in time at any order in perturbation theory [53]

$$\mathcal{F} \equiv \ln \bar{a} + \frac{1}{3} \int^{\bar{\rho}} \frac{d\bar{\rho}'}{(\bar{\rho}' + \bar{P}')}. \quad (8)$$

The perturbation  $\delta \mathcal{F}$  is a gauge-invariant quantity representing the nonlinear extension of the curvature perturbation  $\zeta$  on uniform energy density hypersurfaces on superhorizon scales for adiabatic fluids [53]. Indeed, expanding

it at first and second order one gets the corresponding definition  $\zeta_1 = -\psi_1 - \delta_1\rho/\rho$  and the quantity  $\zeta_2$  introduced in [54]. We will come back to these definitions later. At first order, the adiabaticity condition corresponds to set  $\zeta_{1\gamma} = \zeta_{1m}$  for the curvature perturbations relative to each component. At the nonlinear level the adiabaticity condition generalizes to

$$\frac{1}{3} \int \frac{d\bar{\rho}_m}{\bar{\rho}_m} = \frac{1}{4} \int \frac{d\bar{\rho}_\gamma}{\bar{\rho}_\gamma}, \quad (9)$$

or

$$\ln \bar{\rho}_m = \ln \bar{\rho}_\gamma^{3/4}. \quad (10)$$

To make contact with the standard second-order result, we may expand in (10) the photon energy density perturbations as  $\delta\bar{\rho}_\gamma/\rho_\gamma = \delta_1\rho_\gamma/\rho_\gamma + (1/2)\delta_2\rho_\gamma/\rho_\gamma$ , and similarly for the matter component. We immediately recover the adiabaticity condition

$$\frac{\delta_2\rho_\gamma}{\rho_\gamma} = \frac{4}{3} \frac{\delta_2\rho_m}{\rho_m} + \frac{4}{9} \left( \frac{\delta_1\rho_m}{\rho_m} \right)^2, \quad (11)$$

given in [5].

Next, we need to relate the photon energy density to the gravitational potentials at the nonlinear level. The energy constraint inferred from the (0-0) component of Einstein equations in the matter-dominated era with the ‘‘barred’’ metric (5) is

$$\bar{H}^2 = \frac{8\pi G_N}{3} \bar{\rho}_m. \quad (12)$$

Using (4), the Hubble parameter  $\bar{H}$  reads

$$\bar{H} = \frac{1}{\bar{a}} \frac{d\bar{a}}{dt} = e^{-\Phi_\ell} (H - \dot{\Psi}_\ell), \quad (13)$$

where  $H = d \ln a / dt$  is the Hubble parameter in the ‘‘unbarred’’ metric. Equation (12) thus yields an expression for the energy density of the nonrelativistic matter which is fully nonlinear, being expressed in terms of the gravitational potential  $\Phi_\ell$

$$\bar{\rho}_m = \rho_m e^{-2\Phi_\ell}, \quad (14)$$

where we have dropped  $\dot{\Psi}_\ell$  which is negligible on large scales. By perturbing, (14) we are able to recover in a straightforward way the solutions of the (0-0) component of Einstein equations for a matter-dominated Universe in the large-scale limit obtained at second order in perturbation theory. Indeed, recalling that  $\Phi$  is perturbatively related to the quantity  $\phi = \phi^{(1)} + \phi^{(2)}/2$  used in [5] by  $\phi^{(1)} = \phi^{(1)}$  and  $\phi^{(2)} = \phi^{(2)} - 2(\phi^{(1)})^2$ , one immediately obtains

$$\begin{aligned} \frac{\delta_1\rho_m}{\rho_m} &= -2\phi^{(1)}, \\ \frac{1}{2} \frac{\delta_2\rho_m}{\rho_m} &= -\phi^{(2)} + 4(\phi^{(1)})^2. \end{aligned} \quad (15)$$

The expression for the intrinsic temperature of the photons at the last scattering surface  $\bar{T}_\mathcal{E} \propto \bar{\rho}_\gamma^{1/4}$  follows from (10) and (14)

$$\bar{T}_\mathcal{E} = T_\mathcal{E} e^{-2\Phi_\ell/3}. \quad (16)$$

Plugging (7) and (16) into (6), we are finally able to provide the expression for the CMB temperature which is fully nonlinear and takes into account both the gravitational redshift of the photons due to the metric perturbations at last scattering and the intrinsic temperature anisotropies

$$\bar{T}_\mathcal{O} = \left( \frac{\omega_\mathcal{O}}{\omega_\mathcal{E}} \right) T_\mathcal{E} e^{\Phi_\ell/3}. \quad (17)$$

From (17), we read the *nonperturbative* anisotropy corresponding to the Sachs-Wolfe effect

$$\frac{\delta_{np}\bar{T}_\mathcal{O}}{T_\mathcal{O}} = e^{\Phi_\ell/3} - 1. \quad (18)$$

Equation (18) is one of the main results of this paper and represents *at any order in perturbation theory* the extension of the linear Sachs-Wolfe effect. At first order, one gets

$$\frac{\delta_1 T_\mathcal{O}}{T_\mathcal{O}} = \frac{1}{3} \phi^{(1)}, \quad (19)$$

and at second order

$$\frac{1}{2} \frac{\delta_2 T_\mathcal{O}}{T_\mathcal{O}} = \frac{1}{6} \phi^{(2)} + \frac{1}{18} (\phi^{(1)})^2. \quad (20)$$

This result shows that the CMB anisotropies is nonlinear on large scales and that a source of NG is inevitably sourced by gravity.

### 3. Perturbing Gravity

In this Section, we provide the necessary tools to deal with perturbed gravity, giving the expressions for the Einstein tensor perturbed up to second-order around a flat Friedmann-Robertson-Walker background, and the relevant Einstein equations. In the following, we will adopt the Poisson gauge which eliminates one scalar degree of freedom from the  $g_{0i}$  component of the metric and one scalar and two vector degrees of freedom from  $g_{ij}$ . We rewrite the metric (1) as

$$ds^2 = a^2(\eta) \left[ -e^{2\Phi} d\eta^2 + 2\omega_i dx^i d\eta + \left( e^{-2\Psi} \delta_{ij} + \chi_{ij} \right) dx^i dx^j \right], \quad (21)$$

where  $a(\eta)$  is the scale factor as a function of the conformal time  $\eta$ . As we previously mentioned, for the vector and tensor perturbations, we will neglect linear vector modes since they are not produced in standard mechanisms for the generation of cosmological perturbations (as inflation), and we also neglect tensor modes at linear order, since they give a negligible contribution to second-order perturbations. Therefore, we take  $\omega_i$  and  $\chi_{ij}$  to be second-order vector and tensor perturbations of the metric.

Let us now give our definitions for the connection coefficients and their expressions for the metric (1). The number of spatial dimensions is  $n = 3$ . Greek indices  $(\alpha, \beta, \dots, \mu, \nu, \dots)$  run from 0 to 3, while Latin indices  $(a, b, \dots, i, j, k, \dots, m, n, \dots)$  run from 1 to 3. The total spacetime metric  $g_{\mu\nu}$  has signature  $(-, +, +, +)$ . The connection coefficients are defined as

$$\Gamma_{\beta\gamma}^{\alpha} = \frac{1}{2} g^{\alpha\rho} \left( \frac{\partial g_{\rho\gamma}}{\partial x^{\beta}} + \frac{\partial g_{\beta\rho}}{\partial x^{\gamma}} - \frac{\partial g_{\beta\gamma}}{\partial x^{\rho}} \right). \quad (22)$$

The Riemann tensor is defined as

$$R_{\beta\mu\nu}^{\alpha} = \Gamma_{\beta\nu,\mu}^{\alpha} - \Gamma_{\beta\mu,\nu}^{\alpha} + \Gamma_{\lambda\mu}^{\alpha} \Gamma_{\beta\nu}^{\lambda} - \Gamma_{\lambda\nu}^{\alpha} \Gamma_{\beta\mu}^{\lambda}. \quad (23)$$

The Ricci tensor is a contraction of the Riemann tensor

$$R_{\mu\nu} = R_{\mu\alpha\nu}^{\alpha}, \quad (24)$$

and in terms of the connection coefficient it is given by

$$R_{\mu\nu} = \partial_{\alpha} \Gamma_{\mu\nu}^{\alpha} - \partial_{\mu} \Gamma_{\nu\alpha}^{\alpha} + \Gamma_{\sigma\alpha}^{\alpha} \Gamma_{\mu\nu}^{\sigma} - \Gamma_{\sigma\nu}^{\alpha} \Gamma_{\mu\alpha}^{\sigma}. \quad (25)$$

The Ricci scalar is given by contracting the Ricci tensor

$$R = R_{\mu}^{\mu}. \quad (26)$$

The Einstein tensor is defined as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R. \quad (27)$$

**3.1. The Connection Coefficients.** For the connection coefficients we find

$$\begin{aligned} \Gamma_{00}^0 &= \mathcal{H} + \Phi', \\ \Gamma_{0i}^0 &= \frac{\partial\Phi}{\partial x^i} + \mathcal{H}\omega_i, \\ \Gamma_{00}^i &= \omega^i + \mathcal{H}\omega^i + e^{2\Psi+2\Phi} \frac{\partial\Phi}{\partial x_i}, \\ \Gamma_{ij}^0 &= -\frac{1}{2} \left( \frac{\partial\omega_j}{\partial x^i} + \frac{\partial\omega_i}{\partial x^j} \right) + e^{-2\Psi-2\Phi} (\mathcal{H} - \Psi') \delta_{ij} \\ &\quad + \frac{1}{2} \chi'_{ij} + \mathcal{H}\chi_{ij}, \\ \Gamma_{0j}^i &= (\mathcal{H} - \Psi') \delta_{ij} + \frac{1}{2} \chi'_{ij} + \frac{1}{2} \left( \frac{\partial\omega_i}{\partial x^j} - \frac{\partial\omega_j}{\partial x^i} \right), \\ \Gamma_{jk}^i &= -\mathcal{H}\omega^i \delta_{jk} - \frac{\partial\Psi}{\partial x^k} \delta_j^i - \frac{\partial\Psi}{\partial x^j} \delta_k^i + \frac{\partial\Psi}{\partial x_i} \delta_{jk} \\ &\quad + \frac{1}{2} \left( \frac{\partial\chi_j^i}{\partial x^k} + \frac{\partial\chi_k^i}{\partial x^j} + \frac{\partial\chi_{jk}}{\partial x_i} \right). \end{aligned} \quad (28)$$

**3.2. The Einstein Equations.** The Einstein equations are written as  $G_{\mu\nu} = \kappa^2 T_{\mu\nu}$ , so that  $\kappa^2 = 8\pi G_N$ , where  $G_N$  is the usual Newtonian gravitational constant. They read

$$G_0^0 = -\frac{e^{-2\Phi}}{a^2} \left[ 3\mathcal{H}^2 - 6\mathcal{H}\Psi' + 3(\Psi')^2 - e^{2\Phi+2\Psi} (\partial_i\Psi\partial^i\Psi - 2\nabla^2\Psi) \right] = \kappa^2 T_0^0, \quad (29)$$

$$G_0^i = 2\frac{e^{2\Psi}}{a^2} \left[ \partial^i\Psi' + (\mathcal{H} - \Psi')\partial^i\Phi \right] - \frac{1}{2a^2} \nabla^2\omega^i + \left( 4\mathcal{H}^2 - 2\frac{a''}{a} \right) \frac{\omega^i}{a^2} = \kappa^2 T_0^i, \quad (30)$$

$$G_j^i = \frac{1}{a^2} \left[ e^{-2\Phi} \left( \mathcal{H}^2 - 2\frac{a''}{a} - 2\Psi'\Phi' - 3(\Psi')^2 + 2\mathcal{H}(\Phi' + 2\Psi') + 2\Psi'' \right) + e^{2\Psi} (\partial_k\Phi\partial^k\Phi + \nabla^2\Phi - \nabla^2\Psi) \right] \delta_j^i + \frac{e^{2\Psi}}{a^2} \left( -\partial^i\Phi\partial_j\Phi - \partial^i\partial_j\Phi + \partial^i\partial_j\Psi - \partial^i\Phi\partial_j\Psi + \partial^i\Psi\partial_j\Psi - \partial^i\Psi\partial_j\Phi \right) - \frac{\mathcal{H}}{a^2} (\partial^i\omega_j + \partial_j\omega^i) - \frac{1}{2a^2} (\partial^i\omega_{j'} + \partial_j\omega^{i'}) + \frac{1}{a^2} \left( \mathcal{H}\chi_j^i + \frac{1}{2}\chi_j^{i'} - \frac{1}{2}\nabla^2\chi_j^i \right) = \kappa^2 T_j^i. \quad (31)$$

Taking the traceless part of (31), we find

$$\Psi - \Phi = \mathcal{Q}, \quad (32)$$

where  $\mathcal{Q}$  is defined by  $\nabla^2\mathcal{Q} = -P + 3N$ , with  $P \equiv P^i_i$ ,

$$P_j^i = \partial^i\Phi\partial_j\Psi + \frac{1}{2} (\partial^i\Phi\partial_j\Phi - \partial^i\Psi\partial_j\Psi) + 4\pi G_N a^2 e^{-2\Psi} T_j^i, \quad (33)$$

and  $\nabla^2 N = \partial_i\partial^i P_j^j$ . The trace of (31) gives

$$\begin{aligned} e^{-2\Phi} \left( \mathcal{H}^2 - 2\frac{a''}{a} - 2\Phi'\Psi' - 3(\Psi')^2 + 2\mathcal{H}(3\Psi' - \mathcal{Q}') + 2\Psi'' \right) + \frac{e^{2\Psi}}{3} \left( 2\partial_k\Phi\partial^k\Phi + \partial_k\Psi\partial^k\Psi - 2\partial_k\Phi\partial^k\Psi + 2(P - 3N) \right) &= \frac{8\pi G_N}{3} a^2 T_k^k. \end{aligned} \quad (34)$$

From (30), we may deduce an equation for  $\omega^i$

$$\begin{aligned} & -\frac{1}{2}\nabla^2\omega^i + \left(4\mathcal{H}^2 - 2\frac{a''}{a}\right)\omega^i \\ & = -\left(\delta_j^i - \frac{\partial^i\partial_j}{\nabla^2}\right) \\ & \quad \times \left(2(\partial^j\Psi' + (\mathcal{H} - \Psi')\partial^j\Phi) - 8\pi G_N a^2 e^{-2\Psi} T_0^j\right). \end{aligned} \quad (35)$$

Here,  $T_\nu^\mu$  is the energy momentum tensor accounting for different components, photons, baryons, and dark matter. We will give the expressions later for each component in terms of the distribution functions.

#### 4. Setting the Initial Conditions from the Primordial NG

In this paper we are concerned with the second-order evolution of the cosmological perturbations. This requires that we well define the initial conditions of the cosmological perturbations at second order. These boundary conditions may or may not contain already some level of NG. If they do, we say that there exist some primordial NG. The latter is usually defined in the epoch in which the comoving curvature perturbation remains constant on large superhorizon scales. In the standard single-field inflationary model, the first seeds of density fluctuations are generated on superhorizon scales from the fluctuations of a scalar field, the inflaton [1].

In order to follow the evolution on superhorizon scales of the density fluctuations coming from the various mechanisms, we use the curvature perturbation of uniform density hypersurfaces  $\zeta = \zeta^{(1)} + \zeta^{(2)}/2 + \dots$ , where  $\zeta^{(1)} = -\psi^{(1)} - \mathcal{H}\delta\rho^{(1)}/\rho'$  and the expression for  $\zeta^{(2)}$  is given by [54]

$$\zeta^{(2)} = -\psi^{(2)} - \mathcal{H}\frac{\delta^{(2)}\rho}{\rho'} + \Delta\zeta^{(2)}, \quad (36)$$

with

$$\begin{aligned} \Delta\zeta^{(2)} & = 2\mathcal{H}\frac{\delta^{(1)}\rho'}{\rho'}\frac{\delta^{(1)}\rho}{\rho'} + 2\frac{\delta^{(1)}\rho}{\rho'}(\psi^{(1)'} + 2\mathcal{H}\psi^{(1)}) \\ & \quad - \left(\frac{\delta^{(1)}\rho}{\rho'}\right)^2 \left(\mathcal{H}\frac{\rho''}{\rho} - \mathcal{H}' - 2\mathcal{H}^2\right) + 2\psi^{(1)2}. \end{aligned} \quad (37)$$

The crucial point is that the gauge-invariant curvature perturbation  $\zeta$  remains *constant* on superhorizon scales after it has been generated during a primordial epoch and possible isocurvature perturbations are no longer present. Therefore, we may set the initial conditions at the time when  $\zeta$  becomes constant. In particular,  $\zeta^{(2)}$  provides the necessary information about the ‘‘primordial’’ level of non-Gaussianity generated either during inflation, as in the standard scenario, or immediately after it, as in the curvaton scenario. Different scenarios are characterized by different values of  $\zeta^{(2)}$ . For example, in the standard single-field inflationary model  $\zeta^{(2)} = 2(\zeta^{(1)})^2 + \mathcal{O}(\epsilon, \eta)$  [6, 7], where  $\epsilon$  and  $\eta$  are the standard

slow-roll parameters [1]. In general, we may parametrize the primordial non-Gaussianity level in terms of the conserved curvature perturbation as in [34]

$$\zeta^{(2)} = 2a_{\text{NL}}(\zeta^{(1)})^2, \quad (38)$$

where the parameter  $a_{\text{NL}}$  depends on the physics of a given scenario. For example in the standard scenario  $a_{\text{NL}} \simeq 1$ , while in the curvaton case  $a_{\text{NL}} = (3/4r) - r/2$ , where  $r \simeq (\rho_\sigma/\rho)_D$  is the relative curvaton contribution to the total energy density at curvaton decay [5, 55]. Alternatives to the curvaton model are those models characterized by the curvature perturbation being generated by an inhomogeneity in the decay rate [11–13] or the mass [14] or of the particles responsible for the reheating after inflation. Other opportunities for generating the curvature perturbation occur at the end of inflation [16–18] and during preheating [19–21]. All these models generate a level of NG which is local as the NG part of the primordial curvature perturbation is a local function of the Gaussian part, being generated on superhorizon scales. In momentum space, the three point function, or bispectrum, arising from the local NG is dominated by the so-called ‘‘squeezed’’ configuration, where one of the momenta is much smaller than the other two. Other models, such as DBI inflation [23] and ghost inflation [56], predict a different kind of primordial NG, called ‘‘equilateral’’, because the three-point function for this kind of NG is peaked on equilateral configurations, in which the length of the three wavevectors forming a triangle in Fourier space are equal [25].

One of the best tools to detect or constrain the primordial large-scale non-Gaussianity is through the analysis of the CMB anisotropies, for example by studying the bispectrum [5]. In that case, the standard procedure is to introduce the primordial nonlinearity parameter  $f_{\text{NL}}$  characterizing the primordial non-Gaussianity via the curvature perturbation [5] (the sign convention is such that it follows the WMAP convention for  $f_{\text{NL}}$ , see Section 5)

$$\zeta = \zeta^{(1)} + \frac{3}{5}f_{\text{NL}}(\zeta^{(1)} \star \zeta^{(1)}), \quad (39)$$

where the coefficient 3/5 arises from the first-order relation connecting the comoving curvature perturbation and the gravitational potential,  $\zeta^{(1)} = -5\phi^{(1)}/3$ , and the  $\star$ -product reminds us that one has to perform a convolution product in momentum space and that  $f_{\text{NL}}$  is indeed momentum-dependent (we refer the reader to the end of Section 5 for a detailed discussion about the definition of the nonlinearity parameter). To give the feeling of the resulting size of  $f_{\text{NL}}$  when  $|a_{\text{NL}}| \gg 1$ ,  $f_{\text{NL}} \simeq 5a_{\text{NL}}/3$  (see [5, 34]). Present limits on NG are summarized by  $-4 < f_{\text{NL}}^{\text{loc}} < 80$  [57] and  $-125 < f_{\text{NL}}^{\text{equil}} < 435$  [58] at 95% CL, where  $f_{\text{NL}}^{\text{loc}}$  and  $f_{\text{NL}}^{\text{equil}}$  stand for the nonlinear parameter in the case in which the squeezed and the equilateral configurations dominate, respectively.

## 5. CMB Anisotropies at Second-Order on Large Scales

In this section, we provide the exact expression for large-scale CMB temperature fluctuations at second order in perturbation theory. What this section contains, therefore, should be considered as a more technical elaboration of what the reader can find in Section 2. The final expression, we will find has various virtues. First, from it one can unambiguously extract the exact definition of the nonlinearity parameter  $f_{\text{NL}}$  which is used by the experimental collaborations to pin down the level of NG in the temperature fluctuations. Second, it contains a ‘‘primordial’’ term encoding all the information about the NG generated in primordial epochs, namely during or immediately after inflation, and depends upon the various fluctuation generation mechanisms. As such, the expression neatly disentangles the primordial contribution to the NG from that arising after inflation. Finally, the expression applies to all scenarios for the generation of cosmological perturbations. Third, it is gauge-invariant. For this last point, let us underline however that the gauge-invariant expression that we will provide is just one of the many possible ways in which the CMB temperature fluctuations can be casted in a gauge-invariant form.

In order to obtain our gauge-independent formula for the temperature anisotropies, we again perturb the spatially flat Robertson-Walker background. We expand the metric perturbations (21) in a first and a second-order part as

$$\begin{aligned} g_{00} &= -a^2 \left( 1 + 2\phi^{(1)} + \phi^{(2)} + 2(\phi^{(1)})^2 \right), \\ g_{0i} &= a^2 \left( \omega_i^{(1)} + \frac{1}{2}\omega_i^{(2)} \right), \\ g_{ij} &= a^2 \left[ \left( 1 - 2\psi^{(1)} - \psi^{(2)} + 2(\psi^{(1)})^2 \right) \delta_{ij} + \left( \chi_{ij}^{(1)} + \frac{1}{2}\chi_{ij}^{(2)} \right) \right]. \end{aligned} \quad (40)$$

Again, the functions  $\phi^{(r)}$ ,  $\omega_i^{(r)}$ ,  $\psi^{(r)}$  and  $\chi_{ij}^{(r)}$ , where  $(r) = (1, 2)$ , stand for the  $r$ th-order perturbations of the metric. We can split  $\omega_i^{(r)} = \partial_i \omega^{(r)} + \omega_i^{\perp(r)}$ , where  $\omega^{(r)}$  is the scalar part and  $\omega_i^{\perp(r)}$  is a transverse vector, that is,  $\partial^i \omega_i^{\perp(r)} = 0$ . The metric perturbations will transform according to an infinitesimal change of coordinates. From now on, we limit ourselves to a second-order time shift [51]

$$\eta \longrightarrow \eta - \alpha_{(1)} + \frac{1}{2}(\alpha'_{(1)}\alpha_{(1)} - \alpha_{(2)}), \quad (41)$$

where a prime denotes differentiation w.r.t. conformal time. In general, a gauge corresponds to a choice of coordinates defining a slicing of spacetime into hypersurfaces (at fixed time  $\eta$ ) and a threading into lines (corresponding to fixed spatial coordinates  $x$ ), but in this section only the former is relevant so that gauge-invariant can be taken to mean independent of the slicing. For example, under the time shift, the first-order spatial curvature perturbation  $\psi^{(1)}$  transforms as  $\psi^{(1)} \rightarrow \psi^{(1)} - \mathcal{H}\alpha_{(1)}$  (here  $\mathcal{H} = a'/a$ ), while  $\phi^{(1)} \rightarrow \phi^{(1)} + \alpha'_{(1)} + \mathcal{H}\alpha^{(1)}$ ,  $\omega_i^{(1)} \rightarrow \omega_i^{(1)} - \partial_i \alpha^{(1)}$ , and the traceless part

of the spatial metric  $\chi_{ij}^{(1)}$  turns out to be gauge-invariant. At second order in the perturbations, we just give some useful examples like the transformation of the energy density and the curvature perturbation [51]. (For these two examples, we here report only the expression of the second-order gauge transformations neglecting gradient terms on large-scales. Notice that the expressions of this section relative to linear perturbations are valid on all scales, while for the second-order fluctuations at some point the large-scale limit is considered.)

$$\begin{aligned} \delta^{(2)}\rho &\longrightarrow \delta^{(2)}\rho + \rho' \alpha_{(2)} + \alpha_{(1)} \left( \rho'' \alpha_{(1)} + \rho' \alpha'_{(1)} + 2\delta^{(1)}\rho' \right), \\ \psi^{(2)} &\longrightarrow \psi^{(2)} + 2\alpha_{(1)}\psi^{(1)'} - \mathcal{H}\alpha_{(1)}^2 - \mathcal{H}\alpha_{(1)}\alpha'_{(1)} \\ &\quad - \frac{1}{3}(2\omega_{(1)}^i - \alpha_{(1)}^i)\alpha_{,i}^{(1)} - \mathcal{H}\alpha_{(2)}. \end{aligned} \quad (42)$$

We now construct in a gauge-invariant way temperature anisotropies at second order. Temperature anisotropies beyond the linear regime have been calculated in [32, 51], following the photons path from last scattering to the observer in terms of perturbed geodesics. The linear temperature anisotropies read [32, 51]

$$\frac{\Delta T^{(1)}}{T} = \phi_{\mathcal{E}}^{(1)} - v_{\mathcal{E}}^{(1)i} e_i + \tau_{\mathcal{E}}^{(1)} - \int_{\lambda_0}^{\lambda_{\mathcal{E}}} d\lambda A^{(1)'}, \quad (43)$$

where  $A^{(1)} \equiv \psi^{(1)} + \phi^{(1)} + \omega_i^{(1)} e^i - (1/2)\chi_{ij}^{(1)} e^i e^j$ , the subscript  $\mathcal{E}$  indicates that quantities are evaluated at last-scattering,  $e^i$  is a spatial unit vector specifying the direction of observation and the integral is evaluated along the line-of-sight parametrized by the affine parameter  $\lambda$ . Equation (43) includes the intrinsic fractional temperature fluctuation at emission  $\tau_{\mathcal{E}}$ , the Doppler effect due to emitter’s velocity  $v_{\mathcal{E}}^{(1)i}$  and the gravitational redshift of photons, including the integrated Sachs-Wolfe (ISW) effect. We omitted monopoles due to the observer  $\mathcal{O}$  (e.g., the gravitational potential  $\psi_{\mathcal{O}}^{(1)}$  evaluated at the event of observation), which, being independent of the angular coordinate, can be always recast into the definition of temperature anisotropies. Notice, however, that the physical meaning of each contribution in (43) is not gauge-invariant, as the different terms are gauge-dependent. However, it is easy to show that the whole expression (43) is gauge-invariant. Since the temperature  $T$  is a scalar, the intrinsic temperature fluctuation transforms as  $\tau_{\mathcal{E}}^{(1)} \rightarrow \tau_{\mathcal{E}}^{(1)} + (T'/T)\alpha_{(1)} = \tau_{\mathcal{E}}^{(1)} - \mathcal{H}\alpha_{(1)}$ , having used the fact that the temperature scales as  $T \propto a^{-1}$ . Notice, instead, that the velocity  $v_{\mathcal{E}}^{(1)i}$  does not change. Therefore, using the transformations of metric perturbations we find

$$\begin{aligned} \frac{\Delta T^{(1)}}{T} &\longrightarrow \frac{\Delta T^{(1)}}{T} + \alpha'_{(1)} - \int_{\eta_0}^{\eta_{\mathcal{E}}} d\eta \frac{d\alpha'_{(1)}}{d\eta} \\ &= \frac{\Delta T^{(1)}}{T} + \mathcal{O}, \end{aligned} \quad (44)$$

where we have used the fact that the integral is evaluated along the line-of-sight which can be parametrized by the

background geodesics  $x^{(0)\mu} = (\lambda, (\lambda_{\mathcal{O}} - \lambda_{\mathcal{E}})e^i)$  (with  $d\lambda/d\eta = 1$ ), and the decomposition for the total derivative along the path for a generic function  $f(\lambda, x^i(\lambda))$ ,  $f' = \partial f/\partial\lambda = df/d\lambda + \partial_i f e^i$ . Equation (44) shows that the expression (43) for first-order temperature anisotropies is indeed gauge-invariant (up to monopole terms related to the observer  $\mathcal{O}$ ). Temperature anisotropies can be easily written in terms of particular combinations of perturbations which are manifestly gauge-invariant. For the gravitational potentials, we consider the gauge-invariant definitions  $\psi_{\text{GI}}^{(1)} = \psi^{(1)} - \mathcal{H}\omega^{(1)}$  and  $\phi_{\text{GI}}^{(1)} = \phi^{(1)} + \mathcal{H}\omega^{(1)} + \omega^{(1)'}$ . For the  $(0 - i)$  component of the metric and the traceless part of the spatial metric, we define  $\omega_i^{(1)\text{GI}} = \omega_i^{(1)}$  and  $\chi_{ij}^{(1)\text{GI}} = \chi_{ij}^{(1)}$ . For the matter variables we use a gauge-invariant intrinsic temperature fluctuation  $\tau_{\text{GI}}^{(1)} = \tau^{(1)} - \mathcal{H}\omega^{(1)}$ , while the velocity itself is gauge-invariant  $v_{\text{GI}}^{(1)i} = v^{(1)i}$  under time shifts. Following the same steps leading to (44), one gets the linear temperature anisotropies in (43) in terms of these gauge-invariant quantities

$$\frac{\Delta T_{\text{GI}}^{(1)}}{T} = \phi_{\text{GI}}^{(1)} - v_{\text{GI}}^{(1)i} e_i + \tau_{\text{GI}}^{(1)} - \int_{\lambda_{\mathcal{O}}}^{\lambda_{\mathcal{E}}} d\lambda A_{\text{GI}}^{(1)'}, \quad (45)$$

where  $A_{\text{GI}}^{(1)} = \phi_{\text{GI}}^{(1)} + \psi_{\text{GI}}^{(1)} + \omega_i^{(1)\text{GI}} e_i - (1/2)\chi_{ij}^{(1)\text{GI}} e^i e^j$  and we omitted the subscript  $\mathcal{E}$ . For the primordial fluctuations we are interested in the large-scale modes set by the curvature perturbation  $\zeta^{(1)}$ . Defining a gauge-invariant density perturbation  $\delta^{(1)}\rho_{\text{GI}} = \delta^{(1)}\rho + \rho'\omega^{(1)}$ , we write the curvature perturbation as  $\zeta_{\text{GI}}^{(1)} = -\psi_{\text{GI}}^{(1)} - \mathcal{H}(\delta^{(1)}\rho_{\text{GI}}/\rho')$ . Since for adiabatic perturbations in the radiation ( $y$ ) and matter ( $m$ ) eras  $(1/4)(\delta^{(1)}\rho_y/\rho_y) = (1/3)(\delta^{(1)}\rho_m/\rho_m)$ , one can write the intrinsic temperature fluctuation as  $\tau^{(1)} = (1/4)(\delta^{(1)}\rho_y/\rho_y) = -\mathcal{H}(\delta^{(1)}\rho/\rho')$  and a gauge-invariant definition is  $\tau_{\text{GI}}^{(1)} = -\mathcal{H}(\delta^{(1)}\rho_{\text{GI}}/\rho')$ . In the large-scale limit, from Einstein equations ( $i - i$ ) and continuity equations, in the matter era  $\phi_{\text{GI}}^{(1)} = \psi_{\text{GI}}^{(1)} = -(3/5)\zeta_{\text{GI}}^{(1)}$ . Thus we obtain the large-scale limit of temperature anisotropies (45)  $\Delta T_{\text{GI}}^{(1)}/T = 2\psi_{\text{GI}}^{(1)} + \zeta_{\text{GI}}^{(1)} = \psi_{\text{GI}}^{(1)}/3$ , that is, the usual Sachs-Wolfe effect.

At second order, the procedure is similar to the one described so long, though more lengthy and cumbersome. We only provide the reader with the main steps to get the final expression. The second-order temperature fluctuations in terms of metric perturbations read [32, 59]

$$\begin{aligned} \frac{\Delta T^{(2)}}{T} &= \frac{1}{2}\phi_{\mathcal{E}}^{(2)} - \frac{1}{2}(\phi^{(1)})^2 - \frac{1}{2}v_{\mathcal{E}}^{(2)i} e_i + \frac{1}{2}\tau_{\mathcal{E}}^{(2)} - I_2(\lambda_{\mathcal{E}}) \\ &+ (I_1(\lambda_{\mathcal{E}}) + v_{\mathcal{E}}^{(1)i} e_i) \left( -\phi_{\mathcal{E}}^{(1)} - \tau_{\mathcal{E}}^{(1)} + v_{\mathcal{E}}^{(1)i} e_i + I_1(\lambda_{\mathcal{E}}) \right) \\ &+ x_{\mathcal{E}}^{(1)0} A_{\mathcal{E}}^{(1)'} + \left( x_{\mathcal{E}}^{(1)j} + x_{\mathcal{E}}^{(1)0} e^j \right) \left( \phi_{,j}^{(1)} - v_{i,j}^{(1)} e^i + \tau_{,j}^{(1)} \right)_{\mathcal{E}} \\ &- \frac{1}{2}v_{\mathcal{E}}^{(1)} v_{\mathcal{E}}^{(1)i} + \phi_{\mathcal{E}}^{(1)} \tau_{\mathcal{E}}^{(1)} + \frac{\partial \tau^{(1)}}{\partial d^i} d^{(1)i} - v_{\mathcal{E}}^{(1)i} e_i \phi_{\mathcal{E}}^{(1)} \\ &+ v_{\mathcal{E}}^{(1)} \left( -\omega_{\mathcal{E}}^{(1)i} - I_1^i(\lambda_{\mathcal{E}}) \right). \end{aligned} \quad (46)$$

Here,  $I_2$  is the second-order ISW [59]

$$\begin{aligned} I_2(\lambda_{\mathcal{E}}) &= \int_{\lambda_{\mathcal{O}}}^{\lambda_{\mathcal{E}}} d\lambda \left[ \frac{1}{2}A^{(2)'} - \left( \omega_i^{(1)'} - \chi_{ij}^{(1)'} e^j \right) \left( k^{(1)i} + e^i k^{(1)0} \right) \right. \\ &\quad \left. + 2k^{(1)0} A^{(1)'} + 2\psi^{(1)'} A^{(1)} \right. \\ &\quad \left. + x^{(1)0} A^{(1)'} + x^{(1)i} A_{,i}^{(1)'} \right], \end{aligned} \quad (47)$$

where  $A^{(2)} \equiv \psi^{(2)} + \phi^{(2)} + \omega_i^{(2)} e^i - (1/2)\chi_{ij}^{(2)} e^i e^j$ , while  $k^{(1)0}(\lambda) = -2\phi^{(1)} - \omega^{(1)i} e_i + I_1(\lambda)$  and  $k^{(1)i}(\lambda) = -2\psi^{(1)} e^i - \omega^{(1)i} + \chi^{(1)ij} e_j - I_1^i(\lambda)$  are the photon wave vectors, with  $I_1(\lambda)$  given by the integral in (43) and  $I_1^i(\lambda)$  is obtained from the same integral replacing the time derivative with a spatial gradient. Finally, in (46)

$$\begin{aligned} x^{(1)0}(\lambda) &= \int_{\lambda_{\mathcal{O}}}^{\lambda} d\lambda' \left[ -2\phi^{(1)} - \omega_i^{(1)} e^i + (\lambda - \lambda') A^{(1)'} \right], \\ x^{(1)i}(\lambda) &= - \int_{\lambda_{\mathcal{O}}}^{\lambda} d\lambda' \left[ 2\psi^{(1)} e^i + \omega^{(1)i} - \chi^{(1)ij} e_j + (\lambda - \lambda') A^{(1),i} \right], \end{aligned} \quad (48)$$

are the geodesics at first order, and  $d^{(1)i} = e^i - (e^i - k^{(1)i})/|e^i - k^{(1)i}|$  is the direction of the photon emission evaluated on the hypersurface of constant time of emission  $\eta = \eta_{\mathcal{E}}$ . As usual we have omitted the monopole terms due to the observer. Using the transformation rules of [51], it is possible to check that the expression (46) is gauge-invariant. We can express the second-order anisotropies in terms of explicitly gauge-invariant quantities, whose definition proceeds as for the linear case, by choosing the shifts  $\alpha^{(r)}$  such that  $\omega^{(r)} = 0$ . For example, we consider the gauge-invariant gravitational potential [33]

$$\begin{aligned} \phi_{\text{GI}}^{(2)} &= \phi^{(2)} + \omega^{(1)} \left[ 2 \left( \psi^{(1)'} + 2 \frac{a'}{a} \psi^{(1)} \right) \right. \\ &\quad \left. + \omega^{(1)''} + 5 \frac{a'}{a} \omega^{(1)'} + (\mathcal{H}' + 2\mathcal{H}^2) \omega^{(1)} \right] \\ &+ 2\omega^{(1)'} \left( 2\psi^{(1)} + \omega^{(1)'} \right) + \frac{1}{a} (a\alpha^{(2)}), \end{aligned} \quad (49)$$

where  $\alpha^{(2)} = \omega^{(2)} + \omega^{(1)} \omega^{(1)'} + \nabla^{-2} \partial^i [-4\psi^{(1)} \partial_i \omega^{(1)} - 2\omega^{(1)'} \partial_i \omega^{(1)}]$ . Expressing the second-order temperature anisotropies (46) in terms of our gauge-invariant quantities and taking the large-scale limit, we find

$$\frac{\Delta T_{\text{GI}}^{(2)}}{T} = \left( \frac{1}{2} \right) \phi_{\text{GI}}^{(2)} - \left( \frac{1}{2} \right) (\phi_{\text{GI}}^{(1)})^2 + \left( \frac{1}{2} \right) \tau_{\text{GI}}^{(2)} + \phi_{\text{GI}}^{(1)} \tau_{\text{GI}}^{(1)}, \quad (50)$$

(having dropped the subscript  $\mathcal{E}$ ), and the gauge-invariant intrinsic temperature fluctuation at emission is

$$\tau_{\text{GI}}^{(2)} = \left( \frac{1}{4} \right) \left( \frac{\delta^{(2)} \rho_y^{\text{GI}}}{\rho_y} \right) - 3 \left( \tau_{\text{GI}}^{(1)} \right)^2. \quad (51)$$

We have dropped those terms which represent integrated contributions and other second-order small-scale effects that can be distinguished from the large-scale part through their peculiar scale dependence. At this point, we make use of Einstein's equations. We take the expression for  $\zeta^{(2)}$  in (36) and (37), and we use the  $(0-0)$  component and the traceless part of the  $(i-j)$  Einstein's equations (29) and (31) after having appropriately expanded the exponentials. Thus, on large scales we find that the temperature anisotropies are given by

$$\frac{\Delta T_{\text{GI}}^{(2)}}{T} = \frac{1}{18} (\phi_{\text{GI}}^{(1)})^2 - \frac{\mathcal{K}}{10} - \frac{1}{10} [\zeta_{\text{GI}}^{(2)} - 2(\zeta_{\text{GI}}^{(1)})^2], \quad (52)$$

where we have defined a kernel

$$\mathcal{K} = 10\nabla^{-4} \partial_i \partial_j (\partial^i \psi^{(1)} \partial_j \psi^{(1)}) - \nabla^{-2} \left( \frac{10}{3} \partial^i \psi^{(1)} \partial_i \psi^{(1)} \right). \quad (53)$$

Notice that the factor  $1/18$  nicely matches the corresponding term in (20). Equation (52) is the main result of this section. It clearly shows that there are two contributions to the final nonlinearity in the large-scale temperature anisotropies. The contribution,  $[\zeta_{\text{GI}}^{(2)} - 2(\zeta_{\text{GI}}^{(1)})^2]$ , comes from the ‘‘primordial’’ conditions set during or after inflation. They are encoded in the curvature perturbation  $\zeta$  which remains constant once it has been generated. The remaining part of (52) describes the postinflation processing of the primordial non-Gaussian signal due to the nonlinear gravitational dynamics, including also second-order corrections at last scattering to the Sachs-Wolfe effect [32, 51, 59]. Thus, (52) allows to neatly disentangle the primordial contribution to NG from that coming from that arising after inflation, and from it we can extract the nonlinearity parameter  $f_{\text{NL}}$ , see the expression (39) which is usually adopted to phenomenologically parametrize the NG level of cosmological perturbations and has become the standard quantity to be observationally measured. The definition of  $f_{\text{NL}}$  adopted in the analyses performed in [43] goes through the conventional Sachs-Wolfe formula  $\Delta T/T = -\Phi/3$  where  $\Phi$  is Bardeen's potential, which is conventionally expanded as (up to a constant offset, which only affects the temperature monopole)  $\Phi = \Phi_{\text{L}} + f_{\text{NL}}(\Phi_{\text{L}}) \star (\Phi_{\text{L}})$ . Therefore, using  $\zeta^{(1)} = -(5/3)\psi_{\text{GI}}^{(1)}$  during matter domination, from (52) we read the nonlinearity parameter in momentum space

$$f_{\text{NL}}(\mathbf{k}_1, \mathbf{k}_2) = - \left[ \frac{5}{3} (1 - a_{\text{NL}}) + \frac{1}{6} - \frac{3}{10} \mathcal{K} \right], \quad (54)$$

where

$$\mathcal{K} = \frac{10(\mathbf{k}_1 \cdot \mathbf{k}_3)(\mathbf{k}_2 \cdot \mathbf{k}_3)}{k^4} - \frac{10}{3} \mathbf{k}_1 \cdot \frac{\mathbf{k}_2}{k^2}, \quad (55)$$

with  $\mathbf{k}_3 + \mathbf{k}_1 + \mathbf{k}_2 = 0$  and  $k = |\mathbf{k}_3|$ . To obtain (54) we have made use of the expression (38) to set the initial conditions. In particular, within the standard scenario where cosmological perturbations are due to the inflaton the primordial contribution to NG is given by  $a_{\text{NL}} = 1 - (1/4)(n_\zeta - 1)$  [6, 7], where the spectral index is expressed in

terms of the usual slow-roll parameters as  $n_\zeta - 1 = -6\epsilon + 2\eta$  [1]. The nonlinearity parameter from inflation now reads

$$f_{\text{NL}}^{\text{inf}} = -\frac{5}{12} (n_\zeta - 1) - \frac{1}{6} + \frac{3}{10} \mathcal{K}. \quad (56)$$

Therefore, the main NG contribution comes from the postinflation evolution of the second-order perturbations which give rise to order-one coefficients, while the primordial contribution is proportional to  $|n_\zeta - 1| \ll 1$ . This is true even in the ‘‘squeezed’’ limit first discussed by Maldacena in [7], where one of the wavenumbers is much smaller than the other two, for example  $k_1 \ll k_{2,3}$  and  $\mathcal{K} \rightarrow 0$ .

## 6. CMB Anisotropies at Second-Order at All Scales

As we already mentioned in the Introduction, despite the importance of NG in CMB anisotropies, little effort has been made so far to provide accurate theoretical predictions of it. On the contrary, the vast majority of the literature has been devoted to the computation of the bispectrum of either the comoving curvature perturbation or the gravitational potential on large scales within given inflationary models. These, however, are not the physical quantities which are observed. One should instead provide a full prediction for the second-order radiation transfer function. A preliminary step towards this goal has been taken in [31, 32, 36, 59] (see also [37]) where the full second-order radiation transfer function for the CMB anisotropies on large angular scales in a flat universe filled with matter and cosmological constant was computed, including the second-order generalization of the Sachs-Wolfe effect, both the early and late integrated Sachs-Wolfe (ISW) effects and the contribution of the second-order tensor modes. We have partly reported about these works in the previous sections. (For some recent papers focusing on the generation and evolution of tensor perturbations at second-order see, e.g., [60–62].)

In this section we wish to offer a summary of some of the second-order effects in the CMB anisotropies on small scales. There are many sources of NG in CMB anisotropies, beyond the primordial one. The most relevant sources are the so-called secondary anisotropies, which arise after the last scattering epoch. These anisotropies can be divided into two categories: scattering secondaries, when the CMB photons scatter with electrons along the line of sight, and gravitational secondaries when effects are mediated by gravity [39]. Among the scattering secondaries, we may list the thermal Sunyaev-Zeldovich effect, where hot electrons in clusters transfer energy to the CMB photons, the kinetic Sunyaev-Zeldovich effect produced by the bulk motion of the electrons in clusters, the Ostriker-Vishniac effect, produced by bulk motions modulated by linear density perturbations, and effects due to reionization processes. The scattering secondaries are most significant on small angular scales as density inhomogeneities, bulk and thermal motions grow and become sizeable on small length-scales when structure formation proceeds.

Gravitational secondaries arise from the change in energy of photons when the gravitational potential is time-dependent, the ISW effect, and gravitational lensing. At late times, when the Universe becomes dominated by the dark energy, the gravitational potential on linear scales starts to decay, causing the ISW effect mainly on large angular scales. Other secondaries that result from a time dependent potential are the Rees-Sciama effect, produced during the matter-dominated epoch by the time evolution of the potential on nonlinear scales.

The fact that the potential never grows appreciably means that most second-order effects created by gravitational secondaries are generically small compared to those created by scattering ones. However, when a photon propagates from the last scattering to us, its path may be deflected because of the gravitational lensing. This effect does not create anisotropies, but only modifies existing ones. Since photons with large wavenumbers  $k$  are lensed over many regions ( $\sim k/H$ , where  $H$  is the Hubble rate) along the line of sight, the corresponding second-order effect may be sizeable. The three-point function arising from the correlation of the gravitational lensing and ISW effects generated by the matter distribution along the line of sight [40, 41] and the Sunyaev-Zeldovich effect [42] are large and detectable by Planck [43].

Another relevant source of NG comes from the physics operating at the recombination. A naive estimate would tell that these nonlinearities are tiny being suppressed by an extra power of the gravitational potential. However, the dynamics at recombination is quite involved because all the nonlinearities in the evolution of the baryon-photon fluid at recombination and the ones coming from general relativity should be accounted for. This complicated dynamics might lead to unexpected suppressions or enhancements of the NG at recombination. A step towards the evaluation of the three-point correlation function has been taken in [45] where some effects were taken into account in the in so-called squeezed triangle limit, corresponding to the case when one wavenumber is much smaller than the other two and was outside the horizon at recombination (see however also [37] for a critical reassessment of some of the results contained in [45]. In particular notice that, contrary to what stated in [45], in the result (54), first obtained in [34], the contribution  $(-1/6)$  is actually present and survives in the squeezed limit).

This section, which is based on [46, 47], presents the computation of the full system of Boltzmann equations, describing the evolution of the photon, baryon, and cold dark matter (CDM) fluids, at second order and neglecting polarization. These equations allow to follow the time evolution of the CMB anisotropies at second order on all angular scales from the early epochs, when the cosmological perturbations were generated, to the present time, through the recombination era. These calculations set the stage for the computation of the full second-order radiation transfer function at all scales and for a generic set of initial conditions specifying the level of primordial non-Gaussianity. Of course, on small angular scales, fully nonlinear calculations of specific effects like Sunyaev-Zel'dovich, gravitational lensing, and so forth would provide a more accurate estimate of the resulting CMB anisotropy, however,

as long as the leading contribution to second-order statistics like the bispectrum is concerned, second-order perturbation theory suffices.

*6.1. The Collisionless Boltzmann Equation for Photons.* We are interested in the anisotropies in the cosmic distribution of photons and inhomogeneities in the matter. Photons are affected by gravity and by Compton scattering with free electrons. The latter are tightly coupled to protons. Both are, of course, affected by gravity. The metric which determines the gravitational forces is influenced by all these components plus CDM (and neutrinos). Our plan is to write down Boltzmann equations for the phase-space distributions of each species in the Universe.

The phase-space distribution of particles  $g(x^i, P^\mu, \eta)$  is a function of spatial coordinates  $x^i$ , conformal time  $\eta$ , and momentum of the particle  $P^\mu = dx^\mu/d\lambda$  where  $\lambda$  parametrizes the particle path. Through the constraint  $P^2 \equiv g_{\mu\nu}P^\mu P^\nu = -m^2$ , where  $m$  is the mass of the particle one can eliminate  $P^0$  and  $g(x^i, P^j, \eta)$  gives the number of particles in the differential phase-space volume  $dx^1 dx^2 dx^3 dP^1 dP^2 dP^3$ . In the following, we will denote the distribution function for photons with  $f$ .

The photons' distribution evolves according to the Boltzmann equation

$$\frac{df}{d\eta} = \bar{C}[f], \quad (57)$$

where the collision term is due to the scattering of photons off free electrons. In the following, we will derive the left-hand side of (57) while in the next section, we will compute the collision term.

For photons we can impose  $P^2 \equiv g_{\mu\nu}P^\mu P^\nu = 0$  and using the metric (1) in the conformal time  $\eta$  we find

$$P^2 = a^2 \left[ -e^{2\Phi} (P^0)^2 + \frac{p^2}{a^2} + 2\omega_i P^0 P^i \right] = 0, \quad (58)$$

where we define

$$p^2 = g_{ij}P^i P^j. \quad (59)$$

From the constraint (58)

$$P^0 = e^{-\Phi} \left( \frac{p^2}{a^2} + 2\omega_i P^0 P^i \right)^{1/2}. \quad (60)$$

Notice that we immediately recover the usual zero and first-order relations  $P^0 = p/a$  and  $P^0 = p(1 - \Phi^{(1)})/a$ .

The components  $P^i$  are proportional to  $pn^i$ , where  $n^i$  is a unit vector with  $n^i n_i = \delta_{ij} n^i n^j = 1$ . We can write  $P^i = Cn^i$ , where  $C$  is determined by

$$g_{ij}P^i P^j = C^2 a^2 (e^{-2\Psi} + \chi_{ij} n^i n^j) = p^2, \quad (61)$$

so that

$$P^i = \frac{p}{a} n^i (e^{-2\Psi} + \chi_{km} n^k n^m)^{-1/2} = \frac{p}{a} n^i e^\Psi \left( 1 - \frac{1}{2} \chi_{km} n^k n^m \right), \quad (62)$$

where the last equality holds up to second order in the perturbations. Again, we recover the zero and first-order relations  $P^i = pn^i/a$  and  $P^i = pn^i(1 + \Psi^{(1)})/a$  respectively. Thus, up to second order we can write

$$P^0 = e^{-\Phi} \frac{P}{a} (1 + \omega_i n^i). \quad (63)$$

Equation (62) and (63) allow us to replace  $P^0$  and  $P^i$  in terms of the variables  $p$  and  $n^i$ . Therefore, as it is standard in the literature, from now on we will consider the phase-space distribution  $f$  as a function of the momentum  $\mathbf{p} = pn^i$  with magnitude  $p$  and angular direction  $n^i$ ,  $f \equiv f(x^i, p, n^i, \eta)$ . Notice that our  $P^i$  are expressed via quantities,  $p$  and  $n^i$ , which are different from the ones of [63]. However, the final Boltzmann equations that we obtain agree with the ones of [63]. The reason is that the differences in the evolution for  $p$  emerge only at orders greater than two in the perturbative expansion in the fluctuations (at least if one neglects first-order vector and tensor perturbation modes, as it is done in these computations), while the evolution equation for  $n^i$  at linear-order (which is the one needed here, see below) is the same.

Thus, in terms of these variables, the total time derivative of the distribution function reads

$$\frac{df}{d\eta} = \frac{\partial f}{\partial \eta} + \frac{\partial f}{\partial x^i} \frac{dx^i}{d\eta} + \frac{\partial f}{\partial p} \frac{dp}{d\eta} + \frac{\partial f}{\partial n^i} \frac{dn^i}{d\eta}. \quad (64)$$

In the following, we will compute  $dx^i/d\eta$ ,  $dp/d\eta$  and  $dn^i/d\eta$ .

(a)  $dx^i/d\eta$ : From

$$p^i = \frac{dx^i}{d\lambda} = \frac{dx^i}{d\eta} \frac{d\eta}{d\lambda} = \frac{dx^i}{d\eta} P^0 \quad (65)$$

and from (62) and (63)

$$\frac{dx^i}{d\eta} = n^i e^{\Phi+\Psi} \left( 1 - \omega_j n^j - \frac{1}{2} \chi_{km} n^k n^m \right). \quad (66)$$

(b)  $dp/d\eta$ : For  $dp/d\eta$ , we make use of the time component of the geodesic equation  $dP^0/d\lambda = -\Gamma_{\alpha\beta}^0 P^\alpha P^\beta$ , where  $d/d\lambda = (d\eta/d\lambda) d/d\eta = P^0 d/d\eta$ , and

$$\frac{dP^0}{d\eta} = -\Gamma_{\alpha\beta}^0 \frac{P^\alpha P^\beta}{P^0}, \quad (67)$$

Using the metric (1), we find

$$\begin{aligned} & 2\Gamma_{\alpha\beta}^0 P^\alpha P^\beta \\ &= g^{0\nu} \left[ 2 \frac{\partial g_{\nu\alpha}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu} \right] P^\alpha P^\beta \\ &= 2(\mathcal{H} + \Phi') (P^0)^2 + 4\Phi_{,i} P^0 P^i + 4\mathcal{H} \omega_i P^0 P^i \\ &+ 2e^{-2\Phi} \left[ (\mathcal{H} - \Psi') e^{-2\Psi} \delta_{ij} - \omega_{i,j} + \frac{1}{2} \chi'_{ij} + \mathcal{H} \chi_{ij} \right] P^i P^j. \end{aligned} \quad (68)$$

On the other hand (63) of  $P^0$  in terms of  $p$  and  $n^i$  gives

$$\begin{aligned} \frac{dP^0}{d\eta} &= -\frac{p}{a} \frac{d\Phi}{d\eta} e^{-\Phi} (1 + \omega_i n^i) + e^{-\Phi} (1 + \omega_i n^i) \frac{d(p/a)}{d\eta} \\ &+ \frac{p}{a} e^{-\Phi} \frac{d(\omega_i n^i)}{d\eta}. \end{aligned} \quad (69)$$

Thus, (67) allows us express  $dp/d\eta$  as

$$\frac{1}{p} \frac{dp}{d\eta} = -\mathcal{H} + \Psi' - \Phi_{,i} n^i e^{\Phi+\Psi} - \omega'_i n^i - \frac{1}{2} \chi'_{ij} n^i n^j, \quad (70)$$

where in (68) we have replaced  $P^0$  and  $P^i$  by (63) and (62). Notice that in order to obtain (70), we have used the following expressions for the total time derivatives of the metric perturbations

$$\begin{aligned} \frac{d\Phi}{d\eta} &= \frac{\partial \Phi}{\partial \eta} + \frac{\partial \Phi}{\partial x^i} \frac{dx^i}{d\eta} \\ &= \frac{\partial \Phi}{\partial \eta} + \frac{\partial \Phi}{\partial x^i} n^i e^{\Phi+\Psi} \left( 1 - \omega_j n^j - \frac{1}{2} \chi_{km} n^k n^m \right), \end{aligned} \quad (71)$$

$$\begin{aligned} \frac{d(\omega_i n^i)}{d\eta} &= n^i \left( \frac{\partial \omega_i}{\partial \eta} + \frac{\partial \omega_i}{\partial x^j} \frac{dx^j}{d\eta} \right) \\ &= \frac{\partial \omega_i}{\partial \eta} n^i + \frac{\partial \omega_i}{\partial x^j} n^i n^j, \end{aligned} \quad (72)$$

where we have taken into account that  $\omega_i$  is already a second-order perturbation so that we can neglect  $dn^i/d\eta$  which is at least a first order quantity, and we can take the zero-order expression in (66),  $dx^i/d\eta = n^i$ . In fact, there is also an alternative expression for  $dp/d\eta$  which turns out to be useful later and which can be obtained by applying once more (71)

$$\frac{1}{p} \frac{dp}{d\eta} = -\mathcal{H} - \frac{d\Phi}{d\eta} + \Phi' + \Psi' - \omega'_i n^i - \frac{1}{2} \chi'_{ij} n^i n^j. \quad (73)$$

(c)  $dn^i/d\eta$ : We can proceed in a similar way to compute  $dn^i/d\eta$ . Notice that since in (64) it multiplies  $\partial f/\partial n^i$  which is first order, we need only the first order perturbation of  $dn^i/d\eta$ . We use the spatial components of the geodesic equations  $dP^i/d\lambda = -\Gamma_{\alpha\beta}^i P^\alpha P^\beta$  written as

$$\frac{dP^i}{d\eta} = -\Gamma_{\alpha\beta}^i \frac{P^\alpha P^\beta}{P^0}. \quad (74)$$

For the right-hand side, we find, up to second order,

$$\begin{aligned}
2\Gamma_{\alpha\beta}^i P^\alpha P^\beta &= g^{iv} \left[ \frac{\partial g_{\alpha\nu}}{\partial x^\beta} + \frac{\partial g_{\beta\nu}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\nu} \right] P^\alpha P^\beta \\
&= 4(\mathcal{H} - \Psi') P^i P^0 + 2(\chi_k^i + \omega_{,k}^i - \omega_k^i) P^0 P^k \\
&\quad + \left( 2 \frac{\partial \Phi}{\partial x^i} e^{2\Phi+2\Psi} + 2\omega^i + 2\mathcal{H}\omega^i \right) (P^0)^2 \\
&\quad - 4 \frac{\partial \Psi}{\partial x^k} P^i P^k + 2 \frac{\partial \Psi}{\partial x^i} \delta_{km} P^k P^m \\
&\quad - \left[ 2\mathcal{H}\omega^i \delta_{jk} - \left( \frac{\partial \chi_j^i}{\partial x^k} + \frac{\partial \chi_k^i}{\partial x^j} - \frac{\partial \chi_{jk}}{\partial x_i} \right) \right] P^j P^k,
\end{aligned} \tag{75}$$

while the expression (62) of  $P^i$  in terms of our variables  $p$  and  $n^i$  in the left-hand side of (74) brings

$$\begin{aligned}
\frac{dP^i}{d\eta} &= \frac{p}{a} e^\Psi \left[ \frac{d\Psi}{d\eta} n^i + \frac{a}{p} \frac{d(p/a)}{d\eta} n^i + \frac{dn^i}{d\eta} \right] \left( 1 - \frac{1}{2} \chi_{km} n^k n^m \right) \\
&\quad - \frac{p}{a} n^i e^\Psi \frac{1}{2} \frac{d(\chi_{km} n^k n^m)}{d\eta}.
\end{aligned} \tag{76}$$

Thus, using the expression (62) for  $P^i$  and (60) for  $P^0$  in (75), together with the previous result (70), the geodesic equation (74) gives the following expression  $dn^i/d\eta$  (valid up to first order)

$$\frac{dn^i}{d\eta} = (\Phi_{,k} + \Psi_{,k}) n^k n^i - \Phi^i - \Psi^i. \tag{77}$$

To proceed further we now expand the distribution function for photons around the zero-order value  $f^{(0)}$  which is that of a Bose-Einstein distribution

$$f^{(0)}(p, \eta) = 2 \frac{1}{\exp\{p/(T(\eta))\} - 1}, \tag{78}$$

where  $T(\eta)$  is the average (zero-order) temperature and the factor 2 comes from the spin degrees of photons. The perturbed distribution of photons will depend also on  $x^i$  and on the propagation direction  $n^i$  so as to account for inhomogeneities and anisotropies

$$\begin{aligned}
f(x^i, p, n^i, \eta) &= f^{(0)}(p, \eta) + f^{(1)}(x^i, p, n^i, \eta) \\
&\quad + \frac{1}{2} f^{(2)}(x^i, p, n^i, \eta),
\end{aligned} \tag{79}$$

where we split the perturbation of the distribution function into a first and a second-order part. The Boltzmann equation up to second order can be written in a straightforward way by recalling that the total time derivative of a given  $i$ th perturbation, as for example,  $df^{(i)}/d\eta$  is *at least* a quantity of the  $i$ -th order. Thus, it is easy to realize, looking at (64),

that the left-hand side of Boltzmann equation can be written up to second order as

$$\begin{aligned}
\frac{df}{d\eta} &= \frac{df^{(1)}}{d\eta} + \frac{1}{2} \frac{df^{(2)}}{d\eta} - p \frac{\partial f^{(0)}}{\partial p} \frac{d}{d\eta} \left( \Phi^{(1)} + \frac{1}{2} \Phi^{(2)} \right) \\
&\quad + p \frac{\partial f^{(0)}}{\partial p} \frac{\partial}{\partial \eta} \left( \Phi^{(1)} + \Psi^{(1)} + \frac{1}{2} \Phi^{(2)} + \frac{1}{2} \Psi^{(2)} \right) \\
&\quad - p \frac{\partial f^{(0)}}{\partial p} \frac{\partial \omega_i}{\partial \eta} n^i - \frac{1}{2} p \frac{\partial f^{(0)}}{\partial p} \frac{\partial \chi_{ij}}{\partial \eta} n^i n^j,
\end{aligned} \tag{80}$$

where for simplicity in (80), we have already used the background Boltzmann equation  $(df/d\eta)|^{(0)} = 0$ . In (80) there are all the terms which will give rise to the integrated Sachs-Wolfe effects (corresponding to the terms which explicitly depend on the gravitational perturbations), while other effects, such as the gravitational lensing, are still hidden in the (second-order part) of the first term. In fact in order to obtain (80) we just need for the time being to know the expression for  $dp/d\eta$ , (73).

## 6.2. Collision Term

**6.2.1. The Collision Integral.** In this section, we focus on the collision term due to Compton scattering (notice that in this section all the quantities and their indices are meant to be defined in the local Minkowski frame)

$$e(\mathbf{q})\gamma(\mathbf{p}) \longleftrightarrow e(\mathbf{q}')\gamma(\mathbf{p}'), \tag{81}$$

where we have indicated the momentum of the photons and electrons involved in the collisions. The collision term will be important for small scale anisotropies and spectral distortions. The important point to compute the collision term is that for the epoch of interest very little energy is transferred. Therefore, one can proceed by expanding the right hand side of (57) both in the small perturbation, (79), and in the small energy transfer.

The collision term is given (up to second order) by

$$\bar{C}(\mathbf{p}) = C(\mathbf{p}) a e^\Phi, \tag{82}$$

where  $a$  is the scale factor and

$$\begin{aligned}
C(\mathbf{p}) &= \frac{1}{E(\mathbf{p})} \int \frac{d\mathbf{q}}{(2\pi)^3 2E(\mathbf{q})} \frac{d\mathbf{q}'}{(2\pi)^3 2E(\mathbf{q}')} \frac{d\mathbf{p}'}{(2\pi)^3 2E(\mathbf{p}')} \\
&\quad \times (2\pi)^4 \delta^4(q + p - q' - p') |M|^2 \\
&\quad \times \{g(\mathbf{q}')f(\mathbf{p}') [1 + f(\mathbf{p})] - g(\mathbf{q})f(\mathbf{p}) [1 + f(\mathbf{p}')]\}
\end{aligned} \tag{83}$$

where  $E(\mathbf{q}) = (q^2 + m_e^2)^{1/2}$ ,  $M$  is the amplitude of the scattering process,  $\delta^4(q + p - q' - p') = \delta^3(\mathbf{q} + \mathbf{p} - \mathbf{q}' - \mathbf{p}') \delta(E(\mathbf{q}) + p - E(\mathbf{q}') - p')$  ensures the energy-momentum conservation and  $g$  is the distribution function for electrons. The Pauli suppression factors  $(1 - g)$  have been dropped since for the epoch of interest the density of electrons  $n_e$  is low. The reason why we write the collision term as in (82)

is that the starting point of the Boltzmann equation requires differentiation with respect to an affine parameter  $\lambda$ ,  $df/d\lambda = C'$ . In moving to the conformal time  $\eta$ , one rewrites the Boltzmann equation as  $df/d\eta = C'(P^0)^{-1}$ , with  $P^0 = d\eta/d\lambda$  given by (63). Taking into account that the collision term is at least of first order, (82) then follows. The electrons are kept in thermal equilibrium by Coulomb interactions with protons and they are nonrelativistic, thus we can take a Maxwell-Boltzmann distribution around some bulk velocity  $\mathbf{v}$

$$g(\mathbf{q}) = n_e \left( \frac{2\pi}{m_e T_e} \right)^{3/2} \exp \left\{ -\frac{(\mathbf{q} - m_e \mathbf{v})^2}{2m_e T_e} \right\}. \quad (84)$$

By using the three dimensional delta function the energy transfer is given by  $E(\mathbf{q}) - E(\mathbf{q} + \mathbf{p} - \mathbf{p}')$  and it turns out to be small compared to the typical thermal energies

$$E(\mathbf{q}) - E(\mathbf{q} + \mathbf{p} - \mathbf{p}') \simeq \frac{(\mathbf{p} - \mathbf{p}') \cdot \mathbf{q}}{m_e} = \mathcal{O} \left( \frac{Tq}{m_e} \right). \quad (85)$$

In (85), we have used  $E(\mathbf{q}) = m_e + q^2/2m_e$  and the fact that, since the scattering is almost elastic ( $p \simeq p'$ ),  $(\mathbf{p} - \mathbf{p}')$  is of order  $p \sim T$ , with  $q$  much bigger than  $(\mathbf{p} - \mathbf{p}')$ . In general, the electron momentum has two contributions, the bulk velocity ( $q = m_e v$ ) and the thermal motion ( $q \sim (m_e T)^{1/2}$ ) and thus the parameter expansion  $q/m_e$  includes the small bulk velocity  $\mathbf{v}$  and the ratio  $(T/m_e)^{1/2}$  which is small because the electrons are nonrelativistic.

The expansion of all the quantities entering the collision term in the energy transfer parameter and the integration over the momenta  $\mathbf{q}$  and  $\mathbf{q}'$  is described in details in [64]. It is easy to realize that we just need the scattering amplitude up to first order since at zero-order  $g(\mathbf{q}') = g(\mathbf{q} + \mathbf{p} - \mathbf{p}') = g(\mathbf{q})$  and  $\delta(E(\mathbf{q}) + p - E(\mathbf{q}') - p') = \delta(p - p')$  so that all the zero-order quantities multiplying  $|M|^2$  vanish. To first order

$$|M|^2 = 6\pi\sigma_T m_e^2 \left[ (1 + \cos^2\theta) - 2 \cos\theta(1 - \cos\theta) \mathbf{q} \cdot \frac{(\hat{p} + \hat{p}')}{m_e} \right], \quad (86)$$

where  $\cos\theta = \mathbf{n} \cdot \mathbf{n}'$  is the scattering angle and  $\sigma_T$  the Thompson cross-section. The resulting collision term up to second order is given by [64]

$$C(\mathbf{p}) = \frac{3n_e\sigma_T}{4p} \int dp' p' \frac{d\Omega'}{4\pi} \left[ c^{(1)}(\mathbf{p}, \mathbf{p}') + c_{\Delta}^{(2)}(\mathbf{p}, \mathbf{p}') \right. \\ \left. + c_v^{(2)}(\mathbf{p}, \mathbf{p}') + c_{\Delta v}^{(2)}(\mathbf{p}, \mathbf{p}') \right. \\ \left. + c_{vv}^{(2)}(\mathbf{p}, \mathbf{p}') + c_K^{(2)}(\mathbf{p}, \mathbf{p}') \right], \quad (87)$$

where we arrange the different contributions following [64]. The first order term reads

$$c^{(1)}(\mathbf{p}, \mathbf{p}') \\ = (1 + \cos^2\theta) \\ \times \left[ \delta(p - p') (f^{(1)}(\mathbf{p}') - f^{(1)}(\mathbf{p})) \right. \\ \left. + (f^{(0)}(p') - f^{(0)}(p)) (\mathbf{p} - \mathbf{p}') \cdot \mathbf{v} \frac{\partial \delta(p - p')}{\partial p'} \right], \quad (88)$$

while the second-order terms have been separated into four parts. There is the so-called anisotropy suppression term

$$c_{\Delta}^{(2)}(\mathbf{p}, \mathbf{p}') = \frac{1}{2} (1 + \cos^2\theta) \delta(p - p') (f^{(2)}(\mathbf{p}') - f^{(2)}(\mathbf{p})); \quad (89)$$

a term which depends on the second-order velocity perturbation defined by the expansion of the bulk flow as  $\mathbf{v} = \mathbf{v}^{(1)} + \mathbf{v}^{(2)}/2$

$$c_v^{(2)}(\mathbf{p}, \mathbf{p}') = \frac{1}{2} (1 + \cos^2\theta) (f^{(0)}(p') - f^{(0)}(p)) (\mathbf{p} - \mathbf{p}') \\ \cdot \mathbf{v}^{(2)} \frac{\partial \delta(p - p')}{\partial p'}; \quad (90)$$

a set of terms coupling the photon perturbation to the velocity

$$c_{\Delta v}^{(2)}(\mathbf{p}, \mathbf{p}') \\ = (f^{(1)}(\mathbf{p}') - f^{(1)}(\mathbf{p})) \\ \times \left[ (1 + \cos^2\theta) (\mathbf{p} - \mathbf{p}') \cdot \mathbf{v} \frac{\partial \delta(p - p')}{\partial p'} \right. \\ \left. - 2 \cos\theta(1 - \cos\theta) \delta(p - p') (\mathbf{n} + \mathbf{n}') \cdot \mathbf{v} \right], \quad (91)$$

and a set of source terms quadratic in the velocity

$$c_{vv}^{(2)}(\mathbf{p}, \mathbf{p}') = (f^{(0)}(p') - f^{(0)}(p)) (\mathbf{p} - \mathbf{p}') \cdot \mathbf{v} \\ \times \left[ (1 + \cos^2\theta) \frac{(\mathbf{p} - \mathbf{p}') \cdot \mathbf{v}}{2} \frac{\partial^2 \delta(p - p')}{\partial p'^2} \right. \\ \left. - 2 \cos\theta(1 - \cos\theta) (\mathbf{n} + \mathbf{n}') \cdot \mathbf{v} \frac{\partial \delta(p - p')}{\partial p'} \right]. \quad (92)$$

The last contribution are the Kompaneets terms describing spectral distortions to the CMB

$$\begin{aligned}
& c_K^{(2)}(\mathbf{p}, \mathbf{p}') \\
&= (1 + \cos^2 \theta) \frac{(\mathbf{p} - \mathbf{p}')^2}{2m_e} \\
&\quad \times \left[ (f^{(0)}(p') - f^{(0)}(p)) T_e \frac{\partial^2 \delta(p - p')}{\partial p'^2} \right. \\
&\quad \left. - (f^{(0)}(p') + f^{(0)}(p) + 2f^{(0)}(p')f^{(0)}(p)) \frac{\partial \delta(p - p')}{\partial p'} \right] \\
&\quad + \frac{2(p - p') \cos \theta (1 - \cos^2 \theta)}{m_e} \\
&\quad \times \left[ \delta(p - p') f^{(0)}(p') (1 + f^{(0)}(p)) (f^{(0)}(p') - f^{(0)}(p)) \right. \\
&\quad \left. \times \frac{\partial \delta(p - p')}{\partial p'} \right]. \tag{93}
\end{aligned}$$

Let us make a couple of comments about the various contributions to the collision term. First, notice the term  $c_v^{(2)}(\mathbf{p}, \mathbf{p}')$  due to second-order perturbations in the velocity of electrons which is absent in [64]. In standard cosmological scenarios (like inflation), vector perturbations are not generated at linear order, so that linear velocities are irrotational  $v^{(1)i} = \partial^i v^{(1)}$ . However, at second order vector perturbations are generated after horizon crossing as nonlinear combinations of primordial scalar modes. Thus, we must take into account also a transverse (divergence-free) component,  $v^{(2)i} = \partial^i v^{(2)} + v_T^{(2)i}$  with  $\partial_i v_T^{(2)i} = 0$ . As we will see, such vector perturbations will break azimuthal symmetry of the collision term with respect to a given mode  $\mathbf{k}$ , which instead usually holds at linear order. Secondly, notice that the number density of electrons appearing in (87) must be expanded as  $n_e = \bar{n}_e (1 + \delta_e)$  and then

$$\delta_e^{(1)} c^{(1)}(\mathbf{p}, \mathbf{p}') \tag{94}$$

gives rise to second-order contributions in addition to the list above, where we split  $\delta_e = \delta_e^{(1)} + \delta_e^{(2)}/2$  into a first- and second-order part. In particular, the combination with the term proportional to  $\mathbf{v}$  in  $c^{(1)}(\mathbf{p}, \mathbf{p}')$  gives rise to the so-called Vishniac effect, as discussed in [64].

**6.2.2. Computation of Different Contributions to the Collision Term.** In the integral (87) over the momentum  $\mathbf{p}'$  the first-order term gives the usual collision term

$$\begin{aligned}
& C^{(1)}(\mathbf{p}) \\
&= n_e \sigma_T \left[ f_0^{(1)}(p) + \frac{1}{2} f_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) - f^{(1)} - p \frac{\partial f^{(0)}}{\partial p} \mathbf{v} \cdot \mathbf{n} \right], \tag{95}
\end{aligned}$$

where one uses the decomposition in Legendre polynomials

$$f^{(1)}(\mathbf{x}, p, \mathbf{n}) = \sum_{\ell} (2\ell + 1) f_{\ell}^{(1)}(p) P_{\ell}(\cos \vartheta), \tag{96}$$

where  $\vartheta$  is the polar angle of  $\mathbf{n}$ ,  $\cos \vartheta = \mathbf{n} \cdot \hat{\mathbf{v}}$ .

In the following, we compute the second-order collision term separately for the different contributions, using the notation  $C(\mathbf{p}) = C^{(1)}(\mathbf{p}) + C^{(2)}(\mathbf{p})/2$ . We have not reported the details of the calculation of the first-order term because for its second-order analog,  $c_{\Delta}^{(2)}(\mathbf{p}, \mathbf{p}') + c_v^{(2)}(\mathbf{p}, \mathbf{p}')$ , the procedure is the same. The important difference is that the second-order velocity term includes a vector part, and this leads to a generic angular decomposition of the distribution function (for simplicity drop the time dependence)

$$f^{(i)}(\mathbf{x}, p, \mathbf{n}) = \sum_{\ell} \sum_{m=-\ell}^{\ell} f_{\ell m}^{(i)}(\mathbf{x}, p) (-i)^{\ell} \sqrt{\frac{4\pi}{2\ell + 1}} Y_{\ell m}(\mathbf{n}), \tag{97}$$

such that

$$f_{\ell m}^{(i)} = (-i)^{-\ell} \sqrt{\frac{2\ell + 1}{4\pi}} \int d\Omega f^{(i)} Y_{\ell m}^*(\mathbf{n}). \tag{98}$$

Such a decomposition holds also in Fourier space. The notation at this stage is a bit confusing, so let us restate it: superscripts denote the order of the perturbation; the subscripts refer to the moments of the distribution. Indeed, at first order, one can drop the dependence on  $m$  setting  $m = 0$  using the fact that the distribution function does not depend on the azimuthal angle  $\phi$ . In this case, the relation with  $f_i^{(1)}$  is

$$f_{\ell m}^{(1)} = (-i)^{-\ell} (2\ell + 1) \delta_{m0} f_{\ell}^{(1)}. \tag{99}$$

(a)  $c_{\Delta}^{(2)}(\mathbf{p}, \mathbf{p}')$ :

The integral over  $\mathbf{p}'$  yields

$$\begin{aligned}
C_{\Delta}^{(2)}(\mathbf{p}) &= \frac{3n_e \sigma_T}{4p} \int dp' p' \frac{d\Omega'}{4\pi} c_{\Delta}^{(2)}(\mathbf{p}, \mathbf{p}') \\
&= \frac{3n_e \sigma_T}{4p} \int dp' p' \delta(p - p'), \tag{100}
\end{aligned}$$

To perform the angular integral we write the angular dependence on the scattering angle  $\cos \theta = \mathbf{n} \cdot \mathbf{n}'$  in terms of the Legendre polynomials

$$\begin{aligned}
[1 + (\mathbf{n} \cdot \mathbf{n}')^2] &= \frac{4}{3} \left[ 1 + \frac{1}{2} P_2(\mathbf{n} \cdot \mathbf{n}') \right] \\
&= \left[ 1 + \frac{1}{2} \sum_{m=-2}^2 Y_{2m}(n) Y_{2m}^*(\mathbf{n}') \frac{4\pi}{2\ell + 1} \right], \tag{101}
\end{aligned}$$

where in the last step we used the addition theorem for spherical harmonics

$$P_{\ell} = \frac{4\pi}{2\ell + 1} \sum_{m=-\ell}^{\ell} Y_{\ell m}(\mathbf{n}) Y_{\ell m}^*(\mathbf{n}'). \tag{102}$$

Using the decomposition (98) and the orthonormality of the spherical harmonics, we find

$$C_{\Delta}^{(2)}(\mathbf{p}) = n_e \sigma_T \left[ f_{00}^{(2)}(\mathbf{p}) - f^{(2)}(\mathbf{p}) - \frac{1}{2} \sum_{m=-2}^2 \frac{\sqrt{4\pi}}{5^{3/2}} f_{2m}^{(2)}(p) Y_{2m}(\mathbf{n}) \right]. \quad (103)$$

It is easy to recover the result for the corresponding first-order contribution in (229) by using (205).

(b)  $c_v^{(2)}(\mathbf{p}, \mathbf{p}')$ :

Let us fix for simplicity our coordinates such that the polar angle of  $\mathbf{n}'$  is defined by  $\mu' = \hat{\mathbf{v}}^{(2)} \cdot \mathbf{n}'$  with  $\phi'$  the corresponding azimuthal angle. The contribution of  $c_v^{(2)}(\mathbf{p}, \mathbf{p}')$  to the collision term is then

$$C_v^{(2)}(\mathbf{p}) = \frac{3n_e \sigma_T}{4p} v^{(2)} \int dp' p' [f^{(0)}(p') - f^{(0)}(p)] \times \frac{\partial \delta(p - p')}{\partial p'} \times \int_{-1}^1 \frac{d\mu'}{2} (p\mu - p'\mu') \int_0^{2\pi} \frac{d\phi'}{2\pi} [1 + (\mathbf{p} \cdot \mathbf{p}')^2]. \quad (104)$$

We can use (101) which in our coordinate system reads

$$\frac{4}{3} \left[ 1 + \frac{1}{2} \sum_{m=-2}^m \frac{(2-m)!}{(2+m)!} P_2^m(\mathbf{n} \cdot \hat{\mathbf{v}}^{(2)}) P_2^m(\mathbf{n}' \cdot \hat{\mathbf{v}}^{(2)}) e^{im(\phi' - \phi)} \right], \quad (105)$$

so that

$$\int \frac{d\phi'}{2\pi} P_2(\mathbf{n} \cdot \mathbf{n}') = P_2(\mathbf{n} \cdot \hat{\mathbf{v}}^{(2)}) P_2(\mathbf{n}' \cdot \hat{\mathbf{v}}^{(2)}) = P_2(\mu) P_2(\mu'). \quad (106)$$

By using the orthonormality of the Legendre polynomials and integrating by parts over  $p'$ , we find

$$C_v^{(2)}(\mathbf{p}) = -n_e \sigma_T p \frac{\partial f^{(0)}}{\partial p} \mathbf{v}^{(2)} \cdot \mathbf{n}. \quad (107)$$

As it is clear by the presence of the scalar product  $\mathbf{v}^{(2)} \cdot \mathbf{p}$ , the final result is independent of the coordinates chosen.

(c)  $c_{\Delta v}^{(2)}(\mathbf{p}, \mathbf{p}')$ :

let us consider the contribution from the first term

$$c_{\Delta v(I)}^{(2)}(\mathbf{p}, \mathbf{p}') = (1 + \cos^2 \theta) (f^{(1)}(\mathbf{p}') - f^{(1)}(\mathbf{p})) (\mathbf{p} - \mathbf{p}') \cdot \frac{\mathbf{v} \partial \delta(p - p')}{\partial p'}, \quad (108)$$

where the velocity has to be considered at first order. In the integral (87), it brings

$$\frac{1}{2} C_{\Delta v(I)}^{(2)} = \frac{3n_e \sigma_T v}{4p} \int dp' p' \frac{\partial \delta(p - p')}{\partial p'} \times \int_{-1}^1 \frac{d\mu'}{2} [f^{(1)}(\mathbf{p}') - f^{(1)}(\mathbf{p})] \times (p\mu - p'\mu') \int_0^{2\pi} \frac{d\phi'}{2\pi} (1 + \cos^2 \theta). \quad (109)$$

The procedure to do the integral is the same as above. We use the same relations as in (105) and (106), where now the angles are those taken with respect to the first-order velocity. This eliminates the integral over  $\phi'$ , and integrating by parts over  $p'$  yields

$$\frac{1}{2} C_{\Delta v(I)}^{(2)}(\mathbf{p}) = -\frac{3n_e \sigma_T v}{4p} \int_{-1}^1 \frac{d\mu'}{2} \left[ \frac{4}{3} + \frac{2}{3} P_2(\mu) P_2(\mu') \right] \times \left[ p(\mu - 2\mu') (f^{(1)}(p, \mu') - f^{(1)}(p, \mu)) + p^2 (\mu - \mu') \frac{\partial f^{(1)}(p, \mu')}{\partial p} \right]. \quad (110)$$

We now use the decomposition (96) and the orthonormality of the Legendre polynomials to find

$$\begin{aligned} & \int \frac{d\mu'}{2} \mu' f^{(1)}(p, \mu') P_2(\mu') \\ &= \sum_{\ell} \int \frac{d\mu'}{2} \mu' P_2(\mu') P_{\ell}(\mu') f_{\ell}^{(1)}(p) \\ &= \sum_{\ell} \int \frac{d\mu'}{2} \left[ \frac{2}{5} P_1(\mu') + \frac{3}{5} P_3(\mu') \right] P_{\ell}(\mu') f_{\ell}^{(1)}(p) \\ &= \frac{2}{5} f_1^{(1)}(p) + \frac{3}{5} f_3^{(1)}(p), \end{aligned} \quad (111)$$

where we have used  $\mu'^{P_2}(\mu')P_1(\mu') = (2/5)P_1(\mu') + (3/5)P_3(\mu')$ , with  $P_1(\mu') = \mu'$ . Thus, from (110) we get

$$\begin{aligned} & \frac{1}{2}C_{\Delta v(I)}^{(2)}(\mathbf{p}) \\ &= n_e \sigma_T \left\{ \mathbf{v} \cdot \mathbf{n} \left[ f^{(1)}(\mathbf{p}) - f_0^{(1)}(p) - p \frac{\partial f_0^{(1)}(p)}{\partial p} \right. \right. \\ & \quad \left. \left. - \frac{1}{2}P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \left( f_2^{(1)}(p) + p \frac{\partial f_2^{(1)}(p)}{\partial p} \right) \right] \right. \\ & \quad \left. + v \left[ 2f_1^{(1)}(p) + p \frac{\partial f_1^{(1)}(p)}{\partial p} + \frac{1}{5}P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right. \right. \\ & \quad \left. \left. \times \left( 2f_1^{(1)}(p) + p \frac{\partial f_1^{(1)}(p)}{\partial p} \right. \right. \right. \\ & \quad \left. \left. \left. + 3f_3^{(1)}(p) + \frac{3}{2}p \frac{\partial f_3^{(1)}(p)}{\partial p} \right) \right] \right\}. \end{aligned} \quad (112)$$

In  $c^{(2)}(\mathbf{p}, \mathbf{p}')$ , there is a second term

$$\begin{aligned} c_{\Delta v(II)}^{(2)} &= -2 \cos \theta (1 - \cos \theta) (f^{(1)}(\mathbf{p}') - f^{(1)}(\mathbf{p})) \\ & \quad \times \delta(p - p') (\mathbf{n} + \mathbf{n}') \cdot \mathbf{v}, \end{aligned} \quad (113)$$

whose contribution to the collision term is

$$\begin{aligned} \frac{1}{2}C_{\Delta v(II)}^{(2)}(\mathbf{p}) &= -\frac{3n_e \sigma_T v}{2p} \int dp' p' \delta(p - p') \\ & \quad \times \int_{-1}^1 \frac{d\mu'}{2} (f^{(1)}(\mathbf{p}') - f^{(1)}(\mathbf{p})) (\mu + \mu') \\ & \quad \times \int_0^{2\pi} \frac{d\phi'}{2\pi} \cos \theta (1 - \cos \theta). \end{aligned} \quad (114)$$

This integration proceeds through the same steps as for  $C_{\Delta v(I)}^{(2)}(\mathbf{p})$ . In particular, by noting that  $\cos \theta (1 - \cos \theta) = -1/3 + P_1(\cos \theta) - 2P_3(\cos \theta)/3$ , Equations (105) and (106) allows to compute

$$\begin{aligned} & \int \frac{d\phi'}{2\pi} \cos \theta (1 - \cos \theta) \\ &= -\frac{1}{3} + P_1(\mu)P_1(\mu') - \frac{2}{3}P_2(\mu)P_2(\mu'), \end{aligned} \quad (115)$$

and using the decomposition (96), we arrive at

$$\begin{aligned} & \frac{1}{2}C_{\Delta v(II)}^{(2)}(\mathbf{p}) \\ &= -n_e \sigma_T \left\{ \mathbf{v} \cdot \mathbf{n} f_2^{(1)}(p) (1 - P_2(\hat{\mathbf{v}} \cdot \mathbf{n})) \right. \\ & \quad \left. + v \left[ \frac{1}{5} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) (3f_1^{(1)}(p) - 3f_3^{(1)}(p)) \right] \right\}. \end{aligned} \quad (116)$$

We then obtain

$$\begin{aligned} & \frac{1}{2}C_{\Delta v}^{(2)}(\mathbf{p}) \\ &= n_e \sigma_T \left\{ \mathbf{v} \cdot \mathbf{n} \left[ f^{(1)}(\mathbf{p}) - f_0^{(1)}(p) - p \frac{\partial f_0^{(1)}(p)}{\partial p} - f_2^{(1)}(p) \right. \right. \\ & \quad \left. \left. + \frac{1}{2}P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \left( f_2^{(1)}(p) - p \frac{\partial f_2^{(1)}(p)}{\partial p} \right) \right] \right. \\ & \quad \left. + v \left[ 2f_1^{(1)}(p) + p \frac{\partial f_1^{(1)}(p)}{\partial p} + \frac{1}{5}P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right. \right. \\ & \quad \left. \left. - f_1^{(1)}(p) + p \frac{\partial f_1^{(1)}(p)}{\partial p} \right. \right. \\ & \quad \left. \left. + 6f_3^{(1)}(p) + \frac{3}{2}p \frac{\partial f_3^{(1)}(p)}{\partial p} \right] \right\}. \end{aligned} \quad (117)$$

As far as the remaining terms, these have already been computed in [64] (see also [65]) and here we just report them

(d)  $c_{vv}^{(2)}(\mathbf{p}, \mathbf{p}')$ :

the term proportional to the velocity squared yield a contribution to the collision term

$$\begin{aligned} \frac{1}{2}C_{vv}^{(2)}(\mathbf{p}) &= n_e \sigma_T \left\{ (\mathbf{v} \cdot \mathbf{n})^2 \left[ p \frac{\partial f^{(0)}}{\partial p} + \frac{11}{20}p^2 \frac{\partial^2 f^{(0)}}{\partial p^2} \right] \right. \\ & \quad \left. + v^2 \left[ p \frac{\partial f^{(0)}}{\partial p} + \frac{3}{20}p^2 \frac{\partial^2 f^{(0)}}{\partial p^2} \right] \right\}. \end{aligned} \quad (118)$$

(e)  $c_K^{(2)}(\mathbf{p}, \mathbf{p}')$ :

The terms responsible for the spectral distortions give

$$\frac{1}{2}C_K^{(2)}(\mathbf{p}) = \frac{1}{m_e p^2} \frac{\partial}{\partial p} \left\{ p^4 \left[ T_e \frac{\partial f^{(0)}}{\partial p} + f^{(0)} (1 + f^{(0)}) \right] \right\}. \quad (119)$$

Finally, we write also the part of the collision term coming from (94)

$$\begin{aligned} & \delta_e^{(1)} c^{(1)}(\mathbf{p}, \mathbf{p}') \\ & \rightarrow \delta_e^{(1)} C^{(1)}(\mathbf{p}) = n_e \sigma_T \delta_e^{(1)} \\ & \quad \times \left[ f_0^{(1)}(p) + \frac{1}{2}f_2^{(1)}P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) - f^{(1)} - p \frac{\partial f^{(0)}}{\partial p} \mathbf{v} \cdot \mathbf{n} \right]. \end{aligned} \quad (120)$$

6.2.3. *Final Expression for the Collision Term.* Summing all the terms, we find the final expression for the collision term (87) up to second order

$$C(\mathbf{p}) = C^{(1)}(\mathbf{p}) + \frac{1}{2}C^{(2)}(\mathbf{p}), \quad (121)$$

with

$$C^{(1)}(\mathbf{p}) = n_e \sigma_T \left[ f_0^{(1)}(p) + \frac{1}{2} f_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) f^{(1)} - p \frac{\partial f^{(0)}}{\partial p} \mathbf{v} \cdot \mathbf{n} \right], \quad (122)$$

$$\begin{aligned} \frac{1}{2} C^{(2)}(\mathbf{p}) &= n_e \sigma_T \left\{ \frac{1}{2} f_{00}^{(2)}(p) - \frac{1}{4} \sum_{m=-2}^2 \frac{\sqrt{4\pi}}{5^{3/2}} f_{2m}^{(2)}(p) Y_{2m}(\mathbf{n}) \right. \\ &\quad - \frac{1}{2} f^{(2)}(\mathbf{p}) + \delta_e^{(1)} \left[ f_0^{(1)}(p) + \frac{1}{2} f_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right. \\ &\quad \quad \left. \left. - f^{(1)} - p \frac{\partial f^{(0)}}{\partial p} \mathbf{v} \cdot \mathbf{n} \right] \right. \\ &\quad - \frac{1}{2} p \frac{\partial f^{(0)}}{\partial p} \mathbf{v}^{(2)} \cdot \mathbf{n} + \mathbf{v} \cdot \mathbf{n} \\ &\quad \times \left[ f^{(1)}(\mathbf{p}) - f_0^{(1)}(p) - p \frac{\partial f_0^{(1)}(p)}{\partial p} - f_2^{(1)}(p) \right. \\ &\quad \quad \left. + \frac{1}{2} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \left( f_2^{(1)}(p) - p \frac{\partial f_2^{(1)}(p)}{\partial p} \right) \right] \\ &\quad + v \left[ 2 f_1^{(1)}(p) + p \frac{\partial f_1^{(1)}(p)}{\partial p} + \frac{1}{5} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right. \\ &\quad \quad \times \left( -f_1^{(1)}(p) + p \frac{\partial f_1^{(1)}(p)}{\partial p} \right. \\ &\quad \quad \left. \left. + 6 f_3^{(1)}(p) + \frac{3}{2} p \frac{\partial f_3^{(1)}(p)}{\partial p} \right) \right] \\ &\quad + (\mathbf{v} \cdot \mathbf{n})^2 \left[ p \frac{\partial f^{(0)}}{\partial p} + \frac{11}{20} p^2 \frac{\partial^2 f^{(0)}}{\partial p^2} \right] \\ &\quad + v^2 \left[ p \frac{\partial f^{(0)}}{\partial p} + \frac{3}{20} p^2 \frac{\partial^2 f^{(0)}}{\partial p^2} \right] \\ &\quad \left. + \frac{1}{m_e p^2} \frac{\partial}{\partial p} \left[ p^4 \left( T_e \frac{\partial f^{(0)}}{\partial p} + f^{(0)} (1 + f^{(0)}) \right) \right] \right\}. \quad (123) \end{aligned}$$

Notice that there is an internal hierarchy, with terms which do not depend on the baryon velocity  $\mathbf{v}$ , terms proportional to  $\mathbf{v} \cdot \mathbf{n}$  and then to  $(\mathbf{v} \cdot \mathbf{n})^2$ ,  $v$  and  $v^2$  (apart from the Kompaneets terms). In particular, notice the term proportional to  $\delta_e^{(1)} \mathbf{v} \cdot \mathbf{n}$  is the one corresponding to the Vishniac effect. We point out that we have kept all the terms up to second order in the collision term. In [64, 65], many terms coming from  $c_{\Delta v}^{(2)}$  have been dropped mainly because

these terms are proportional to the photon distribution function  $f^{(1)}$  which on very small scales (those of interest for reionization) is suppressed by the diffusion damping. Here, we want to be completely general and we have to keep them.

### 6.3. The Brightness Equation

6.3.1. *First Order.* The Boltzmann equation for photons is obtained by combining (80) with (122)-(123). At first order, the left-hand side reads

$$\begin{aligned} \frac{df}{d\eta} &= \frac{df^{(1)}}{d\eta} - p \frac{\partial f^{(0)}}{\partial p} \frac{\partial \Phi^{(1)}}{\partial x^i} \frac{dx^i}{d\eta} \\ &\quad + p \frac{\partial f^{(0)}}{\partial p} \frac{\partial \Psi^{(1)}}{\partial \eta}. \end{aligned} \quad (124)$$

At first order, it is useful to characterize the perturbations to the Bose-Einstein distribution function (78) in terms of a perturbation to the temperature as

$$f(x^i, p, n^i, \eta) = 2 \left[ \exp \left\{ \frac{p}{T(\eta)(1 + \Theta^{(1)})} \right\} - 1 \right]^{-1}. \quad (125)$$

Thus, it turns out that

$$f^{(1)} = -p \frac{\partial f^{(0)}}{\partial p} \Theta^{(1)}, \quad (126)$$

where we have used the fact that  $\partial f / \partial \Theta|_{\Theta=0} = -p \partial f^{(0)} / \partial p$ . In terms of this variable  $\Theta^{(1)}$  the linear collision term (122) will now become proportional to  $-p \partial f^{(0)} / \partial p$  which contains the only explicit dependence on  $p$ , and the same happens for the left-hand side, (124). This is telling us that at first order  $\Theta^{(1)}$  does not depend on  $p$  but only on  $x^i, n^i, \eta$ ,  $\Theta^{(1)} = \Theta^{(1)}(x^i, n^i, \tau)$ . This is well known and the physical reason is that at linear order the energy transfer in Compton collisions between photons and electrons is almost elastic, so that no spectral distortions are generated (in the sense that the temperature fluctuations do not depend on  $p$ ). Therefore, the Boltzmann equation for  $\Theta^{(1)}$  reads

$$\begin{aligned} \frac{\partial \Theta^{(1)}}{\partial \eta} + n^i \frac{\partial \Theta^{(1)}}{\partial x^i} + \frac{\partial \Phi^{(1)}}{\partial x^i} n^i - \frac{\partial \Psi^{(1)}}{\partial \eta} \\ = n_e \sigma_T a \left[ \Theta_0^{(1)} + \frac{1}{2} \Theta_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) - \Theta^{(1)} + \mathbf{v} \cdot \mathbf{n} \right], \end{aligned} \quad (127)$$

where we made us of  $f_e^{(1)} = -p \partial f^{(0)} / \partial p \Theta_e^{(1)}$ , according to the decomposition of (96), and we have taken the zero-order expressions for  $dx^i / d\eta$ , dropping the contribution from  $dn^i / d\eta$  in (64) since it is already first order.

Notice that since  $\Theta^{(1)}$  is independent of  $p$ , it is equivalent to consider the quantity

$$\Delta^{(1)}(x^i, n^i, \tau) = \frac{\int dp p^3 f^{(1)}}{\int dp p^3 f^{(0)}}, \quad (128)$$

being  $\Delta^{(1)} = 4\Theta^{(1)}$  at this order. The physical meaning of  $\Delta^{(1)}$  is that of a fractional energy perturbation (in a given direction). From (80), another way to write an equation for  $\Delta^{(1)}$ —the so-called brightness equation—is

$$\begin{aligned} & \frac{d}{d\eta} \left[ \Delta^{(1)} + 4\Phi^{(1)} \right] - 4 \frac{\partial}{\partial \eta} \left( \Phi^{(1)} + \Psi^{(1)} \right) \\ &= n_e \sigma_T a \left[ \Delta_0^{(1)} + \frac{1}{2} \Delta_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) - \Delta^{(1)} + 4\mathbf{v} \cdot \mathbf{n} \right]. \end{aligned} \quad (129)$$

6.3.2. *Second Order.* The previous results show that at linear order the photon distribution function has a Planck spectrum with the temperature that at any point depends on the photon direction. At second order, one could characterize the perturbed photon distribution function in a similar way as in (125)

$$f(x^i, p, n^i, \eta) = 2 \left[ \exp \left\{ \frac{p}{T(\eta) e^\Theta} - 1 \right\} \right]^{-1}, \quad (130)$$

where by expanding  $\Theta = \Theta^{(1)} + \Theta^{(2)}/2 + \dots$  as usual one recovers the first-order expression. For example, in terms of  $\Theta$ , the perturbation of  $f^{(1)}$  is given by (126), while at second order

$$\frac{f^{(2)}}{2} = -\frac{p}{2} \frac{\partial f^{(0)}}{\partial p} \Theta^{(2)} + \frac{1}{2} \left( p^2 \frac{\partial^2 f^{(0)}}{\partial p^2} + p \frac{\partial f^{(0)}}{\partial p} \right) (\Theta^{(1)})^2. \quad (131)$$

However, as discussed in details in [64, 65], now the second-order perturbation  $\Theta^{(2)}$  will not be momentum independent because the collision term in the equation for  $\Theta^{(2)}$  does depend explicitly on  $p$  (defining the combination  $-(p \partial f^{(0)}/\partial p)^{-1} f^{(2)}$  does not lead to a second-order momentum independent equation as above). Such dependence is evident, for example, in the terms of  $C^{(2)}(\mathbf{p})$ , (123), proportional to  $v$  or  $v^2$ , and in the Kompaneets terms. The physical reason is that nonlinearities in the energy exchange between photons and electrons during Compton collisions bring spectral distortions. For example, in the isotropic limit, only the Kompaneets terms survive giving rise to the Sunyaev-Zeldovich distortions. As discussed in [65], the Sunyaev-Zeldovich distortions can also be obtained with the correct coefficients by replacing the average over the direction electron  $\langle v^2 \rangle$  with the mean squared thermal velocity  $\langle v_{th}^2 \rangle = 3T_e/m_e$  in (123). This is due simply to the fact that the distinction between thermal and bulk velocity of the electrons is just for convenience. This fact also shows that spectral distortions due to the bulk flow (kinetic Sunyaev-Zeldovich) has the same form as the thermal effect. Thus spectral distortions can be in general described by a global Compton  $y$ -parameter (see [65] for a full discussion of spectral distortions and also [66]). However, in the following we will not be interested in the frequency dependence but only in the anisotropies of the radiation distribution.

Therefore we can integrate over the momentum  $p$  and define [64, 65]

$$\Delta^{(2)}(x^i, n^i, \tau) = \frac{\int dp p^3 f^{(2)}}{\int dp p^3 f^{(0)}}, \quad (132)$$

as in (128).

Integration over  $p$  of (80)–(123) is straightforward using the following relations:

$$\begin{aligned} & \int dp p^3 p \frac{\partial f^{(0)}}{\partial p} = -4N; \\ & \int dp p^3 p^2 \frac{\partial^2 f^{(0)}}{\partial p^2} = 20N; \\ & \int dp p^3 f^{(1)} = N \Delta^{(1)}; \\ & \int dp p^3 p \frac{\partial f^{(1)}}{\partial p} = -4N \Delta^{(1)}. \end{aligned} \quad (133)$$

Here,  $N = \int dp p^3 f^{(0)}$  is the normalization factor (it is just proportional the background energy density of photons  $\bar{\rho}_\gamma$ ). At first order one recovers (129). At second, order we find

$$\begin{aligned} & \frac{1}{2} \frac{d}{d\eta} \left[ \Delta^{(2)} + 4\Phi^{(2)} \right] + \frac{d}{d\eta} \left[ \Delta^{(1)} + 4\Phi^{(1)} \right] \\ & - 4\Delta^{(1)} \left( \Psi^{(1)'} - \Phi_{,i}^{(1)} n^i \right) - 2 \frac{\partial}{\partial \eta} \left( \Psi^{(2)} + \Phi^{(2)} \right) \\ & + 4 \frac{\partial \omega_i}{\partial \eta} n^i + 2 \frac{\partial \chi_{ij}}{\partial \eta} n^i n^j \\ & = -\frac{\tau'}{2} \left[ \Delta_{00}^{(2)} - \Delta^{(2)} - \frac{1}{2} \sum_{m=-2}^2 \frac{\sqrt{4\pi}}{5^{3/2}} \Delta_{2m}^{(2)} \right. \\ & \quad \times Y_{2m}(\mathbf{n}) + 2 \left( \delta_e^{(1)} + \Phi^{(1)} \right) \\ & \quad \times \left( \Delta_0^{(1)} + \frac{1}{2} \Delta_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) - \Delta^{(1)} + 4\mathbf{v} \cdot \mathbf{n} \right) \\ & \quad + 4\mathbf{v}^{(2)} \cdot \mathbf{n} + 2(\mathbf{v} \cdot \mathbf{n}) \\ & \quad \times \left[ \Delta^{(1)} + 3\Delta_0^{(1)} - \Delta_2^{(1)} \left( 1 - \frac{5}{2} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right) \right] \\ & \quad \left. - v \Delta_1^{(1)} (4 + 2P_2(\hat{\mathbf{v}} \cdot \mathbf{n})) + 14(\mathbf{v} \cdot \mathbf{n})^2 - 2v^2 \right], \end{aligned} \quad (134)$$

where we have expanded the angular dependence of  $\Delta$  as in (97)

$$\Delta^{(i)}(\mathbf{x}, \mathbf{n}) = \sum_{\ell} \sum_{m=-\ell}^{\ell} \Delta_{\ell m}^{(i)}(\mathbf{x}) (-i)^\ell \sqrt{\frac{4\pi}{2\ell+1}} Y_{\ell m}(\mathbf{n}), \quad (135)$$

with

$$\Delta_{\ell m}^{(i)} = (-i)^{-\ell} \sqrt{\frac{2\ell+1}{4\pi}} \int d\Omega \Delta^{(i)} Y_{\ell m}^*(\mathbf{n}), \quad (136)$$

where we recall that the superscript stands by the order of the perturbation. When going to Fourier space, some convolutions will appear and the coefficients  $\Delta_{\ell m}^{(1)}(\mathbf{k}', \eta)$  are related to the more familiar ones  $\Delta_{\ell}^{(1)}(\mathbf{k}', \eta)$  as

$$\Delta_{\ell m}^{(1)}(\mathbf{k}') = i^{\ell} \sqrt{4\pi(2\ell+1)} Y_{\ell m}^*(\hat{\mathbf{k}}') \Delta_{\ell}^{(1)}(\mathbf{k}'). \quad (137)$$

Equation (137) is obtained for linear perturbations by inserting into (136) the expansion

$$\Delta^{(1)}(\mathbf{k}, \mathbf{n}) = \sum_{l'} (2l'+1) \Delta_{l'}^{(1)}(\mathbf{k}) P_{l'}(\mu), \quad (138)$$

and expanding the Legendre polynomials as

$$P_{l'}(\mu) = \frac{4\pi}{2l'+1} \sum_{m'=-l'}^{m'=l'} Y_{l'm'}(\mathbf{n}) Y_{l'm'}^*(\hat{\mathbf{k}}). \quad (139)$$

Notice that the expansion (138) is valid only for linear scalar perturbations, which is consistent with our treatment where at linear order we neglect vector and tensor perturbation modes.

In (134), we have introduced the differential optical depth

$$\tau' = -\bar{n}_e \sigma_T a. \quad (140)$$

It is understood that on the left-hand side of (134) one has to pick up for the total time derivatives only those terms which contribute to second order. Thus, we have to take

$$\begin{aligned} & \frac{1}{2} \frac{d}{d\eta} \left[ \Delta^{(2)} + 4\Phi^{(2)} \right] + \frac{d}{d\eta} \left[ \Delta^{(1)} + 4\Phi^{(1)} \right] \Big|^{(2)} \\ &= \frac{1}{2} \left( \frac{\partial}{\partial \eta} + n^i \frac{\partial}{\partial x^i} \right) \left( \Delta^{(2)} + 4\Phi^{(2)} \right) + n^i \left( \Phi^{(1)} + \Psi^{(1)} \right) \\ & \quad \times \partial_i \left( \Delta^{(1)} + 4\Phi^{(1)} \right) + \left[ \left( \Phi_{,j}^{(1)} + \Psi_{,j}^{(1)} \right) n^i n^j - \left( \Phi^{,i} + \Psi^{,i} \right) \right] \\ & \quad \times \frac{\partial \Delta^{(1)}}{\partial n^i}, \end{aligned} \quad (141)$$

where we used (66) and (77).

**6.3.3. Hierarchy Equations for Multipole Moments.** Let us now move to Fourier space. In the following, for a given wave-vector  $\mathbf{k}$ , we will choose the coordinate system such that  $\mathbf{e}_3 = \hat{\mathbf{k}}$  and the polar angle of the photon momentum is  $\vartheta$ , with  $\mu = \cos \vartheta = \hat{\mathbf{k}} \cdot \mathbf{n}$ . Then, (134) can be written as

$$\Delta^{(2)'} + ik\mu\Delta^{(2)} - \tau'\Delta^{(2)} = S(\mathbf{k}, \mathbf{n}, \eta), \quad (142)$$

where  $S(\mathbf{k}, \mathbf{n}, \eta)$  can be easily read off (134). We now expand the temperature anisotropy in the multipole moments  $\Delta_{\ell m}^{(2)}$

order to obtain a system of coupled differential equations. By applying the angular integral of (136) to (142), we find

$$\begin{aligned} \Delta_{\ell m}^{(2)'}(\mathbf{k}, \eta) &= k \left[ \frac{\kappa_{\ell m}}{2\ell-1} \Delta_{\ell-1, m}^{(2)} - \frac{\kappa_{\ell+1, m}}{2\ell+3} \Delta_{\ell+1, m}^{(2)} \right] \\ & \quad + \tau' \Delta_{\ell m}^{(2)} + S_{\ell m} \end{aligned} \quad (143)$$

where the first term on the right-hand side of (143) has been obtained by using the relation

$$\begin{aligned} & i\mathbf{k} \cdot \mathbf{n} \Delta^{(2)}(\mathbf{k}) \\ &= \sum_{\ell m} \Delta_{\ell m}^{(2)}(\mathbf{k}) \frac{k}{2\ell+1} \left[ \kappa_{\ell m} \tilde{G}_{\ell-1, m} - \kappa_{\ell+1, m} \tilde{G}_{\ell+1, m} \right] \\ &= k \sum_{\ell m} \left[ \frac{\kappa_{\ell m}}{2\ell-1} \Delta_{\ell-1, m}^{(2)} - \frac{\kappa_{\ell m}}{2\ell+3} \Delta_{\ell+1, m}^{(2)} \right] \tilde{G}_{\ell m}, \end{aligned} \quad (144)$$

where  $\tilde{G}_{\ell m} = (-i)^{\ell} \sqrt{4\pi/(2\ell+1)} Y_{\ell m}(\mathbf{n})$  is the angular mode for the decomposition (135) and  $\kappa_{\ell m} = \sqrt{l^2 - m^2}$ . This relation has been discussed in [67, 68] and corresponds to the term  $n^i \partial \Delta^{(2)}/\partial x^i$  in (134). In (143)  $S_{\ell m}$  are the multipole moments of the source term according to the decomposition (136). We do not show its complete expression here, since it is very long, and we prefer to make some general comments (for some specific examples about the terms  $S_{\ell m}$  see Section 10). As it should be already clear from the calculations involving second-order perturbations performed so far, (143) has the same functional form as the Boltzmann equation at linear order (just replace the order of the linear perturbations (1) with that of intrinsically second-order terms (2) with the exception of the source terms  $S_{\ell m}$  which now contain both intrinsic second-order perturbations and also products of first-order perturbations. Therefore, as expected, at second order we recover some intrinsic effects which are characteristic of the linear regime. In (143), (144) represents the free streaming effect: when the radiation undergoes free-streaming, the inhomogeneities of the photon distribution are seen by the observer as angular anisotropies. At first order, it is responsible for the hierarchy of Boltzmann equations coupling the different  $\ell$  modes, and it represents a projection effect of fluctuations on a scale  $k$  onto the angular scale  $\ell \sim k\eta$ . The term  $\tau' \Delta_{\ell m}^{(2)}$  causes an exponential suppression of anisotropies in the absence of the source term  $S_{\ell m}$ . The source term contains additional scattering processes and gravitational effects. The intrinsically second-order part of the source term just reproduces the expression of the first order case. Of course the dynamics of the second-order metric and baryon-velocity perturbations which appear will be different and governed by the second-order Einstein equations and continuity equations. The remaining terms in the source are second-order effects generated as nonlinear combinations of the primordial (first-order) perturbations. Notice in particular that they involve the first-order anisotropies  $\Delta_{\ell}^{(1)}$  and as a consequence such terms contribute to generate the hierarchy of equations (apart from the free-streaming effect). On large scales (above the horizon at recombination), we can say that the main effects are due to gravity, and they include

the Sachs-Wolfe and the (late and early) Sachs-Wolfe effect due to the redshift photons suffer when travelling through the second-order gravitational potentials. These, together with the contribution due to the second-order tensor modes, have been already studied in details in [31] (see also [37]). Another important gravitational effect is that of lensing of photons as they travel from the last scattering surface to us. A contribution of this type is given by the last term of (141). On the other hand, examples of second-order scattering effects are the terms proportional to the square of the baryon-velocity fluid ( $v^2$ ), giving rise to the quadratic Doppler effect (like those in the last line of (134)), or those coupling the photon fluctuations to the baryon velocity (second-line from the bottom of (134)). The Vishniac effect corresponds to the terms proportional to  $\delta_e^{(1)}$ . Finally, notice that in the Boltzmann equation (134) the second-order baryon velocity appears. At linear order, the baryon velocity is irrotational, meaning that it is the gradient of a potential, and thus in Fourier space it is parallel to  $\hat{\mathbf{k}}$ , and following the same notation of [69], we write

$$\mathbf{v}^{(1)}(\mathbf{k}) = -iv_0^{(1)}(\mathbf{k})\hat{\mathbf{k}}. \quad (145)$$

The second-order velocity perturbation will contain a transverse (divergence-free) part whose components are orthogonal to  $\hat{\mathbf{k}} = \mathbf{e}_3$ , and we can write

$$\mathbf{v}^{(2)}(\mathbf{k}) = -iv_0^{(2)}(\mathbf{k})\mathbf{e}_3 + \sum_{m=\pm 1} v_m^{(2)} \frac{\mathbf{e}_2 \mp i\mathbf{e}_1}{\sqrt{2}}, \quad (146)$$

where  $\mathbf{e}_i$  form an orthonormal basis with  $\hat{\mathbf{k}}$ . The second-order perturbation  $\omega_i$  is decomposed in a similar way, with  $\omega_{\pm 1}$  the corresponding components (in this case, in the Poisson gauge, there is no scalar component). Similarly for the tensor perturbation  $\chi_{ij}$  we have indicated its amplitudes as  $\chi_{\pm 2}$  in the decomposition [68]

$$\chi_{ij} = \sum_{m=\pm 2} -\sqrt{\frac{3}{8}}\chi_m(\mathbf{e}_1 \pm i\mathbf{e}_2)_i(\mathbf{e}_1 \pm i\mathbf{e}_2)_j. \quad (147)$$

In computing the source term  $S_{lm}$ , one has to take into account that in the gravitational part of the Boltzmann equation and in the collision term there are some terms, like  $\delta_e^{(1)}\mathbf{v}$ , which still can be decomposed in the scalar and transverse parts in Fourier space as in (146). For a generic quantity  $f(\mathbf{x})\mathbf{v}$ , one can indicate the corresponding scalar and vortical components with  $(f\mathbf{v})_m$  and their explicit expression is easily found by projecting the Fourier modes of  $f(\mathbf{x})\mathbf{v}$  along the  $\hat{\mathbf{k}} = \mathbf{e}_3$  and  $(\mathbf{e}_2 \mp i\mathbf{e}_1)$  directions

$$(f\mathbf{v})_m(\mathbf{k}) = \int \frac{d^3k_1}{(2\pi)^3} v_0^{(1)}(\mathbf{k}_1) f(\mathbf{k}_2) Y_{1m}^*(\hat{\mathbf{k}}_1) \sqrt{\frac{4\pi}{3}}. \quad (148)$$

Similarly, for a term like  $f(\mathbf{x})\nabla g(\mathbf{x})$  one can use the notation

$$(f\nabla g)_m(\mathbf{k}) = - \int \frac{d^3k_1}{(2\pi)^3} k_1 g(\mathbf{k}_1) f(\mathbf{k}_2) Y_{1m}^*(\hat{\mathbf{k}}_1) \sqrt{\frac{4\pi}{3}}. \quad (149)$$

**6.3.4. Integral Solution of the Second-Order Boltzmann Equation.** As in linear theory, one can derive an integral solution of the Boltzmann equation (134) in terms of the source term  $S$ . Following the standard procedure (see, e.g., [5]) for linear perturbations, we write the left-hand side as  $\Delta^{(2)'} + ik_\mu \Delta^{(2)} - \tau' \Delta^{(2)} = e^{-ik_\mu \eta + \tau} d[\Delta^{(2)} e^{ik_\mu \eta - \tau}] / d\eta$  in order to derive the integral solution

$$\Delta^{(2)}(\mathbf{k}, \mathbf{n}, \eta_0) = \int_0^{\eta_0} d\eta S(\mathbf{k}, \mathbf{n}, \eta) e^{ik_\mu(\eta - \eta_0) - \tau}, \quad (150)$$

where  $\eta_0$  stands by the present time. The expression of the photon moments  $\Delta_{\ell m}^{(2)}$  can be obtained as usual from (136). In the previous section we have already found the coefficients for the decomposition of source term  $S$

$$S(\mathbf{k}, \mathbf{n}, \eta) = \sum_{\ell} \sum_{m=-\ell}^{\ell} S_{\ell m}(\mathbf{k}, \eta) (-i)^{\ell} \sqrt{\frac{4\pi}{2\ell+1}} Y_{\ell m}(\mathbf{n}). \quad (151)$$

In (150), there is an additional angular dependence in the exponential. It is easy to take it into account by recalling that

$$e^{ik \cdot \mathbf{x}} = \sum_{\ell} (i)^{\ell} (2\ell+1) j_{\ell}(kx) P_{\ell}(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}}). \quad (152)$$

Thus the angular integral (136) is computed by using the decomposition of the source term (151) and (152)

$$\begin{aligned} \Delta_{\ell m}^{(2)}(\mathbf{k}, \eta_0) &= (-1)^{-m} (-i)^{-\ell} (2\ell+1) \int_0^{\eta_0} d\eta e^{-\tau(\eta)} \\ &\times \sum_{\ell_2} \sum_{m_2=-\ell_2}^{\ell_2} (-i)^{\ell_2} S_{\ell_2 m_2} \sum_{\ell_1} i^{\ell_1} j_{\ell_1}(k(\eta - \eta_0)) \\ &\times (2\ell_1+1) \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell \\ 0 & m_2 & -m \end{pmatrix}, \end{aligned} \quad (153)$$

where the Wigner 3- $j$  symbols appear because of the Gaunt integrals

$$\begin{aligned} \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} &\equiv \int d^2\hat{\mathbf{n}} Y_{\ell_1 m_1}(\hat{\mathbf{n}}) Y_{\ell_2 m_2}(\hat{\mathbf{n}}) Y_{\ell_3 m_3}(\hat{\mathbf{n}}) \\ &= \sqrt{\frac{(2\ell_1+1)(2\ell_2+1)(2\ell_3+1)}{4\pi}} \\ &\times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}, \end{aligned} \quad (154)$$

Since the second of the Wigner 3- $j$  symbols in (153) is nonzero only if  $m = m_2$ , our solution can be rewritten to recover the corresponding expression found for linear anisotropies in [68]

$$\frac{\Delta_{\ell m}^{(2)}(\mathbf{k}, \eta_0)}{2\ell + 1} = \int_0^{\eta_0} d\eta e^{-\tau(\eta)} \sum_{\ell_2} S_{\ell_2 m} j_{\ell}^{(\ell_2, m)}[k(\eta_0 - \eta)], \quad (155)$$

where  $j_{\ell}^{(\ell_2, m)}[k(\eta_0 - \eta)]$  are the so called radial functions. Of course the main information at second order is included in the source term containing different effects due to the nonlinearity of the perturbations. In the total angular momentum method of [68], (155) is interpreted just as the intergration over the radial coordinate ( $\chi = \eta_0 - \eta$ ) of the projected source term. Another important comment is that, as in linear theory, the integral solution (153) is in fact just a formal solution, since the source term  $S$  contains itself the (intrinsically) second-order photon moments up to  $l = 2$  (and the products of first-order perturbations in principle may contain  $l$ -modes up to infinity). This means that one has anyway to resort to the hierarchy equations for photons, (143), to solve for these moments. Once the moments entering in the source function are computed the higher moments are obtained from the integral solution.

**6.4. The Boltzmann Equation for Baryons and Cold Dark Matter.** In this section we will derive the Boltzmann equation for massive particles, which is the case of interest for baryons and dark matter. These equations are necessary to find the time evolution of number densities and velocities of the baryon fluid which appear in the brightness equation, thus allowing to close the system of equations. Let us start from the baryon component. Electrons are tightly coupled to protons via the Coulomb interactions. This forces the relative energy density contrasts and the velocities to a common value,  $\delta_e = \delta_p \equiv \delta_b$  and  $v_e = v_p \equiv v$ , so that we can identify electrons and protons collectively as ‘‘baryonic’’ matter.

To derive the Boltzmann equation for baryons let us first focus on the collisionless equation and compute therefore  $dg/d\eta$ , where  $g$  is the distribution function for a massive species with mass  $m$ . One of the differences with respect to photons is just that baryons are nonrelativistic for the epochs of interest. Thus, the first step is to generalize the formulae in Section 4 up to (77) to the case of a massive particle. In this case, one enforces the constraint  $Q^2 = g_{\mu\nu} Q^\mu Q^\nu = -m^2$  and it is also useful to use the particle energy  $E = \sqrt{q^2 + m^2}$ , where  $q$  is defined as in (59). Moreover, in this case it is very convenient to take the distribution function as a function of the variables  $q^i = qn^i$ , the position  $x^i$  and time  $\eta$ , without using the explicit splitting into the magnitude of the momentum  $q$  (or the energy  $E$ ) and its direction  $n^i$ . Thus, the total time derivative of the distribution functions reads

$$\frac{dg}{d\eta} = \frac{\partial g}{\partial \eta} + \frac{\partial g}{\partial x^i} \frac{dx^i}{d\eta} + \frac{\partial g}{\partial q^i} \frac{dq^i}{d\eta}. \quad (156)$$

We will not give the details of the calculation since we just need to replicate the same computation we did for the photons. For the four momenta of the particle, notice that  $Q^i$  has the same form as (62), while for  $Q^0$ , we find

$$Q^0 = \frac{e^{-\Phi}}{a} E \left( 1 + \omega_i \frac{q^i}{E} \right). \quad (157)$$

In the following, we give the expressions for  $dx^i/d\eta$  and  $dq^i/d\eta$ .

(a) As in (66)  $dx^i/d\eta = Q^i/Q^0$  and it turns out to be

$$\frac{dx^i}{d\eta} = \frac{q}{E} n^i e^{\Phi+\Psi} \left( 1 - \omega_i n^i \frac{q}{E} \right) \left( 1 - \frac{1}{2} \chi_{km} n^k n^m \right). \quad (158)$$

(b) For  $dq^i/d\eta$  we need the expression of  $Q^i$  which is the same as that of (62)

$$Q^i = \frac{q^i}{a} e^\Psi \left( 1 - \frac{1}{2} \chi_{km} n^k n^m \right). \quad (159)$$

The spatial component of the geodesic equation, up to second order, reads

$$\begin{aligned} \frac{dQ^i}{d\eta} &= -2(\mathcal{H} - \Psi') \left( 1 - \frac{1}{2} \chi_{km} n^k n^m \right) \frac{q}{a} n^i e^\Psi + e^{\Phi+2\Psi} \\ &\times \left( \frac{\partial \Psi}{\partial x^k} \frac{q^2}{aE} (2n^i n^k - \delta^{ik}) - \frac{\partial \Phi}{\partial x^i} \frac{E}{a} \right) \\ &- \frac{E}{a} \left[ \omega^i + \mathcal{H} \omega^i + q^k (\chi_{k}^i + \omega_{k}^i - \omega_k^i) \right] \\ &+ \left[ \mathcal{H} \omega^i \delta_{jk} - \frac{1}{2} (\chi_{j,k}^i + \chi_{k,j}^i - \chi_{jk}^i) \right] \frac{q^j q^k}{Ea}. \end{aligned} \quad (160)$$

Proceeding as in the massless case, we now take the total time derivative of (159) and using (160), we find

$$\begin{aligned} \frac{dq^i}{d\eta} &= -(\mathcal{H} - \Psi') q^i + \Psi_{,k} \frac{q^i q^k}{E} e^{\Phi+\Psi} - \Phi^i E e^{\Phi+\Psi} \\ &- \Psi_{,i} \frac{q^2}{E} e^{\Phi+\Psi} - E (\omega^i + \mathcal{H} \omega^i) - (\chi_{k}^i + \omega_{k}^i - \omega_k^i) E q^k \\ &+ \left[ \mathcal{H} \omega^i \delta_{jk} - \frac{1}{2} (\chi_{j,k}^i + \chi_{k,j}^i - \chi_{jk}^i) \right] \frac{q^j q^k}{E}. \end{aligned} \quad (161)$$

We can now write the total time derivative of the distribution function as

$$\begin{aligned} & \frac{dg}{d\eta} \\ &= \frac{\partial g}{\partial \eta} + \frac{q}{E} n^i e^{\Phi+\Psi} \left( 1 - \omega_i n^i - \frac{1}{2} \chi_{km} n^k n^m \right) \frac{\partial g}{\partial x^i} \\ &+ \left[ -(\mathcal{H} - \Psi') q^i + \Psi_{,k} \frac{q^i q^k}{E} e^{\Phi+\Psi} - \Phi^{,i} E e^{\Phi+\Psi} \right. \\ &\quad - \Psi_{,i} \frac{q^2}{E} e^{\Phi+\Psi} - E(\omega^i + \mathcal{H} \omega^i) - (\chi_k^i + \omega_k^i - \omega_k^i) E q^k \\ &\quad \left. + \left( \mathcal{H} \omega^i \delta_{jk} - \frac{1}{2} (\chi_{j,k}^i + \chi_{k,j}^i - \chi_{jk}^i) \right) \frac{q^j q^k}{E} \right] \frac{\partial g}{\partial q^i}. \end{aligned} \quad (162)$$

This equation is completely general since we have just solved for the kinematics of massive particles. As far as the collision terms are concerned, for the system of electrons and protons, we consider the Coulomb scattering processes between the electrons and protons and the Compton scatterings between photons and electrons

$$\frac{dg_e}{d\eta}(\mathbf{x}, \mathbf{q}, \eta) = \langle c_{ep} \rangle_{QQ'q'} + \langle c_{e\gamma} \rangle_{pp'q'}, \quad (163)$$

$$\frac{dg_p}{d\eta}(\mathbf{x}, \mathbf{Q}, \eta) = \langle c_{ep} \rangle_{qq'Q'}, \quad (164)$$

where we have adopted the same formalism of [70] with  $\mathbf{p}$  and  $\mathbf{p}'$  the initial and final momenta of the photons,  $\mathbf{q}$  and  $\mathbf{q}'$  the corresponding quantities for the electrons and for protons  $\mathbf{Q}$  and  $\mathbf{Q}'$ . The integral over different momenta is indicated by

$$\langle \dots \rangle_{pp'q'} \equiv \int \frac{d^3 p}{(2\pi)^3} \int \frac{d^3 p'}{(2\pi)^3} \int \frac{d^3 q'}{(2\pi)^3} \dots, \quad (165)$$

and thus one can read  $c_{e\gamma}$  as the unintegrated part of (82), and similarly for  $c_{ep}$  (with the appropriate amplitude  $|M|^2$ ). In (163) Compton scatterings between protons and photons can be safely neglected because the amplitude of this process has a much smaller amplitude than Compton scatterings with electrons being weighted by the inverse squared mass of the particles.

At this point for the photons we considered the perturbations around the zero-order Bose-Einstein distribution function (which are the unknown quantities). For the electrons (and protons) we can take the thermal distribution described by (84). Moreover, we will take the moments of (163)-(164) in order to find the energy-momentum continuity equations.

**6.4.1. Energy Continuity Equations.** We now integrate (162) over  $d^3 q / (2\pi)^3$ . Let us recall that in terms of the distribution function the number density  $n_e$ , and the bulk velocity  $\mathbf{v}$  are given by

$$n_e = \int \frac{d^3 q}{(2\pi)^3} g, \quad (166)$$

$$v^i = \frac{1}{n_e} \int \frac{d^3 q}{(2\pi)^3} g \frac{q^i}{E}, \quad (167)$$

where one can set  $E \simeq m_e$  since we are considering nonrelativistic particles. We will also make use of the following relations when integrating over the solid angle  $d\Omega$

$$\int d\Omega n^i = \int d\Omega n^i n^j n^k = 0, \quad \int \frac{d\Omega}{4\pi} n^i n^j = \frac{1}{3} \delta^{ij}. \quad (168)$$

Finally, notice that  $dE/dq = q/E$  and  $\partial g / \partial q = (q/E) \partial g / \partial E$ .

Thus, the first two integrals just brings  $n_e'$  and  $(n_e v^i)_{,i}$ . Notice that all the terms proportional to the second-order vector and tensor perturbations of the metric give a vanishing contribution at second order since in this case, we can take the zero-order distribution functions which depends only on  $\eta$  and  $E$ , integrate over the direction and use the fact that  $\delta^{ij} \chi_{ij} = 0$ . The trick to solve the remaining integrals is an integration by parts over  $q^i$ . We have an integral like (the one multiplying  $(\Psi' - \mathcal{H})$ )

$$\int \frac{d^3 q}{(2\pi)^3} q^i \frac{\partial g}{\partial q^i} = -3 \int \frac{d^3 q}{(2\pi)^3} g = -3n_e, \quad (169)$$

after an integration by parts over  $q^i$ . The remaining integrals can be solved still by integrating by parts over  $q^i$ . The integral proportional to  $\Phi^{,i}$  in (162) gives

$$\int \frac{d^3 q}{(2\pi)^3} E = -v_i n_e, \quad (170)$$

where we have used the fact that  $dE/dq^i = q^i/E$ . For the integral

$$\int \frac{d^3 q}{(2\pi)^3} \frac{q^i q^k}{E} \frac{\partial g}{\partial q^i}, \quad (171)$$

the integration by parts brings two pieces, one from the derivation of  $q^i q^k$  and one from the derivation of the energy  $E$

$$\begin{aligned} & -4 \int \frac{d^3 q}{(2\pi)^3} g \frac{q^k}{E} + \int \frac{d^3 q}{(2\pi)^3} g \frac{q^2}{E} \frac{q^k}{E} \\ &= -4v^k n_e + \int \frac{d^3 q}{(2\pi)^3} g \frac{q^2}{E^2} \frac{q^k}{E}. \end{aligned} \quad (172)$$

The last integral in (172) can indeed be neglected. To check this one makes use of the explicit expression (84) for the distribution function  $g$  to derive

$$\frac{\partial g}{\partial v^i} = g \frac{q_i}{T_e} - \frac{m_e}{T_e} v_i g, \quad (173)$$

$$\int \frac{d^3 q}{(2\pi)^3} g q^i q^j = \delta^{ij} n_e m_e T_e + n_e m_e^2 v^i v^j. \quad (174)$$

Thus, it is easy to compute

$$\frac{\Psi_{,k}}{m_e^3} \int \frac{d^3q}{(2\pi)^3} g q^2 q^k = -\Psi_{,k} v^2 \frac{T_e}{m_e} + 3\Psi_{,k} v_k n_e \frac{T_e}{m_e} + \Psi_{,k} v_k v^2, \quad (175)$$

which is negligible taking into account that  $T_e/m_e$  is of the order of the thermal velocity squared.

With these results, we are now able to compute the left-hand side of the Boltzmann equation (163) integrated over  $d^3q/(2\pi)^3$ . The same operation must be done for the collision terms on the right hand side. For example, for the first of the equations in (163) this brings to the integrals  $\langle c_{ep} \rangle_{QQ'qq'} + \langle c_{ey} \rangle_{pp'qq'}$ . However, looking at (83) one realizes that  $\langle c_{ey} \rangle_{pp'qq'}$  vanishes because the integrand is antisymmetric under the change  $\mathbf{q} \leftrightarrow \mathbf{q}'$  and  $\mathbf{p} \leftrightarrow \mathbf{p}'$ . In fact this is simply a consequence of the fact that the electron number is conserved for this process. The same argument holds for the other term  $\langle c_{ep} \rangle_{QQ'qq'}$ . Therefore, the right-hand side of (163) integrated over  $d^3q/(2\pi)^3$  vanishes and we can give the evolution equation for  $n_e$ . Collecting the results of (169) to (175), we find

$$\begin{aligned} \frac{\partial n_e}{\partial \eta} + e^{\Phi+\Psi} \frac{\partial(v^i n_e)}{\partial x^i} + 3(\mathcal{H} - \Psi') n_e \\ + e^{\Phi+\Psi} v^k n_e (\Phi_{,k} - 2\Psi_{,k}) = 0. \end{aligned} \quad (176)$$

Similarly, for CDM particles, we find

$$\begin{aligned} \frac{\partial n_{\text{CDM}}}{\partial \eta} + e^{\Phi+\Psi} \frac{\partial(v^i n_{\text{CDM}})}{\partial x^i} + 3(\mathcal{H} - \Psi') n_{\text{CDM}} \\ + e^{\Phi+\Psi} v_{\text{CDM}}^k n_{\text{CDM}} (\Phi_{,k} - 2\Psi_{,k}) = 0. \end{aligned} \quad (177)$$

**6.4.2. Momentum Continuity Equations.** Let us now multiply (162) by  $(q^i/E)/(2\pi)^3$  and integrate over  $d^3q$ . In this way, we will find the continuity equation for the momentum of baryons. The first term just gives  $(n_e v^i)'$ . The second integral is of the type

$$\frac{\partial}{\partial x^j} \int \frac{d^3q}{(2\pi)^3} g \frac{q n^j}{E} \frac{q n^i}{E} = \frac{\partial}{\partial x^j} \left( n_e \frac{T_e}{m_e} \delta^{ij} + n_e v^i v^j \right), \quad (178)$$

where we have used (174) and  $E = m_e$ . The third term proportional to  $(\mathcal{H} - \Psi')$  is

$$\int \frac{d^3q}{(2\pi)^3} q^k \frac{\partial g}{\partial q_k} \frac{q^i}{E} = 4n_e + \int \frac{d^3q}{(2\pi)^3} g \frac{q^2}{E^2} \frac{q^i}{E}, \quad (179)$$

where we have integrated by parts over  $q^i$ . Notice that the last term in (179) is negligible being the same integral we discussed above in (175). By the same arguments that lead to neglect the term of (175) it is easy to check that all the remaining integrals proportional to the gravitational potentials are negligible except for

$$-e^{\Phi+\Psi} \Phi_{,k} \int \frac{d^3q}{(2\pi)^3} \frac{\partial g}{\partial q_k} q^i = n_e e^{\Phi+\Psi} \Phi^i. \quad (180)$$

The integrals proportional to the second-order vector and tensor perturbations vanish as vector and tensor perturbations are traceless and divergence-free. The only one which survives is the term proportional to  $\omega^i + \mathcal{H}\omega^i$  in (162).

Therefore, for the integral over  $d^3qq^i/E$  of the left-hand side of the Boltzmann equation (162) for a massive particle with mass  $m_e(m_p)$  and distribution function (84), we find

$$\begin{aligned} \int \frac{d^3q}{(2\pi)^3} \frac{q^i}{E} \frac{dg_e}{d\eta} = \frac{\partial(n_e v^i)}{\partial \eta} + 4(\mathcal{H} - \Psi') n_e v^i + \Phi^i e^{\Phi+\Psi} n_e \\ + e^{\Phi+\Psi} \left( n_e \frac{T_e}{m_e} \right)^i + e^{\Phi+\Psi} \frac{\partial}{\partial x^j} (n_e v^j v^i) + \frac{\partial \omega^i}{\partial \eta} n_e + \mathcal{H} \omega^i n_e. \end{aligned} \quad (181)$$

Now, in order to derive the momentum conservation equation for baryons, we take the first moment of both (163) and (164) multiplying them by  $\mathbf{q}$  and  $\mathbf{Q}$ , respectively, and integrating over the momenta. Since previously we integrated the left-hand side of these equations over  $d^3qq^i/E$ , we just need to multiply the previous integrals by  $m_e$  for the electrons and for  $m_p$  for the protons. Therefore, if we sum the first moment of (163) and (164) the dominant contribution on the left-hand side will be that of the protons

$$\int \frac{d^3Q}{(2\pi)^3} Q^i \frac{dg_p}{d\eta} = \langle c_{ep} (q^i + Q^i) \rangle_{QQ'qq'} + \langle c_{ey} q^i \rangle_{pp'qq'}. \quad (182)$$

Notice that the integral of the Coulomb collision term  $c_{ep}(q^i + Q^i)$  over all momenta vanishes simply because of momentum conservation (due to the Dirac function  $\delta^4(q + Q - q' - Q')$ ). As far as the Compton scattering is concerned, we have that, following [70],

$$\langle c_{ey} q^i \rangle_{pp'qq'} = -\langle c_{ey} p^i \rangle_{pp'qq'}, \quad (183)$$

still because of the total momentum conservation. Therefore what we can compute now is the integral over all momenta of  $c_{ey} p^i$ . Notice however that this is equivalent just to multiply the Compton collision term  $C(\mathbf{p})$  of (83) by  $p^i$  and integrate over  $d^3p/(2\pi)^3$

$$\langle c_{ey} p^i \rangle_{pp'qq'} = a e^{\Phi} \int \frac{d^3p}{(2\pi)^3} p^i C(\mathbf{p}). \quad (184)$$

where  $C(\mathbf{p})$  has been already computed in (122) and (123).

We will do the integral (184) in the following. First, let us introduce the definition of the velocity of photons in terms of the distribution function

$$(\rho_\gamma + p_\gamma) v_\gamma^i = \int \frac{d^3p}{(2\pi)^3} f p^i, \quad (185)$$

where  $p_\gamma = \rho_\gamma/3$  is the photon pressure and  $\rho_\gamma$  the energy density. At first order, we get

$$\frac{4}{3} v_\gamma^{(1)i} = \int \frac{d\Omega}{4\pi} \Delta^{(1)i} n^i, \quad (186)$$

where  $\Delta$  is the photon distribution anisotropies defined in (132). At second order, we instead find

$$\frac{4}{3} \frac{v_y^{(2)i}}{2} = \frac{1}{2} \int \frac{d\Omega}{4\pi} \Delta^{(2)} n^i - \frac{4}{3} \delta_y^{(1)} v_y^{(1)i}. \quad (187)$$

Therefore, the terms in (122) and (123) proportional to  $f^{(1)}(\mathbf{p})$  and  $f^{(2)}(\mathbf{p})$  will give rise to terms containing the velocity of the photons. On the other hand the terms proportional to  $f_0^{(1)}(p)$  and  $f_{00}^{(2)}(p)$ , once integrated, vanish because of the integral over the momentum direction  $n^i$ ,  $\int d\Omega n^i = 0$ . Also, the integrals involving  $P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) = [3(\hat{\mathbf{v}} \cdot \mathbf{n})^2 - 1]/2$  in the first line of (122) and (123) vanish since

$$\int d\Omega P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) n^i = \hat{v}^k \hat{v}^j \int d\Omega n_k n_j n^i = 0, \quad (188)$$

where we are using the relations (168). Similarly, all the terms proportional to  $v$ ,  $(\mathbf{v} \cdot \mathbf{n})^2$  and  $v^2$  do not give any contribution to (184) and, in the second-order collision term, one can check that  $\int d\Omega Y_2(\mathbf{n}) n^i = 0$ . Then there are terms proportional to  $(\mathbf{v} \cdot \mathbf{n}) f^{(0)}(p)$ ,  $(\mathbf{v} \cdot \mathbf{n}) p \partial f^{(0)}/\partial p$  and  $(\mathbf{v} \cdot \mathbf{n}) p \partial f_0^{(1)}/\partial p$  for which we can use the rules (133) when integrating over  $p$  while the integration over the momentum direction is

$$\int \frac{d\Omega}{4\pi} (\mathbf{v} \cdot \mathbf{n}) n^i = v_k \int \frac{d\Omega}{4\pi} n^k n^i = \frac{1}{3} v^i. \quad (189)$$

Finally, from the second line of (123), we get three integrals. One is

$$\int \frac{d^3 p}{(2\pi)^3} p^i (\mathbf{v} \cdot \mathbf{n}) f^{(1)}(\mathbf{p}) = \bar{\rho}_\gamma \int \frac{d\Omega}{4\pi} \Delta^{(1)} (\mathbf{v} \cdot \mathbf{n}) n^i, \quad (190)$$

where  $\bar{\rho}_\gamma$  is the background energy density of the photons. The second comes from

$$\begin{aligned} & \frac{1}{2} \int \frac{d^3 p}{(2\pi)^3} p^i (\mathbf{v} \cdot \mathbf{n}) P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \left( f_2^{(1)}(p) - p \frac{\partial f_2^{(1)}(p)}{\partial p} \right) \\ &= \frac{5}{4} \bar{\rho}_\gamma \Delta_2^{(1)} \left[ 3v_j \hat{v}_k \hat{v}_l \int \frac{d\Omega}{4\pi} n^i n^j n^k n^l - v_j \int \frac{d\Omega}{4\pi} n^i n^j \right] \\ &= \frac{1}{3} \bar{\rho}_\gamma \Delta_2^{(1)} \hat{v}^i, \end{aligned} \quad (191)$$

where we have used the rules (133), Equation (168) and  $\int (d\Omega/4\pi) n^i n^j n^k n^l = (\delta^{ij} \delta^{kl} + \delta^{ik} \delta^{lj} + \delta^{il} \delta^{jk})/15$ . In fact, the third integral

$$- \int \frac{d^3 p}{(2\pi)^3} p^i (\mathbf{v} \cdot \mathbf{n}) f_2^{(1)}(p), \quad (192)$$

exactly cancels the previous one. Summing the various integrals, we find

$$\begin{aligned} \int \frac{d\mathbf{p}}{(2\pi)^3} C(\mathbf{p}) \mathbf{p} &= n_e \sigma_T \bar{\rho}_\gamma \left[ \frac{4}{3} (\mathbf{v}^{(1)} - \mathbf{v}_y^{(1)}) - \int \frac{d\Omega}{4\pi} \frac{\Delta^{(2)}}{2} \mathbf{n} \right. \\ &\quad + \frac{4}{3} \frac{\mathbf{v}^{(2)}}{2} + \frac{4}{3} \delta_e^{(1)} (\mathbf{v}^{(1)} - \mathbf{v}_y^{(1)}) \\ &\quad \left. + \int \frac{d\Omega}{4\pi} \Delta^{(1)} (\mathbf{v} \cdot \mathbf{n}) \mathbf{n} + \Delta_0^{(1)} \mathbf{v} \right]. \end{aligned} \quad (193)$$

Equation (193) can be further simplified. Recalling that  $\delta_y^{(1)} = \Delta_0^{(1)}$ , we use (187) and notice that

$$\int \frac{d\Omega}{4\pi} \Delta^{(1)} (\mathbf{v} \cdot \mathbf{n}) n^i = v_j^{(1)} \Pi_\gamma^{ij} + \frac{1}{3} v^i \Delta_0^{(1)}, \quad (194)$$

where the photon quadrupole  $\Pi_\gamma^{ij}$  is defined as

$$\Pi_\gamma^{ij} = \int \frac{d\Omega}{4\pi} \left( n^i n^j - \frac{1}{3} \delta^{ij} \right) \left( \Delta^{(1)} + \frac{\Delta^{(2)}}{2} \right). \quad (195)$$

Thus, our final expression for the integrated collision term (184) reads

$$\begin{aligned} & \int \frac{d^3 p}{(2\pi)^3} C(\mathbf{p}) p^i \\ &= n_e \sigma_T \bar{\rho}_\gamma \left[ \frac{4}{3} (v^{(1)i} - v_y^{(1)i}) + \frac{4}{3} \left( \frac{v^{(2)i}}{2} - \frac{v_y^{(2)i}}{2} \right) \right. \\ &\quad \left. + \frac{4}{3} (\delta_e^{(1)} + \Delta_0^{(1)}) (v^{(1)i} - v_y^{(1)i}) + v_j^{(1)} \Pi_\gamma^{ij} \right]. \end{aligned} \quad (196)$$

We are now able to give the momentum continuity equation for baryons by combining  $m_p dg_p/d\eta$  from (181) with the collision term (184)

$$\begin{aligned} & \frac{\partial(\rho_b v^i)}{\partial \eta} + 4(\mathcal{H} - \Psi') \rho_b v^i + \Phi^i e^{\Phi+\Psi} \rho_b + e^{\Phi+\Psi} \left( \rho_b \frac{T_b}{m_p} \right)^i \\ &+ e^{\Phi+\Psi} \frac{\partial}{\partial x^j} (\rho_b v^j v^i) + \frac{\partial \omega^i}{\partial \eta} \rho_b + \mathcal{H} \omega^i \rho_b \\ &= -n_e \sigma_T a \bar{\rho}_\gamma \left[ \frac{4}{3} (v^{(1)i} - v_y^{(1)i}) + \frac{4}{3} \left( \frac{v^{(2)i}}{2} - \frac{v_y^{(2)i}}{2} \right) \right. \\ &\quad \left. + \frac{4}{3} (\delta_b^{(1)} + \Delta_0^{(1)} + \Phi^{(1)}) (v^{(1)i} - v_y^{(1)i}) + v_j^{(1)} \Pi_\gamma^{ij} \right], \end{aligned} \quad (197)$$

where  $\rho_b$  is the baryon energy density and, as we previously explained, we took into account that to a good approximation the electrons do not contribute to the mass of baryons. In the following, we will expand explicitly at first and second-order (197).

**6.4.3. Second-Order Momentum Continuity Equation for Baryons.** At first order, we find

$$\frac{\partial v^{(1)i}}{\partial \eta} + \mathcal{H} v^{(1)i} + \Phi^{(1),i} = \frac{4}{3} \tau' \frac{\bar{\rho}_\gamma}{\bar{\rho}_b} (v^{(1)i} - v_y^{(1)i}). \quad (198)$$

At second order, there are various simplifications. In particular, notice that the term on the right-hand side of (197) which is proportional to  $\delta_b$  vanishes when matched to expansion of the left-hand side by virtue of the first-order

equation (198). Thus, at the end, we find a very simple equation

$$\begin{aligned} & \frac{1}{2} \left( \frac{\partial v^{(2)i}}{\partial \eta} + \mathcal{H} v^{(2)i} + 2 \frac{\partial \omega^i}{\partial \eta} + 2 \mathcal{H} \omega_i + \Phi^{(2),i} \right) - \frac{\partial \Psi^{(1)}}{\partial \eta} v^{(1)i} \\ & + v^{(1)j} \partial_j v^{(1)i} + (\Phi^{(1)} + \Psi^{(1)}) \Phi^{(1),i} + \left( \frac{T_b}{m_p} \right)^i \\ & = \frac{4}{3} \tau' \frac{\bar{\rho}_y}{\bar{\rho}_b} \times \left[ \left( \frac{v^{(2)i}}{2} - \frac{v_y^{(2)i}}{2} \right) + (\Delta_0^{(1)} + \Phi^{(1)}) \right. \\ & \quad \left. \times (v^{(1)i} - v_y^{(1)i}) + \frac{3}{4} v_j^{(1)} \Pi_y^{ji} \right], \end{aligned} \quad (199)$$

with  $\tau' = -\bar{n}_e \sigma_T a$ .

**6.4.4. Second-Order Momentum Continuity Equation for CDM.** Since CDM particles are collisionless, at first order, we find

$$\frac{\partial v_{\text{CDM}}^{(1)i}}{\partial \eta} + \mathcal{H} v_{\text{CDM}}^{(1)i} + \Phi^{(1),i} = 0. \quad (200)$$

At second order we find

$$\begin{aligned} & \frac{1}{2} \left( \frac{\partial v_{\text{CDM}}^{(2)i}}{\partial \eta} + \mathcal{H} v_{\text{CDM}}^{(2)i} + 2 \frac{\partial \omega^i}{\partial \eta} + 2 \mathcal{H} \omega_i + \Phi^{(2),i} \right) - \frac{\partial \Psi^{(1)}}{\partial \eta} v_{\text{CDM}}^{(1)i} \\ & + v_{\text{CDM}}^{(1)j} \partial_j v_{\text{CDM}}^{(1)i} + (\Phi^{(1)} + \Psi^{(1)}) \Phi^{(1),i} + \left( \frac{T_{\text{CDM}}}{m_{\text{CDM}}} \right)^i = 0. \end{aligned} \quad (201)$$

## 7. CMB Anisotropies at Second-Order at All Scales: Analytical Approach

**7.1. Towards a Second-Order CMB Radiation Transfer Function.** As pointed out in Section 6 various nonprimordial sources of non-Gaussianity for the CMB anisotropies can arise from the nonlinear evolution of the cosmological perturbations, including the Sunyaev-Z'eldovich, ISW (Rees-Sciama) effects, and the gravitational lensing, with possible correlations between these contributions. Here we will focus on another relevant source of non-Gaussianity: the nonlinear effects operating at the recombination epoch. The dynamics at recombination is quite involved because all the nonlinearities in the evolution of the baryon-photon fluid at recombination and the ones coming from general relativity should be accounted for.

The following section can be considered as an application of the Boltzmann equations found previously. Despite they depend on a high numbers of terms, because of the appearance within second-order perturbation theory of products of first-order perturbations, it is remarkable that an analytical study of the complicated recombination dynamics is possible. This allows to account for those effects that at the last scattering surface produce a non-Gaussian contribution

to the CMB anisotropies that add to the primordial one. Such a contribution is so relevant because it represents a major part of the second-order radiation transfer function which must be determined in order to have a complete control of both the primordial and nonprimordial part of NG in the CMB anisotropies and to gain from the theoretical side the same level of precision that could be reached experimentally in the near future [4, 5, 71].

In order to achieve this goal, we consider the Boltzmann equations derived in Section 6 at second-order describing the evolution of the photon, baryon and CDM fluids, and we manipulate them further under the assumption of tight coupling between photons and baryons. This leads to the generalization at second-order of the equations for the photon energy density and velocity perturbations which govern the acoustic oscillations of the photon-baryon fluid for modes that are inside the horizon at recombination. The evolution is that of a damped harmonic oscillator, with a source term which is given by the gravitational potentials generated by the different species. An interesting result is that, unlike the linear case, at second-order the quadrupole moment of the photons is not suppressed in the tight coupling limit and it must be taken into account.

The analytical solutions for the acoustic oscillations of the photon-baryon fluid at second order are derived adopting some simplifications which are also standard for an analytical treatment of the linear CMB anisotropies, and which nonetheless allow to catch most of the physics at recombination. One of these simplifications is to study separately two limiting regimes: intermediate scales which enter the horizon in between the equality epoch ( $\eta_{\text{eq}}$ ) and the recombination epoch ( $\eta_r$ ), with  $\eta_r^{-1} \ll k \ll \eta_{\text{eq}}^{-1}$ , and shortwave perturbations, with  $k \gg \eta_{\text{eq}}^{-1}$ , which enter the horizon before the equality epoch. Here, our main concern is to provide a simple estimate of the quantitative behaviour of the nonlinear evolution taking place at recombination, offering at the same time all the tools for a more accurate computation. We find that the second-order CMB anisotropies generated on the last scattering surface do not reduce only to the energy density and velocity perturbations of the photons evaluated at recombination, but a number of second-order corrections at last scattering arise from the Boltzmann equations of [46] in the form of products of first-order perturbations.

We will see that the dynamics at recombination is indeed dominated on small scales by the nonlinear evolution of the second-order gravitational potentials fed by the cold dark matter density perturbations. The gravitational potentials determine the energy density fluctuations of the photons at recombination and their effects show up in the CMB anisotropies as

$$\frac{\Delta T}{T} = \frac{1}{4} \Delta_{00}^{(2)} + \Phi^{(2)}, \quad (202)$$

which is the usual term due to the intrinsic fractional temperature fluctuation  $\Delta_{00}^{(2)}/4$  on the last scattering surface ( $\Delta_{00}^{(2)}$  is the monopole of the photon distribution function) and the gravitational redshift due to the gravitational potential. However, the analysis of the remaining contributions that

come in the form of products of first-order perturbations is equally important. The reason is that one of the central quantities we are interested in is the *contamination* to the primordial non-Gaussianity that is produced by the secondary effects. In that respect, the reasonable question to ask is which kind of primordial non-Gaussianity a given secondary effect can contaminate most, whether it is of the so called “local type” or of the “equilateral type”, for example. Therefore, it might well be the case that, even if some secondary effects appear to be the dominant ones, they might give a high contamination to a given type of primordial non-Gaussianity, but a low contamination to a different kind of primordial non-Gaussianity, for which “subdominant” terms, on the other hand, represent the strongest contaminant. This section and the following two provide a clear example: we anticipate here that while the dominant contribution on small scales to the term in (202) will mainly mimic an equilateral type of primordial non-Gaussianity, second-order effects that come as products of first-order times first-order perturbations actually mainly contaminate local primordial non-Gaussianity.

Notice that the case  $k \gg \eta_{\text{eq}}^{-1}$  has been treated in two steps. First, we just assume a radiation dominated universe, and then we provide a much more complete analysis by solving the evolution of the perturbations from the equality epoch onwards taking into account that the dark matter perturbations around the equality epoch tend to dominate the second-order gravitational potentials. As a byproduct, this last step provides the Meszaros effect at second-order. In deriving the analytical solutions, we have accurately accounted for the initial conditions set on superhorizon scales by the primordial non-Gaussianity. In fact, the primordial contribution is always transferred linearly, while the real new contribution to the radiation transfer function is given by all the additional terms provided in the source functions of the equations. Let us stress here that the analysis of the CMB bispectrum performed so far, as for example in [72], adopt just the linear radiation transfer function (unless the bispectrum originated by specific secondary effects, such as Rees-Sciama or Sunyaev-Zel’dovich effects, are considered, see on this Section 8). The formalism to provide in a systematic way the second-order CMB radiation transfer function will be reviewed in Section 10.

### 7.2. The Boltzmann Equations in the Tightly Coupled Limit.

We now derive the moments of the Boltzmann equations for photons in the limit when the photons are tightly coupled to the baryons (the electron-proton system) due to Compton scattering. This leads to the governing equations for the acoustic oscillations of the photon-baryon fluid. The well-known computation at linear order is briefly reviewed in (C.1) under some standard simplifying assumptions. Here, we will focus on the derivation of the equations at second-order in the perturbations, pointing out some interesting differences with respect to the linear case. In particular, note that, while we already know that the L.H.S. of the equations at second-order will have the same form as for the linear case, the source term on the R.H.S. of the moments of the

Boltzmann equations will also consist of first-order squared terms.

**7.2.1. Energy Continuity Equation.** Let us start by integrating (134) over  $d\Omega_n/4\pi$  to get the evolution equation for the second-order photon energy density perturbations  $\Delta_{00}^{(2)}$

$$\begin{aligned} \Delta_{00}^{(2)'} + \frac{4}{3}\partial_i v_\gamma^{(2)i} + \frac{8}{3}\partial_i (\Delta_{00}^{(1)} v_\gamma^{(1)i}) - 4\Psi^{(2)'} + \frac{8}{3}(\Phi^{(1)} + \Psi^{(1)}) \\ \times \partial_i v_\gamma^{(1)i} + 2 \int \frac{d\Omega_n}{4\pi} \\ \times \left[ (\Phi_{,j}^{(1)} + \Psi_{,j}^{(1)}) n^i n^j - (\Phi^{(1)} + \Psi^{(1)}) \right] \frac{\partial \Delta^{(1)}}{\partial n^i} \\ - 8\Psi^{(1)'} \Delta_{00}^{(1)} + \frac{32}{3}\Phi_{,i}^{(1)} v_\gamma^{(1)i} = -\frac{8}{3}\tau' v_i^{(1)} (v^{(1)i} - v_\gamma^{(1)i}), \end{aligned} \quad (203)$$

where we have used the explicit definition for the second-order velocity of the photons [46]

$$\frac{4}{3} \frac{v_\gamma^{(2)i}}{2} = \frac{1}{2} \int \frac{d\Omega}{4\pi} \Delta^{(2)} n^i - \frac{4}{3} \delta_\gamma^{(1)} v_\gamma^{(1)i}. \quad (204)$$

We can now make use of the tight coupling expansion to simplify (203). In the L.H.S., we use  $\partial_i v_\gamma^{(1)i} = \partial_i v^{(1)i} = 3\Psi^{(1)'} - \delta_b^{(1)'} = 3\Psi^{(1)'} - 3\Delta_{00}^{(1)'}/4$  obtained in the tightly coupled limit from (C.7) and (C.9). On the other hand, in the R.H.S. of (203)

$$\begin{aligned} (v^{(1)i} - v_\gamma^{(1)i}) &= \frac{R}{\tau'} (v_\gamma^{(1)i'} + \mathcal{H} v_\gamma^{(1)i} + \Phi^{(1),i}) \\ &= \frac{R}{\tau'} \left( \frac{\mathcal{H}}{1+R} v_\gamma^{(1)i} - \frac{1}{4} \frac{\Delta_{00}^{(1),i}}{1+R} \right), \end{aligned} \quad (205)$$

by using (C.4) and the evolution equation for the photon velocity (C.5). Here, we introduce the baryon-photon ratio

$$R = \frac{3\rho_b}{4\rho_\gamma}. \quad (206)$$

We thus arrive at the following equation:

$$\Delta_{00}^{(2)'} + \frac{4}{3}\partial_i v_\gamma^{(2)i} - 4\Psi^{(2)'} = \mathcal{S}_\Delta, \quad (207)$$

where the source term is given by

$$\begin{aligned} \mathcal{S}_\Delta &= (\Delta_{00}^{(12)})' - 2(\Phi^{(1)} + \Psi^{(1)}) (4\Psi^{(1)'} - \Delta_{00}^{(1)'}) \\ &\quad - \frac{8}{3} v_\gamma^{(1)i} (\Delta_{00}^{(1)} + 4\Phi^{(1)})_{,i} \\ &\quad + \frac{16}{3} (\Phi^{(1)} + \Psi^{(1)})^{,i} v_i \\ &\quad - \frac{8}{3} R \left( \frac{\mathcal{H}}{1+R} v_\gamma^{(1)2} - \frac{1}{4} \frac{v_{\gamma i}^{(1)} \Delta_{00}^{(1),i}}{1+R} \right). \end{aligned} \quad (208)$$

Notice that the integral in the second line of (203) has been computed by expanding the linear anisotropies as  $\Delta^{(1)} = \sum_{\ell} (2\ell + 1) \Delta_{\ell}^{(1)} P_{\ell}(\mathbf{v} \cdot \mathbf{n})$  and expressing explicitly the dependence on  $n^i$  as  $P_1(\mathbf{v} \cdot \mathbf{n}) = \hat{v}^i n_i$ ,  $P_2(\mathbf{v} \cdot \mathbf{n}) = (3(\mathbf{v} \cdot \mathbf{n})^2 - 1)/2$  and so on. This allows to perform the derivative  $\partial \Delta^{(1)}/\partial n^i$ . It turns out that the term proportional to the first-order quadrupole vanishes, and higher order terms can be neglected because of the tight coupling approximation. Therefore, the only term that is non negligible is the dipole term and using  $\Delta_1^{(1)} = 4\nu/3$  in the tight coupling limit, one obtains that the integral is equal to  $(16/3)(\Phi^{(1)} + \Psi^{(1)})^i \nu_i$ .

**7.2.2. Velocity Continuity Equation.** We now derive the second moment of the Boltzmann equation (134) and then we take its tight coupling limit. The integration of (134) over  $d\Omega_n n^i/4\pi$  yields the continuity equation for the photon velocity

$$\begin{aligned}
& \frac{4}{3} \frac{\nu_y^{(2)i'}}{2} + \frac{1}{2} \partial_j \Pi_y^{(2)ji} + \frac{1}{3} \frac{\Delta_{00}^{(2),i}}{2} + \frac{2}{3} \Phi^{(2),i} + \frac{4}{3} \omega^i \\
&= -\frac{4}{3} (\Delta_{00}^{(1)} \nu_y^{(1)i})' + \frac{16}{3} \Psi^{(1)'} \nu_y^{(1)i} - 4 \Phi_{,j}^{(1)} \Pi_y^{(1)ji} \\
&\quad - \frac{4}{3} \Phi^{(1),i} \Delta_{00}^{(1)} - (\Phi^{(1)} + \Psi^{(1)}) \partial_j \Pi_y^{(1)ji} \\
&\quad - \frac{1}{3} (\Phi^{(1)} + \Psi^{(1)}) \Delta_{00}^{(1),i} + \int \frac{d\Omega_{\mathbf{n}}}{4\pi} \\
&\quad \times n^i \left[ (\Phi_{,j}^{(1)} + \Psi_{,j}^{(1)}) n^j n^i - (\Phi^{,i} + \Psi^{,i}) \right] \frac{\partial \Delta^{(1)}}{\partial n^i} \\
&\quad - \frac{\tau'}{2} \left[ \frac{4}{3} (\nu^{(2)i} - \nu_y^{(2)i}) + \frac{8}{3} (\delta_b^{(1)} + \Phi^{(1)} + \Delta_{00}^{(1)}) \right. \\
&\quad \quad \left. \times (\nu^{(1)i} - \nu_y^{(1)}) + 2 \nu_j^{(1)} \Pi_y^{(1)ji} \right] \\
&\quad - \frac{4}{3} \Phi^{(1),i} (\Phi^{(1)} + \Psi^{(1)}). \tag{209}
\end{aligned}$$

The difference between the second-order baryon and photon velocities  $(\nu^{(2)i} - \nu_y^{(2)i})$  appearing in (209) is obtained from the baryon continuity equation which can be written as (see [46])

$$\begin{aligned}
\nu^{(2)i} &= \nu_y^{(2)i} + \frac{R}{\tau'} \left[ (\nu^{(2)i'}) + \mathcal{H} \nu^{(2)i} + 2\omega^i + 2\mathcal{H} \omega^i + \Phi^{(2),i} \right] \\
&\quad - 2\Psi^{(1)'} \nu^{(1)i} \partial_i \nu^{(1)2} + 2\Phi^{(1),i} (\Phi^{(1)} + \Psi^{(1)}) \\
&\quad - \frac{3}{2} \nu_j^{(1)} \Pi_y^{(1)ji} - 2(\Delta_{00}^{(1)} + \Phi^{(1)}) (\nu^{(1)i} - \nu_y^{(1)i}). \tag{210}
\end{aligned}$$

We want now to reduce (209) in the tightly coupled limit. We first insert the expression (210) in (209). Notice that the last three terms in (210) will cancel out. On the other hand, in the

tight coupling limit expansion one can set  $\nu^{(1)i} = \nu_y^{(1)i}$  and  $\nu^{(2)i} = \nu_y^{(2)i}$  in the remaining terms on the R.H.S. of (210). Thus, (209) becomes

$$\begin{aligned}
& (\nu_y^{(2)i} + 2\omega^i)' + \mathcal{H} \frac{R}{1+R} (\nu_y^{(2)i} + 2\omega^i) + \frac{1}{4} \frac{\Delta_{00}^{(2),i}}{1+R} + \Phi^{(2),i} \\
&= -\frac{3}{4(1+R)} \partial_j \Pi_y^{(2)ji} - \frac{2}{1+R} (\Delta_{00}^{(1)} \nu_y^{(1)i})' + \frac{8}{1+R} \Psi^{(1)'} \nu_y^{(1)i} \\
&\quad - \frac{2}{1+R} \Phi^{(1),i} \Delta_{00}^{(1)} - \frac{2}{1+R} \Phi^{(1),i} (\Phi^{(1)} + \Psi^{(1)}) \\
&\quad + 2 \frac{R}{1+R} \Psi^{(1)'} \nu_y^{(1)i} - \frac{1}{2(1+R)} (\Phi^{(1)} + \Psi^{(1)}) \Delta_{00}^{(1),i} \\
&\quad - \frac{R}{1+R} \partial^i \nu_y^{(1)2} - 2 \frac{R}{1+R} (\Phi^{(1)} + \Psi^{(1)}) \Phi^{(1),i} \\
&\quad - \tau' \frac{2}{1+R} \delta_b^{(1)} (\nu^{(1)i} - \nu_y^{(1)i}), \tag{211}
\end{aligned}$$

where in the tightly coupled limit, we are neglecting the first-order quadrupole and (higher-order moments) of the photon distribution since it is suppressed by  $1/\tau$  with respect to the other terms. Moreover, the integral in the fourth line of (209) has been computed following the same steps used for the analogous integral in (203). In this case, it turns out that the contribution from the first-order dipole term vanishes, while the one from the quadrupole is non zero, but this and higher-order terms can be neglected because of the tight coupling approximation, so that the integral is in fact negligible. Next, for the term like  $\tau' \delta_b^{(1)} (\nu^{(1)i} - \nu_y^{(1)i})$  we employ the relation previously derived in (99) with  $\delta_b^{(1)} = 3\Delta_{00}^{(1)}/4$  and we use the first order tight coupling equations (C.1) and (C.5) in order to further simplify (211). We finally obtain

$$\nu_y^{(2)i'} + \mathcal{H} \frac{R}{1+R} \nu_y^{(2)i} + \frac{1}{4} \frac{\Delta_{00}^{(2),i}}{1+R} + \Phi^{(2),i} = \mathcal{S}_V^i, \tag{212}$$

where

$$\begin{aligned}
\mathcal{S}_V^i &= -\frac{3}{4(1+R)} \partial_j \Pi_y^{(2)ji} - 2\omega_i' - 2\mathcal{H} \frac{R}{1+R} \omega^i \\
&\quad + 2 \frac{\mathcal{H}R}{(1+R)^2} \Delta_{00}^{(1)} \nu_y^{(1)i} + \frac{1}{4(1+R)^2} (\Delta_{00}^{(1)2})^i \\
&\quad + \frac{8}{3(1+R)} \nu_y^{(1)i} \partial_j \nu_y^{(1)j} + 2 \frac{R}{1+R} \Psi^{(1)'} \nu_y^{(1)i} \\
&\quad - 2(\Phi^{(1)} + \Psi^{(1)}) \Phi^{(1),i} - \frac{1}{2(1+R)} (\Phi^{(1)} + \Psi^{(1)}) \Delta_{00}^{(1),i} \\
&\quad - \frac{R}{1+R} \partial^i \nu_y^{(1)2} - \frac{3}{2} \frac{R}{1+R} \Delta_{00}^{(1)} \left( \frac{\mathcal{H}}{1+R} \nu_y^{(1)i} - \frac{1}{4} \frac{\Delta_{00}^{(1),i}}{1+R} \right). \tag{213}
\end{aligned}$$

We have spent some time in giving the details of the computation for the photon Boltzmann equations at second-order in the perturbations. As a summary of the results obtained so far, we refer the reader to (207) and (212) as our master equations which we will solve in the next sections. In particular (212) is the second-order counterpart of (C.5) for the photon velocity in the tight coupling regime. Notice that there are two important differences with respect to the linear case. One is that, in (212), there will be a contribution not only from scalar perturbations but also from vector modes which, at second-order, are inevitably generated as nonlinear combinations of first-order scalar perturbations. In particular, we have included the vector metric perturbations  $\omega^i$  in the source term. Second, and most important, we have also kept in the source term the second-order quadrupole of the photon distribution  $\Pi_y^{(2)ij}$ . At linear order we can neglect it together with higher order moments of the photons since they turn out to be suppressed with respect to the first two moments in the tight coupling limit by increasing powers of  $1/\tau$ . However, in the next section we will show that at second order this does not hold anymore, as the photon quadrupole is no longer suppressed.

Finally, following the same steps leading to (C.6) at linear order we can derive a similar equation for the second-order photon energy density perturbation  $\Delta_{00}^{(2)}$  which now will be characterized by the source terms  $\mathcal{S}_\Delta$  and  $\mathcal{S}_V^i$

$$\begin{aligned} & (\Delta_{00}^{(2)''} - 4\Psi^{(2)''}) + \mathcal{H} \frac{R}{1+R} (\Delta_{00}^{(2)'} - 4\Psi^{(2)'}) \\ & - c_s^2 \nabla^2 (\Delta_{00}^{(2)} - 4\Psi^{(2)}) \\ & = \frac{4}{3} \nabla^2 \left( \Phi^{(2)} + \frac{\Psi^{(2)}}{1+R} \right) + \mathcal{S}'_\Delta + \mathcal{H} \frac{R}{1+R} \mathcal{S}_\Delta - \frac{4}{3} \partial_i \mathcal{S}_V^i, \end{aligned} \quad (214)$$

where we have introduced the photon-baryon fluid sound speed  $c_s = 1/\sqrt{3(1+R)}$ .

**7.2.3. Second-Order Quadrupole Moment of the Photons in the Tight Coupling Limit.** Let us now consider the quadrupole moment of the photon distribution defined in (195) and show that at second-order it cannot be neglected in the tightly coupled limit, unlike for the linear case. We first integrate the R.H.S. of (134) over  $d\Omega_n (n^i n^j - \delta^{ij}/3)/4\pi$  and then we set it to be vanishing in the limit of tight coupling.

The integration involves various pieces to compute. For clarity we will consider each of them separately. The term  $\Delta_{00}^{(2)}$  does not contribute. For the third term we can write, from (135)

$$\begin{aligned} & - \sum_{m=-2}^{m=2} \frac{\sqrt{\pi}}{5^{3/2}} \Delta_{2m}^{(2)} Y_{2m} \\ & = \frac{\Delta^{(2)}}{10} - \frac{1}{10} \sum_{\ell \neq 2} \sum_{m=-\ell}^{\ell} \Delta_{\ell m}^{(2)} (-i)^\ell \sqrt{\frac{4\pi}{2\ell+1}} Y_{\ell m}, \end{aligned} \quad (215)$$

so that the integral just brings  $\Pi_y^{(2)ij}/10$ , since the only contribution in (215) comes from  $\Delta^{(2)}/10$  with all the other terms vanishing. The following nontrivial integral is

$$\begin{aligned} & \int \frac{d\Omega}{4\pi} \left( n^i n^j - \frac{1}{3} \delta^{ij} \right) \Delta_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \\ & = \hat{v}_k \hat{v}_l \int \frac{d\Omega}{4\pi} \left( n^i n^j - \frac{1}{3} \delta^{ij} \right) \Delta_2^{(1)} \times \left( \frac{3}{2} n^k n^l - \frac{1}{2} \right) \\ & = \frac{\Delta_2^{(1)}}{5} \left( \hat{v}^i \hat{v}^j - \frac{1}{3} \delta^{ij} \right), \end{aligned} \quad (216)$$

where the baryon velocity appearing in  $P_2(\hat{\mathbf{v}} \cdot \mathbf{n})$  is first order and we make use of the relations (168) together with

$$\int \frac{d\Omega}{4\pi} n^i n^j n^k n^l = \frac{1}{15} (\delta^{ij} \delta^{kl} + \delta^{ik} \delta^{lj} + \delta^{il} \delta^{jk}). \quad (217)$$

The integrals of  $\delta_e^{(1)} \Delta_0^{(1)}$ ,  $\delta_e^{(1)} (\mathbf{v} \cdot \mathbf{n})$  and  $\mathbf{v}^{(2)} \cdot \mathbf{n}$  vanish and

$$\begin{aligned} & v \Delta_1^{(1)} \int \frac{d\Omega}{4\pi} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \left( n^i n^j - \frac{1}{3} \delta^{ij} \right) \\ & = \frac{1}{5} v \Delta_1^{(1)} \left( \hat{v}^i \hat{v}^j - \frac{1}{3} \delta^{ij} \right) = \frac{4}{15} \left( v^i v^j - \frac{1}{3} \delta^{ij} v^2 \right), \end{aligned} \quad (218)$$

where in the last step, we take  $\Delta_1^{(1)} = 4v/3$  in the tight coupling limit. Similarly, the integral of  $14(\mathbf{v} \cdot \mathbf{n})^2$  brings

$$14v^k v^\ell \int \frac{d\Omega}{4\pi} n_k n_\ell \left( n^i n^j - \frac{1}{3} \delta^{ij} \right) = \frac{28}{15} \left( v^i v^j - \frac{1}{3} \delta^{ij} v^2 \right). \quad (219)$$

The integral of  $2(\mathbf{v} \cdot \mathbf{n}) \Delta^{(1)}$  can be performed by expanding the linear anisotropies as  $\Delta^{(1)} = \sum_\ell (2\ell+1) \Delta_\ell^{(1)} P_\ell(\hat{\mathbf{v}} \cdot \mathbf{n})$ . We thus find

$$\begin{aligned} & v^k \hat{v}^m \int \frac{d\Omega}{4\pi} n_k \left( n^i n^j - \frac{1}{3} \delta^{ij} \right) n_m \Delta_1^{(1)} + \mathcal{O}_{\ell>2} \\ & = \frac{8}{15} \left( v^i v^j - \frac{1}{3} \delta^{ij} v^2 \right), \end{aligned} \quad (220)$$

where we have used (168) and  $\mathcal{O}_{\ell>2}$  indicates all the integrals coming from the multipoles  $\ell > 2$  in the expansion (for  $\ell = 0$  and  $\ell = 2$  they vanish.) In fact we have dropped the  $\mathcal{O}(\ell > 2)$  since they are proportional to first-order photon moments  $\ell > 2$  which turn out to be suppressed in the tight coupling limit. Finally, the term proportional to  $(\mathbf{v} \cdot \mathbf{n}) \Delta_2^{(1)} (1 - P_2(\hat{\mathbf{v}} \cdot \mathbf{n}))/5$  gives a vanishing contribution.

Collecting all the various pieces we find that the third moment of the R.H.S. of (134) is given by

$$-\frac{\tau'}{2} \left[ -\Pi_y^{(2)ij} + \frac{1}{10} \Pi_y^{(2)ij} + 2\delta_e^{(1)} \left( -\Pi_y^{(1)ij} + \frac{1}{10} \Delta_2^{(1)} \right) \times \left( \hat{v}^i \hat{v}^j - \frac{1}{3} \delta^{ij} \right) + \frac{12}{5} \left( v^i v^j - \frac{1}{3} \delta^{ij} v^2 \right) \right]. \quad (221)$$

Therefore, in the limit of tight coupling, when the interaction rate is very high, the second-order quadrupole moment turns out to be

$$\Pi_y^{(2)ij} \simeq \frac{8}{3} \left( v^i v^j - \frac{1}{3} \delta^{ij} v^2 \right), \quad (222)$$

by setting (221) to be vanishing (the term multiplying  $\delta_e^{(1)}$  goes to zero in the tight coupling limit since it just comes from the first-order collision term). At linear order, one would simply get the term  $9\tau' \Pi_y^{(1)ij}/10$  implying that, in the limit of a high scattering rate  $\tau'$ ,  $\Pi_y^{(1)ij}$  goes to zero. However at second-order the quadrupole is not suppressed in the tight coupling limit because it turns out to be sourced by the linear velocity squared. Indeed (222) has a very clear physical interpretation. If one describes the photons as a perfect fluid, at second-order in the perturbations the quadrupole one would get (by taking the traceless part of the energy momentum tensor) would be exactly given by (222) (see, e.g., [33]). Of course, this does not come as a surprise here, because the tight coupling limit exactly corresponds to a regime where the photons behave as a perfect fluid.

**7.3. Second-Order CMB Anisotropies Generated at Recombination.** The previous equations allow us to follow the evolution of the monopole and dipole of CMB photons at recombination. As at linear order, they will appear in the expression for the CMB anisotropies today  $\Delta^{(2)}(k, n, \eta_0)$  together with various integrated effects. Our focus now will be to obtain an expression for the second-order CMB anisotropies today  $\Delta^{(2)}(k, n, \eta_0)$  from which we can extract all those contributions generated specifically at recombination due to the nonlinear dynamics of the photon-baryon fluid. This expression will not only relate the moments  $\Delta_{\ell m}^{(2)}$  today to the second-order monopole and dipole at recombination as it happens at linear order, but one has to properly account also for additional first-order squared contributions. Let us see how to achieve this goal in some details.

As we have seen previously, it is possible to write down an integral solution of the photon Boltzmann equation (134) in Fourier space. We wrote

$$\Delta^{(2)'} + ik\mu\Delta^{(2)} - \tau'\Delta^{(2)} = e^{-ik\mu+\tau} \frac{d}{d\eta} \left[ \Delta^{(2)} e^{ik\mu\eta-\tau} \right] \quad (223)$$

$$= S(\mathbf{k}, \mathbf{n}, \eta)$$

in order to derive a solution of the form

$$\Delta^{(2)}(\mathbf{k}, \mathbf{n}, \eta_0) = \int_0^{\eta_0} d\eta S(\mathbf{k}, \mathbf{n}, \eta) e^{ik\mu(\eta-\eta_0)} e^{-\tau}. \quad (224)$$

Here,  $\mu = \cos\vartheta = \hat{\mathbf{k}} \cdot \mathbf{n}$  is the polar angle of the photon momentum in a coordinate system such that  $\mathbf{e}_3 = \hat{\mathbf{k}}$ . At second-order the source term has been computed in [46] and can be read off (134) and (141) to be

$$S = -\tau' \Delta_{00}^{(2)} - 4n^i \Phi_{,i}^{(2)} + 4\Psi^{(2)'} - 8\omega'_i n^i - 4\chi'_{ij} n^i n^j$$

$$+ 8\Delta^{(1)} \left( \Psi^{(1)'} - n^i \Phi_{,i}^{(1)} \right) - 2n^i \left( \Phi^{(1)} + \Psi^{(1)} \right) \left( \Delta^{(1)} + 4\Phi^{(1)} \right)_{,i}$$

$$- 2 \left[ \left( \Phi^{(1)} + \Psi^{(1)} \right)_{,j} n^i n^j - \left( \Phi^{(1)} + \Psi^{(1)} \right)^{,i} \right] \frac{\partial \Delta^{(1)}}{\partial n^i}$$

$$- \tau' \left[ -\frac{1}{2} \sum_{m=-2}^2 \frac{\sqrt{4\pi}}{5^{3/2}} \Delta_{2m}^{(2)} Y_{2m}(\mathbf{n}) + 4\mathbf{v}^{(2)} \cdot \mathbf{n} \right.$$

$$+ 2\delta_e^{(1)} \left( \Delta_0^{(1)} - \Delta^{(1)} + 4\mathbf{v} \cdot \mathbf{n} + \frac{1}{2} \Delta_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right)$$

$$+ 2(\mathbf{v} \cdot \mathbf{n}) \left[ \Delta^{(1)} + 3\Delta_0^{(1)} - \Delta_2^{(1)} \left( 1 - \frac{5}{2} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right) \right]$$

$$\left. - v\Delta_1^{(1)} (4 + 2P_2(\hat{\mathbf{v}} \cdot \mathbf{n})) + 14(\mathbf{v} \cdot \mathbf{n})^2 - 2v^2 \right]. \quad (225)$$

The key point here is to isolate all those terms that multiply the differential optical depth  $\tau'$ . The reason is that in this case in the integral (224) one recognizes the visibility function  $g(\eta) = -e^{-\tau}\tau'$  which is sharply peaked at the time of recombination and whose integral over time is normalized to unity. Thus, for these terms the integral just reduces to the remaining integrand (apart from the visibility function) evaluated at recombination. The standard example that one encounters also at linear order is given by the first term appearing in the source  $S$ , (225), that is  $-\tau'\Delta_{00}^{(2)}$ . The contribution of this term to the integral (224) just reduces to

$$\Delta^{(2)}(\mathbf{k}, \mathbf{n}, \eta_0)$$

$$= \int_0^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)} e^{-\tau} (-\tau') \Delta_{00}^{(2)} \simeq e^{ik\mu(\eta_*-\eta_0)} \Delta_{00}^{(2)}(\eta_*), \quad (226)$$

where  $\eta_*$  is the epoch of recombination and, in the multipole decomposition (136), (226) brings the standard result

$$\Delta_{\ell m}^{(2)}(\eta_0) \propto \Delta_{00}^{(2)}(\eta_*) j_\ell(k(\eta_* - \eta_0)), \quad (227)$$

having used the Legendre expansion  $e^{i\mathbf{k} \cdot \mathbf{x}} = \sum_\ell (i)^\ell (2\ell + 1) j_\ell(kx) P_\ell(\hat{\mathbf{k}} \cdot \hat{\mathbf{x}})$ . In (227), the monopole at recombination is found by solving the Boltzmann equations (207)–(212) derived in tight coupling limit.

Looking at (225) we recognize immediately some terms which multiply explicitly  $\tau'$  (the first one discussed in the example above and the last two lines of (225)). However, it is easy to realize from the standard procedure adopted at the linear-order that such terms are not the only ones. This is clear by focusing, as an example, on the term  $-4n^i\Phi_{,i}^{(2)}$  in the source  $S$  which appears in the same form also at linear order. In Fourier space one can replace the angle  $\mu$  with a time derivative and thus this term gives rise to

$$\begin{aligned} & -4ik \int_0^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)} e^{-\tau} \mu \Phi^{(2)} \\ &= -4 \int_0^{\eta_0} d\eta \Phi^{(2)} e^{-\tau} \frac{d}{d\eta} \left( e^{ik\mu(\eta-\eta_0)} \right) \\ &= 4 \int_0^{\eta_0} d\eta e^{ik\mu(\eta-\eta_0)} e^{-\tau} \left( \Phi^{(2)'} - \tau' \Phi^{(2)} \right), \end{aligned} \quad (228)$$

where, in the last step, we have integrated by parts. In (228) the time derivative of the gravitational potential contributes to the Integrated Sachs-Wolfe effect, but also also a  $\tau'$  results implying that we have also to evaluate  $\Phi^{(2)}$  at recombination. Thus, in the following, we look for those terms in the source (225) which give rise to a  $\tau'$  factor in the same way as for  $-4n^i\Phi_{,i}^{(2)}$ . In particular, let us consider the combination in(225)

$$\begin{aligned} C &\equiv 8\Delta^{(1)} \left( \Psi^{(1)'} - n^i \Phi_{,i}^{(1)} \right) - 2n^i \left( \Phi^{(1)} + \Psi^{(1)} \right) \left( \Delta^{(1)} + 4\Phi^{(1)} \right)_{,i} \\ &= 8\Delta^{(1)} \Psi^{(1)'} - 8n^i \left( \Delta^{(1)} \Phi^{(1)} \right)_{,i} + 4\Phi^{(1)} n^i \Delta_{,i}^{(1)} - 8n^i \left( \Phi^{(1)2} \right)_{,i}, \end{aligned} \quad (229)$$

where for simplicity we are setting  $\Phi^{(1)} \simeq \Psi^{(1)}$ . We already recognize terms of the form  $n^i \partial_i(\cdot)$ . Moreover we can use the Boltzmann equation (129) to replace  $n^i \Delta_{,i}^{(1)}$  in (95). This brings

$$\begin{aligned} C &= 8\Delta^{(1)} \Psi^{(1)'} - 4\Psi^{(1)} \Delta^{(1)'} - 8 \left( \Psi^{(1)2} \right)' \\ &\quad - 8n^i \left( \Delta^{(1)} \Phi^{(1)} + 2\Phi^{(1)2} \right)_{,i} \\ &\quad - 4\tau' \Psi^{(1)} \left[ \Delta_{00}^{(1)} - \Delta^{(1)} + 4\mathbf{v}^{(1)} \cdot \mathbf{n} + \frac{1}{2} \Delta_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right]. \end{aligned} \quad (230)$$

In fact, we will not be interested for our purposes in the first three terms of (230), since they will not contribute to the anisotropies generated at recombination.

Therefore, as a result of (225), (228) and (230), we can rewrite the source term (225) as

$$S = S_* + S' \quad (231)$$

where

$$\begin{aligned} S_* &= -\tau' \left[ \Delta_{00}^{(2)} + 4\Phi^{(2)} - \frac{1}{2} \sum_{m=-2}^2 \frac{\sqrt{4\pi}}{5^{3/2}} \Delta_{2m}^{(2)} Y_{2m}(\mathbf{n}) \right. \\ &\quad \left. + 4\mathbf{v}^{(2)} \cdot \mathbf{n} + 2\delta_e^{(1)} \right] \\ &\quad \times \left( \Delta_0^{(1)} - \Delta^{(1)} + 4\mathbf{v} \cdot \mathbf{n} + \frac{1}{2} \Delta_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right) \\ &\quad + 2(\mathbf{v} \cdot \mathbf{n}) \left[ \Delta^{(1)} + 3\Delta_0^{(1)} - \Delta_2^{(1)} \left( 1 - \frac{5}{2} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right) \right] \\ &\quad - \nu \Delta_1^{(1)} (4 + 2P_2(\hat{\mathbf{v}} \cdot \mathbf{n})) + 14(\mathbf{v} \cdot \mathbf{n})^2 \\ &\quad - 2\nu^2 + 8\Delta^{(1)} \Phi^{(1)} + 16\Phi^{(1)2} + 4\Psi^{(1)} \\ &\quad \times \left[ \Delta_0^{(1)} - \Delta^{(1)} + 4\mathbf{v}^{(1)} \cdot \mathbf{n} + \frac{1}{2} \Delta_2^{(1)} P_2(\hat{\mathbf{v}} \cdot \mathbf{n}) \right], \end{aligned} \quad (232)$$

$$\begin{aligned} S' &= 4 \left( \Phi^{(2)} + \Psi^{(2)} \right)' - 8\omega'_{,i} n^i - 4\chi'_{ij} n^i n^j \\ &\quad - 2 \left[ \left( \Phi^{(1)} + \Psi^{(1)} \right)_{,j} n^i n^j - \left( \Phi^{(1)} + \Psi^{(1)} \right)^{,i} \right] \frac{\partial \Delta^{(1)}}{\partial n^i} \\ &\quad + 8 \left( \Delta^{(1)} \Phi^{(1)} \right)' + 8\Delta^{(1)} \Psi^{(1)'} - 4\Psi^{(1)} \Delta^{(1)} + 16\Psi^{(1)} \Psi^{(1)'} \end{aligned} \quad (233)$$

In (231),  $S_*$  contains the contribution to the second-order CMB anisotropies created on the last scattering surface at recombination, while  $S'$  includes all those effects which are integrated in time from the last scattering surface up to now, including the second-order integrated Sachs-Wolfe effect and the second-order lensing effect. Since the main concern of this Section is the CMB anisotropies generated at last scattering, from now on we will focus only on the contribution from the last scattering surface  $S_*$ . In particular, notice that following the same steps that lead to (226), the first two terms of (232) give rise to the CMB temperature anisotropies as written in (202).

**7.4. Tightly Coupled Solutions for the Second-Order Perturbations.** In this section, we will solve the tightly coupled limit of the Boltzmann equations (207) and (209) at second-order in perturbation theory. We will proceed as for the linear case, focusing on the two limiting cases of perturbation modes entering the horizon, respectively, much before and much after the time of equality. The solution of (214) can be

written as

$$\begin{aligned}
& [1 + R(\eta)]^{1/4} (\Delta_{00}^{(2)} - 4\Psi^{(2)}) \\
& = A \cos[kr_s(\eta)] + B \sin[kr_s(\eta)] \\
& \quad - 4 \frac{k}{\sqrt{3}} \int_0^\eta d\eta' [1 + R(\eta')]^{3/4} \\
& \quad \times \left( \Phi^{(2)}(\eta') + \frac{\Psi^{(2)}(\eta')}{1 + R} \right) \sin_k[\eta, \eta'] \\
& \quad + \frac{\sqrt{3}}{k} \int_0^\eta d\eta' [1 + R(\eta')]^{3/4} \\
& \quad \times \left( \mathcal{H}'_\Delta + \frac{\mathcal{H}R}{1 + R} \mathcal{H}_\Delta - \frac{4}{3} ik_i \mathcal{H}_V^i \right) \sin_k[\eta, \eta'],
\end{aligned} \tag{234}$$

where for simplicity, we use the shorthand notation  $\sin_k[\eta, \eta'] \equiv \sin[k(r_s(\eta) - r_s(\eta'))]$ , and  $r_s(\eta)$  is the sound horizon

$$r_s(\eta) = \int_0^\eta d\eta' c_s(\eta'). \tag{235}$$

The source terms are given in (208) and (213). Notice that we can write  $\mathcal{H}'_\Delta + (\mathcal{H}R/(1 + R))\mathcal{H}_\Delta = (\mathcal{H}_\Delta(1 + R))'/1 + R$  so that we can perform an integration by parts in (234) leading to

$$\begin{aligned}
& [1 + R(\eta)]^{1/4} (\Delta_{00}^{(2)} - 4\Psi^{(2)}) \\
& = A \cos[kr_s(\eta)] + B \sin[kr_s(\eta)] \\
& \quad - 4 \frac{k}{\sqrt{3}} \int_0^\eta d\eta' [1 + R(\eta')]^{3/4} \left( \Phi^{(2)}(\eta') + \frac{\Psi^{(2)}(\eta')}{1 + R} \right) \\
& \quad \times \sin_k[\eta, \eta'] + \int_0^\eta d\eta' \mathcal{H}'_{\Delta}(\eta') (1 + R(\eta'))^{1/4} \cos_k[\eta, \eta'] \\
& \quad - \frac{\sqrt{3}}{k} \mathcal{H}_\Delta(0) \sin[kr_s(\eta)] - \frac{4}{\sqrt{3}} \frac{ik_i}{k} \int_0^\eta d\eta' \mathcal{H}_V^i(\eta') \\
& \quad \times (1 + R(\eta'))^{3/4} \sin_k[\eta, \eta'] + \frac{\sqrt{3}}{4k} \int_0^\eta d\eta' \mathcal{H}'_{\Delta}(\eta') \\
& \quad \times (1 + R(\eta'))^{-1/4} R'(\eta') \sin_k[\eta, \eta'].
\end{aligned} \tag{236}$$

In order to give an analytical solution that catches most of the physics underlying (234) and which remains at the same time very simple to treat, we will make some simplifications which are also usually adopted in linear theory, see for example, [73]. In particular, we will treat  $R = R_*$  as a constant evaluated at the time of recombination. In this way, with the presence of  $R$  in the varying cosines and sines, we can keep track of a damping of the photon velocity amplitude with respect to the case  $R = 0$  which prevents the acoustic peaks in the power-spectrum to disappear. Treating  $R$  as a constant is justified by the fact that for modes within the horizon the time scale of the oscillations is much shorter than the time

scale on which  $R$  varies. If  $R$  is a constant the sound speed is just a constant

$$c_s = \frac{1}{\sqrt{3(1 + R_*)}}, \tag{237}$$

and the sound horizon is simply  $r_s(\eta) = c_s \eta$ .

There is also another good reason not to forget about  $R$ , even though  $R < 1$ , which is particularly relevant when estimating the second-order contributions to the CMB anisotropies in (202). As we will see in Section 9, that combination on small scales comes actually as proportional to  $R$ , so that neglecting  $R$  would lead to miss one of the most significant CMB contribution at second order.

Another simplification consists in solving the evolutions of the perturbations in two well distinguished limiting regimes. One regime is for those perturbations which enter the Hubble radius when matter is the dominant component, that is at times much bigger than the equality epoch, with  $k \ll k_{\text{eq}} \sim \eta_{\text{eq}}^{-1}$ , where  $k_{\text{eq}}$  is the wavenumber of the Hubble radius at the equality epoch. The other regime is for those perturbations with much smaller wavelengths which enter the Hubble radius when the universe is still radiation dominated, that is perturbations with wavenumbers  $k \gg k_{\text{eq}} \sim \eta_{\text{eq}}^{-1}$ . In fact, we are interested in perturbation modes which are within the horizon by the time of recombination  $\eta_*$ . Therefore, we will further suppose that  $\eta_* \gg \eta_{\text{eq}}$  in order to study such modes in the first regime. Even though  $\eta_* \gg \eta_{\text{eq}}$  is not the real case, it allows to obtain some analytical expressions.

*7.4.1. Setting the Initial Conditions: Primordial Non-Gaussianity.* The integration constants  $A$  and  $B$  are fixed according to the initial conditions for the second-order cosmological perturbations. These refer to the values of the perturbations on superhorizon scales deep in the radiation dominated period. We will consider the case of initial adiabatic perturbations, for which there exist some useful conserved quantities on large scales which as such carry directly the information about the initial conditions.

As explained in Section 4, such a conserved quantity is given by the curvature perturbation  $\zeta$  and its conserved value allows to set the initial conditions for the metric and matter perturbations accounting for the primordial contributions. At linear order during the radiation-dominated epoch and on large scales  $\zeta^{(1)} = -2\Psi^{(1)}/3$ . On the other hand, after some calculations, one can easily compute  $\Delta\zeta^{(2)}$  for a radiation dominated epoch

$$\Delta\zeta^{(2)} = \frac{7}{2} (\Psi^{(1)})^2, \tag{238}$$

where in (37) one uses that on large scales  $\delta^{(1)}\rho_\gamma/\rho_\gamma = -2\Psi^{(1)}$  and the energy continuity equation  $\delta^{(1)'}\rho_\gamma + 4\mathcal{H}\delta^{(1)}\rho_\gamma - 4\Psi^{(1)'}\rho_\gamma = 0$ . Therefore, we find

$$\zeta^{(2)} = -\Psi^{(2)} + \frac{\Delta_{00}^{(2)}}{4} + \frac{7}{2}\Psi^{(1)2}(0), \tag{239}$$

where we are evaluating the quantities in the large scale limit for  $\eta \rightarrow 0$ . Using the parametrization (38) at the initial times the quantity  $\Delta_{00}^{(2)} - 4\Psi^{(2)}$  is given by

$$\Delta_{00}^{(2)} - 4\Psi^{(2)} = 2(9a_{\text{NL}} - 7)\Psi^{(1)2}(0). \quad (240)$$

Since for adiabatic perturbations such a quantity is conserved on superhorizon scales, it follows that the constant  $B = 0$  and  $A = 2(9a_{\text{NL}} - 7)\Psi^{(1)2}(0)$ .

Equations (234) and (236) are analytical expressions describing the acoustic oscillations of the photon-baryon fluid induced at second-order for perturbation modes within the horizon at recombination. In the following, we will adopt similar simplifications already used for the linear case in order to provide some analytical solutions. In particular, if in (236) we treat  $R$  as a constant we can write, using the initial conditions determined above,

$$\begin{aligned} & (\Delta_{00}^{(2)} - 4\Psi^{(2)}) \\ &= 2(9a_{\text{NL}} - 7)\Psi^{(1)2}(0) \cos[kr_s(\eta)] \\ & \quad - 4kc_s \int_0^\eta d\eta' \left( \Phi^{(2)}(\eta') + \Psi^{(2)}(\eta') + R\Phi^{(2)} \right) \sin k[\eta, \eta'] \\ & \quad + \int_0^\eta d\eta' \delta_\Delta(\eta') \cos k[\eta, \eta'] - \frac{\sqrt{3}}{k} \delta_\Delta(0) \sin[kr_s(\eta)] \\ & \quad - \frac{4}{3} \frac{ik_i}{kc_s} \int_0^\eta d\eta' \delta_V^i(\eta') \sin k[\eta, \eta']. \end{aligned} \quad (241)$$

**7.5. Perturbation Modes with  $k \gg k_{\text{eq}}$ .** In order to study the contribution to the second-order CMB anisotropies coming from perturbation modes that enter the horizon during the radiation dominated epoch, we will assume that the second-order gravitational potentials are the ones of a pure radiation dominated universe throughout the evolution. Though not strictly correct, this approximation will give us the basic picture of the acoustic oscillations for the baryon-photon fluid occurring for these modes. Also for the second-order case, in Section 7.7 we will provide the appropriate corrections accounting for the transition from radiation to matter domination which is indeed (almost) achieved by the recombination epoch. Before moving into the details a note of caution is in order here. At second order in the perturbations all the relevant quantities are expressed as convolutions of linear perturbations, bringing to a mode-mode mixing. In some cases, in our treatment for a given regime under analysis ( $k \gg k_{\text{eq}}$  or  $k \ll k_{\text{eq}}$ ) we use for the first-order perturbations the solutions corresponding to that particular regime, while the mode-mode mixing would require to consider in the convolutions (where one is integrating over all the wavenumbers) a more general expression for the first-order perturbations (which analytically does not exist anyway). For the computation of the CMB bispectrum this would be equivalent to consider just some specific scales, that is, all the three scales involved in the bispectrum should correspond approximately to wavenumbers  $k \gg k_{\text{eq}}$  or  $k \ll k_{\text{eq}}$ , and not a combination of

the two regimes (a step towards the evaluation of the three-point correlation function has been taken on [45] where it was computed in the in so-called squeezed triangle limit, when one mode has a wavelength much larger than the other two and is outside the horizon).

In Appendix C, we show how to solve the Boltzmann equations at the linear level in the various regimes. In fact, for  $k \gg k_{\text{eq}}$  the acoustic oscillations can be solved in an alternative way (see [47]) for this procedure at linear order. One can start directly from (207), where we can neglect the gravitational potential term  $\Psi^{(2)}$ . The reason is that, as it happens at linear order, the second-order gravitational potentials decay at late times as  $\eta^{-2}$ , while the second-order velocity  $v_y^{(2)i}$  oscillates in time. Let us now see that in some details.

The evolution equation for the gravitational potential  $\Psi^{(2)}$  is given by (B.13) and is characterized by the source term  $S_y$ , (B.15). In particular, the source term contains the second-order quadrupole moment of the photons  $\Pi_y^{(2)ij}$ . We saw in Section 7.2.3 that at second-order the quadrupole moment is not suppressed in the tight coupling limit, being fed by the nonlinear combination of the first-order velocities, (222). For the perturbation modes, we are considering here the velocity at late times is oscillating being given in Appendix C in Fourier space. Since the linear gravitational potential decays in time and for a radiation dominated period  $\mathcal{H} = 1/\eta$ , it is easy to check that the dominant contribution at late times to the source term  $S_y$  simply reduces to

$$\begin{aligned} S_y &\simeq \frac{3}{2} \mathcal{H}^2 \frac{\partial_i \partial^j}{\nabla^2} \Pi_{yj}^{(2)i} \\ &\equiv \frac{F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})}{\eta^2} C \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0) \sin(k_1 c_s \eta) \sin(k_2 c_s \eta), \end{aligned} \quad (242)$$

where

$$C = -\frac{9}{c_s^2 k_1 k_2} \quad (243)$$

and the sound speed is  $c_s = 1/\sqrt{3(1+R)}$ . Before proceeding further let us explain the notation that we are using. The equivalence symbol will be used to indicate that we are evaluating the expression in Fourier space. At second-order in perturbation theory most of the Fourier transforms reduce to some convolutions. We will not indicate these convolutions explicitly but just through their kernel. For example in (242) by  $F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})$ , we actually indicate the convolution operator

$$F \equiv \frac{1}{2\pi^3} \int d^3 k_1 d^3 k_2 \delta^{(1)}(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}) F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}). \quad (244)$$

In the specific case of (242) the kernel is given by

$$F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) = \frac{(\mathbf{k} \cdot \mathbf{k}_1)(\mathbf{k} \cdot \mathbf{k}_2)}{k^2} - \frac{1}{3} \mathbf{k}_1 \cdot \mathbf{k}_2. \quad (245)$$

The choice of these conventions is due not only for simplicity and to keep our expressions shorter, but also because at the

end we will be interested to the bispectrum of the CMB anisotropies generated at recombination, and the relevant expressions entering in the bispectrum are just the kernels of the convolution integrals.

Having determined the leading contribution to the source term at late times, we can now solve the evolution equation (B.13). Since the source, term scales like  $\eta^{-2}$ , it is useful to introduce the rescaled variable  $\chi = \eta^2 \Psi^{(2)}$ . Equation (B.13) then reads

$$\chi'' + \left( k^2 c_s^2 - \frac{2}{\eta^2} \right) \chi = \eta^2 S_y. \quad (246)$$

For perturbation modes which are subhorizon with  $k\eta \gg 1$  the solution of the homogeneous equation is given by

$$\chi_{\text{hom.}} = A \cos(kc_s\eta) + B \sin(kc_s\eta), \quad (247)$$

from which we can build the general solution

$$\begin{aligned} \chi = \chi_{\text{hom.}} + \chi_+ \int_0^\eta d\eta' \frac{\chi_-(\eta')}{W(\eta')} S_y(\eta') \\ - \chi_- \int_0^\eta d\eta' \frac{\chi_+(\eta')}{W(\eta')} S_y(\eta'), \end{aligned} \quad (248)$$

where  $W = -kc_s$  is the Wronskian, and  $\chi_+ = \cos(kc_s\eta)$ ,  $\chi_- = \sin(kc_s\eta)$ . Using (242), the integrals involve products of sines and cosines which can be performed giving

$$\begin{aligned} \chi = \chi_{\text{hom.}} - \frac{FC}{c_s^2} \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0) \\ \times (k[2k_1k_2 \cos(k_1c_s\eta) \cos(k_2c_s\eta) - 2k_1k_2 \\ \times \cos(kc_s\eta) + (k_1^2 + k_2^2 - k^2) \sin(k_1c_s\eta) \sin(k_2c_s\eta)]) \\ / (k_1^4 + k_2^4 + k^4 - 2k_1^2k_2^2 - 2k_1^2k^2 - 2k_2^2k^2). \end{aligned} \quad (249)$$

Thus, the gravitational potential  $\Psi^{(2)}$  at late times is given by

$$\begin{aligned} \Psi_{\mathbf{k}}^{(2)}(\eta) \\ = -3\Psi^{(2)}(0) \frac{\cos(kc_s\eta)}{(kc_s\eta)^2} - \frac{FC}{\eta^2 c_s^2} \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0) \\ \times ([2k_1k_2 \cos(k_1c_s\eta) \cos(k_2c_s\eta) - 2k_1k_2 \cos(kc_s\eta) \\ + (k_1^2 + k_2^2 - k^2) \sin(k_1c_s\eta) \sin(k_2c_s\eta)]) \\ / (k_1^4 + k_2^4 + k^4 - 2k_1^2k_2^2 - 2k_1^2k^2 - 2k_2^2k^2), \end{aligned} \quad (250)$$

where we have set the integration constant  $B = 0$  and  $A = -3\Psi^{(2)}(0)/(kc_s)^2$  in order to match the homogeneous solution at late times. Here,  $\Psi^{(2)}(0)$  is the initial condition for the gravitational potential taken on large scales deep in the radiation dominated era which will be determined in Section 7.5.2.

Equation (250) shows the result that we anticipated: also at second order the gravitational potential varies in time oscillating with an amplitude that decays as  $\eta^{-2}$ . Let us then take the divergence of the  $(i-0)$  Einstein equation (30) expanded at second order

$$\begin{aligned} \partial_i \left[ \frac{1}{2} \partial^i \Psi^{(2)'} + \frac{\mathcal{H}}{2} \partial^i \Phi^{(2)} + 2\Psi^{(1)} \partial^i \Psi^{(1)'} \right. \\ \left. + 2\mathcal{H} \Psi^{(1)} \partial^i \Phi^{(1)} - \Psi^{(1)'} \partial^i \Phi^{(1)} \right] \\ = -2\mathcal{H}^2 \partial_i \left[ \frac{1}{2} v_y^{(2)i} + (\Phi^{(1)} + \Psi^{(1)}) v_y^{(1)i} + \Delta_{00}^{(1)} v_y^{(1)i} \right], \end{aligned} \quad (251)$$

which, using the first-order  $(i-0)$  Einstein equation and  $\Phi^{(1)} \simeq \Psi^{(1)}$ , reduces to

$$\begin{aligned} \partial_i \left[ \frac{1}{2} \partial^i \Psi^{(2)'} + \frac{\mathcal{H}}{2} \partial^i \Phi^{(2)} - \Psi^{(1)'} \partial^i \Psi^{(1)} \right] \\ = -2\mathcal{H}^2 \partial_i \left[ \frac{1}{2} v_y^{(2)i} + \Delta_{00}^{(1)} v_y^{(1)i} \right]. \end{aligned} \quad (252)$$

Since  $\Psi^{(1)}$  during a radiation dominated period is given by (B.11) and at late times it decays oscillating, it is easy to see that  $(\Psi^{(1)'} \partial^i \Psi^{(1)})$  will be oscillating and decaying as  $\eta^{-4}$  and thus can be neglected with respect to  $\Psi^{(2)'}$ , which oscillates with an amplitude decaying as  $\eta^{-2}$ . Also,  $\mathcal{H} \Phi^{(2)}$  turns out to be subdominant. Recall that  $\Phi^{(2)} = \Psi^{(2)} - Q^{(2)}$  and  $Q^{(2)}$  is dominated by the second-order quadrupole of the photons in (B.15), so that  $\Phi^{(2)}$  scales like  $\Psi^{(2)}$  but there is the additional damping factor of the Hubble rate  $\mathcal{H} = 1/\eta$ . Thus, the dominant terms give

$$\partial_i v_y^{(2)i} \simeq -\frac{1}{2\mathcal{H}^2} \nabla^2 \Psi^{(2)'} - 2\partial_i (\Delta_{00}^{(1)} v_y^{(1)i}). \quad (253)$$

Equation (253) allows to proceed further in a similar way as for the linear case by using the results found so far, (250) and (253), in the energy continuity equation (207). In (207) the first- and second-order gravitational potentials can be neglected with respect to the remaining terms given by  $\Delta_{00}^{(1)}$  and  $v_y^{(1)i}$  which oscillate in time. Thus, replacing the divergence of the second-order velocity by the expression (253), (207) becomes

$$\Delta_{00}^{(2)'} = \frac{2}{3\mathcal{H}^2} \nabla^2 \Psi^{(2)'} + \frac{8}{3} \partial_i v_y^{(1)i} \Delta_{00}^{(1)} + (\Delta_{00}^{(1)2})', \quad (254)$$

which, using the first-order equation (C.1), further simplifies to

$$\Delta_{00}^{(2)'} = \frac{2}{3\mathcal{H}^2} \nabla^2 \Psi^{(2)'}, \quad (255)$$

where we have kept only the dominant terms at late times.

The gravitational potential  $\Psi^{(2)}$  is given in (250), so the integration of (255) gives

$$\begin{aligned} \Delta_{00}^{(2)} &= 6\Psi^{(2)}(0) \cos(kc_s\eta) + 2\frac{FC}{3c_s^2}\Psi_{k_1}^{(1)}(0)\Psi_{k_2}^{(1)}(0)k^2 \\ &\times ([2k_1k_2 \cos(k_1c_s\eta) \cos(k_2c_s\eta) - 2k_1k_2 \cos(kc_s\eta)] \\ &\quad + (k_1^2 + k_2^2 - k^2) \sin(k_1c_s\eta) \sin(k_2c_s\eta)) \\ &/ (k_1^4 + k_2^4 + k^4 - 2k_1^2k_2^2 - 2k_1^2k^2 - 2k_2^2k^2). \end{aligned} \quad (256)$$

Needless to say, modes for  $k \gg k_D$ , where  $k_D^{-1}$  indicates the usual damping length, are supposed to be multiplied by an exponential  $e^{-(k/k_D)^2}$  (see, e.g., [70]).

**7.5.1. Vector Perturbations.** So far, we have discussed only scalar perturbations. However, at second-order in perturbation theory an unavoidable prediction is that also vector (and tensor) perturbation modes are produced dynamically as nonlinear combination of first-order scalar perturbations. In particular notice that the second-order velocity appearing in (232), giving rise to a second-order Doppler effect at last scattering, will contain a scalar and a vector (divergence free) part. Equation (253) provides the scalar component of the second-order velocity. We now derive an expression for the velocity that includes also the vector contribution.

The (second-order) vector metric perturbation  $\omega^i$  when radiation dominates can be obtained from (B.17)

$$-\frac{1}{2}\nabla^2\omega^i + 3\mathcal{H}^2\omega^i = -4\mathcal{H}^2\left(\delta_j^i - \frac{\partial^i\partial_j}{\nabla^2}\right)\left(\frac{v_\gamma^{(2)j}}{2} + \Delta_{00}^{(1)}v_\gamma^{(1)j}\right), \quad (257)$$

where we have dropped the gravitational potentials  $\Psi^{(1)} \simeq \Phi^{(1)}$  which are subdominant at late times. On the other hand, from the velocity continuity equation (212), we get

$$\begin{aligned} v_\gamma^{(2)i} + \frac{1}{4}\Delta_{00}^{(2),i} \\ = \frac{1}{4}\left(\Delta_{00}^{(1)2}\right)^i + \frac{8}{3}v_\gamma^{(1)i}\partial_j v_\gamma^{(1)j} - 2\omega^i - \frac{3}{4}\partial_k \Pi_\gamma^{(2)ki}, \end{aligned} \quad (258)$$

neglecting the term proportional to  $R$  and the decaying gravitational potentials. Using the tight coupling equations at first order and integrating over time, one finds

$$v_\gamma^{(2)i} + 2\left(v_\gamma^{(1)i}\Delta_{00}^{(1)}\right) = -2\omega^i - \frac{1}{4}\int d\eta' \Delta_{00}^{(2),i} - \frac{3}{4}\int d\eta' \partial_k \Pi_\gamma^{(2)ki}. \quad (259)$$

We can thus plug (259) into (257) to find that at late times (for  $k\eta \gg 1$ )

$$\nabla^2\omega^i = -3\mathcal{H}^2\left(\delta_j^i - \frac{\partial^i\partial_j}{\nabla^2}\right)\int d\eta' \partial_k \Pi_\gamma^{(2)kj}. \quad (260)$$

We will come later to the explicit expression for the term on the R.H.S. of (260). Here, it is enough to notice that the

second-order quadrupole oscillate in time and thus  $\omega^i$  will decay in time as  $\mathcal{H}^2 = 1/\eta^2$ . This shows that  $\omega^i$  in (259) can be in fact neglected with respect to the other terms giving

$$v_\gamma^{(2)i} = -2\left(v_\gamma^{(1)i}\Delta_{00}^{(1)}\right) - \frac{1}{4}\int d\eta' \Delta_{00}^{(2),i} - \frac{3}{4}\int d\eta' \partial_k \Pi_\gamma^{(2)ki}. \quad (261)$$

It can be useful to compute the combination on the R.H.S. of (260)  $(\delta_j^i - \partial^i\partial_j/\nabla^2)\partial_k \Pi_\gamma^{(2)kj}$ . The second-order quadrupole moment of the photons in the tightly coupled limit is given by (222), and

$$\partial_k \Pi_\gamma^{(2)kj} = \frac{8}{3}\left[\partial_k(v^k v^j) - 2v^k \partial^j v_k\right] = \frac{8}{3}\left[v^j \partial_k v^k - v^k \partial^j v_k\right], \quad (262)$$

where in the last step, we have used that the linear velocity is the gradient of a scalar perturbation. We thus find

$$\begin{aligned} \left(\delta_j^i - \frac{\partial^i\partial_j}{\nabla^2}\right)\partial_k \Pi_\gamma^{(2)kj} \\ = \frac{8}{3}\left(v^i \partial_k v^k - v^k \partial^i v_k\right) - \frac{8}{3}\frac{\partial^i}{\nabla^2} \\ \times \left[\left(\partial_k v^k\right)^2 + v^j \partial_j \partial_k v^k - \partial_j v^k \partial^j v_k - v^k \nabla^2 v_k\right]. \end{aligned} \quad (263)$$

Notice that if we split the quadrupole moment into a scalar, vector (divergence-free) and tensor (divergence-free and traceless) parts as

$$\Pi_\gamma^{(2)kj} = \Pi_\gamma^{(2),kj} - \frac{1}{3}\nabla^2 \delta^{kj} \Pi_\gamma^{(2)} + \Pi_\gamma^{(2),k,j} + \Pi_\gamma^{(2),j,k} + \Pi_{\gamma T}^{(2)kj}, \quad (264)$$

then it turns out that

$$\left(\delta_j^i - \frac{\partial^i\partial_j}{\nabla^2}\right)\partial_k \Pi_\gamma^{(2)kj} = \nabla^2 \Pi_\gamma^{(2)i}, \quad (265)$$

where  $\Pi_\gamma^{(2)i}$  is the vector part of the quadrupole moment. Therefore, one can rewrite (260) as

$$\omega^i = -3\mathcal{H}^2 \int d\eta' \Pi_\gamma^{(2)i}. \quad (266)$$

**7.5.2. Initial Conditions for the Second-Order Gravitational Potentials.** In order to complete the study of the CMB anisotropies at second-order for modes  $k \gg k_{\text{eq}}$ , we have to specify the initial conditions  $\Psi^{(2)}(0)$  appearing in (256). These are set on superhorizon scales deep in the standard radiation dominated epoch (for  $\eta \rightarrow 0$ ) by exploiting the conservation in time of the curvature perturbation  $\zeta$ . On superhorizon scales,  $\zeta^{(2)}$  is given by (239) during the radiation dominated epoch and, using the (0-0) Einstein equation in the large scale limit  $\Delta_{00}^{(2)} = -2\Phi^{(2)} + 4\Phi^{(1)2}$ , we find

$$\zeta^{(2)} = -\frac{3}{2}\Psi^{(2)}(0) - \frac{1}{2}\left(\Phi^{(2)}(0) - \Psi^{(2)}(0)\right) + \frac{9}{2}\Psi^{(1)2}(0). \quad (267)$$

The conserved value of  $\zeta^{(2)}$  is parametrized by  $\zeta^{(2)} = 2a_{\text{NL}}\zeta^{(1)2}$ , where, as explained in Section 7.4 the parameter  $a_{\text{NL}}$  specifies the level of primordial non-Gaussianity depending on the particular scenario for the generation of the cosmological perturbations. On the other hand, during radiation-domination

$$\Phi^{(2)}(0) - \Psi^{(2)}(0) = -Q^{(2)}(0), \quad (268)$$

where

$$\begin{aligned} Q^{(2)}(0) &= -2\nabla^{-2}\partial_k\Phi^{(1)}(0)\partial^k\Phi^{(1)}(0) \\ &+ 6\frac{\partial_i\partial^j}{\nabla^4}\left(\partial^i\Phi^{(1)}(0)\partial_j\Phi^{(1)}(0)\right) + \frac{9}{2}\mathcal{H}^2\frac{\partial_i\partial^j}{\nabla^4}\Pi_{\gamma j}^{(2)i}, \end{aligned} \quad (269)$$

where we are evaluating (B.15) in the limit  $k\eta \ll 1$ . The contribution from the second-order quadrupole moment in this limit reads

$$\begin{aligned} &\frac{9}{2}\mathcal{H}^2\frac{\partial_i\partial^j}{\nabla^4}\Pi_{\gamma j}^{(2)i} \\ &= \frac{9}{2\eta^2}\frac{8}{3}\frac{\partial_i\partial^j}{\nabla^4}\left(v^i v^j - \frac{1}{3}\delta_j^i v^2\right) \\ &\equiv -3\frac{FC}{k^2\eta^2}\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0)\sin(k_1 c_s \eta)\sin(k_2 c_s \eta) \\ &\longrightarrow \frac{27}{k^2}F\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0), \end{aligned} \quad (270)$$

where  $F$  and  $C$  are defined in (244) and (243). Therefore, we find that in Fourier space

$$Q^{(2)}(0) = 33\frac{F}{k^2}\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0), \quad (271)$$

and from (267) we read off the initial condition as (convolution products are understood)

$$\Psi^{(2)}(0) = \left[-3(a_{\text{NL}} - 1) + 11\frac{F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})}{k^2}\right]\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0). \quad (272)$$

**7.6. Perturbation Modes with  $k \ll k_{\text{eq}}$ .** Let us consider the photon perturbations which enter the horizon between the equality epoch and the recombination epoch, with wavelengths  $\eta_*^{-1} < k < \eta_{\text{eq}}^{-1}$ . In fact, in order to find some analytical solutions, we will assume that by the time of recombination the universe is matter dominated  $\eta_{\text{eq}} \ll \eta_*$ . In this case, the gravitational potentials are sourced by the dark matter component and their evolution is given in Section B. At linear order, the gravitational potentials remain constant in time, while at second-order they are given by (B.6). In turn, the gravitational potentials act as an external force on the CMB photons as in the equation (214) describing the CMB energy density evolution in the tightly coupled regime.

For the regime of interest it proves convenient to use the solution of (214) found in (241). The source functions  $\mathcal{S}_\Delta$  and  $\mathcal{S}_V^i$  are given by (208) and (213), respectively. In particular  $\mathcal{S}_\Delta$  at early times— $\mathcal{S}_\Delta(0)$  appearing in (241)—vanishes. For a matter-dominated period

$$\begin{aligned} \mathcal{S}_\Delta(R=0) &= \left(\Delta_{00}^{(1)2}\right)' - \frac{16}{3}\Psi^{(1)}\partial_i v_\gamma^{(1)i} + \frac{16}{3}\left(v_\gamma^2\right)' \\ &+ \frac{32}{3}\partial^i\Psi^{(1)}v_i, \end{aligned} \quad (273)$$

where we have used the linear evolution equations (C.1) and (C.5) with  $\Phi^{(1)} = \Psi^{(1)}$ , and

$$\begin{aligned} \mathcal{S}_V^i(R=0) &= \frac{8}{3}v_\gamma^{(1)i}\partial_j v_\gamma^{(1)j} + \frac{1}{4}\partial^i\Delta_{00}^{(1)2} - 2\partial^i\Psi^{(1)2} \\ &- \Psi^{(1)}\partial^i\Delta_{00}^{(1)} - 2\omega^{i'} - \frac{3}{4}\partial_j\Pi^{(2)ij}. \end{aligned} \quad (274)$$

As at linear order, we are evaluating these expressions in the limit  $R = 3\rho_b/4\rho_\gamma \rightarrow 0$ , while retaining a non-vanishing and constant value for  $R$  in the expression for the photon-baryon fluid sound speed entering in the sines and cosines. In fact, this approximation gives the dominant contributions to the source terms  $\mathcal{S}_\Delta$  and  $\mathcal{S}_V^i$  made by first-order squared terms. Using the linear solutions (C.15) and (C.17) for the energy density and velocity of photons, the source functions in Fourier space read

$$\begin{aligned} \mathcal{S}_\Delta(R=0) &= \left[-2\left(\frac{6}{5}\right)^2 k_2 c_s \cos(k_1 c_s \eta)\sin(k_2 c_s \eta) + \frac{108}{25}k_2 c_s \right. \\ &\times \sin(k_2 c_s \eta) - \frac{32}{3}\left(\frac{9}{10}\right)^2 \frac{\mathbf{k}_1}{k_1} \cdot \mathbf{k}_2 c_s^3 \\ &\times \sin(k_1 c_s \eta)\cos(k_2 c_s \eta) + \left.\frac{32}{3}\left(\frac{9}{10}\right)^2 \right. \\ &\times \left.\frac{\mathbf{k}_1}{k_1} \cdot \mathbf{k}_2 c_s \sin(k_1 c_s \eta)\right]\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0), \end{aligned} \quad (275)$$

$$\begin{aligned} \mathcal{S}_V^i(R=0) &= \left[-i\frac{2}{3}\left(\frac{9}{10}\right)^2 c_s^2 \frac{k_1^i}{k_1} k_2 \sin(k_1 c_s \eta)\sin(k_2 c_s \eta) \right. \\ &+ \frac{i}{4}k^i\left(\frac{6}{5}\cos(k_1 c_s \eta) - \frac{18}{5}\right)\left(\frac{6}{5}\cos(k_2 c_s \eta) - \frac{18}{5}\right) \\ &- 2ik^i\left(\frac{9}{10}\right)^2 - i\frac{9}{10}k_2^i\left(\frac{6}{5}\cos(k_2 c_s \eta) - \frac{18}{5}\right) - 2\omega^{i'} \\ &+ \left.\frac{2}{3}\left(\frac{9}{10}\right)^2 c_s^2 \frac{\mathbf{k}_2}{k_2} \cdot \mathbf{k}_1 \frac{k_1^i}{k_1} \sin(k_1 c_s \eta)\sin(k_2 c_s \eta)\right] \\ &\times \Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (276)$$

In  $\mathcal{S}_V^i$ , we have used the expression (222) for the second-order quadrupole moment  $\Pi_\gamma^{(2)ij}$  of the photons in the

tight coupling limit, with the velocity  $v^{(1)} = v_\gamma^{(1)}$ . Notice that, for the modes crossing the horizon at  $\eta > \eta_{\text{eq}}$ , we have expressed the gravitational potential during the matter-dominated period in terms of the initial value on superhorizon scales deep in the radiation dominated epoch as  $\Psi^{(1)} = 9\Psi^{(1)}(0)/10$ .

As for the second-order gravitational potentials, we have to compute the combination  $\Phi^{(2)} + \Psi^{(2)}$  appearing in (241). The gravitational potential  $\Psi^{(2)}$  is given by (B.6), while  $\Phi^{(2)}$  is given by

$$\Phi^{(2)} = \Psi^{(2)} - Q^{(2)}, \quad (277)$$

according to the relation (32), where for a matter-dominated period

$$Q^{(2)} = 5\nabla^{-4}\partial_i\partial_j(\partial^i\Psi^{(1)}\partial_j\Psi^{(1)}) - \frac{5}{3}\nabla^{-2}(\partial_k\Psi^{(1)}\partial^k\Psi^{(1)}). \quad (278)$$

We thus find

$$\begin{aligned} \Phi^{(2)} + \Psi^{(2)} &= -5\nabla^{-4}\partial_i\partial^j(\partial^i\Psi^{(1)}\partial_j\Psi^{(1)}) \\ &+ \frac{5}{3}\nabla^{-2}(\partial_k\Psi^{(1)}\partial^k\Psi^{(1)}) + 2\Psi_m^{(2)}(0) \\ &- \frac{1}{7}\left(\partial_k\Psi^{(1)}\partial^k\Psi^{(1)} - \frac{10}{3}\nabla^{-2}\partial_i\partial^j(\partial^i\Psi^{(1)}\partial_j\Psi^{(1)})\right)\eta^2, \end{aligned} \quad (279)$$

which in Fourier space reads

$$\begin{aligned} \Phi^{(2)} + \Psi^{(2)} &= 2\Psi_m^{(2)}(0) + \left[\frac{1}{7}G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})\eta^2 - \frac{5}{k^2}F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})\right] \\ &\times \left(\frac{9}{10}\right)^2\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0), \end{aligned} \quad (280)$$

where the kernels of the convolutions are given by (245) and

$$G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) = \mathbf{k}_1 \cdot \mathbf{k}_2 - \frac{10}{3}\frac{(\mathbf{k} \cdot \mathbf{k}_1)(\mathbf{k} \cdot \mathbf{k}_2)}{k^2}. \quad (281)$$

In (279),  $\Psi_m^{(2)}(0)$  is the initial condition for the gravitational potential fixed at some time  $\eta_i > \eta_{\text{eq}}$ . For the regime of interest it corresponds to the value of the gravitational potential on superhorizon scales during the matter-dominated epoch.

Notice a property that will be useful later on. By looking at (278) and the explicit solution for  $\Psi^{(2)}$  (B.6) which grows as  $\eta^2$ , it is easy to realize that on very small scales, for  $k\eta \gg 1$ , the two gravitational potentials are equal with

$$\Phi^{(2)} \simeq \Psi^{(2)} = \Psi_m^{(2)}(0) + \frac{1}{14}G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})\eta^2. \quad (282)$$

We are now able to compute the integrals entering in the solution (241). The one involving the second-order gravitational potentials is straightforward to compute

$$\begin{aligned} &-4kc_s \int_0^\eta d\eta' (\Phi^{(2)} + \Psi^{(2)} + R\Phi^{(2)}) \sin[kc_s(\eta - \eta')] \\ &= -8\Psi_m^{(2)}(0)(1 - \cos(kc_s\eta)) \\ &-4 \left[ -\frac{5F}{k^2}(1 - \cos(kc_s\eta)) + \frac{1}{7k^2c_s^2} \right. \\ &\quad \left. \times G(-2 + (kc_s\eta)^2 + 2\cos(kc_s\eta)) \right] \\ &\times \left(\frac{9}{10}\right)^2 \Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0) - 4R\Psi_m^{(2)}(0)(1 - \cos(kc_s\eta)) \\ &-4R \left[ -\frac{5F}{k^2}(1 - \cos(kc_s\eta)) + \frac{1}{14k^2c_s^2} \right. \\ &\quad \left. \times G(-2 + (kc_s\eta)^2 + 2\cos(kc_s\eta)) \right] \\ &\times \left(\frac{9}{10}\right)^2 \Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0), \end{aligned} \quad (283)$$

where for brevity  $F$  and  $G$  stand for the convolutions (245) and (281). Notice also that we have isolated the terms proportional to  $R$  as it will be useful later on.

For the two remaining integrals, in the following we will show only the terms that in the final expression for  $\Delta_{00}^{(2)}$  and the second-order velocity  $v_\gamma^{(2)i}$  give the dominant contributions for  $k\eta \gg 1$ , even though we have performed a fully computation. The integral over the source function  $\mathcal{S}_\Delta$  yields a sum of oscillating functions (cosines) which turn out to be subdominant, so we do not display the full result. For the last integral we find ((1  $\leftrightarrow$  2) stands by an exchange of indices)

$$\begin{aligned} &-\frac{4}{3}\frac{ik_i}{kc_s} \int_0^\eta d\eta' s_v' \sin[kc_s(\eta - \eta')] \\ &= \left[ \frac{27}{25}\frac{2\mathbf{k} \cdot \mathbf{k}_2 + k^2}{k^2c_s^2} + (1 \leftrightarrow 2) \right] \Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0), \end{aligned} \quad (284)$$

where the terms that have been dropped vary in time as sines and cosines. We have written the contribution in (284) because, upon integration over time, it will give a nonnegligible contribution to the velocity  $v_\gamma^{(2)i}$ .

From the general solution (241) and the expression (B.6) for the second-order gravitational potential  $\Psi^{(2)}$ , we thus obtain

$$\begin{aligned} \Delta_{00}^{(2)} = & -4(1+R)\Psi_m^{(2)}(0) \\ & + \left[ 2(9a_{\text{NL}} - 7)\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0) + (8 + 4R)\Psi_m^{(2)}(0) \right] \\ & \times \cos(kc_s\eta) - \frac{2}{7}G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})\eta^2(1+R)\left(\frac{9}{10}\right)^2 \\ & \times \Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (285)$$

We warn the reader that in writing (285), we have kept all those terms that contain the primordial non-Gaussianity parametrized by  $a_{\text{NL}}$ , and the terms which dominate at late times for  $k\eta \gg 1$ .

**7.6.1. Initial Conditions for the Second-Order Gravitational Potentials.** The initial condition  $\Psi_m^{(2)}(0)$  for the modes that cross the horizon after the equality epoch is fixed by the value of the gravitational potential on superhorizon scales during the matter dominated epoch. To compute this value we use the conservation on superhorizon scales of the curvature perturbation  $\zeta^{(2)}$  defined in (36). For a matter-dominated period, the curvature perturbation on large-scales turns out to be

$$\zeta^{(2)} = -\Psi_m^{(2)}(0) + \frac{1}{3}\frac{\delta^{(2)}\rho_m}{\rho_m} + \frac{38}{9}\Psi_m^{(1)2}(0), \quad (286)$$

where we used the energy continuity equation  $\delta^{(1)}\rho'_m + 3\mathcal{H}\delta^{(1)}\rho_m - 3\rho_m\Psi^{(1)'} = 0$  and the  $(0-0)$  Einstein equation  $\delta^{(1)}\rho_m/\rho_m = -2\Psi^{(1)}$  in the superhorizon limit.

From the  $(0-0)$  Einstein equation on large scales  $\delta^{(2)}\rho_m/\rho_m = -2\Phi^{(2)} + 4\Phi^{(1)2}$  bringing

$$\zeta^{(2)} = -\frac{5}{3}\Psi_m^{(2)}(0) - \frac{2}{3}\left(\Phi_m^{(2)}(0) - \Psi_m^{(2)}(0)\right) + \frac{50}{9}\Psi_m^{(1)2}(0). \quad (287)$$

The conserved value of  $\zeta^{(2)}$  is parametrized as in (38),  $\zeta^{(2)} = 2a_{\text{NL}}\zeta^{(1)2} = (50a_{\text{NL}}/9)\Psi^{(1)2}$ , with  $\zeta^{(1)} = -5\Psi^{(1)}/3$  on large scales after the equality epoch. At second order, the two gravitational potentials in a matter dominated epoch differ according to (278) and using (287), we find

$$\begin{aligned} \Psi_m^{(2)}(0) = & -\frac{27}{10}(a_{\text{NL}} - 1)\left(\Psi^{(1)}(0)\right)^2 + \left(\frac{9}{10}\right)^2 \\ & \times \left[ 2\nabla^{-4}\partial_i\partial_j\left(\partial^i\Psi^{(1)}(0)\partial_j\Psi^{(1)}(0)\right) \right. \\ & \left. - \frac{2}{3}\nabla^{-2}\left(\partial_k\Psi^{(1)}(0)\partial^k\Psi^{(1)}(0)\right) \right], \end{aligned} \quad (288)$$

we have expressed the gravitational potential during the matter dominated period  $\Psi^{(1)}$  in terms of the initial value on superhorizon scales after the equality epoch as  $\Psi^{(1)} = 9\Psi^{(1)}(0)/10$ . In Fourier space, (288) becomes

$$\begin{aligned} \Psi_m^{(2)}(0) = & \left[ -\frac{27}{10}(a_{\text{NL}} - 1) + 2\left(\frac{9}{10}\right)^2\frac{F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})}{k^2} \right] \\ & \times \Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0), \end{aligned} \quad (289)$$

where  $F$  is the kernel defined in (245).

We can use the explicit expression for  $\Psi_m^{(2)}(0)$  in (285), still keeping only the terms that contain the primordial non-Gaussianity parametrized by  $a_{\text{NL}}$ , and the terms which dominate at late times for  $k\eta \gg 1$  to find

$$\begin{aligned} \Delta_{00}^{(2)} = & \left[ \frac{54}{5}(1+R)(a_{\text{NL}} - 1) - \frac{2}{5}(9a_{\text{NL}} - 19) \right. \\ & \times \cos(kc_s\eta) + \frac{54}{5}R(a_{\text{NL}} - 1)\cos(kc_s\eta) \\ & \left. - \frac{2}{7}\left(\frac{9}{10}\right)^2G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})\eta^2(1+R) \right] \\ & \times \Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (290)$$

**7.6.2. Second-Order Photon Velocity Perturbation.** The second-order velocity of the photons can be obtained from (212)

$$v_y^{(2)i} \simeq \int_0^\eta d\eta' \left( g_V^i - \partial^i\Phi^{(2)} - \frac{1}{4}\partial^i\Delta_{00}^{(2)} \right). \quad (291)$$

Notice that for simplicity in writing (291) we have dropped off the dependence on  $R$ . In fact, the main conclusions of this Subsection remains unchanged. The second-order gravitational potential in matter-dominated universe can be obtained from (277)-(278) and (B.6) as

$$\begin{aligned} \Phi^{(2)} = & \Psi_m^{(2)}(0) - \frac{1}{14}\left(\partial_k\Psi^{(1)}\partial^k\Psi^{(1)} - \frac{10}{3}\nabla^{-2}\partial_i\partial^i\right. \\ & \left.\times\left(\partial^i\Psi^{(1)}\partial_j\Psi^{(1)}\right)\right)\eta^2 - 5\nabla^{-4}\partial_i\partial^i \\ & \times\left(\partial^i\Psi^{(1)}\partial_j\Psi^{(1)}\right) + \frac{5}{3}\nabla^{-2}\left(\partial_k\Psi^{(1)}\partial^k\Psi^{(1)}\right). \end{aligned} \quad (292)$$

In Fourier space, this becomes

$$\begin{aligned} \Phi^{(2)} = & \Psi_m^{(2)}(0) + \left[ \frac{1}{14}G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})\eta^2 - \frac{5}{k^2}F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) \right] \\ & \times \left(\frac{9}{10}\right)^2\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0), \end{aligned} \quad (293)$$

where the kernels of the convolutions are given by (245) and (281). The integral over  $\Phi^{(2)}$  in (291) is then easily computed

$$\begin{aligned} & - \int_0^\eta d\eta' \partial^i \Phi^{(2)} \\ & \equiv -ik^i \left[ \Psi_m^{(2)}(0)\eta \right. \\ & \quad + \left( \frac{1}{42} G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) \eta^3 - \frac{5}{k^2} F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) \eta \right) \\ & \quad \left. \times \left( \frac{9}{10} \right)^2 \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0) \right], \end{aligned} \quad (294)$$

where as usual the equivalence symbol means that we are evaluating a given expression in Fourier space. For the integral over the source function  $\mathcal{S}_V^i$  we use its expression in Fourier space, (276), and the dominant terms for  $k\eta \gg 1$  are

$$\begin{aligned} & \int_0^\eta d\eta' \mathcal{S}_V^i \\ & \equiv \left[ -2ik^i \left( \frac{9}{10} \right)^2 + ik^i \frac{81}{25} + i \frac{8}{3} \left( \frac{9}{10} \right)^2 \frac{1}{k^4} (k_2^2 - k_1^2) \mathbf{k} \cdot \mathbf{k}_1 k_2^i \right] \eta. \end{aligned} \quad (295)$$

Notice that, in order to compute this integral, we must know the second-order vector metric perturbation  $\omega^i$ . This is easily obtained for a matter-dominated universe from (B.7). Using (B.4) and (B.2) one finds

$$\begin{aligned} \omega^i &= -\frac{4}{3} \left( \frac{9}{10} \right)^2 \nabla^{-4} \partial_j \\ & \quad \times \left[ \partial^i \nabla^2 \Psi^{(1)}(0) \partial^j \Psi^{(1)}(0) - \partial^j \nabla^2 \Psi^{(1)}(0) \partial^i \Psi^{(1)}(0) \right] \eta, \end{aligned} \quad (296)$$

giving rise to the third term in (295).

Finally, for the integral over  $\Delta_{00}^{(2)}$  some caution is needed. Since in the final expression for  $v_y^{(2)i}$  the dominant terms at late times turn out to be proportional  $\eta$ , one has to use an expression for  $\Delta_{00}^{(2)}$  that keep track of all those contributions that, upon integration, scale like  $\eta$ . Thus, we must use the expression written in (285), plus (284), and some terms of (283) that have been previously neglected in (285). Then, we find for  $k\eta \gg 1$

$$\begin{aligned} & + \frac{1}{4} \int_0^\eta d\eta' \partial^i \Delta_{00}^{(2)} \\ & \equiv \frac{ik^i}{4} \left[ -4\Psi_m^{(2)}(0)\eta + \left( 2(9a_{\text{NL}} - 7)\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0) \right. \right. \\ & \quad \left. \left. + 8\Psi_m^{(2)}(0) \frac{\sin(kc_s\eta)}{kc_s} \right) \right] \\ & + \frac{ik^i}{4} \left[ -\frac{2}{21} \left( \frac{9}{10} \right)^2 G \eta^3 \right. \\ & \quad + \left( \frac{20}{3c_s^2} \left( \frac{9}{10} \right)^2 \frac{F}{k^2} + \frac{8}{21c_s^4} \left( \frac{9}{10} \right)^2 \frac{G}{k^2} \right. \\ & \quad \left. \left. + \frac{54}{25c_s^2} \frac{k^2 + \mathbf{k} \cdot (\mathbf{k}_1 + \mathbf{k}_2)}{k^2} \right) \eta \right] \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (297)$$

Using (294), (295), and (297) we get

$$\begin{aligned} v_y^{(2)i} &= \left[ i \frac{k^i}{k} \frac{1}{10c_s} (9a_{\text{NL}} - 19) \sin(kc_s\eta) \right. \\ & \quad + \left( -i \frac{2}{21c_s^4} k^i \left( \frac{9}{10} \right)^2 \frac{G}{k^2} - 2i \left( \frac{9}{10} \right)^2 k^i \right. \\ & \quad + i \frac{81}{50} (k_2^i + k_1^i) - i \frac{27}{50c_s^2} \frac{k^2 + \mathbf{k} \cdot (\mathbf{k}_1 + \mathbf{k}_2)}{k^2} k^i \\ & \quad \left. \left. + i \frac{4}{3} \left( \frac{9}{10} \right)^2 \frac{k_2^2 - k_1^2}{k^4} (\mathbf{k} \cdot \mathbf{k}_1 k_2^i - \mathbf{k} \cdot \mathbf{k}_2 k_1^i) \right) \eta \right] \\ & \quad \times \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (298)$$

To obtain (298), we have also used the explicit expression (289) for  $\Psi_m^{(2)}(0)$  and we have kept the terms depending on  $a_{\text{NL}}$  parametrizing the primordial non-Gaussianity and the terms that dominate at late times for  $k\eta \gg 1$ .

*7.7. Perturbation Modes with  $k \gg k_{\text{eq}}$ : Cold Dark Matter Perturbations at Second Order and Improved Analytical Solutions for CMB Anisotropies.* In Section 7.5, we have computed the perturbations of the CBM photons at last scattering for the modes that cross the horizon at  $\eta < \eta_{\text{eq}}$  under the approximation that the universe is radiation-dominated. However, around the equality epoch, through recombination, the dark matter component will start to dominate. In this section, we will account for its contribution to the gravitational potential and for the resulting perturbations of the photons from the equality epoch onwards. This leads to a more realistic and accurate analytical solutions for the acoustic oscillations of the photon-baryon fluid for the modes of interest.

The starting point is to consider the density perturbation in the dark matter component for subhorizon modes during the radiation dominated epoch. Its value at the equality epoch will fix the magnitude of the gravitational potential at  $\eta_{\text{eq}}$  and hence the initial conditions for the subsequent evolution of the photons fluctuations during the matter dominated period. At linear order, the procedure is standard (see, e.g [70]), and we will use a similar one at second-order in the perturbations. Thus, this section serves also as a guide through the evolution of the CDM density perturbations at second-order accounting for those modes that enter the horizon during the radiation dominated epoch. This allows to determine the second-order transfer function for the density perturbations with a generalization at second-order of the Meszaros effect.

*7.7.1. Subhorizon Evolution of CDM Perturbations for  $\eta < \eta_{\text{eq}}$ .* From the energy and velocity continuity equations for CDM, it is possible to isolate an evolution equation for the density perturbation  $\delta_d = \delta\rho_d/\rho_d$ , where the subscript  $d$  stands for cold dark matter. In [46], we have obtained the Boltzmann

equations up to second-order for CDM. The number density of CDM evolves according to [46]

$$\begin{aligned} \frac{\partial n_d}{\partial \eta} + e^{\Phi+\Psi} \frac{\partial (v_d^i n_d)}{\partial x^i} + 3(\mathcal{H} - \Psi') n_d - 2e^{\Phi+\Psi} \Psi_{,k} v_d^k n_d \\ + e^{\Phi+\Psi} \Phi_{,k} v_d^k n_d = 0. \end{aligned} \quad (299)$$

At linear order,  $n_d = \bar{n}_d + \delta^{(1)} n_d$  and one recovers the usual energy continuity equation

$$\delta_d^{(1)'} + v_{d,i}^{(1)i} - 3\Psi^{(1),i} = 0, \quad (300)$$

with  $\delta_d^{(1)} = \delta^{(1)} \rho_d / \bar{\rho}_d = \delta^{(1)} n_d / \bar{n}_d$ . The CDM velocity at the same order of perturbation obeys [46]

$$v_d^{(1)i'} + \mathcal{H} v_d^{(1)i} = -\Phi^{(1),i}. \quad (301)$$

Perturbing  $n_{\text{CDM}}$  up to second-order, we find

$$\begin{aligned} \delta_d^{(2)'} + v_{d,i}^{(2)i} - 3\Psi^{(2)'} \\ = -2\left(\Phi^{(1)} + \Psi^{(1)}\right) v_{d,i}^{(1)i} - 2v_{d,i}^{(1)i} \delta_d^{(1)} - 2v_d^{(1)i} \delta_{d,i}^{(1)} \\ + 6\Psi^{(1)'} \delta_d^{(1)} + \left(4\Psi_{,k}^{(1)} - 2\Phi_{,k}^{(1)}\right) v_d^{(1)k}. \end{aligned} \quad (302)$$

The R.H.S. of this equation can be further manipulated by using the linear equation (300) to replace  $v_{d,i}^{(1)i}$  yielding

$$\begin{aligned} \delta_d^{(2)'} + v_{d,i}^{(2)i} - 3\Psi^{(2)'} \\ = 4\delta_d^{(1)'} \Psi^{(1)} - 6\left(\Psi^{(1)2}\right)' + \left(\delta_d^{(1)2}\right)' \\ - 2v_d^{(1)i} \delta_{d,i}^{(1)} + 2\Psi_{,k}^{(1)} v_d^{(1)k}, \end{aligned} \quad (303)$$

where we use  $\Phi^{(1)} = \Psi^{(1)}$ . In [46] the evolution equation for the second-order CDM velocity perturbation has been already obtained

$$\begin{aligned} v_d^{(2)i'} + \mathcal{H} v_d^{(2)i} + 2\omega^i + 2\mathcal{H}\omega^i + \Phi^{(2),i} \\ = 2\Psi^{(1)'} v_d^{(1)i} - 2v_d^{(1)j} \partial_j v_d^{(1)i} - 4\Phi^{(1)} \Phi^{(1),i}. \end{aligned} \quad (304)$$

At linear order, we can take the divergence of (301) and, using (300) to replace the velocity perturbation, we obtain a differential equation for the CDM density contrast

$$\left[ a\left(3\Psi^{(1)'} - \delta_d^{(1)'}\right) \right]' = -a\nabla^2 \Phi^{(1)}, \quad (305)$$

which can be rewritten as

$$\delta_d^{(1)''} + \mathcal{H} \delta_d^{(1)'} = S^{(1)}, \quad (306)$$

where

$$S^{(1)} = 3\Psi^{(1)''} + 3\mathcal{H}\Psi^{(1)'} + \nabla^2 \Phi^{(1)}. \quad (307)$$

When the radiation is dominating the gravitational potential is mainly due to the perturbations in the photons, and  $a(\eta) \propto \eta$ . For subhorizon scales, (306) can be solved following the procedure introduced in [74]. Using the Green method, the general solution to (306) (in Fourier space) is given by

$$\begin{aligned} \delta_d^{(1)}(\mathbf{k}, \eta) = C_1 + C_2 \ln(\eta) \\ - \int_0^\eta d\eta' S^{(1)}(\eta') \eta' (\ln(k\eta') - \ln(k\eta)), \end{aligned} \quad (308)$$

where the first two terms correspond to the solution of the homogeneous equation. At early times, the density contrast is constant with

$$\delta_d^{(1)}(0) = \frac{3}{4} \Delta_{00}^{(1)}(0) = -\frac{3}{2} \Phi_{\mathbf{k}}^{(1)}(0), \quad (309)$$

having used the adiabaticity condition, and thus we fix the integration constant as

$$C_1 = -\frac{3\Phi_{\mathbf{k}}^{(1)}(0)}{2}, \quad (310)$$

and  $C_2 = 0$ . The gravitational potential during the radiation-dominated epoch starts to decay as a given mode enters the horizon. Therefore, the source term  $S^{(1)}$  behaves in a similar manner and this implies that the integrals over  $\eta'$  reach asymptotically a constant value. Once the mode has crossed the horizon we can thus write the solution as

$$\delta_d^{(1)}(\mathbf{k}, \eta) = A^{(1)} \Phi^{(1)}(0) \ln[B^{(1)} k\eta], \quad (311)$$

where the constants  $A^{(1)}$  and  $B^{(1)}$  are defined as

$$A^{(1)} \Phi^{(1)}(0) = \int_0^\infty d\eta' S^{(1)}(\eta') \eta', \quad (312)$$

$$A^{(1)} \Phi^{(0)} \ln(B^{(1)}) = -\frac{3}{2} \Phi^{(1)}(0) - \int_0^\infty d\eta' S^{(1)}(\eta') \eta' \ln(k\eta'). \quad (313)$$

The upper limit of the integrals can be taken to infinity because the main contribution comes from when  $k\eta \sim 1$  and once the mode has entered the horizon the result will change by a very small quantity. Performing the integrals in (312) and (313), one finds that  $A^{(1)} = -9.6$  and  $B^{(1)} \simeq 0.44$ .

Before moving to the second-order case, a useful quantity to compute is the CDM velocity in a radiation-dominated epoch. From (301), it is given by

$$\begin{aligned} v_d^{(1)i} - \frac{1}{a} \int_0^\eta d\eta' \partial^i \Phi^{(1)} a(\eta') \\ \equiv -3\left(ik^i\right) \Phi^{(1)}(0) \frac{kc_s \eta - \sin(kc_s \eta)}{k^3 c_s^3 \eta^2}, \end{aligned} \quad (314)$$

where the last equality holds in Fourier space, and we have used (B.11) (and the fact that  $a(\eta) \propto \eta$  when radiation dominates).

Combining (303) and (304), we get the analogue of (306) at second-order in perturbation theory

$$\delta_d^{(2)''} + \mathcal{H}\delta_d^{(2)'} = S^{(2)}, \quad (315)$$

where the source function is

$$\begin{aligned} S^{(2)} = & 3\Psi^{(2)''} + 3\mathcal{H}\Psi^{(2)'} + \nabla^2\Phi^{(2)} - 2\partial_i(\Psi^{(1)}v_d^{(1)i}) \\ & + \nabla^2v_d^{(1)2} + 2\nabla^2\Phi^{(1)2} \\ & + \frac{1}{a}\left[ a\left(4\delta_d^{(1)'}\Psi^{(1)} - 6(\Psi^{(1)2})' + (\delta_d^{(1)2})'\right. \right. \\ & \left. \left. - 2v_d^{(1)i}\delta_{d,i}^{(1)} + 2\Psi_{,k}^{(1)}v_d^{(1)k}\right) \right]'. \end{aligned} \quad (316)$$

In fact, we write (315) in a more convenient way as

$$\delta_d^{(2)''} - 3\Psi^{(2)''} - s_1' + \mathcal{H}(\delta_d^{(2)'} - 3\Psi^{(2)'} - s_1) = s_2, \quad (317)$$

where for simplicity, we have introduced the two functions

$$\begin{aligned} s_1 = & 4\delta_d^{(1)'}\Psi^{(1)} - 6(\Psi^{(1)2})' + (\delta_d^{(1)2})' \\ & - 2v_d^{(1)i}\delta_{d,i}^{(1)} + 2\Psi_{,k}^{(1)}v_d^{(1)k}, \end{aligned} \quad (318)$$

$$\begin{aligned} s_2 = & \nabla^2\Phi^{(2)} - 2\partial_i(\Psi^{(1)'}v_d^{(1)i}) \\ & + \nabla^2v_d^{(1)2} + 2\nabla^2\Phi^{(1)2}. \end{aligned} \quad (319)$$

In this way, we get an equation of the same form as (306) in the variable  $[\delta_d^{(2)} - 3\Psi^{(2)} - \int_0^\eta d\eta' s_1(\eta')]$  with source  $s_2$  on the R.H.S.. Its solution in Fourier space, therefore, is just as (308)

$$\begin{aligned} \delta_d^{(2)} - 3\Psi^{(2)} - \int_0^\eta d\eta' s_1(\eta') \\ = C_1 + C_2 \ln(\eta) \\ - \int_0^\eta d\eta' s_2(\eta') \eta' [\ln(k\eta') - \ln(k\eta)]. \end{aligned} \quad (320)$$

As we will see, (320) provides the generalization of the Meszaros effect at second order in perturbation theory.

**7.7.2. Initial Conditions.** In the next two sections, we will compute explicitly the expression (320) for the second-order CDM density contrast on subhorizon scales during the radiation dominated era. First, let us fix the constants  $C_1$  and  $C_2$  by appealing to the initial conditions. At  $\eta \rightarrow 0$  the L.H.S. of (320) is constant, as one can check by using the results of Section 7.5.2 and the condition of adiabaticity at second-order (see, e.g., [5, 75]) which relates the CDM density contrast at early times on superhorizon scales to the energy density fluctuations of photons by

$$\begin{aligned} \delta_d^{(2)}(0) &= \frac{3}{4}\Delta_{00}^{(2)}(0) - \frac{1}{3}(\delta_d^{(1)}(0))^2 \\ &= \frac{3}{4}\Delta_{00}^{(2)}(0) - \frac{3}{4}(\Phi^{(1)}(0))^2, \end{aligned} \quad (321)$$

where in the last step we have used (309). Therefore, we can fix  $C_2 = 0$  and

$$C_1 = \delta_d^{(2)}(0) - 3\Psi^{(2)}(0). \quad (322)$$

Equation (240) gives  $\Delta_{00}^{(2)}(0) - 4\Psi^{(2)}(0)$  in terms of the primordial non-Gaussianity parametrized by  $a_{\text{NL}}$ , and the expression for  $\Psi^{(2)}(0)$  have been already computed in (272). Thus, we find (in Fourier space)

$$\Delta_{00}^{(2)} = \left[ 2(3a_{\text{NL}} - 1) + 44 \frac{F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})}{k^2} \right] \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0), \quad (323)$$

and from (321) we derive the initial density contrast for CDM at second-order

$$\delta_d^{(2)}(0) = \left[ \frac{3}{2}(3a_{\text{NL}} - 1) + 33 \frac{F(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})}{k^2} - \frac{3}{4} \right] \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \quad (324)$$

Equation (324) together with (272) allows to compute the constant  $C_1$  as

$$\begin{aligned} C_1 &= \delta_d^{(2)}(0) - 3\Psi^{(2)}(0) \\ &= \left[ \frac{27}{2}(a_{\text{NL}} - 1) + \frac{9}{4} \right] \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (325)$$

**7.7.3. Meszaros Effect at Second Order.** We now compute the integrals over the functions  $s_1$  and  $s_2$  appearing in (320). Let us first focus on the integral  $\int_0^\eta d\eta' s_1(\eta')$ .

Notice that using the linear (300) and (301) for the CDM density and velocity perturbations, the function  $s_1(\eta')$  can be written in a more convenient way as

$$s_1(\eta) = -6\Psi^{(1)}v_{d,i}^{(1)i} + (\delta_d^{(1)2})' - 2v_d^{(1)i}\delta_{d,i}^{(1)} + 2(\Psi^{(1)}v_d^{(1)k})_{,k}, \quad (326)$$

and then

$$\begin{aligned} \int_0^\eta d\eta' s_1(\eta') \\ = & (\delta_d^{(1)}(\eta))^2 - (\delta_d^{(1)}(0))^2 \\ & + \int_0^\eta d\eta' \left[ -2v_d^{(1)i}\delta_{d,i}^{(1)} + 2(\Psi^{(1)}v_d^{(1)k})_{,k} - 6\Psi^{(1)}v_{d,i}^{(1)i} \right]. \end{aligned} \quad (327)$$

In (327), all the quantities are known being first-order perturbations: the linear gravitational potential  $\Psi^{(1)}$  for a radiation dominated era is given in (B.11), the CDM velocity perturbation corresponds to (314) and the CDM density contrast is given by (311). Thus, the integral in (327) reads (in Fourier space)

$$\begin{aligned} \int_0^\eta d\eta' \left[ -3A^{(1)} \mathbf{k}_1 \cdot \mathbf{k}_2 \frac{k_1 c_s \eta' - \sin(k_1 c_s \eta')}{k_1^3 c_s^3 \eta'^2} \ln(B^{(1)} k_2 \eta') \right. \\ + (9(\mathbf{k} \cdot \mathbf{k}_1) - 27k_1^2) \frac{k_1 c_s \eta' - \sin(k_1 c_s \eta')}{k_1^3 c_s^3 \eta'^2} \\ \left. \times \frac{\sin(k_s c_s \eta') - k_2 c_s \eta' \cos(k_2 c_s \eta')}{k_2^3 c_s^3 \eta'^3} \right] \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (328)$$

Let us recall that we are interested in the evolution of the CDM second-order density contrast on subhorizon scales during the radiation dominated epoch. Therefore, once we compute the integrals we are interested in the limit of their expression for late times ( $k\eta \gg 1$ ). For the first contribution to (328), we find that at late times it is well approximated by the expression

$$\begin{aligned} & \int_0^\eta d\eta' 3A^{(1)}(\mathbf{k}_1 \cdot \mathbf{k}_2) \frac{k_1 c_s \eta' - \sin(k_1 c_s \eta')}{k_1^3 c_s^3 \eta'^2} \ln(B^{(1)} k_2 \eta') \\ & \simeq 3A^{(1)} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2 c_s^2} \left[ 2.2 \left( -\frac{1.2}{2} [\ln(k_1 c_s \eta)]^2 \right. \right. \\ & \quad \left. \left. + \ln(B^{(1)} k_2 \eta) \ln(k_1 c_s \eta) \right) \right]. \end{aligned} \quad (329)$$

We have computed also the remaining integral in (328), but it turns out to be negligible compared to (329). The reason is that the integrand oscillates with an amplitude decaying in time as  $\eta^{-3}$ , and thus it leads just to a constant (the argument is the same we used at linear order to compute the integrals in (308)). Thus, we can write

$$\begin{aligned} & \int_0^\eta d\eta' s_1(\eta') \\ & = \left( \delta_d^{(1)}(\eta) \right)^2 - \left( \delta_d^{(1)}(0) \right)^2 - 3A^{(1)} \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2 c_s^2} \\ & \quad \times \left[ 2.2 \left( -\frac{1.2}{2} [\ln(k_1 c_s \eta)]^2 + \ln(B^{(1)} k_2 \eta) \ln(k_1 c_s \eta) \right) \right] \\ & \quad \times \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (330)$$

We now compute the integrals over the function  $s_2(\eta)$  given in (319). Since at late times  $\phi^{(1)2} \sim 1/\eta^4$  and  $(\Psi^{(1)'} v_d^{(1)i})_{,i} \sim 1/\eta^3$  the main contribution to the integral will come from the two remaining terms,  $\Phi^{(2)}$  and  $v_d^{(1)2}$ , whose amplitudes scale at late times as  $1/\eta^2$

$$s_2 \simeq \nabla^2 \Phi^{(2)} + \nabla^2 v_d^{(1)2}. \quad (331)$$

Two are the integrals that we have to compute

$$\int_0^\eta d\eta' s_2(\eta') \eta' \ln(k\eta'), \quad (332)$$

and the one multiplying  $\ln(k\eta)$

$$\int_0^\eta d\eta' s_2(\eta') \eta'. \quad (333)$$

Let us first consider the contributions from  $\nabla^2 v_d^{(1)2}$ . The second integral is easily computed using the expression (314) for the linear CDM velocity. We find that at late times the dominant term is (for  $k\eta \gg 1$ )

$$\begin{aligned} & - \int_0^\eta d\eta' \nabla^2 v_d^{(1)2} \eta' \equiv - \left[ \frac{9}{c_s^4} k^2 \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2 k_2^2} \ln(k\eta) \right] \\ & \quad \times \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (334)$$

The first integral can be computed by making the following approximation: we split the integral between  $0 < k\eta < 1$  and  $k\eta > 1$  and for  $0 < k\eta < 1$ , we use the asymptotic expression

$$v_d^{(1)i} \simeq -\frac{1}{2} i k^i \eta \Psi_{\mathbf{k}}^{(1)}(0) \quad (k\eta \ll 1), \quad (335)$$

while for  $k\eta > 1$  we use the limit

$$v_d^{(1)i} \simeq -i \frac{3}{c_s^2} \frac{k^i}{k} \frac{1}{k\eta} \Psi_{\mathbf{k}}^{(1)}(0) \quad (k\eta \gg 1). \quad (336)$$

The the integral for  $0 < k\eta < 1$  just gives a constant, while the integral for  $k\eta > 1$  brings the dominant contribution at late times being proportional to  $[\ln(k\eta)]^2$  so that we can write ( $k\eta \gg 1$ )

$$\begin{aligned} & \int_0^\eta d\eta' \nabla^2 v_d^{(1)2} \eta' \ln(k\eta') \\ & = \frac{9}{2c_s^4} k^2 \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2 k_2^2} [\ln(k\eta)]^2 \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (337)$$

As far as the contribution to the integrals (332) and (333) due to  $\nabla^2 \Phi^{(2)}$  is concerned we have just to keep track of the initial condition provided by the primordial non-Gaussianity. We have verified that all the other terms give a negligible contribution. This is easy to understand: the integrand function on large scale is a constant while at late times it oscillates with decreasing amplitudes as  $\eta^{-2}$ , and thus the integrals will tend asymptotically to a constant. We find that

$$\int_0^\eta d\eta' \nabla^2 \Phi^{(2)} \eta' \simeq -9\Phi^{(2)}(0), \quad (338)$$

$$\int_0^\eta d\eta' \nabla^2 \Phi^{(2)} \eta' \ln(k\eta') \simeq \left( -9 + 9\gamma - 9\frac{\ln 3}{2} \right) \Phi^{(2)}(0),$$

where  $\gamma = 0.577\dots$  is the Euler constant, and  $\Phi^{(2)}(0)$  is given by (268).

Therefore, from (337), (334), and (338) we find that for  $k\eta \gg 1$

$$\begin{aligned} & \int_0^\eta d\eta' s_2(\eta') \eta' [\ln(k\eta') - \ln(k\eta)] \\ & = -\frac{9}{2c_s^4} k^2 \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2 k_2^2} [\ln(k\eta)]^2 \\ & \quad \times \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0) + 9\Phi^{(2)} \\ & \quad \times (0) \left( -9 + 9\gamma - 9\frac{\ln(3)}{2} \right) \ln(k\eta) \Phi^{(2)}(0). \end{aligned} \quad (339)$$

Let us collect the results of (325), (330), and (339) into (320). We find that for  $k\eta \gg 1$

$$\begin{aligned} \delta_d^{(2)}(k\eta \gg 1) &= \left[ -3(a_{\text{NL}} - 1)A_1 \ln(B_1 k\eta) + A_1^2 \ln(B_1 k_1 \eta) \ln(B_1 k_2 \eta) \right. \\ &\quad \left. + \left[ -\frac{3}{2}A_1 \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2 c_s^2} 2.2 \left( -\frac{1.2}{2} [\ln(k_1 c_s \eta)]^2 \right. \right. \right. \\ &\quad \left. \left. \left. + \ln(B_1 k_2 \eta) \ln(k_1 c_s \eta) \right) \right. \right. \\ &\quad \left. \left. + (1 \leftrightarrow 2) \right] + \frac{9}{2c_s^4} k^2 \frac{\mathbf{k}_1 \cdot \mathbf{k}_2}{k_1^2 k_2^2} [\ln(k\eta)]^2 \right] \\ &\quad \times \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (340)$$

Notice that in (320) we have neglected  $\Psi^{(2)}$ , which decays on subhorizon scales during the radiation dominated epoch (see (250), and we have used (309) and (311). Equation (340) represents the second-order Meszaros effect: the CDM density contrast on small scales (inside the horizon) slowly grows starting from the initial conditions that, at second-order, are set by the primordial non-Gaussianity parameter  $a_{\text{NL}}$ . As one could have guessed the primordial non-Gaussianity is just transferred linearly. The other terms scale in time as a logarithm squared. We stress that the computation of these terms allows one to derive the full transfer function for the matter perturbations at second order accounting for the dominant second-order corrections. In the next section, we will use (340) to fix the initial conditions for the evolution on subhorizon scales of the photons density fluctuations  $\Delta_{00}^{(2)}$  after the equality epoch.

*7.7.4. Second-Order CMB Anisotropies for Modes Crossing the Horizon During the Radiation Epoch.* In this section, we derive the energy density perturbations  $\Delta_{00}^{(2)}$  of the photons during the matter dominated epoch, for the modes that cross the horizon before equality. In Section 7.6, we have already solved the problem assuming matter domination for modes crossing the horizon after equality. Thus, it is sufficient to take the result (285) and replace the initial conditions

$$\begin{aligned} \Delta_{00}^{(2)} &= -4(1+R)\Psi_m^{(2)}(0) \\ &\quad + \left[ A + (8+4R)\Psi_m^{(2)}(0) \right] \cos(kc_s \eta) + B \sin(kc_s \eta) \\ &\quad - \frac{2}{7} G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) \eta^2 (1+R) \left( \frac{9}{10} \right)^2 \Psi_{\mathbf{k}_1}^{(1)}(0) \Psi_{\mathbf{k}_2}^{(1)}(0). \end{aligned} \quad (341)$$

where we have restored the generic integration constants  $A$  and  $B$ ,  $\Psi^{(1)}$  is the linear gravitational potential (which is constant for the matter era) and  $\Psi_m^{(2)}(0)$  represents the initial condition for the second-order gravitational potential fixed at some time  $\eta_i > \eta_{\text{eq}}$ . Equation (340) allows to fix the proper initial conditions for the gravitational potentials

on subhorizon scales (accounting for the fact that around the equality epoch they are mainly determined by the CDM density perturbations). At linear order this is achieved by solving the equation for  $\delta_d^{(1)}$  which is obtained from (305) and the (0-0) Einstein equation which reads (see (29))

$$3\mathcal{H}\Psi^{(1)'} + 3\mathcal{H}^2\Psi^{(1)} - \nabla^2\Psi^{(1)} = -\frac{3}{2}\mathcal{H}^2\left(\frac{\rho_d}{\rho}\delta_d^{(1)} + \frac{\rho_\gamma}{\rho}\Delta_{00}^{(1)}\right). \quad (342)$$

On small scales one neglects the time derivatives of the gravitational potential in (305) and (342) to obtain

$$\delta_d^{(1)''} + \mathcal{H}\delta_d^{(1)'} = \frac{3}{2}\mathcal{H}^2\delta_d^{(1)}, \quad (343)$$

where we have also dropped the contribution to the gravitational potential from the radiation component. The solution of this equation is matched to the value that  $\delta_d^{(1)}$  has during the radiation dominated epoch on subhorizon scales, (311), and one finds that for  $\eta \gg \eta_{\text{eq}}$  on subhorizon scales the gravitational potential remains constant with

$$\Psi_{\mathbf{k}}^{(1)}(\eta > \eta_{\text{eq}}) = \frac{\ln(0.15k\eta_{\text{eq}})}{(0.27k\eta_{\text{eq}})^2} \Psi_{\mathbf{k}}^{(1)}(0). \quad (344)$$

We skip the details of the derivation of (344) since it is a standard computation that the reader can find, for example, in [70]. Since around  $\eta_{\text{eq}}$  the dark matter begins to dominate, an approximation to the result (344) can be simply achieved by requiring that during matter domination the gravitational potential remains constant to a value determined by the density contrast (311) at the equality epoch

$$\nabla^2\Psi^{(1)} \Big|_{\eta_{\text{eq}}} \simeq \frac{3}{2}\mathcal{H}^2\delta_d^{(1)} \Big|_{\eta_{\text{eq}}}, \quad (345)$$

from (342) on small scales, leading to

$$\begin{aligned} \Psi_{\mathbf{k}}^{(1)}(\eta > \eta_{\text{eq}}) &\simeq -\frac{6}{(k\eta_{\text{eq}})^2} \delta_d^{(1)} \Big|_{\eta_{\text{eq}}} \\ &= \frac{\ln(B_1 k\eta_{\text{eq}})}{(0.13k\eta_{\text{eq}})^2} \Psi_{\mathbf{k}}^{(1)}(0), \end{aligned} \quad (346)$$

where we used  $a(\eta) \propto \eta^2$  during matter domination and (311) with  $A_1 = -9.6$  and  $B_1 = 0.44$ .

At second order, we follow a similar approximation. The general solution for the evolution of the the second-order gravitational potential  $\Psi^{(2)}$  for  $\eta > \eta_{\text{eq}}$  is given by (B.6). We have to determine the initial conditions for those modes that

cross the horizon during the radiation epoch. The (0 - 0) Einstein equation reads

$$\begin{aligned}
& -\frac{3}{2}\mathcal{H}^2\left(\frac{\rho_d}{\rho}\delta_d^{(2)}+\frac{\rho_\gamma}{\rho}\Delta_{00}^{(2)}\right) \\
& = 3\mathcal{H}\Psi^{(2)'}+3\mathcal{H}^2\Phi^{(2)}-\nabla^2\Psi^{(2)} \\
& - 6\mathcal{H}^2\left(\Phi^{(1)}\right)^2-12\mathcal{H}\Phi^{(1)}\Psi^{(1)'} \\
& - 3\left(\Psi^{(1)'}\right)^2+\partial_i\Psi^{(1)}\partial^i\Psi^{(1)}-4\Psi^{(1)}\nabla^2\Psi^{(1)}.
\end{aligned} \tag{347}$$

We fix the initial conditions with the matching at equality (neglecting the radiation component)

$$\begin{aligned}
& \nabla^2\Psi^{(2)}-\partial_i\Psi^{(1)}\partial^i\Psi^{(1)}+4\Psi^{(1)}\nabla^2\Psi^{(1)}\Big|_{\eta_{\text{eq}}} \\
& \simeq\frac{3}{2}\mathcal{H}^2\delta_d^{(2)}\Big|_{\eta_{\text{eq}}},
\end{aligned} \tag{348}$$

where for small scales we neglected the time derivatives in (347). Using (340) to evaluate  $\delta_d^{(2)}|_{\eta_{\text{eq}}}$  and (344) to evaluate  $\Psi_{\mathbf{k}}^{(1)}(\eta_{\text{eq}})$ , we find in Fourier space

$$\begin{aligned}
& \Psi^{(2)}(\eta_{\text{eq}}) \\
& = \left[ -3(a_{\text{NL}}-1)\frac{\ln(B_1k\eta_{\text{eq}})}{(0.13k\eta_{\text{eq}})^2}+\left(\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k^2}-4\right) \right. \\
& \quad \times\frac{\ln(0.15k_1\eta_{\text{eq}})}{(0.27k_1\eta_{\text{eq}})^2}\frac{\ln(0.15k_2\eta_{\text{eq}})}{(0.27k_2\eta_{\text{eq}})^2} \\
& \quad +A_1\ln(B_1k_1\eta_{\text{eq}})\frac{\ln(B_1k_2\eta_{\text{eq}})}{(0.13k\eta_{\text{eq}})^2}-\frac{27}{c_s^4}k^2 \\
& \quad \times\frac{\mathbf{k}_1\cdot\mathbf{k}_2}{k_1^2k_2^2}\frac{[\ln(k\eta_{\text{eq}})]^2}{(k\eta_{\text{eq}})^2}+\frac{3\mathbf{k}_1\cdot\mathbf{k}_2}{2c_s^2k_1^2} \\
& \quad \left. \times 2.2\left[\frac{1.2}{2}\frac{[\ln(k_1c_s\eta_{\text{eq}})]^2}{(0.13k\eta_{\text{eq}})^2}-\ln(k_1c_s\eta_{\text{eq}}) \right. \right. \\
& \quad \left. \left. \times\frac{\ln(B_1k_2\eta_{\text{eq}})}{(0.13\eta_{\text{eq}})^2}+(1\leftrightarrow 2)\right]\right]\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0).
\end{aligned} \tag{349}$$

In (341), the initial condition  $\Psi_m^{(2)}(0)$  is given by (349) and  $\Psi^{(1)}$  is given by (344). The integration constants can be fixed by comparing at  $\eta \simeq \eta_{\text{eq}}$  the oscillating part of (341) to the solution  $\Delta_{00}^{(2)}$  obtained for modes crossing the horizon before

equality and for  $\eta < \eta_{\text{eq}}$ , (256). Thus, for  $\eta \gg \eta_{\text{eq}}$  and  $k \gg \eta_{\text{eq}}^{-1}$ , we find that

$$\begin{aligned}
\Delta_{00}^{(2)} & = -4(1+R)\Psi^{(2)}(\eta_{\text{eq}})+\bar{A}\cos(kc_s\eta) \\
& -\frac{2}{7}(1+R)G(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k})\eta^2\Psi_{\mathbf{k}_1}^{(1)}(\eta_{\text{eq}})\Psi_{\mathbf{k}_2}^{(1)}(\eta_{\text{eq}}),
\end{aligned} \tag{350}$$

where

$$\begin{aligned}
\bar{A} & = 6\Psi^{(2)}(0)-\frac{6(\mathbf{k}\cdot\mathbf{k}_1)(\mathbf{k}\cdot\mathbf{k}_2)}{c_s^4k_1k_2\cos(kc_s\eta_{\text{eq}})}\Psi_{\mathbf{k}_1}^{(1)}(0)\Psi_{\mathbf{k}_2}^{(1)}(0) \\
& \times\left[2k_1k_2\cos(k_1c_s\eta_{\text{eq}})\cos(k_2c_s\eta_{\text{eq}})-2k_1k_2\cos(kc_s\eta_{\text{eq}}) \right. \\
& \quad \left. +(k_1^2+k_2^2-k^2)\sin(k_1c_s\eta_{\text{eq}})\sin(k_2c_s\eta_{\text{eq}})\right] \\
& \times\frac{1}{k_1^4+k_2^4+k^4-2k_1^2k_2^2-2k_1^2k^2-2k_2^2k^2},
\end{aligned} \tag{351}$$

and  $\Psi^{(2)}(0)$  is given in (272).

## 8. Secondary Effects and Contamination to Primordial NG

Given that a detection of a sizable primordial non-Gaussianity (and its shape) would represent a real breakthrough into the understanding of the dynamics of the universe during its very first stages, it is crucial that all sources of contamination to the primordial signal are well understood and kept under control. In fact, any nonlinearities can make initially Gaussian perturbations non-Gaussian. Such nonprimordial effects can thus complicate the extraction of the primordial non-Gaussianity: we have to be sure we are not ascribing a primordial origin to a signal that is extracted from the CMB (or LSS) data using estimators of non-Gaussianity when that signal has a different origin. Moreover, as stressed in Section 7.1, we must always specify of which primordial non-Gaussianity we study the contamination from the nonprimordial sources (e.g., primordial non-Gaussianity of ‘‘local’’, ‘‘equilateral’’, or ‘‘folded’’ shape).

Broadly speaking, nonprimordial sources of non-Gaussianity can be classified into four categories: instrumental systematic effects; residual foregrounds and unresolved point sources; some well-known secondary CMB anisotropies, such as Sunyaev-Zel’dovich (SZ) effect, gravitational lensing, Rees-Sciama effect, and finally previously unknown effects coming from nonlinearities in the Boltzmann equations, which are related to the nonlinear nature of General Relativity and to the nonlinear dynamics of the photon-baryon system. This paper focuses on the last category, but in this section, we offer also a summary of some recent results of the other types of possible contaminations. Before doing that, however, it is important to be more precise about how secondary effects may impact the extraction of a primordial non-Gaussian signal. Mainly, they act in two ways. They might ‘‘mimic’’ a three-point correlation function

similar in shape to the primordial one. This would produce a *bias* or a *contamination* to the estimator of primordial non-Gaussianity. On the other, hand secondary effects might increase the variance of the estimator without contributing to the signal-to-noise ratio. In other words, in this case they degrade the signal-to-noise ratio. Let us define in a quantitative way how to characterize these effects.

*8.1. Signal-to-Noise Ratio, Shapes, and Contamination to Primordial Non-Gaussianity.* The angular bispectrum,  $B_{\ell_1\ell_2\ell_3}$ , the harmonic transform of the angular three-point function, of the CMB anisotropies is often used to measure non-Gaussianity (see, e.g., [5]). From the usual spherical harmonic coefficients of the temperature anisotropies

$$a_{\ell m} = \int d^2\mathbf{n} Y_{\ell m}^*(\mathbf{n}) \frac{\Delta T(\mathbf{n})}{T}, \quad (352)$$

the angle-averaged bispectrum is given by (for more details see, e.g., [43])

$$B_{\ell_1\ell_2\ell_3} \equiv \sum_{\text{all } m} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle, \quad (353)$$

where the brackets indicate ensemble average.

We shall quantify the degree to which the primordial and the secondary bispectra are correlated, as well as their expected signal-to-noise ratio following a method that as now become standard [41, 43, 76]. Namely, in the limit of weak non-Gaussianity, one can introduce the Fisher matrix for the amplitudes of the bispectra,  $F_{ij}$ , given by

$$F_{ij} \equiv \sum_{2 \leq \ell_1 \leq \ell_2 \leq \ell_3} \frac{B_{\ell_1\ell_2\ell_3}^{(i)} B_{\ell_1\ell_2\ell_3}^{(j)}}{\sigma_{\ell_1\ell_2\ell_3}^2}, \quad (354)$$

where the variance of the bispectrum is

$$\sigma_{\ell_1\ell_2\ell_3}^2 \equiv \langle B_{\ell_1\ell_2\ell_3}^2 \rangle - \langle B_{\ell_1\ell_2\ell_3} \rangle^2 \approx C_{\ell_1} C_{\ell_2} C_{\ell_3} \Delta_{\ell_1\ell_2\ell_3}, \quad (355)$$

and  $\Delta_{\ell_1\ell_2\ell_3}$  takes values 1, 2, and 6 when all  $\ell$ 's are different, two of them are equal and all are the same, respectively. The power spectrum,  $C_\ell$ , is the sum of the theoretical CMB and the detector noise.

The signal-to-noise ratio is given by

$$\left(\frac{S}{N}\right)_i = \frac{1}{\sqrt{F_{ii}^{-1}}}, \quad (356)$$

and we define the cross-correlation coefficient between different shapes  $i$  and  $j$ ,  $r_{ij}$ , as

$$r_{ij} \equiv \frac{F_{ij}^{-1}}{\sqrt{F_{ii}^{-1} F_{jj}^{-1}}}, \quad (357)$$

which does not depend on the amplitudes of the different bispectra, but on the shapes, thus measuring how similar are two bispectra  $i$  and  $j$ . One can also define a degradation parameter  $d_i = F_{ii} F_{ii}^{-1}$  for a degraded  $(S/N)$  due to  $r_{ij}$ .

If a secondary bispectra has some correlation with the primordial one, one expects that the signal-to-noise ratio will be degraded: if one does not account for any secondary bispectra  $(S/N)_0 = \sqrt{F_{\text{prim,prim}}}$  while, marginalizing over a (single) secondary bispectrum,  $(S/N)_{\text{prim}}$  gets modified from its zero-order value to  $(S/N)_{\text{prim}} = (S/N)_0 (1 - r_{ij}^2)^{1/2}$ . This means that the minimum detectable value of the primordial nonlinearity parameter  $f_{\text{NL}}$  which is obtained by imposing  $(S/N)_{\text{prim}} = 1$  (or the  $1-\sigma$  uncertainty on  $f_{\text{NL}}$  given by  $\delta f_{\text{NL}} = \sqrt{(F_{\text{prim,prim}}^{-1})_{f_{\text{NL}}=1}}$ ) gets shifted by a quantity  $\Delta f_{\text{NL}} / (f_{\text{NL}})_0 = (1 - r_{ij}^2)^{-1/2} - 1$ .

How much does a given secondary bispectra contaminate the extraction of the primordial bispectrum? If, for example, the predicted shape of the secondary bispectra is sufficiently different from that of the primordial bispectrum, then one would hope that the contamination would be minimal. We can quantify the contamination of the primordial bispectrum as follows: we fit the primordial bispectrum template to the second-order bispectrum, and find the best-fitting  $f_{\text{NL}}^{\text{con}}$  ("con" stands for contamination) by minimizing the  $\chi^2$  given by

$$\chi^2 = \sum_{2 \leq \ell_1 \leq \ell_2 \leq \ell_3} \frac{(f_{\text{NL}} B_{\ell_1\ell_2\ell_3}^{\text{prim}} - B_{\ell_1\ell_2\ell_3}^{\text{2nd}})^2}{\sigma_{\ell_1\ell_2\ell_3}^2}, \quad (358)$$

with respect to  $f_{\text{NL}}$ . Here,  $B_{\ell_1\ell_2\ell_3}^{\text{prim}}$  is the primordial bispectrum with amplitude  $f_{\text{NL}} = 1$  [43]. We obtain

$$f_{\text{NL}}^{\text{con}} = \frac{1}{N} \sum_{2 \leq \ell_1 \leq \ell_2 \leq \ell_3} \frac{B_{\ell_1\ell_2\ell_3}^{\text{2nd}} B_{\ell_1\ell_2\ell_3}^{\text{prim}}}{\sigma_{\ell_1\ell_2\ell_3}^2}, \quad (359)$$

$$N = \sum_{2 \leq \ell_1 \leq \ell_2 \leq \ell_3} \frac{(B_{\ell_1\ell_2\ell_3}^{\text{prim}})^2}{\sigma_{\ell_1\ell_2\ell_3}^2}.$$

This is the value of  $f_{\text{NL}}$  one would find, if one did not know that the primordial bispectrum did not exist but there was only the secondary bispectrum. In other words, the effective nonlinearity parameter  $f_{\text{NL}}^{\text{con}}$  gives the value that a bispectrum estimator designed for constraining primordial non-Gaussianity would measure due to the presence of the secondary bispectrum  $B_{\ell_1\ell_2\ell_3}^{\text{2nd}}$ . As one can see  $f_{\text{NL}}^{\text{con}}$  turns out to be proportional to the mixed entry of the Fisher matrix, since this governs the correlation between different types of bispectra on the basis of their shape dependence.

*8.2. Some "well-known" Secondary Sources of Non-Gaussianity.* Instrumental systematic effects are one of the principal issues in the search for primordial non-Gaussianity. Since they depend on the specific instrument, here we do not intend to go into any detail, but we refer the reader to [77] as an interesting example of this kind of effects for the currently flying mission of the satellite *Planck*. Residual foregrounds and unresolved point sources represent other important possible contaminants and particular attention have been devoted to them in the analysis of the WMAP data [44, 57, 72]. For an experiment like WMAP secondary

effects such as Sunyaev-Zel'dovich effect, gravitational lensing, Integrated-Sachs-Wolfe (or Rees-Sciama) effect should be not so relevant for  $\ell < 500$  with a bias of just 1.5 to 2 [78] for local primordial non-Gaussianity (see also [43, 44]). (In order to understand the relative importance of the bias from a given secondary effect let us recall that, including the information from polarization, the minimum detectable (local)  $f_{\text{NL}}^{\text{loc}}$  for an experiment like WAMP is of the order of 10 while for *Planck* it is of the order of 3; the minimum detectable value for an equilateral primordial non-Gaussianity is  $f_{\text{NL}}^{\text{eq}} \simeq 30$  for *Planck*, see, for example, [4, 43, 71, 79].) However, for higher resolution and more sensitive experiments like *Planck*, such effects must be taken into account with care. Here, we report the results of some recent analyses about the contamination to primordial non-Gaussianity from some secondary effects. Such a brief summary is by no means exhaustive. An interesting case is given by the cross-correlation of the ISW and the lensing effects, see [41]. In the detailed analysis of [80], it is shown that the ISW-lensing bispectrum produces a bias of  $f_{\text{NL}}^{\text{cont}} \simeq 9$  to local primordial non-Gaussianity (see also [78]), while the bias to equilateral primordial non-Gaussianity is negligible (see also [44, 78]). The impact is mainly on the local type of primordial non-Gaussianity because the ISW-lensing correlation correlates the large-scale gravitational potential fluctuations sourcing the ISW effect with the small scale lensing effects of the CMB, thus producing a bispectrum which peaks in the squeezed configurations as the local shape. As far as the increase of the variance of the estimator of primordial non-Gaussianity induced by CMB lensing is concerned it turns out to be of the order of 20% for an experiment like *Planck* (see [79, 80]). Notice that interestingly enough, gravitational lensing can also have another peculiar effect on the extraction of primordial non-Gaussianity: it can modify its shape by smoothing its acoustic features. However, according to the analysis of [80] (and contrary to the results of [81]) such a “distorsion” of the shape should be small, of the order of 10% for  $l < 2000$ .

A similar analysis has been performed looking also at scales where nonlinear effects can become relevant. In this case a bispectrum from the correlation of the Rees-Sciama and lensing effects has been studied in [82], and it has been found that in this case the bias to a local primordial non-Gaussianity corresponds to  $f_{\text{NL}}^{\text{cont}} \simeq 10$ , agreeing with the results of [80] where the two analyses can be compared.

Another secondary source of contamination that has been studied quite recently in [83] deals with bispectra generated by correlations of number density and lensing magnification of radio and SZ point sources with the ISW effect. Also, in this case the large-scale modulation of small-scale number density due to fluctuations which source the ISW generates a bispectrum which peaks on squeezed configurations and it has been found that it corresponds to a contamination of local primordial non-Gaussianity of  $f_{\text{NL}}^{\text{cont}} \simeq 1.5$ , which should be not so relevant for an experiment like *Planck*. The novelty of this effect is that it is different from the usual bispectrum of point sources when treated with a Poisson distribution (see, e.g., [43]).

Other secondary effects include, for example, the SZ-SZ-SZ bispectrum [84] and non-Gaussianities from the kinetic SZ and Ostriker-Vishniac effect [76, 85]. However, the SZ-SZ-SZ bispectrum can be considered as an extra Poisson contribution to the unresolved point sources for  $l < 1500$  [78]. Moreover, recently an analysis of the bispectrum generated from inhomogeneous reionization has been performed in [86], where it has been shown that it can give rise to a very small contamination to the local primordial non-Gaussianity of  $f_{\text{NL}}^{\text{cont}} \simeq -0.1$ .

Finally let us also recall some recent studies of a specific effect arising at second-order in perturbation theory, when looking at the fluctuations in the Boltzmann equations for the photon-baryon fluid. In [63, 87–89], the perturbations to the phase of recombination between electrons and baryons has been considered (“inhomogeneous recombination”). It turns out that the electron density perturbation can be of the order of 5 times bigger than the one in the density of the baryons. This gives rise to the prospect of a non-Gaussianity which corresponds to  $|f_{\text{NL}}^{\text{cont}}| \simeq 5$ . In fact, the detailed study of this effect shows that the contamination is much smaller. The bispectrum from the perturbed recombination peaks in the squeezed configuration with a contamination to the primordial non-Gaussianity of the local type of  $|f_{\text{NL}}^{\text{cont}}| \simeq 0.7$ .

Let us conclude this section by mentioning that the nonlinearities emerging from secondary effects can be very interesting by themselves, since they carry a lot of information about the evolution of the universe and the growth of structures after recombination till very low redshift. Detailed examples and reviews about this aspect can be found, for example, in [90–92].

## 9. How to Analytically Estimate the Signal-to-Noise Ratio from the NG at Recombination and Its Contamination to the Primordial NG

As we have seen in Section 7, the dynamics at recombination is quite involved because all the nonlinearities in the evolution of the baryon-photon fluid at recombination and the ones coming from general relativity should be accounted for. Such a contribution is so relevant because it represents a major part of the second-order radiation transfer function which must be determined in order to have a complete control of both the primordial and nonprimordial part of NG in the CMB anisotropies. The NG generated at the surface of last scattering comprises various effects. In this section, we devote our attention to one particular relevant contribution, the one coming from the nonlinear evolution of the second-order gravitational potential which grows in time on small scales. The analysis of this contribution offers the opportunity to understand how some secondary effects can contaminate mostly a given type of primordial NG, while leaving almost untouched other different forms of primordial NG. In this case, this effect is a causal one, developing on small scales, so we expect that the NG it generates will be mainly of the equilateral type, rather than of local type. Therefore, a reasonable question is to

which extent the NG from recombination alters the possible detection of the primordial NG of the equilateral type. The goal of this Section is therefore to illustrate how to estimate in a semianalytical way the contribution to NG from recombination.

*9.1. Signal-to-Noise Ratio for the Primordial Equilateral Bispectrum.* In this subsection we wish to recover the estimate for the signal-to-noise ratio ( $S/N$ ) given in [44] for the primordial bispectra of “equilateral” type [93] by adopting a simple model. In other words, we test the goodness of the semi-analytical model we will be using in the next Section to estimate the bispectrum from the recombination era.

Our starting point is the primordial equilateral bispectrum [93]

$$\begin{aligned} & \langle \Phi(\mathbf{k}_1)\Phi(\mathbf{k}_2)\Phi(\mathbf{k}_3) \rangle \\ & = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{\text{equil}}(k_1, k_2, k_3), \end{aligned} \quad (360)$$

where

$$\begin{aligned} & B_{\text{equil}}(k_1, k_2, k_3) \\ & = f_{\text{NL}}^{\text{equil}} \cdot 6A^2 \cdot \left( -\frac{1}{k_1^3 k_2^3} - \frac{1}{k_1^3 k_3^3} - \frac{1}{k_2^3 k_3^3} \right. \\ & \quad \left. - \frac{2}{k_1^2 k_2^2 k_3^2} + \frac{1}{k_1 k_2^2 k_3^3} + (5 \text{ perm.}) \right), \end{aligned} \quad (361)$$

and the permutations act only on the last term in parentheses. The parameter  $f_{\text{NL}}^{\text{equil}}$  quantifies the level of NG while  $A = 17.46 \times 10^{-9}$  is the amplitude of the primordial gravitational potential power spectrum computed at first order

$$\langle \Phi^{(1)}(\mathbf{k}_1)\Phi^{(1)}(\mathbf{k}_2) \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2) P(k_1), \quad (362)$$

with  $P(k) = A/k^3$ . Since the signal-to-noise ratio ( $S/N$ ) will be some function of the maximum multipole a given experiment can reach,  $\ell_{\text{max}} \gg 1$ , we can use the flat-sky approximation [79, 94] and write for the bispectrum

$$\langle a(\vec{\ell}_1) a(\vec{\ell}_2) a(\vec{\ell}_3) \rangle = (2\pi)^2 \delta^{(2)}(\vec{\ell}_{123}) B(\ell_1, \ell_2, \ell_3), \quad (363)$$

where  $\vec{\ell}_{123} = \vec{\ell}_1 + \vec{\ell}_2 + \vec{\ell}_3$ , with [25]

$$\begin{aligned} & B_{\text{equil}}(\ell_1, \ell_2, \ell_3) \\ & = \frac{(\eta_0 - \eta_r)^2}{(2\pi)^2} \int dk_1^z dk_2^z dk_3^z \delta^{(1)} \\ & \quad \times (k_{123}^z) B_{\text{equil}}(k_1', k_2', k_3') \tilde{\Delta}^T(\ell_1, k_1^z) \tilde{\Delta}^T(\ell_2, k_2^z) \tilde{\Delta}^T(\ell_3, k_3^z), \end{aligned} \quad (364)$$

where  $k'$  means  $k$  evaluated such that  $\vec{k}^{\parallel} = \vec{\ell}/(\eta_0 - \eta_r)$  and

$$\begin{aligned} \tilde{\Delta}^T(\ell, k^z) & = \int_0^{\eta_0} \frac{d\eta}{(\eta_0 - \eta)^2} \\ & \quad \times S\left(\sqrt{(k^z)^2 + \ell^2/(\eta_0 - \eta)^2}, \eta\right) e^{ik^z(\eta_r - \eta)}, \end{aligned} \quad (365)$$

is the radiation transfer function defined by the CMB source function  $S(k, \eta)$ . In this notation,  $\eta_0$  and  $\eta_r$  represent the present-day and the recombination conformal time, respectively and  $k^z$  and  $\vec{k}^{\parallel}$  are the momentum components perpendicular and parallel, respectively, to the plane orthogonal to the line-of-sight.

The ( $S/N$ ) ratio in the flat-sky formalism is [79, 94]

$$\begin{aligned} \left(\frac{S}{N}\right)^2 & = \frac{f_{\text{sky}}}{\pi} \frac{1}{(2\pi)^2} \int d^2\ell_1 d^2\ell_2 d^2\ell_3 \delta^{(2)} \\ & \quad \times \left(\vec{\ell}_{123}\right) \frac{B_{\text{equil}}^2(\ell_1, \ell_2, \ell_3)}{6 C(\ell_1) C(\ell_2) C(\ell_3)}, \end{aligned} \quad (366)$$

where  $f_{\text{sky}}$  stands for the portion of the observed sky. In order to compute the bispectrum  $B_{\text{equil}}(\ell_1, \ell_2, \ell_3)$  and the power spectrum  $C(\ell)$ , we adopt the following model

$$a(\vec{\ell}) = \int \frac{dk^z}{2\pi} e^{ik^z(\eta_0 - \eta_r)} \Phi(\mathbf{k}') \tilde{\Delta}^T(\ell, k^z) \quad (367)$$

where we mimic the effects of the transfer function on small scales as

$$\tilde{\Delta}^T(\ell, k^z) = a (\eta_0 - \eta_r)^{-2} e^{-1/2(\ell/\ell_*)^{1.2}} e^{-1/2(|k^z/k_*|)^{1.2}}, \quad (368)$$

That is, a simple exponential and a normalization coefficient  $a$  to be determined to match the amplitude of the angular power spectrum at the characteristic scale  $\ell \simeq \ell_* = k_*(\eta_0 - \eta_r)$ . (We could equally choose a transfer function as  $\tilde{\Delta}^T(\ell, k^z) = a(\eta_0 - \eta_r)^{-2} e^{-1/2(\ell/\ell_*)^{1.2}} \theta(k_* - |k^z|)$ , the relevant approximation being that the integral over  $k^z$  is cut at the scale  $k_*$ .) It is important to make clear what are the reasons underlying the choice of such a model. When computing the ( $S/N$ ), (367) with  $\ell_* = k_*(\eta_0 - \eta_r) \simeq 750$  and  $a \simeq 3$  is able to account for the combined effects of “radiation driving”, which occurs at  $\ell > \ell_{\text{eq}} \simeq 160$  and boosts the angular power spectrum with respect to the Sachs-Wolfe plateau, and the effects of Silk damping which tend to suppress the CMB anisotropies for scales  $\ell > \ell_D \simeq 1300$ . The combination of these effects produces a decrease in the angular power spectrum from a scale  $\ell_* \simeq 750$ . (The choice of the exponent 1.2 derives from the study of the diffusion damping envelope in [95]). The power spectrum in the flat-sky approximation is given by  $\langle a(\vec{\ell}_1) a(\vec{\ell}_2) \rangle = (2\pi)^2 \delta^{(2)}(\vec{\ell}_{12}) C(\ell_1)$  with

$$C(\ell) = \frac{(\eta_0 - \eta_r)^2}{(2\pi)} \int dk^z \left| \tilde{\Delta}^T(\ell, k^z) \right|^2 P(k). \quad (369)$$

The exponential of the transfer function for (367) allows to cut off the integral for  $k \simeq k_*$  and one finds (see also [79])

$$C(\ell) = a^2 \frac{A}{\pi \ell^2} \frac{e^{-(\ell/\ell_*)^{1.2}}}{\sqrt{1 + \ell^2/\ell_*^2}} \simeq a^2 \frac{A}{\pi} \frac{\ell_*}{\ell^3} e^{-(\ell/\ell_*)^{1.2}}, \quad (370)$$

where the last equality holds for  $\ell \gg \ell_*$ . To compute the bispectrum we proceed in a similar way. One first uses the Dirac deltas,  $\delta^{(1)}(k_{123}^z)$  and  $\delta^{(2)}(\vec{\ell}_{123})$ . Then, it proves to be useful the change of variable  $k_1^z = x_1 \ell_1 / (\eta_0 - \eta_r)$ ,  $k_2^z = x_2 \ell_2 / (\eta_0 - \eta_r)$ . In this way, the transfer functions become  $\tilde{\Delta}^T(\ell_i, k_i^z) \propto e^{-1/2(|x_i| \ell_i / \ell_*)^{1.2}}$  which allows to cut the integrals over  $x_i$  ( $i = 1, 2$ ) at  $\ell_*/\ell_i$ . Now, as a good approximation to see the effects of the transfer functions, we can take  $\ell \gg \ell_*$  and thus the integral over  $x_i$  can be easily computed by just evaluating the integrand in  $x_i = 0$  times  $4(\ell_*/\ell_i)(\ell_*/\ell_2)$ . With this approximation the integral in  $k_i^z$  is easily obtained and we get for the bispectrum

$$\begin{aligned} B_{\text{equil}}(\ell_1, \ell_2, \ell_3) &= \frac{24 f_1}{(2\pi)^2} f_{\text{NL}}^{\text{equil}} a^3 A^2 e^{-(\ell_1^2 + \ell_2^2 + \ell_3^2)/2\ell_*^{1.2}} \ell_*^2 \\ &\times \left( -\frac{1}{\ell_1^3 \ell_2^3} - \frac{1}{\ell_1^3 \ell_3^3} - \frac{1}{\ell_2^3 \ell_3^3} - \frac{2}{\ell_1^2 \ell_2^2 \ell_3^2} \right. \\ &\quad \left. + \frac{1}{\ell_1 \ell_2^2 \ell_3^3} + (5 \text{ perm.}) \right), \end{aligned} \quad (371)$$

where

$$\ell_3^2 = \ell_1^2 + \ell_2^2 + 2 \vec{\ell}_1 \cdot \vec{\ell}_2. \quad (372)$$

The coefficient  $f_1 \simeq 1/1.4 = 0.7$  is a fudge factor that improves the matching between our approximation for the bispectrum and numerical results that have been consistently checked. Notice that according to our approximation, the equilateral structure of (361) is preserved in  $\ell$  space. The expression (371) can be also written as  $B_{\text{equil}}(\ell_1, \ell_2, \ell_3) = (2\pi)^{-2} 48 f_1 f_{\text{NL}}^{\text{equil}} a^3 A^2 \ell_*^2 e^{-(\ell_1^2 + \ell_2^2 + \ell_3^2)/2\ell_*^{1.2}} (1 + \cos \theta)(\ell_1 + \ell_2 - \ell_3)/\ell_1^2 \ell_2^2 \ell_3^3$ ,  $\theta$  being the angle between  $\vec{\ell}_1$  and  $\vec{\ell}_2$ . In computing the signal-to-noise ratio, consistency with our approximation (371) requires that we integrate over  $\ell_1, \ell_2$  starting from a minimum  $\ell_{\text{min}} > \ell_*$  up to  $\ell_{\text{max}}$  and paying attention to the fact that even  $\ell_3$  in (372) must be larger than  $\ell_{\text{min}}$ . The scaling with  $\ell_{\text{max}}$  with respect to the case of a local type bispectrum turns out to be much milder. While for the local type  $(S/N)^2 \propto \ell_{\text{max}}^2$  [79], for the equilateral bispectrum (360), we find  $(S/N)^2 \propto \ell_{\text{max}}$  and, setting  $\ell_* = 750$  and  $\ell_{\text{min}} \simeq 1200$ ,

$$\begin{aligned} \left( \frac{S}{N} \right)_{\text{equil}}^2 &= 0.48 \times 10^5 \frac{f_{\text{sky}}}{2^5 \pi^3 6} A \left( f_{\text{NL}}^{\text{equil}} \right)^2 \ell_{\text{max}} \\ &\simeq 8 f_{\text{sky}} A \left( f_{\text{NL}}^{\text{equil}} \right)^2 \ell_{\text{max}}. \end{aligned} \quad (373)$$

(these scalings can be easily understood by analyzing the expressions  $(S/N)^2$  for the local and equilateral primordial NG. In the local case,  $(S/N)^2$  is proportional to  $\int d^2 \ell_1 d^2 \ell_2 d^2 \ell_3 \delta^{(2)}(\vec{\ell}_{123}) (\ell_1^3 + \ell_2^3 + \ell_3^3)^2 / (\ell_1 \ell_2 \ell_3)^3$  [79]; since the squeezed configuration, e.g.,  $\ell_1 \ll \ell_2, \ell_3$ , is dominating the local bispectrum, the integral becomes proportional to  $\int d\ell_1 d\ell_2 (\ell_2/\ell_1^2) \propto \ell_{\text{max}}^2$ . In the equilateral case, however,  $(S/N)^2$  receives contributions from the configuration which is peaked at  $\ell_1 \sim \ell_2 \sim \ell_3$  and therefore it can be written as  $\int d^2 \ell_1 d^2 \ell_2 \delta(\ell_1 - \ell_2) / \ell_1^2 \propto \ell_{\text{max}}$ . By choosing  $f_{\text{sky}} = 0.8$  and  $\ell_{\text{max}} = 2000$  we find a minimum detectable

$$f_{\text{NL}}^{\text{equil}} \simeq 66, \quad (374)$$

obtained imposing  $(S/N)_{\text{equil}} = 1$ . Both the estimate of the minimum value of  $f_{\text{NL}}^{\text{equil}}$  and the scaling  $(S/N)^2 \propto \ell_{\text{max}}$  are in remarkable agreement with the result obtained in [44] where the full transfer function is used and a value of  $f_{\text{NL}}^{\text{equil}} = 67$  is obtained. (We thank M. Liguori for discussions about the minimum value of  $f_{\text{NL}}^{\text{equil}}$  detectable by Planck and for its scaling with  $\ell_{\text{max}}$ .) Notice that our estimate is independent from the coefficient  $a$  and the exponential  $e^{-1/2(\ell/\ell_*)^{1.2}}$  introduced below (367) to mimic the full transfer function. This is because there is an equal number of transfer functions in the numerator and denominator of the expression (366) for the signal-to-noise ratio and their effect tend to cancel despite they are not simple multiplicative factors (see discussion in [79]).

**9.2. Non-Gaussianity from Recombination.** Comforted by the goodness of our model, in this section, we wish to estimate the level of NG generated at the recombination era. One can check that on small scales the second-order anisotropies are dominated by the second-order gravitational potential  $\Phi^{(2)}$  which grows as  $\eta^2$ , as first pointed out in [48]. Let us first briefly show how this result can be obtained using the calculations of Section 7. Consider the tight coupling regime, and small scales, well inside the diffusion Silk-damping scale  $\lambda_D$ . Then, look at (232). One realizes that, apart from the combination  $\Delta_{00}^{(2)} + 4\Phi^{(2)}$ , all the other terms scale at most as  $\eta$  or they are suppressed either because in the tight coupling regime or because of the damping diffusion on small scales. On the other hand, one finds that the main contribution to the bispectrum generated at recombination comes from

$$\Theta_{\text{SW}}^{(2)} = \frac{1}{4} \Delta_{00}^{(2)} + \Phi^{(2)}, \quad (375)$$

which is the usual term appearing in the CMB anisotropies due to the intrinsic photon energy density fluctuations  $\Delta_{00}^{(2)}$  and the gravitational redshift due to the potential. In fact,

such a term on large scales reduces to the Sachs-Wolfe effect, while on small scales at recombination, using (282) and (285), we find

$$\begin{aligned} \frac{1}{4}\Delta_{00}^{(2)} + \Phi^{(2)} &\simeq -R\Phi^{(2)} \\ &= -\frac{R\eta_r^2}{14}G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})T(k_1)\Phi^{(1)} \\ &\quad \times (\mathbf{k}_1)T(k_2)\Phi^{(1)}(\mathbf{k}_2), \end{aligned} \quad (376)$$

where we have evaluated the expression at the recombination time  $\eta_r$  and  $R = 3\rho_b/4\rho_\gamma$ , is the baryon-to-photon energy density ratio. In writing (376), we have dropped off the contribution from primordial non-Gaussianity (which, anyway, always propagates linearly) since here we are focusing on the relevant secondary effects that can constitute a source of contamination to it (see (202) and the discussion below). Notice also that this expression has been obtained assuming all the momenta much larger than  $k_{\text{eq}}$ .

Equation (376) has been confirmed numerically in [48]. There is also another way to obtain the analytical expression in the first line of (376), as discussed in details in [48], which exploits some well-known results at linear order extending them at second-order. At second order, in the perturbations the solution for the acoustic oscillations will have a form very similar to the linear solution written in (C.18), except, as usual, of some source terms  $S$  made by products of first-order perturbations.

$$\frac{\Delta_{00}^{(2)}}{4} = \left[ \frac{\Delta_{00}^{(2)}}{4}(0) + (1+R)\Phi^{(2)} \right] \cos(kr_s) - (1+R)\Phi^{(2)} + S. \quad (377)$$

Now on sufficiently small scales the products of first-order terms are indeed suppressed, due either to Silk damping, or because gravitational potentials are suppressed as  $(k/k_{\text{eq}})^2$ . Recall that only acoustic oscillations are damped by Silk damping diffusion, so that the cosine in (377) is multiplied by  $e^{-(k/k_D)^2} k \gg k_D$ , where  $k_D^{-1} = \lambda_D$  indicates damping length, but the baryon induced shift  $(1+R)\Phi^{(1)}$  is left untouched. Therefore, on small scales the combination of the damping effects and the growth of the second-order gravitational potential  $\Phi^{(2)}$  as  $\eta^2$  single out the dominant effect in (377) to be  $\Theta_{\text{SW}}^{(2)} \simeq -R\Phi^{(2)}$ .

*9.2.1. Bispectrum from Recombination due to the Nonlinear Growth of the Gravitational Potential.* Let us start to go into some details of the form of the bispectrum induced by the contribution (376). In (376), the kernel  $G$  is given by

$$G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}) = \mathbf{k}_1 \cdot \mathbf{k}_2 - \frac{10}{3} \frac{(\mathbf{k} \cdot \mathbf{k}_1)(\mathbf{k} \cdot \mathbf{k}_2)}{k^2}. \quad (378)$$

From the form of the kernel we see that the NG at recombination is dominated by an equilateral configuration, as expected from the fact that its origin is gravitational. Here, and in the following, we are implicitly assuming that a convolution is acting on the kernel as

$$\begin{aligned} &\frac{1}{(2\pi)^3} \int d\mathbf{k}_1 d\mathbf{k}_2 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) \\ &\quad \times G(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k})T(k_1)\Phi^{(1)}(\mathbf{k}_1)T(k_2)\Phi^{(1)}(\mathbf{k}_2). \end{aligned} \quad (379)$$

The reader should remember that, at first order in perturbation theory, the combination  $\Theta_{\text{SW}}^{(1)} + R\Phi^{(1)}$  is exponentially suppressed by the Silk damping, but still greater than the term  $R\Phi^{(1)}$  (which does not suffer the damping) for the maximum multipole of interest,  $\ell_{\text{max}} \sim 2000$ . This is mainly due to the fact that the first-order gravitational potential rapidly decays on small scales. On the contrary, at second-order in perturbation theory, the gravitational potential grows like the scale factor on small scales and it turns out that the  $R\Phi^{(2)}$  dominates on small scales (see [48]).

The gravitational potential at linear order can be expressed as usual in terms of the transfer function  $T(k)$

$$T(k) \approx 12 \left( \frac{k_{\text{eq}}}{k} \right)^2 \ln[k/8k_{\text{eq}}], \quad (380)$$

where the last step is an approximation valid on scales smaller than the equivalence scale,  $k \gg k_{\text{eq}}$ . In the following, we will account for the logarithmic growth just with a coefficient  $T_0(k) = 12 \ln[k/8k_{\text{eq}}] \approx 11$  for the scales of interest.

In the flat-sky approximation, one arrives at an expression similar to (364), where now one of the linear transfer functions must be replaced by a transfer function at second-order. Specifically, one finds

$$\begin{aligned} &B_{\text{rec}}(\ell_1, \ell_2, \ell_3) \\ &= \frac{(\eta_0 - \eta_r)^2}{(2\pi)^2} \int dk_1^z dk_2^z dk_3^z \delta^{(1)}(k_{123}^z) \\ &\quad \times \left[ G(\mathbf{k}'_1, \mathbf{k}'_2, \mathbf{k}'_3)T(k'_1)T(k'_2)P(k'_1)P(k'_2) \right. \\ &\quad \left. \times \tilde{\Delta}^T(\ell_1, k_1^z) \tilde{\Delta}^T(\ell_2, k_2^z) \tilde{\Delta}^{T(2)}(\ell_3, k_3^z) + \text{cyclic} \right]. \end{aligned} \quad (381)$$

By using our model (367) and

$$\tilde{\Delta}^{T(2)}(\ell, k^z) = -\frac{R}{14} \frac{\tau_r^2}{(\eta_0 - \eta_r)^2}, \quad (382)$$

for the second-order radiation transfer function, we find

$$\begin{aligned}
& B_{\text{rec}}(\ell_1, \ell_2, \ell_3) \\
&= -\frac{R}{14} \frac{(\eta_0 - \eta_r)^{-4}}{(2\pi)^2} k_{\text{eq}}^4 \eta_r^2 A^2 a^2 T_0^2 e^{-1/2(\ell_1/\ell_*)^{1.2}} e^{-1/2(\ell_2/\ell_*)^{1.2}} \\
&\quad \times \int dk_1^z dk_2^z dk_3^z \delta^{(1)}(k_{123}^z) \\
&\quad \times \left[ G(\mathbf{k}'_1, \mathbf{k}'_2, \mathbf{k}'_3) \frac{1}{k_1'^5 k_2'^5} e^{-1/2(|k_{1z}|/k_*)^{1.2}} \right. \\
&\quad \quad \left. \times e^{-1/2(|k_{2z}|/k_*)^{1.2}} + \text{cyclic} \right]. \tag{383}
\end{aligned}$$

At this point, we proceed further by employing the same approximation described after (370). We use the Dirac delta to replace the variable  $k_{3z}$ , and the exponential allow us to evaluate the integral for  $k_{1z} = k_{2z} = 0$ , for scales  $\ell_i \gg \ell_*$ . This leads to

$$\begin{aligned}
& B_{\text{rec}}(\ell_1, \ell_2, \ell_3) \\
&= -\frac{4f_2}{(2\pi)^2} \frac{R}{14} A^2 a^2 T_0^2 (k_{\text{eq}} \eta_r)^2 \ell_{\text{eq}}^2 \ell_*^2 e^{-1/2(\ell_1/\ell_*)^{1.2}} e^{-1/2(\ell_2/\ell_*)^{1.2}} \\
&\quad \times \frac{1}{\ell_1^5 \ell_2^5} \left[ \vec{\ell}_1 \cdot \vec{\ell}_2 - \frac{10}{3} \frac{(\vec{\ell}_3 \cdot \vec{\ell}_1)(\vec{\ell}_3 \cdot \vec{\ell}_2)}{\ell_3^2} \right] + \text{cyclic}. \tag{384}
\end{aligned}$$

Again, here  $f_2$  is a coefficient to better calibrate our approximations with numerical results that we have performed in order to test the validity of our approach. Not surprisingly, it turns out that  $f_2 \simeq f_1 \simeq 1/1.4$ .

### 9.3. Contamination to Primordial Non-Gaussianity from Recombination: Fisher Matrices

**9.3.1. Contamination to Primordial Non-Gaussianity of the Equilateral Type.** Our goal now is to quantify the level of NG coming from the recombination era and to estimate the level of degradation it causes on the possible measurement of the equilateral primordial bispectrum. The reader should keep in mind that, given the form of the kernel function (378), the NG from recombination is expected to be of the equilateral type. Following Section 8.1, a rigorous procedure is to define the Fisher matrix (in flat-sky approximation) as

$$F_{ij} = \int d^2\ell_1 d^2\ell_2 d^2\ell_3 \delta^{(2)}(\vec{\ell}_{123}) \frac{B^i(\ell_1, \ell_2, \ell_3) B^j(\ell_1, \ell_2, \ell_3)}{6C(\ell_1)C(\ell_2)C(\ell_3)}, \tag{385}$$

where  $i$  (or  $j$ ) = (rec, equil), and to define the signal-to-noise ratio for a component  $i$ ,  $(S/N)_i = 1/\sqrt{F_{ii}^{-1}}$ , and the degradation parameter  $d_i = F_{ii} F_{ii}^{-1}$  due to the correlation between the different components  $r_{ij} = F_{ij}^{-1}/\sqrt{F_{ii}^{-1} F_{jj}^{-1}}$ . The first entry  $F_{\text{rec, equil}}$  of the Fisher matrix corresponds to the

$(S/N)^2$  ratio computed in (373) which does not account for any kind of cross-correlation. Due to the equilateral form of the NG generated at recombination, we expect that the minimum value detectable for  $f_{\text{NL}}^{\text{equil}}$  will be higher than the one reported in (374). For the mixed entry, we find

$$\begin{aligned}
F_{\text{rec, equil}} &= \frac{f_{\text{sky}}}{\pi} \frac{1}{(2\pi)^2} \int d^2\ell_1 d^2\ell_2 d^2\ell_3 \delta^{(2)}(\vec{\ell}_{123}) \\
&\quad \times \frac{B_{\text{rec}}(\ell_1, \ell_2, \ell_3) B_{\text{equil}}(\ell_1, \ell_2, \ell_3)}{6C(\ell_1)C(\ell_2)C(\ell_3)} \\
&= -3f_1 f_2 \frac{f_{\text{sky}}}{\pi^3} \frac{4R}{14} \frac{48}{2^5 6} \frac{T_0^2}{a} (k_{\text{eq}} \eta_r)^2 \ell_{\text{eq}}^2 \ell_* A f_{\text{NL}}^{\text{equil}} \\
&\quad \times \int d\ell_1 d\ell_2 (1 + \vec{\ell}_1 \cdot \vec{\ell}_2 / \ell_1 \ell_2) e^{1/2(\ell_3/\ell_*)^{1.2}} \\
&\quad \times \frac{1}{\ell_1^3 \ell_2^3} (\ell_1 + \ell_2 - \ell_3) \\
&\quad \times \left[ \vec{\ell}_1 \cdot \vec{\ell}_2 - \frac{10}{3} \frac{(\vec{\ell}_3 \cdot \vec{\ell}_1)(\vec{\ell}_3 \cdot \vec{\ell}_2)}{\ell_3^2} \right], \tag{386}
\end{aligned}$$

where  $\ell_3$  is given by (372). The factor 3 in front of this expression comes from cyclic permutations. The integral can be performed numerically and, integrating from a minimum  $\ell_{\text{min}} \simeq 1200$  up to  $\ell_{\text{max}} = 2000$ , and by taking  $R \simeq 0.3$  when evaluated at recombination,  $a \simeq 3$ ,  $T_0 \simeq 11$ ,  $(k_{\text{eq}} \eta_r)^2 \simeq 26$ ,  $\ell_{\text{eq}} = 150$ ,  $\ell_* = 750$ , we find  $F_{\text{rec, equil}} \simeq 9.4 \times 10^{-4}$ . Finally, for the entry  $F_{\text{rec, rec}}$ , we get

$$\begin{aligned}
F_{\text{rec, rec}} &= \frac{f_{\text{sky}}}{\pi} \frac{1}{(2\pi)^2} \int d^2\ell_1 d^2\ell_2 d^2\ell_3 \delta^{(2)}(\vec{\ell}_{123}) \\
&\quad \times \frac{B_{\text{rec}}^2(\ell_1, \ell_2, \ell_3)}{6C(\ell_1)C(\ell_2)C(\ell_3)} \\
&= f_2^2 \frac{f_{\text{sky}}}{2^5 \pi^3 6} \left( \frac{4R}{14} \right)^2 \left( \frac{T_0^2}{a} \right)^2 (k_{\text{eq}} \eta_r)^4 \ell_{\text{eq}}^4 \ell_* A \\
&\quad \times \left[ 3 \int d\ell_1 d\ell_2 e^{(\ell_3/\ell_*)^{1.2}} \right. \\
&\quad \quad \times \frac{\ell_3^3}{\ell_1^6 \ell_2^6} \left( \vec{\ell}_1 \cdot \vec{\ell}_2 - \frac{10}{3} \frac{(\vec{\ell}_3 \cdot \vec{\ell}_1)(\vec{\ell}_3 \cdot \vec{\ell}_2)}{\ell_3^2} \right)^2 \\
&\quad \quad + 6 \int d\ell_1 d\ell_2 e^{1/2(\ell_3/\ell_*)^{1.2}} e^{1/2(\ell_2/\ell_*)^{1.2}} \\
&\quad \quad \times \frac{1}{\ell_1^6 \ell_2^6 \ell_3^2} \left( \vec{\ell}_1 \cdot \vec{\ell}_2 - \frac{10}{3} \frac{(\vec{\ell}_3 \cdot \vec{\ell}_1)(\vec{\ell}_3 \cdot \vec{\ell}_2)}{\ell_3^2} \right) \\
&\quad \quad \left. \times \left( \vec{\ell}_1 \cdot \vec{\ell}_3 - \frac{10}{3} \frac{(\vec{\ell}_2 \cdot \vec{\ell}_3)(\vec{\ell}_1 \cdot \vec{\ell}_2)}{\ell_2^2} \right) \right], \tag{387}
\end{aligned}$$

and we find a value  $F_{\text{rec,rec}} \simeq 0.014$ . We are now able to compute the entries of inverse of the Fisher matrix,  $F_{ij}^{-1}$ . In the following, we report our results for the signal-to-noise ratios and the degradation parameters

$$\left(\frac{S}{N}\right)_{\text{equil}} = \frac{1}{\sqrt{F_{\text{equil,equil}}^{-1}}} \simeq 12.6 \times 10^{-3} f_{\text{NL}}^{\text{equil}}, \quad (388)$$

$$\left(\frac{S}{N}\right)_{\text{rec}} = \frac{1}{\sqrt{F_{\text{rec,rec}}^{-1}}} \simeq 0.1, \quad (389)$$

$$r_{\text{rec,equil}} = \frac{F_{\text{rec,equil}}^{-1}}{\sqrt{F_{\text{equil,equil}}^{-1} F_{\text{rec,rec}}^{-1}}} \simeq -0.53, \quad (390)$$

$$d_{\text{rec}} = F_{\text{rec,rec}} F_{\text{rec,rec}}^{-1} \simeq 1.4, \quad (391)$$

$$d_{\text{equil}} = F_{\text{equil,equil}} F_{\text{equil,equil}}^{-1} \simeq 1.4. \quad (392)$$

As a confirmation of our expectations, we find that the NG of the type given by (378) has a quite high correlation with an equilateral primordial bispectrum. This translates into a degradation in the minimum detectable value for  $f_{\text{NL}}^{\text{equil}}$  with respect to the value given in (374). In fact, from the signal-to-noise ratio (388), we find a minimum detectable value of

$$f_{\text{NL}}^{\text{equil}} \simeq 79, \quad (393)$$

imposing that  $(S/N)_{\text{equil}} = 1$  with an increase of  $\Delta f_{\text{NL}}^{\text{equil}} = \mathcal{O}(10)$ . This corresponds to an increase of the 1- $\sigma$  uncertainty on  $f_{\text{NL}}^{\text{equil}}$  of 20% (see Section 8.1). (Due to a non-vanishing correlation,  $r_{ij}$ ,  $(S/N)$  gets modified from its zero-order value to  $(S/N) = (S/N)_0 (1 - r_{ij}^2)^{1/2}$ , so that the minimum detectable value of  $f_{\text{NL}}^{\text{equil}}$  gets shifted by a quantity  $\Delta f_{\text{NL}}^{\text{equil}} / (f_{\text{NL}}^{\text{equil}})_0 = (1 - r_{ij}^2)^{-1/2} - 1$ ).

Following Section 8.1, in order to measure the contamination to the primordial bispectra one can define that effective nonlinearity parameter  $f_{\text{NL}}^{\text{con}}$  which minimizes the  $\chi^2$  defined as

$$\chi^2 = \int d^2 \ell_1 d^2 \ell_2 d^2 \ell_3 \delta^{(2)}(\vec{\ell}_{123}) \times \frac{\left(f_{\text{NL}}^{\text{equil}} B_{\text{eq}}(\ell_1, \ell_2, \ell_3; f_{\text{NL}}^{\text{equil}} = 1) - B_{\text{rec}}(\ell_1, \ell_2, \ell_3)\right)^2}{6 C(\ell_1) C(\ell_2) C(\ell_3)}, \quad (394)$$

to find

$$f_{\text{NL}}^{\text{con}} = \frac{F_{\text{rec,equil}}}{F_{\text{equil,equil}}} \Big|_{f_{\text{NL}}^{\text{equil}}=1}, \quad (395)$$

and an analogous expression to compute the contamination to the local primordial bispectrum. In this case we find an effective nonlinearity parameter

$$f_{\text{NL}}^{\text{cont}} \simeq 5. \quad (396)$$

(Notice that sometimes in the literature one can find also an effective nonlinearity parameter  $f_{\text{NL}}^{\text{eff}}$  defined in such a way

that the equilateral (or the local) bispectrum has the same Fisher matrix errors as the recombination bispectrum

$$f_{\text{NL}}^{\text{eff}} = \frac{\sqrt{F_{\text{rec,rec}}}}{\sqrt{F_{\text{equil,equil}}}} \Big|_{f_{\text{NL}}^{\text{equil}}=1}. \quad (397)$$

However this is not the proper quantity to define the contamination level to primordial NG; here we are just comparing signal-to-noise ratios, while the contamination  $f_{\text{NL}}^{\text{con}}$  defined in (395) contains a somewhat richer information, since we are asking what is the value of equilateral (local)  $f_{\text{NL}}$  which best mimics the bispectrum from recombination).

**9.3.2. Contamination to Primordial Non-Gaussianity of the Local Type.** Similarly, we can compute the Fisher matrix accounting for the NG generated at recombination and the primordial NG of the local type

$$\langle \Phi(\mathbf{k}_1) \Phi(\mathbf{k}_2) \Phi(\mathbf{k}_3) \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{\text{loc}}(k_1, k_2, k_3), \quad (398)$$

where

$$B_{\text{loc}}(k_1, k_2, k_3) = f_{\text{NL}}^{\text{loc}} \cdot 2A^2 \cdot \left( \frac{1}{k_1^3 k_2^3} + \frac{1}{k_1^3 k_3^3} + \frac{1}{k_2^3 k_3^3} \right). \quad (399)$$

The bispectrum and the signal-to-noise ratio as defined in (366) have already been computed in the flat-sky approximation in [79]. The result is that  $(S/N)_{\text{loc}}^2 = 4\pi^{-2} f_{\text{sky}}(\ell_*/\ell_{\text{min}})(f_{\text{NL}}^{\text{loc}})^2 A \ell_{\text{max}}^2$ , corresponding to a minimum detectable value of  $f_{\text{NL}}^{\text{loc}} = \mathcal{O}(7)$  for  $\ell_{\text{max}} = 2000$  (when other possible sources of NG are ignored). We can compute the off-diagonal entry of the Fisher matrix in a similar way to what we have described in this section, and we get  $F_{\text{rec,loc}} \simeq 8 \times 10^{-3} f_{\text{NL}}^{\text{loc}}$ . Finally, the entry  $F_{\text{rec,rec}} \simeq 0.014$  has already been computed above. From inverting the Fisher matrix, we get the following signal-to-noise ratios and the degradation parameters

$$\left(\frac{S}{N}\right)_{\text{loc}} = \frac{1}{\sqrt{F_{\text{loc,loc}}^{-1}}} \simeq 14 \times 10^{-2} f_{\text{NL}}^{\text{loc}}, \quad (400)$$

$$\left(\frac{S}{N}\right)_{\text{rec}} = \frac{1}{\sqrt{F_{\text{rec,rec}}^{-1}}} \simeq 0.1, \quad (401)$$

$$r_{\text{rec,loc}} = \frac{F_{\text{rec,loc}}^{-1}}{\sqrt{F_{\text{loc,loc}}^{-1} F_{\text{rec,rec}}^{-1}}} \simeq -0.44, \quad (402)$$

$$d_{\text{rec}} = F_{\text{rec,rec}} F_{\text{rec,rec}}^{-1} \simeq 1.2, \quad (403)$$

$$d_{\text{loc}} = F_{\text{equil,equil}} F_{\text{equil,equil}}^{-1} \simeq 1.2. \quad (404)$$

In particular, from (400) we see that now the minimum detectable value of  $f_{\text{NL}}^{\text{loc}}$  remains basically unchanged in the presence of the recombination signal. Similarly, the effective  $f_{\text{NL}}^{\text{con}}$  reads

$$f_{\text{NL}}^{\text{con}} = \frac{F_{\text{rec,loc}}}{F_{\text{loc,loc}}} \Big|_{f_{\text{NL}}^{\text{loc}}=1} \simeq 0.3, \quad (405)$$

which is much smaller than the effective nonlinearity parameter (405) for the equilateral case. We have also checked the cross-correlation between the primordial local and equilateral bispectra finding a value of  $r_{\text{loc, equil}} \simeq 0.23$ , which is in agreement with the value reported in [44]. This reflects the fact that the primordial local and equilateral signals are not fully uncorrelated. The reason is due to the fact that the equilateral and local bispectrum (399) and (361) approach the same shape in the equilateral configuration. This is also the reason why the cross-correlation between the primordial local and recombination bispectra is not so small.

## 10. How to Perform a Numerical Analysis of the CMB Bispectrum Produced by Second-Order Perturbations

In the previous section, through a specific example, we have learned some basic principles to determine a reasonable and quite fast analytic estimate for the contamination to the primordial NG. In this section, we provide all the tools necessary to a full numerical implementation of the second-order Boltzmann equations which allow to obtain a systematic numerical evaluation of the CMB angular bispectrum produced by second-order cosmological perturbations. This section is mainly based on the results of [96].

In particular, we will apply this formalism to numerically evaluate the contamination to primordial NG from the second-order fluctuations of the Boltzmann equations that come as products of the first-order perturbations, and ignore the intrinsically second-order terms, or the effects of the perturbed recombination [63, 87, 89]. (At the moment of writing this review the calculations that also include the intrinsically second-order terms and the perturbed recombination are in progress and will be presented in [97]). Here, therefore, we come to the discussion of the second example mentioned in Section 7.1. As anticipated there we will see that, unlike the intrinsically second-order contribution considered in the previous section, the products of the first-order perturbations give a CMB bispectrum that peaks in the squeezed configuration. Therefore, we will study the contamination of the primordial NG of the local type.

### 10.1. CMB Bispectrum from Second-Order Perturbations

*10.1.1. Definitions.* Again, we expand the temperature fluctuation into the linear (first-order) part and the second-order part as

$$\frac{\Delta T(\hat{\mathbf{n}})}{T} = \frac{\Delta T^{(1)}(\hat{\mathbf{n}})}{T} + \frac{\Delta T^{(2)}(\hat{\mathbf{n}})}{T} + \dots \quad (406)$$

The spherical harmonic coefficients of temperature anisotropy,  $a_{\ell m} = T^{-1} \int d^2 \hat{\mathbf{n}} Y_{\ell m}^*(\hat{\mathbf{n}}) \Delta T(\hat{\mathbf{n}})$ , are therefore expanded as

$$a_{\ell m} = a_{\ell m}^{(1)} + a_{\ell m}^{(2)} + \dots \quad (407)$$

Recall that to expand the Boltzmann equation up to the second order in perturbations, we had to expand the distribution function,

$$f(\mathbf{x}, p, \hat{\mathbf{n}}, \eta) = 2 \left[ \exp \left\{ \frac{p}{T(\eta) e^{\Theta(\mathbf{x}, \hat{\mathbf{n}}, \eta)}} \right\} - 1 \right]^{-1}, \quad (408)$$

up to the second order in perturbations:  $\Theta = \Theta^{(1)} + \Theta^{(2)}/2 + \dots$ , and accordingly  $f = f^{(0)} + f^{(1)} + f^{(2)}/2 + \dots$ . Notice that here  $\Theta$  does not depend on  $p$ . This means that in the following, we will just account for those contributions that do not come from terms related to spectral distortions. Such terms will be accounted for in the full computation of [97]. We can then compute the fractional perturbation in photon's energy density at the  $i$ -th order in perturbations,  $\Delta^{(i)}$ , by multiplying  $f^{(i)}$  by  $p$ , and integrating over  $p^2 dp$ ,

$$\Delta^{(i)} \equiv \frac{\int dp p^3 f^{(i)}}{\int dp p^3 f^{(0)}}. \quad (409)$$

At the linear order, we recovered the usual relation between the linear fractional temperature fluctuation,  $\Theta^{(1)} = \Delta T^{(1)}/T$ , and the linear fractional energy density perturbation,  $\Delta^{(1)} = \delta \rho_\gamma^{(1)}/\rho_\gamma$ , that is,  $\Delta^{(1)} = 4\Theta^{(1)}$ . At the second order, we have

$$\Delta^{(2)} = 4\Theta^{(2)} + 16[\Theta^{(1)}]^2, \quad (410)$$

which is related to the second-order temperature fluctuation as

$$\begin{aligned} \frac{\Delta T^{(2)}}{T} &= \frac{1}{8} (\Delta^{(2)} - \langle \Delta^{(2)} \rangle) - \frac{3}{2} \left( [\Theta^{(1)}]^2 - \langle [\Theta^{(1)}]^2 \rangle \right) \\ &= \frac{1}{2} \left( \Theta^{(2)} - \langle \Theta^{(2)} \rangle + [\Theta^{(1)}]^2 - \langle [\Theta^{(1)}]^2 \rangle \right), \end{aligned} \quad (411)$$

where we have subtracted the average of the temperature fluctuation so that the average of  $\Delta T^{(2)}/T$  vanishes. Notice that the temperature we are interested in is the temperature of the blackbody that has the same energy as the actual distribution, see for a discussion on this [66].

We compute  $a_{\ell m}^{(2)}$  from  $\Delta T^{(2)}/T$  using

$$\begin{aligned} a_{\ell m}^{(2)} &= \int d^2 \hat{\mathbf{n}} Y_{\ell m}^*(\hat{\mathbf{n}}) \frac{\Delta T^{(2)}}{T} \\ &= \tilde{a}_{\ell m}^{(2)} - \frac{3}{2} \sum_{\ell' m'} \sum_{\ell'' m''} (-1)^m \mathcal{G}_{\ell \ell' \ell''}^{-m m' m''} \\ &\quad \times \left( a_{\ell' m'}^{(1)} a_{\ell'' m''}^{(1)} - \langle a_{\ell' m'}^{(1)} a_{\ell'' m''}^{(1)} \rangle \right), \end{aligned} \quad (412)$$

where we define

$$\tilde{a}_{\ell m}^{(2)} \equiv \frac{1}{8} \int d^2 \hat{\mathbf{n}} Y_{\ell m}^*(\hat{\mathbf{n}}) \left( \Delta^{(2)}(\hat{\mathbf{n}}) - \langle \Delta^{(2)}(\hat{\mathbf{n}}) \rangle \right), \quad (413)$$

$$\begin{aligned} \mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} &\equiv \int d^2 \hat{\mathbf{n}} Y_{\ell_1 m_1}(\hat{\mathbf{n}}) Y_{\ell_2 m_2}(\hat{\mathbf{n}}) Y_{\ell_3 m_3}(\hat{\mathbf{n}}) \\ &= \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \\ &\times \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix}. \end{aligned} \quad (414)$$

Here, the matrix is the Wigner  $3j$  symbol.

The CMB angular-averaged bispectrum,  $B_{\ell_1 \ell_2 \ell_3}$ , is related to the ensemble average of  $a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3}$ , as

$$B_{\ell_1 \ell_2 \ell_3} \equiv \sum_{\text{all } m} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle. \quad (415)$$

This definition guarantees rotational invariance for the bispectrum, and the Wigner  $3j$  symbol ensures that the bispectrum must satisfy triangle conditions:  $\ell_i - \ell_j \leq \ell_k \leq \ell_i + \ell_j$  for all permutations of indices, and selection rules:  $m_1 + m_2 + m_3 = 0$ .

The ensemble average is given by

$$\begin{aligned} &\langle a_{\ell_1 m_1} a_{\ell_2 m_2} a_{\ell_3 m_3} \rangle \\ &= \langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} a_{\ell_3 m_3}^{(2)} \rangle + \text{cyclic} \\ &= \langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} \tilde{a}_{\ell_3 m_3}^{(2)} \rangle - \frac{3}{2} \sum_{\ell_3' m_3'} \sum_{\ell_3'' m_3''} (-1)^{m_3} \mathcal{G}_{\ell_3 \ell_3' \ell_3''}^{-m_3 m_3' m_3''} \\ &\times \left( \langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} a_{\ell_3' m_3'}^{(1)} a_{\ell_3'' m_3''}^{(1)} \rangle \right. \\ &\quad \left. - \langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} \rangle \langle a_{\ell_3' m_3'}^{(1)} a_{\ell_3'' m_3''}^{(1)} \rangle \right) + \text{cyclic}, \end{aligned} \quad (416)$$

where cyclic means that we have to sum the cyclic permutations of (416) for indices  $(1, 2, 3) \rightarrow (3, 1, 2) \rightarrow (2, 3, 1)$ .

As we assume that  $a_{\ell m}^{(1)}$ 's are Gaussian random variables, the four-point function of  $a_{\ell m}^{(1)}$ 's in (416) is given by the sum of products of all possible pairs. Each pair gives the angular power spectrum,  $C_\ell$ :

$$\langle a_{\ell m}^{(1)} a_{\ell' m'}^{(1)} \rangle = (-1)^m C_\ell \delta_{\ell \ell'} \delta_{-m m'}. \quad (417)$$

We obtain

$$\begin{aligned} &\langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} a_{\ell_3' m_3'}^{(1)} a_{\ell_3'' m_3''}^{(1)} \rangle - \langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} \rangle \langle a_{\ell_3' m_3'}^{(1)} a_{\ell_3'' m_3''}^{(1)} \rangle \\ &= (-1)^{m_1 + m_2} C_{\ell_1} C_{\ell_2} [\delta_{\ell_1 \ell_3'} \delta_{-m_1 m_3'} \delta_{\ell_2 \ell_3''} \delta_{-m_2 m_3''} + (1 \leftrightarrow 2)]. \end{aligned} \quad (418)$$

Substituting the right hand side of equation (418) for the second term of equation (416), and using  $\ell_1 + \ell_2 + \ell_3 = \text{even}$ , we obtain the angular averaged bispectrum,

$$B_{\ell_1 \ell_2 \ell_3} = \tilde{B}_{\ell_1 \ell_2 \ell_3} - 3I_{\ell_1 \ell_2 \ell_3} (C_{\ell_1} C_{\ell_2} + \text{cyclic}), \quad (419)$$

where we have defined the quantities,

$$I_{\ell_1 \ell_2 \ell_3} \equiv \sqrt{\frac{(2\ell_1 + 1)(2\ell_2 + 1)(2\ell_3 + 1)}{4\pi}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix},$$

$$\tilde{B}_{\ell_1 \ell_2 \ell_3} = \sum_{\text{all } m} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} \tilde{a}_{\ell_3 m_3}^{(2)} \rangle + \text{cyclic}. \quad (420)$$

### 10.1.2. The Second-Order CMB Radiation Transfer Function.

In this section, we show how it is possible to define in a rigorous way the radiation transfer function for CMB anisotropies at second-order in the perturbations. It is a generalization of the well-known quantity used at linear-order and as such it allows us to develop a systematic numerical analysis of the angular bispectrum produced by second-order perturbations in the very same way as at linear-order various numerical codes are nowadays available for the computation of the CMB power spectrum, such as CMBFAST or CAMB.

The starting point is the Boltzmann equation that governs the evolution of  $\Delta^{(1)}(k, \mu, \eta)$  and  $\Delta^{(2)}(\mathbf{k}, \hat{\mathbf{n}}, \eta)$ , where  $\mu = \hat{\mathbf{k}} \cdot \hat{\mathbf{n}}$  and  $\mathbf{n}$  is the direction of propagation of photons. Note that for the linear perturbation there is azimuthal symmetry such that  $\Delta^{(1)}$  depends only on the angle between  $\mathbf{k}$  and  $\mathbf{n}$ ; however, for the second-order perturbation there is no such symmetry. We write again the Boltzmann equations in Fourier space

$$\begin{aligned} \Delta^{(1)'} + ik\mu\Delta^{(1)} - \tau' \Delta^{(1)} &= S^{(1)}(k, \mu, \eta), \\ \Delta^{(2)'} + ik\mu\Delta^{(2)} - \tau' \Delta^{(2)} &= S^{(2)}(\mathbf{k}, \hat{\mathbf{n}}, \eta), \end{aligned} \quad (421)$$

where the primes denote derivatives with respect to the conformal time  $\partial/\partial\eta$ ,  $S^{(1)}$  and  $S^{(2)}$  are the source functions at the first and the second orders, respectively, and  $\tau'$  is the differential optical depth which is defined by using the mean electron number density,  $\bar{n}_e$ , the Thomson scattering cross-section,  $\sigma_T$ , and the scale factor,  $a$ , as

$$\tau' = -\bar{n}_e \sigma_T a. \quad (422)$$

We expand again the angular dependence of  $\Delta^{(i)}$  as

$$\Delta_{\ell m}^{(i)}(\mathbf{k}, \eta) = i^\ell \sqrt{\frac{2\ell + 1}{4\pi}} \int d^2 \hat{\mathbf{n}} Y_{\ell m}^*(\hat{\mathbf{n}}) \Delta^{(i)}(\mathbf{k}, \hat{\mathbf{n}}, \eta), \quad (423)$$

and that of the source terms as

$$S_{\ell m}^{(i)}(\mathbf{k}, \eta) = i^\ell \sqrt{\frac{2\ell + 1}{4\pi}} \int d^2 \hat{\mathbf{n}} Y_{\ell m}^*(\hat{\mathbf{n}}) S^{(i)}(\mathbf{k}, \hat{\mathbf{n}}, \eta), \quad (424)$$

where  $i = 1, 2$ .

The source functions relate the observed  $a_{\ell m}$ 's to the primordial curvature perturbations in comoving gauge,  $\zeta(\mathbf{k})$ . The relations contain the linear radiation transfer function,  $g_\ell(k)$ , and the second-order radiation transfer function,  $F_{\ell m}^{\ell' m'}(k)$ , and are given by

$$\begin{aligned} a_{\ell m}^{(1)} &= 4\pi(-i)^\ell \int \frac{d^3 k}{(2\pi)^3} g_\ell(k) Y_{\ell m}^*(\hat{\mathbf{k}}) \zeta(\mathbf{k}), \\ \tilde{a}_{\ell m}^{(2)} &= \frac{4\pi}{8} (-i)^\ell \int \frac{d^3 k}{(2\pi)^3} \int \frac{d^3 k'}{(2\pi)^3} \int d^3 k'' \delta^{(3)}(\mathbf{k}', \mathbf{k}'', \mathbf{k}) \\ &\quad \times \sum_{\ell' m'} F_{\ell m}^{\ell' m'}(\mathbf{k}', \mathbf{k}'', \mathbf{k}) Y_{\ell' m'}^*(\hat{\mathbf{k}}) \zeta(\mathbf{k}') \zeta(\mathbf{k}''). \end{aligned} \quad (425)$$

The linear transfer function is given by

$$\begin{aligned} g_\ell(k) &= \int_0^{\eta_0} d\eta e^{-\tau} \left[ S_{00}^{(1)}(k, \eta) + S_{10}^{(1)}(k, \eta) \frac{d}{du} + S_{20}^{(1)}(k, \eta) \right. \\ &\quad \left. \times \left( \frac{3}{2} \frac{d^2}{du^2} + \frac{1}{2} \right) \right] j_\ell(u), \end{aligned} \quad (426)$$

where  $u \equiv k(\eta_0 - \eta)$  and  $S_{\ell m}^{(1)}$  is the standard linear source function

$$\begin{aligned} S_{00}^{(1)}(k, \eta) &= 4\Psi^{(1)'}(k, \eta) - \tau' \Delta_0^{(1)}(k, \eta), \\ S_{10}^{(1)}(k, \eta) &= 4k\Phi^{(1)}(k, \eta) - 4\tau' v_0^{(1)}(k, \eta), \\ S_{20}^{(1)}(k, \eta) &= \frac{\tau'}{2} \Delta_2^{(1)}(k, \eta), \end{aligned} \quad (427)$$

where  $\Phi^{(1)}(k, \eta)$  and  $\Psi^{(1)}(k, \eta)$  are the metric perturbations at the linear order in the longitudinal gauge and  $\Delta_0^{(1)}(k, \eta)$ ,  $\Delta_1^{(1)}(k, \eta)$ , and  $\Delta_2^{(1)}(k, \eta)$  are the coefficients of the expansion in Legendre polynomials of  $\Delta^{(1)}(k, \mu, \eta)$ . These coefficients  $\Delta_\ell^{(1)}(k, \eta)$  are related to  $\Delta_{\ell m}^{(1)}$  (423) via

$$\Delta_{\ell m}^{(1)}(\mathbf{k}) = i^\ell \sqrt{\frac{4\pi}{2\ell+1}} Y_{\ell m}^*(\hat{\mathbf{k}}) \Delta_\ell^{(1)}(\mathbf{k}) (2\ell+1). \quad (428)$$

The first-order velocity perturbation,  $v_0^{(1)}(k, \eta)$ , is the irrotational part of the baryon velocity defined by  $\mathbf{v}(\mathbf{k}) = -iv_0(k)\hat{\mathbf{k}}$ .

The new piece, *the second-order radiation transfer function*, is the line-of-sight integral of the second-order source terms in the Boltzmann equation:

$$\begin{aligned} F_{\ell m}^{\ell' m'}(\mathbf{k}', \mathbf{k}'', \mathbf{k}) &= i^\ell \sum_{\lambda \mu} (-1)^m (-i)^{\lambda-\ell'} \mathcal{G}_{\ell' \ell' \lambda}^{-mm' \mu} \sqrt{\frac{4\pi}{2\lambda+1}} \\ &\quad \times \int_0^{\eta_0} d\eta e^{-\tau} \mathcal{S}_{\lambda \mu}^{(2)}(\mathbf{k}', \mathbf{k}'', \mathbf{k}, \eta) j_\ell[k(\eta - \eta_0)]. \end{aligned} \quad (429)$$

Here, we have introduced a new function,  $\mathcal{S}_{\ell m}^{(2)}(\mathbf{k}', \mathbf{k}'', \mathbf{k}, \eta)$ , which is defined by the following equation:

$$\begin{aligned} \mathcal{S}_{\ell m}^{(2)}(\mathbf{k}, \eta) &= \int \frac{d\mathbf{k}'}{(2\pi)^3} d\mathbf{k}'' \delta^{(3)}(\mathbf{k}', \mathbf{k}'', \mathbf{k}) \mathcal{S}_{\ell m}^{(2)} \\ &\quad \times (\mathbf{k}', \mathbf{k}'', \mathbf{k}, \eta) \zeta(\mathbf{k}') \zeta(\mathbf{k}''). \end{aligned} \quad (430)$$

Basically,  $\mathcal{S}_{\ell m}^{(2)}(\mathbf{k}', \mathbf{k}'', \mathbf{k}, \eta)$  is the second-order source function divided by  $\zeta(\mathbf{k}')\zeta(\mathbf{k}'')$ . The explicit expression of the second-order source function can be read from (134).

Using (425), we calculate the first term in (419),  $\tilde{B}_{\ell_1 \ell_2 \ell_3}$ , as follows:

$$\begin{aligned} \langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} \tilde{a}_{\ell_3 m_3}^{(2)} \rangle &= \frac{(-i)^{\ell_1+\ell_2+\ell_3}}{(2\pi)^3} \sum_{\ell_3 M_3} \prod_i \int d^3 k_i \delta^{(3)}(\mathbf{k}_{123}) Y_{\ell_1 m_1}^*(\hat{\mathbf{k}}_1) \\ &\quad \times Y_{\ell_2 m_2}^*(\hat{\mathbf{k}}_2) Y_{\ell_3 M_3}^*(\hat{\mathbf{k}}_3) g_{\ell_1}(k_1) g_{\ell_2}(k_2) P_\zeta(k_1) P_\zeta(k_2) \\ &\quad \times \left\{ F_{\ell_3 M_3}^{\ell_3 M_3}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) + F_{\ell_3 M_3}^{\ell_3 M_3}(\mathbf{k}_2, \mathbf{k}_1, \mathbf{k}_3) \right\}, \end{aligned} \quad (431)$$

where, as in the previous section, we will sometimes use the notation  $\mathbf{k}_{123} = \mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ .  $P_\zeta(k)$  is the power spectrum of  $\zeta$  given by the usual definition

$$\langle \zeta(\mathbf{k}_1) \rangle = 0, \quad \langle \zeta(\mathbf{k}_1) \zeta(\mathbf{k}_2) \rangle = (2\pi)^3 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2) P_\zeta(k_1). \quad (432)$$

In order to perform the integral over angles,  $\hat{\mathbf{k}}$ , we expand the three-dimensional  $\delta$ -function using Rayleigh's formula,

$$\begin{aligned} \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) &= 8 \sum_{\text{all } \ell' m'} i^{\ell'+\ell_2'+\ell_3'} \mathcal{G}_{\ell' \ell_2' \ell_3'}^{m_1' m_2' m_3'} Y_{\ell' m_1'}(\hat{\mathbf{k}}_1) Y_{\ell_2' m_2'}(\hat{\mathbf{k}}_2) \\ &\quad \times Y_{\ell_3' m_3'}(\hat{\mathbf{k}}_3) \int dr r^2 j_{\ell'}(rk_1) j_{\ell_2'}(rk_2) j_{\ell_3'}(rk_3), \end{aligned} \quad (433)$$

and also expand the angular dependence of  $\mathcal{S}_{\ell m}^{(2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \eta)$  by introducing the transformed source function,  $\mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}^{\mu_1 \mu_2 \mu_3}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \eta)$ , as

$$\begin{aligned} \mathcal{S}_{\lambda_3 \mu_3}^{(2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \eta) &= \sum_{\lambda_1 \mu_1} \sum_{\lambda_2 \mu_2} (-i)^{\lambda_1+\lambda_2} \sqrt{\frac{4\pi}{2\lambda_1+1}} \sqrt{\frac{4\pi}{2\lambda_2+1}} \\ &\quad \times \mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}^{\mu_1 \mu_2 \mu_3}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \eta) Y_{\lambda_1 \mu_1}(\hat{\mathbf{k}}_1) Y_{\lambda_2 \mu_2}(\hat{\mathbf{k}}_2). \end{aligned} \quad (434)$$

This result shows that  $\mathcal{S}_{\lambda_3 \mu_3}^{(2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \eta) = \mathcal{S}_{\lambda_3 \mu_3}^{(2)}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3, \eta)$ , and thus  $F_{\ell m}^{\ell' m'}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$  follows (see (429)).

Now, we can perform the angular integration of (431) to obtain

$$\begin{aligned}
& \tilde{B}_{\ell_1 \ell_2 \ell_3} \\
&= \frac{4}{\pi^2} (-i)^{\ell_1 + \ell_2 + \ell_3} \sum_{\text{all } m} \sum_{\ell' \lambda'} \sum_{\text{all } \mu} i^{\ell'_1 + \ell'_2 + \ell'_3 - \lambda_1 - \lambda_2 - \lambda_3} \\
&\quad \times \sqrt{\frac{4\pi}{(2\lambda_1 + 1)(2\lambda_2 + 1)(2\lambda_3 + 1)}} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \\
&\quad \times \mathcal{G}_{\ell'_1 \ell'_2 \ell'_3}^{m'_1 m'_2 m'_3} \mathcal{G}_{\ell'_1 \ell'_1 \lambda_1}^{m'_1 - m_1 \mu_1} \mathcal{G}_{\ell'_2 \ell'_2 \lambda_2}^{m'_2 - m_2 \mu_2} \mathcal{G}_{\ell'_3 \ell'_3 \lambda_3}^{m'_3 - m_3 \mu_3} \\
&\quad \times \int drr^2 \prod_{i=1}^3 \int k_i^2 dk_i j_{\ell_i}(rk_i) g_{\ell_i}(k_i) g_{\ell_i}(k_2) P_\zeta(k_1) P_\zeta(k_2) \\
&\quad \times i^{\ell_3 + \ell'_3} \int d\eta e^{-\tau} \left\{ \mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}^{\mu_1 \mu_2 \mu_3}(k_1, k_2, k_3, \eta) + 1 \longleftrightarrow 2 \right\} \\
&\quad \times j_{\ell'_3}[k_3(\eta - \eta_0)] + \text{cyclic}, \tag{435}
\end{aligned}$$

where for brevity we indicate  $P_{\lambda_1 \lambda_2 \lambda_3} = \sqrt{4\pi/[(2\lambda_1 + 1)(2\lambda_2 + 1)(2\lambda_3 + 1)]}$ , and we have used the following relation of the Wigner 9j symbol,

$$\begin{aligned}
& (-1)^{\ell'_1 + \ell'_2 + \ell'_3} \sum_{\text{all } mm'} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \mathcal{G}_{\ell'_1 \ell'_2 \ell'_3}^{m'_1 m'_2 m'_3} \mathcal{G}_{\ell'_1 \ell'_1 \lambda_1}^{m'_1 - m_1 \mu_1} \\
&\quad \times \mathcal{G}_{\ell'_2 \ell'_2 \lambda_2}^{m'_2 - m_2 \mu_2} \mathcal{G}_{\ell'_3 \ell'_3 \lambda_3}^{m'_3 - m_3 \mu_3} \\
&= (-1)^R I_{\ell'_1 \ell'_2 \ell'_3} I_{\ell_1 \ell'_1 \lambda_1} I_{\ell_2 \ell'_2 \lambda_2} I_{\ell_3 \ell'_3 \lambda_3} \\
&\quad \times \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell'_1 & \ell'_2 & \ell'_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{Bmatrix} \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix}, \tag{436}
\end{aligned}$$

with  $R \equiv \ell_1 + \ell_2 + \ell_3 + \ell'_1 + \ell'_2 + \ell'_3 + \lambda_1 + \lambda_2 + \lambda_3$ . The Wigner 9j symbols have the permutation symmetry,

$$\begin{aligned}
(-1)^R \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell'_1 & \ell'_2 & \ell'_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{Bmatrix} &= \begin{Bmatrix} \ell_2 & \ell_1 & \ell_3 \\ \ell'_2 & \ell'_1 & \ell'_3 \\ \lambda_2 & \lambda_1 & \lambda_3 \end{Bmatrix} = \begin{Bmatrix} \ell_1 & \ell_3 & \ell_2 \\ \ell'_1 & \ell'_3 & \ell'_2 \\ \lambda_1 & \lambda_3 & \lambda_2 \end{Bmatrix} \\
&= \begin{Bmatrix} \ell'_1 & \ell'_2 & \ell'_3 \\ \ell_1 & \ell_2 & \ell_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{Bmatrix} = \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \\ \ell'_1 & \ell'_2 & \ell'_3 \end{Bmatrix}, \tag{437}
\end{aligned}$$

and the coefficients  $I_{\ell'_1 \ell'_2 \ell'_3}$ ,  $I_{\ell_1 \ell'_1 \lambda_1}$ ,  $I_{\ell_2 \ell'_2 \lambda_2}$ , and  $I_{\ell_3 \ell'_3 \lambda_3}$ , ensure  $\ell'_1 + \ell'_2 + \ell'_3 = \text{even}$ ,  $\ell_1 + \ell'_1 + \lambda_1 = \text{even}$ ,  $\ell_2 + \ell'_2 + \lambda_2 = \text{even}$ , and  $\ell_3 + \ell'_3 + \lambda_3 = \text{even}$ , respectively, which gives  $R = \text{even}$ . Hence, the Wigner 9j coefficients are invariant under the permutations.

Finally, we obtain the angular averaged bispectrum,

$$\begin{aligned}
& \tilde{B}_{\ell_1 \ell_2 \ell_3} \\
&= \frac{4}{\pi^2} \sum_{\text{all } \ell' \lambda'} \sqrt{\frac{4\pi}{(2\lambda_1 + 1)(2\lambda_2 + 1)(2\lambda_3 + 1)}} i^{\ell_3 - \ell'_3 + R} \\
&\quad \times I_{\ell'_1 \ell'_2 \ell'_3} I_{\ell_1 \ell'_1 \lambda_1} I_{\ell_2 \ell'_2 \lambda_2} I_{\ell_3 \ell'_3 \lambda_3} \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell'_1 & \ell'_2 & \ell'_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{Bmatrix} \\
&\quad \times \int dr r^2 \prod_{i=1}^2 \int dk_i k_i^2 P_\zeta(k_i) g_{\ell_i}(k_i) j_{\ell'_i}(rk_i) \int dk_3 k_3^2 j_{\ell'_3}(rk_3) \\
&\quad \times \int dr' e^{-\tau(r')} j_{\ell'_3}(r'k_3) \mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, k_3, r') + \text{perm.}, \tag{438}
\end{aligned}$$

where  $r' \equiv (\eta_0 - \eta)$  and we have used the relation of the spherical Bessel function,  $j_\ell(-x) = (-1)^\ell j_\ell(x)$ . We also define the ‘‘angular-averaged source function,’’

$$\begin{aligned}
& \mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, k_3, r) \\
&\equiv \sum_{\text{all } \mu} \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} \mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}^{\mu_1 \mu_2 \mu_3}(k_1, k_2, k_3, r) \\
&= i^{\lambda_1 + \lambda_2} \sqrt{\frac{2\lambda_1 + 1}{4\pi}} \sqrt{\frac{2\lambda_2 + 1}{4\pi}} \sum_{\text{all } \mu} \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} \\
&\quad \times \int d^2 \hat{\mathbf{k}}_1 d^2 \hat{\mathbf{k}}_2 Y_{\lambda_1 \mu_1}^*(\hat{\mathbf{k}}_1) Y_{\lambda_2 \mu_2}^*(\hat{\mathbf{k}}_2) \mathcal{S}_{\lambda_3 \mu_3}(\mathbf{k}_1, \mathbf{k}_2, r), \tag{439}
\end{aligned}$$

where we have used the inverse relation of (434). Note that cyclic terms in (435) have become permutations because of invariance of the Wigner 9j coefficients under the permutations.

The final analytic formula (438), we have obtained is a general formula which can be applied to any second-order perturbations. The information about the specific second-order terms is contained in the angular-averaged source term,  $\mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}$  (see (439) and (434) for the definition).

For terms that are products of the first-order perturbations, there is indeed a further simplification. One can show that  $\mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, k_3, \eta)$  does not depend on  $k_3$ , that is,  $\mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, k_3, \eta) = \mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, \eta)$ . This property enables us to integrate (438) over  $k_3$ . We obtain

$$\begin{aligned}
& \tilde{B}_{\ell_1 \ell_2 \ell_3} = \frac{2}{\pi} \sum_{\text{all } \ell' \lambda'} \sqrt{\frac{4\pi}{(2\lambda_1 + 1)(2\lambda_2 + 1)(2\lambda_3 + 1)}} i^{\ell_3 - \ell'_3 + R} \\
&\quad \times I_{\ell'_1 \ell'_2 \ell'_3} I_{\ell_1 \ell'_1 \lambda_1} I_{\ell_2 \ell'_2 \lambda_2} I_{\ell_3 \ell'_3 \lambda_3} \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell'_1 & \ell'_2 & \ell'_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{Bmatrix} \\
&\quad \times \int dr e^{-\tau} \prod_{i=1}^2 \int dk_i k_i^2 P_\zeta(k_i) j_{\ell'_i}(rk_i) g_{\ell_i} \\
&\quad \times (k_i) \mathcal{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, r) + \text{perm.}, \tag{440}
\end{aligned}$$

where  $r \equiv (\eta_0 - \eta)$ ,  $R = \ell_1 + \ell_2 + \ell_3 + \ell'_1 + \ell'_2 + \ell'_3 + \lambda_1 + \lambda_2 + \lambda_3$ , and we have used

$$\int dk_3 k_3^2 j_{\ell'_3}(rk_3) j_{\ell_3}(r'k_3) = \frac{\pi}{2r^2} \delta(r - r'). \quad (441)$$

Finally, by adding the remaining term in the full bispectrum, (419), we obtain

$$\begin{aligned} & B_{\ell_1 \ell_2 \ell_3} \\ &= \frac{2}{\pi} \sum_{\text{all } \ell' \lambda} \sqrt{\frac{4\pi}{(2\lambda_1 + 1)(2\lambda_2 + 1)(2\lambda_3 + 1)}} i^{\ell_3 - \ell'_3 + R} \\ & \times I_{\ell'_1 \ell'_2 \ell'_3} I_{\ell_1 \ell'_1 \lambda_1} I_{\ell_2 \ell'_2 \lambda_2} I_{\ell_3 \ell'_3 \lambda_3} \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell'_1 & \ell'_2 & \ell'_3 \\ \lambda_1 & \lambda_2 & \lambda_3 \end{Bmatrix} \\ & \times \int dr e^{-\tau} \prod_{i=1}^2 \int dk_i k_i^2 P_\zeta(k_i) j_{\ell'_i}(rk_i) g_{\ell_i}(k_i) \mathfrak{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, r) \\ & - \frac{3}{2} I_{\ell_1 \ell_2 \ell_3} C_{\ell_1} C_{\ell_2} + \text{perm.} \end{aligned} \quad (442)$$

The remaining task is to calculate the angular-averaged source term,  $\mathfrak{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, k_3, r)$  according to (439). Up to now our scope has been to offer the reader all the general tools that allow a systematic treatment of the CMB angular bispectrum from second-order perturbations. In the following, we will give just some examples of how such tools can be implemented in order to obtain numerical results about signal-to-noise ratios and contamination to primordial non-Gaussianity.

*10.1.3. Source Term: An Example.* The explicit expression for  $S_{\ell m}^{(2)}(\mathbf{k}, \eta)$  can be obtained from (134). Here, we do not want to report the complete expression  $S_{\ell m}^{(2)}(\mathbf{k}, \eta)$  of the source function at second-order, rather, for the goal of this paper, we think it is more instructive to show explicitly how the calculation of  $\mathfrak{S}_{\lambda_1 \lambda_2 \lambda_3}$  proceeds focusing on just one simple example.

Therefore, consider, for example, the term of the second-order source term from (134) given by

$$2 \tau' \delta_e^{(1)} \times \Delta^{(1)}. \quad (443)$$

First, we compute the multipole coefficients of the source term  $S_{\ell m}^{(2)}(k)$  from this contribution, as defined in (424). They are given by the convolution

$$S_{\ell m}^{(2)}(\mathbf{k}) = \int \frac{d^3 k'}{(2\pi)^3} 2\tau' \delta_e^{(1)}(\mathbf{k} - \mathbf{k}') \Delta_{\ell m}^{(1)}(\mathbf{k}'). \quad (444)$$

Now, for  $\Delta_{\ell m}^{(1)}(\mathbf{k})$ , we use

$$\Delta_{\ell m}^{(1)}(\mathbf{k}) = i^\ell \sqrt{\frac{4\pi}{2\ell + 1}} Y_{\ell m}^*(\hat{\mathbf{k}}) \Delta_\ell^{(1)}(\mathbf{k}) (2\ell + 1), \quad (445)$$

so that

$$\begin{aligned} S_{\ell m}^{(2)}(\mathbf{k}) &= \int \frac{d^3 k_1}{(2\pi)^3} \int d^3 k_2 \delta^{(3)}(\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}) \\ & \times i^\ell \sqrt{\frac{4\pi}{2\ell + 1}} (2\ell + 1) 2\tau' \delta_e(\mathbf{k}_2) \Delta_\ell^{(1)}(\mathbf{k}_1) Y_{\ell m}^*(\hat{\mathbf{k}}_1). \end{aligned} \quad (446)$$

We now compute the corresponding ‘‘angular-averaged source function’’ coefficients  $\mathfrak{S}_{\lambda_1 \lambda_2 \lambda_3}(k_1, k_2, r)$  defined by (439). From (446) you read the kernel defined in (430)

$$\begin{aligned} & \mathfrak{S}_{\ell m}^{(2)}(\mathbf{k}_1, \mathbf{k}_2, \eta) \\ &= i^\ell \sqrt{\frac{4\pi}{2\ell + 1}} (2\ell + 1) 2\tau' \delta_e(\mathbf{k}_2) \Delta_\ell^{(1)}(\mathbf{k}_1) Y_{\ell m}^*(\hat{\mathbf{k}}_1). \end{aligned} \quad (447)$$

From here

$$\begin{aligned} & S_{\lambda_1 \lambda_2 \lambda_3}^{(2)}(k_1, k_2, r) \\ &= i^{\lambda_1 + \lambda_2} \sqrt{\frac{2\lambda_1 + 1}{4\pi}} \sqrt{\frac{2\lambda_2 + 1}{4\pi}} \sum_{\text{all } \mu} \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} \\ & \times \int d^2 \hat{\mathbf{k}}_1 \int d^2 \hat{\mathbf{k}}_2 Y_{\lambda_1 \mu_1}^*(\hat{\mathbf{k}}_1) Y_{\lambda_2 \mu_2}^*(\hat{\mathbf{k}}_2) S_{\lambda_3 \mu_3}^{(2)}(\mathbf{k}_1, \mathbf{k}_2, r), \\ &= i^{\lambda_1 + \lambda_2 + \lambda_3} \sqrt{\frac{2\lambda_1 + 1}{4\pi}} \sqrt{\frac{2\lambda_2 + 1}{4\pi}} \sqrt{\frac{4\pi}{2\lambda_3 + 1}} (2\lambda_3 + 1) \\ & \times \sum_{\text{all } \mu} \begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix} \int d^2 \hat{\mathbf{k}}_1 \int d^2 \hat{\mathbf{k}}_2 Y_{\lambda_1 \mu_1}^*(\hat{\mathbf{k}}_1) \\ & \times Y_{\lambda_2 \mu_2}^*(\hat{\mathbf{k}}_2) Y_{\lambda_3 \mu_3}^*(\hat{\mathbf{k}}_2) 2\tau' \delta_e(k_1) \Delta_{\lambda_3}^{(1)}(k_2) \\ &= i^{\lambda_2 + \lambda_3} \delta_{\lambda_1 0} \delta_{\lambda_2 \lambda_3} \sqrt{\frac{2\lambda_2 + 1}{4\pi}} \sqrt{\frac{4\pi}{2\lambda_3 + 1}} (2\lambda_3 + 1) \\ & \times 2\tau' \delta_e(k_1) \Delta_{\lambda_3}^{(1)}(k_2) \sum_{\mu_2} (-1)^{\mu_2} \begin{pmatrix} 0 & \lambda_2 & \lambda_2 \\ 0 & \mu_2 & -\mu_2 \end{pmatrix}, \end{aligned} \quad (448)$$

where we have used  $Y_{\ell m}^* = (-1)^{-m} Y_{\ell -m}$ , and the orthonormality of the spherical harmonics. Using the property of the Wigner  $3j$  symbols

$$\begin{aligned} \sum_{\mu_2} (-1)^{\mu_2} \begin{pmatrix} 0 & \lambda_2 & \lambda_2 \\ 0 & \mu_2 & -\mu_2 \end{pmatrix} &= \sum_{\mu_2} \frac{(-1)^{\lambda_2}}{\sqrt{2\lambda_2 + 1}} \\ &= (-1)^{\lambda_2} \sqrt{2\lambda_2 + 1}, \end{aligned} \quad (449)$$

we find

$$\mathfrak{S}_{\lambda_1 \lambda_2 \lambda_3}^{(2)}(k_1, k_2, r) = 2\tau' (2\lambda_2 + 1)^{(3/2)} \delta_e(k_1) \Delta_{\lambda_3}^{(1)}(k_2, \eta) \delta_{\lambda_1 0} \delta_{\lambda_2 \lambda_3}. \quad (450)$$

This term therefore corresponds to coefficients  $\mathfrak{S}_{0\lambda_2\lambda_2}(k_1, k_2, r)$ .

In general, the perturbation variables of the source term can be split into two parts (see, e.g., (134)). A part containing perturbations that are intrinsically second-order (these perturbations have superscripts (2), and  $\omega_m$  and  $\chi_m$  are also intrinsically second-order). Solving for these terms requires solving the full second-order Boltzmann equations coupled with the Einstein equations.

Another part contains terms that are products of two linear variables, as the example that we have just considered. Evaluation of these terms is much easier than that of the intrinsically second-order terms, as the first-order variables have already been calculated using the standard linearized Boltzmann code such as CMBFAST.

## 10.2. Second-Order Bispectrum from Products of the First-Order Terms

**10.2.1. A Worked Example.** We shall now focus only on the products of the first-order perturbations and we choose to analyze just some of these contributions to offer the reader an example of the analysis one can perform. We warn the reader that the full analysis is under progress and that the full results will be presented in [97], including the intrinsically second-order perturbations which are equally important, and the contribution from perturbing the recombination history [63, 87, 89].

For the products of the first-order perturbations, from now on we will consider the following nonzero four cases for the source terms,  $\mathcal{S}_{\lambda_1\lambda_2\lambda_3}$ , which have been analyzed in [96] (for notational simplicity we shall omit the superscripts (1)). (The following terms are included in the source term  $S_{\ell m}^{(2)}(k, \eta)$  as given in [96, Equation (3.2)]. However notice that that expression does not include all the products of first-order perturbations; the complete expression for the source term will be given in [97]).

$$\begin{aligned}
\mathcal{S}_{000} &= 4i\tau'v_0(k_1)\Delta_1(k_2) \\
&\quad + [2\tau'(\delta_e + \Phi)(k_1) + 8\Psi'(k_1)]\Delta_0(k_2), \\
\mathcal{S}_{110} &= -\frac{20}{\sqrt{3}}\tau'v_0(k_1)v_0(k_2), \\
\mathcal{S}_{101} &= 2i\sqrt{3}\{\tau'v_0(k_1)(4\delta_e + 4\Phi + 2\Delta_0 - \Delta_2)(k_2) + 4k_1\Phi(k_1) \\
&\quad \times (\Delta_0 - \Psi)(k_2) + k_1\Delta_0(k_1)(\Psi + \Phi)(k_2)\}, \\
\mathcal{S}_{112} &= 14\sqrt{\frac{10}{3}}\tau'v_0(k_1)v_0(k_2).
\end{aligned} \tag{451}$$

In particular, it is recognizable the contribution to  $\mathcal{S}_{000}$  from the example just discussed. From these results we find that  $\mathcal{S}_{\lambda_1\lambda_2\lambda_3}$  does not depend on  $k_3$ , that is,  $\mathcal{S}_{\lambda_1\lambda_2\lambda_3} = \mathcal{S}_{\lambda_1\lambda_2\lambda_3}(k_1, k_2, r)$ . Note also that  $\mathcal{S}_{011}(k_1, k_2, r) = \mathcal{S}_{101}(k_2, k_1, r)$ .

**10.2.2. Bispectrum from Products of the First-Order Terms.** For the four nonvanishing combinations of  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_3$  in (451), we rewrite the expression for the bispectrum, (442), as

$$\begin{aligned}
B_{\ell_1\ell_2\ell_3} &= \sum_{\lambda_1\lambda_2\lambda_3} B_{\ell_1\ell_2\ell_3}^{(\lambda_1,\lambda_2,\lambda_3)} + B_{\ell_1\ell_2\ell_3}^{Cl} \\
&= B_{\ell_1\ell_2\ell_3}^{(0,0,0)} + B_{\ell_1\ell_2\ell_3}^{(1,1,0)} + 2B_{\ell_1\ell_2\ell_3}^{(1,0,1)} + B_{\ell_1\ell_2\ell_3}^{(1,1,2)} + B_{\ell_1\ell_2\ell_3}^{Cl},
\end{aligned} \tag{452}$$

where we have used  $B_{\ell_1\ell_2\ell_3}^{(0,1,1)} = B_{\ell_1\ell_2\ell_3}^{(1,0,1)}$ , and defined

$$\begin{aligned}
B_{\ell_1\ell_2\ell_3}^{Cl} &\equiv -3I_{\ell_1\ell_2\ell_3}C_{\ell_1}C_{\ell_2} + \text{cyclic}, \\
B_{\ell_1\ell_2\ell_3}^{(\lambda_1,\lambda_2,\lambda_3)} &\equiv \frac{2}{\pi} \sum_{\text{all } l'} \sqrt{\frac{4\pi}{(2\lambda_1+1)(2\lambda_2+1)(2\lambda_3+1)}} i^{\ell_3-\ell_3'+R} \\
&\quad \times I_{\ell_1'\ell_2'\ell_3'} I_{\ell_1\ell_1'\lambda_1} I_{\ell_2\ell_2'\lambda_2} I_{\ell_3\ell_3'\lambda_3} \begin{Bmatrix} \ell_1 & \ell_2 & \ell_3 \\ \ell_1' & \ell_2' & \ell_3' \\ \lambda_1 & \lambda_2 & \lambda_3 \end{Bmatrix} \\
&\quad \times \int dr e^{-\tau} \prod_{i=1}^2 \int dk_i k_i^2 P_{\zeta}(k_i) j_{l_i}(rk_i) g_{l_i} \\
&\quad \times (k_i) \mathcal{S}_{\lambda_1\lambda_2\lambda_3}(k_1, k_2, r) + \text{perm.}
\end{aligned} \tag{453}$$

To proceed further, it turns out to be useful to simplify the expression by introducing the following notation for the integral over  $k$  that appears many times:

$$[x]_{\ell\ell'}^{(n)}(r) \equiv \frac{2}{\pi} \int dk k^{2+n} P_{\zeta}(k) j_{\ell'}(rk) g_{\ell}(k) x(k, r). \tag{454}$$

This function corresponds to the existing functions in the literature in the appropriate limits. For example, for  $x(k, r) = \pi/2$ , this function is the same as  $\beta_{\ell\ell'}^{(n)}(r)$  introduced in [98]. In fact, we find that an order-of-magnitude estimate of  $[x]_{\ell\ell'}^{(n)}(r)$  is given by  $[x]_{\ell\ell'}^{(n)}(r) \sim 2\beta_{\ell\ell'}^{(n)}(r)/\pi \times x(k = \ell'/r, r)$  for a smooth function of  $x(k, r)$ . As  $\beta_{\ell\ell'}^{(n)}(r)$  is a sharply peaked function at the decoupling epoch,  $r = r_*$ , we find that  $[x]_{\ell\ell'}^{(n)}(r)$  is also sharply peaked at  $r = r_*$ . With these tools in hand, one can calculate  $B_{\ell_1\ell_2\ell_3}^{(0,0,0)}$ ,  $B_{\ell_1\ell_2\ell_3}^{(1,1,0)}$ ,  $B_{\ell_1\ell_2\ell_3}^{(1,0,1)}$ , and  $B_{\ell_1\ell_2\ell_3}^{(1,1,2)}$ . We do not bother here the reader with the details of this computation whose details can be found in [96]. Rather here, we prefer to go straight to the results of the analysis of these bispectra which can be particularly instructive as far as their shape and their contamination to primordial non-Gaussianity are concerned.

**10.3. Shape and Signal-to-Noise of the Second-Order Bispectrum from Products of the First-Order Terms.** One of the motivations for calculating the second-order bispectrum is to see how much the second-order effects in gravity and the photon-baryon fluid contaminate the extraction of the primordial bispectrum. If, for example, the predicted shape of the second-order bispectrum is sufficiently different from that of the primordial bispectrum, then one would hope that the contamination would be minimal. To investigate

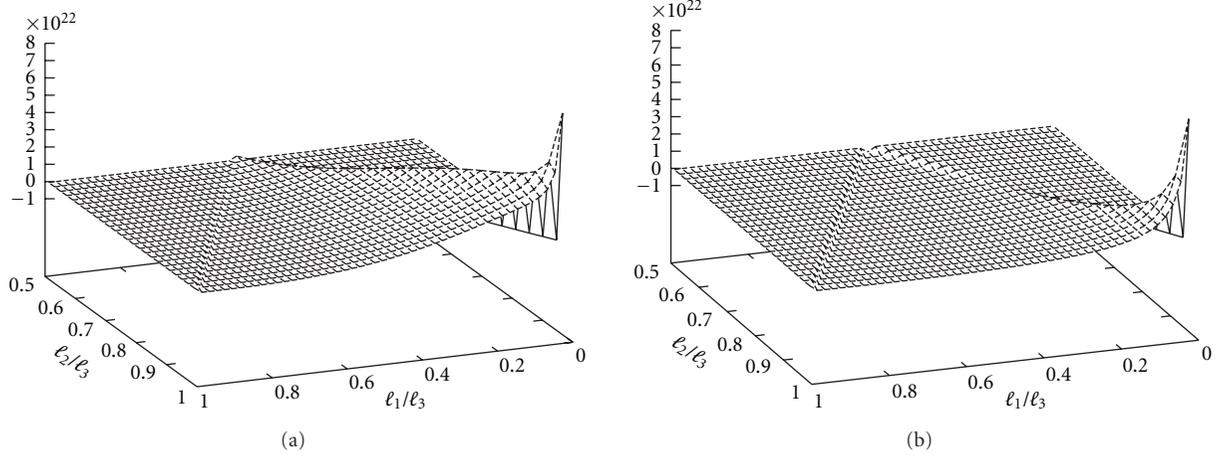


FIGURE 1: Shape dependence of the second-order bispectrum from products of the first-order terms (top) and that of the local primordial bispectrum (bottom). We show  $\ell_1 \ell_2 \langle a_{\ell_1 m_1}^{(1)} a_{\ell_2 m_2}^{(1)} a_{\ell_3 m_3}^{(2)} \rangle (\mathcal{G}_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3})^{-1} / (2\pi)^2 \times 10^{22}$  as a function of  $\ell_1/\ell_3$  and  $\ell_2/\ell_3$  where  $\ell_3 = 200$ . Both shapes have the largest signals in the squeezed triangles,  $\ell_1 \ll \ell_2 \approx \ell_3$ .

this, we shall compare the numerical results of the second-order bispectrum with the so-called ‘‘local’’ model of the primordial bispectrum.

We extract the first-order perturbations from the CMBFAST code. We use the following cosmological parameters:  $\Omega_\Lambda = 0.72$ ,  $\Omega_m = 0.23$ ,  $\Omega_b = 0.046$ ,  $h = 0.70$ , and assume a power law spectrum,  $P_\zeta \propto k^{n-4}$ , with  $n = 1$ . We determine the decoupling time,  $\eta_*$ , from the peak of the visibility function. In this model, we have  $c\eta_0 = 14.9$  Gpc and  $c\eta_* = 288$  Mpc. While the most of the signal is generated in the region of the decoupling epoch, in the low- $\ell$  regime we must also take into account the late time contribution due to the late integrated Sachs-Wolfe effect; thus, we integrate over the line-of-sight,  $r$ , in the following regions:  $c(\eta_0 - 5\eta_*) < r < c(\eta_0 - 0.7\eta_*)$  for  $\ell > 100$ , and  $0 < r < c(\eta_0 - 0.7\eta_*)$  for  $\ell \leq 100$ . The step size is  $\Delta r = 0.1\eta_*$  around the decoupling epoch, and we use the same time steps used by CMBFAST after the decoupling epoch.

The local primordial bispectrum is given by [43]

$$B_{\ell_1 \ell_2 \ell_3} = 2I_{\ell_1 \ell_2 \ell_3} \int_0^\infty r^2 dr b_{\ell_1}^L(r) b_{\ell_2}^L(r) b_{\ell_3}^{\text{NL}}(r) + \text{cyclic}, \quad (455)$$

where

$$\begin{aligned} b_\ell^L(r) &\equiv \frac{2}{\pi} \int_0^\infty k^2 dk P_\Phi(k) g_{T\ell}^{\text{KS}}(k) j_\ell(kr), \\ b_\ell^{\text{NL}}(r) &\equiv \frac{2}{\pi} \int_0^\infty k^2 dk f_{\text{NL}} g_{T\ell}^{\text{KS}}(k) j_\ell(kr). \end{aligned} \quad (456)$$

Note that our linear transfer function,  $g_\ell(k)$ , is related to that of [43],  $g_{T\ell}^{\text{KS}}(k)$ , by  $g_\ell(k) = (3/5)g_{T\ell}^{\text{KS}}(k)$ .

Figure 1 shows the shape of the bispectrum generated by the products of the first-order terms  $f$  selected in (451), and compares it to the primordial local bispectrum, for  $\ell_3 = 200$ . Both shapes (second-order and primordial) have the largest signals in the squeezed triangles,  $\ell_1 \ll \ell_2 \approx \ell_3$ . This is an expected result: both the local primordial

bispectrum and the second-order bispectrum that we have computed here arise from the products of the first-order terms, also products in position space. However, these two shapes are slightly different when  $\ell_1/\ell_3$  is not so small ( $\ell_1/\ell_3 = \mathcal{O}(0.1)$ ): the ways in which the radiation transfer function (which gives the acoustic oscillations) enters into the bispectrum are different for the products of the first-order terms and the primordial bispectrum. The primordial bispectrum contains  $j_\ell(kr_*)g_\ell(k)$  whereas the second-order bispectrum contains  $j_\ell(kr_*)g_\ell(k)x(k, r_*)$ , where  $x = \Delta_0, \nu_0$ , and so forth, also has the oscillations. Therefore, the second-order bispectrum has more interferences between multiple radiation transfer functions. Moreover, the second-order effects contain derivatives that the local primordial effects do not have, which also makes the details of the two shapes different.

Notice, in particular, that most of these gradients in the source term, (134), are contracted with the direction vector,  $\hat{\mathbf{n}}$ . There is only one term that has a scalar product of two wave-vectors,  $\mathbf{k}_1 \cdot \mathbf{k}_2$ , which vanishes in the squeezed limit. The resulting bispectrum, (452), resembles that of a local form, except for the extra powers of  $k$  coming from the derivatives. These extra powers of  $k$  will affect the scale-dependence of the bispectrum, that is, the second-order bispectrum is no longer scale-invariant. Nevertheless, the largest signal of the bispectrum still comes from the squeezed configurations, as the number of extra powers of  $k$  from the derivatives in the source term is not large enough to change the fact that we have the largest contribution when one of  $k_1, k_2$ , and  $k_3$  is very small. In other words, schematically the bispectrum looks like  $B(k_1, k_2, k_3) \sim (k_1^{m_1} k_2^{m_1}) / (k_1^3 k_2^3) + \text{cyclic}$ , where  $m_1$  and  $m_2$  are the extra powers of  $k$  from the derivatives. Therefore, the largest contribution is in the squeezed configurations as long as  $m_i < 3$ . Figure 2 shows the same for  $\ell_3 = 1000$ . The results are similar to those for  $\ell_3 = 200$ , but the acoustic oscillations are more clearly visible.

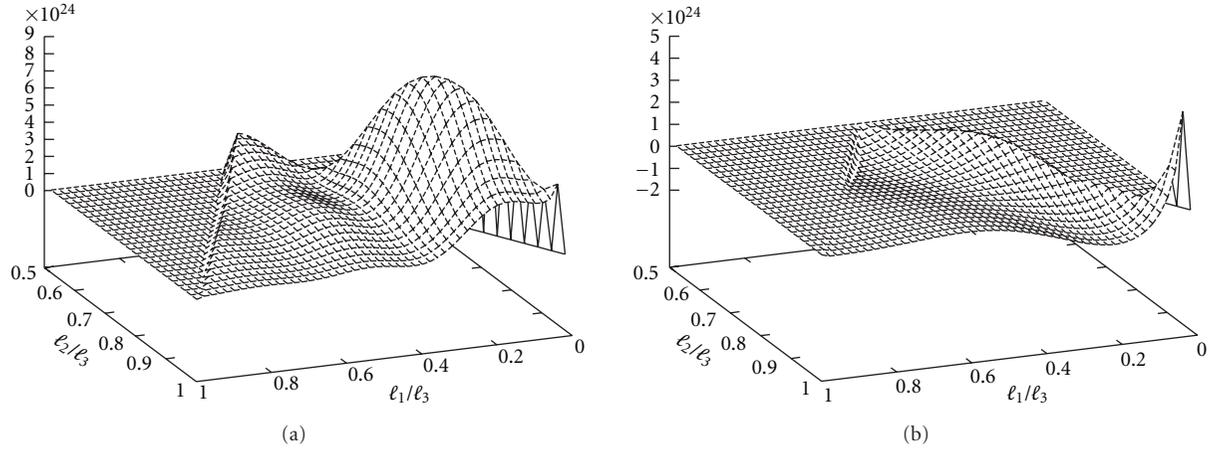


FIGURE 2: Same as Figure 1 for  $\ell_3 = 1000$ . The acoustic oscillations are clearly seen.

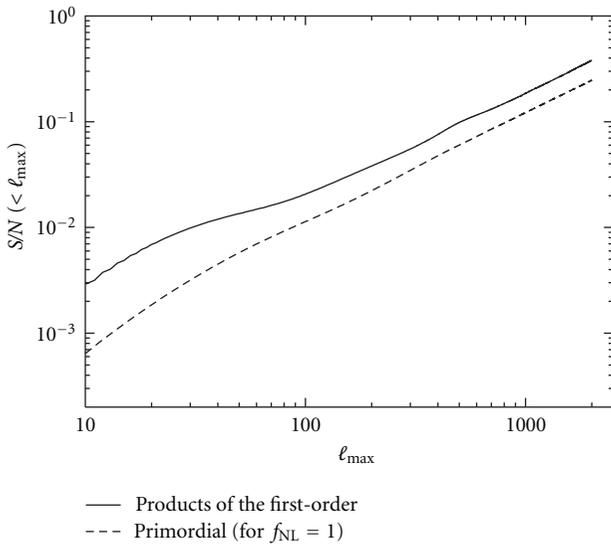


FIGURE 3: Signal-to-noise ratios for the local primordial bispectrum for  $f_{\text{NL}} = 1$  (dashed), and the second-order bispectrum from the products of the first-order terms (solid), for an ideal full-sky and cosmic-variance-limited (noiseless) experiment.

We can quantify the degree to which the second-order and the primordial bispectra are correlated, as well as the expected signal-to-noise ratio of the second-order bispectrum, following the general definitions summarized in Section 8. Notice that we are ignoring the noise contribution. In other words, we shall only consider ideal cosmic-variance limited experiments with full sky coverage, which however, at least for the multipole maximum multipole of  $l_{\text{max}} \sim 2000$  we will consider, is a good reference for an experiment like *Planck* (see, e.g., [43]).

The signal-to-noise ratio is given by (356). In Figure 3 we show the cumulative signal-to-noise ratio, summed up to a maximum multipole of  $l_{\text{max}}$ , of the primordial bispectrum, assuming  $f_{\text{NL}} = 1$  and ignoring the second-order bispectrum, that is,  $(S/N)_{\text{prim}} = (F_{\text{prim,prim}})^{1/2}$ , as well as

that of the second-order bispectrum, ignoring the primordial bispectrum, that is,  $(S/N)_{\text{2nd}} = (F_{\text{2nd,2nd}})^{1/2}$ . In both cases  $S/N$  increases roughly as  $S/N \propto \ell_{\text{max}}$  (or  $\propto \sqrt{N_{\text{pix}}}$  where  $N_{\text{pix}}$  is the number of independent pixels in the map). A larger contribution to the second-order bispectrum at  $\ell \leq 50$  comes from the terms involving the Integrated Sachs-Wolfe effect. The signal-to-noise ratio of the second-order bispectrum reaches  $\sim 0.4$  at  $\ell_{\text{max}} = 2000$ ; thus, this signal is undetectable. While our calculation includes the temperature anisotropy only, including polarization would increase the signal-to-noise by a factor of two at most, which would not be enough to push the signal-to-noise above unity.

How similar are the second-order and the primordial bispectra? In Figure 4, we show the cross-correlation coefficient, between the local bispectrum and second-order bispectrum from the products of the first-order terms given in (451). The cross-correlation coefficient (see (357)), reaches  $\sim 0.5$  for  $\ell_{\text{max}} = 200$ , and the shapes for  $\ell_3 = 200$  are shown in Figure 1. After  $\ell_{\text{max}} = 200$  the correlation weakens, and reaches  $\sim 0.35$  at  $\ell_{\text{max}} = 1000$ , and the shapes for  $\ell_3 = 1000$  are shown in Figure 2. These results show that the second-order bispectrum from the products of the first-order perturbations and the local primordial bispectrum are fairly similar, with a sizable correlation coefficient.

How large is the contamination of the primordial bispectrum? The level of contamination is measured by the effective nonlinearity parameter given by (359).

In Figure 5, we show  $f_{\text{NL}}^{\text{con}}$  as a function of the maximum multipoles,  $\ell_{\text{max}}$ . We find that  $f_{\text{NL}}^{\text{con}}$  reaches the maximum value,  $\sim 0.9$ , when the correlation coefficient reaches the maximum at  $\ell_{\text{max}} \sim 200$ , but then decreases to  $\sim 0.5$  at  $\ell_{\text{max}} \sim 2000$ . Therefore, we conclude that the contamination of the primordial bispectrum due to the second-order bispectrum from the terms in (451) is negligible for CMB experiments.

Finally, one can also calculate the  $1\text{-}\sigma$  uncertainty of  $f_{\text{NL}}$ ,  $\delta f_{\text{NL}}$ , with the second-order bispectrum marginalized over. This is given by  $\delta f_{\text{NL}} = \sqrt{(F)_{\text{prim,prim}}^{-1}|_{f_{\text{NL}}=1}}$ . Figure 6 shows that an increase in the uncertainty of  $f_{\text{NL}}$  due to marginalization is totally negligible.

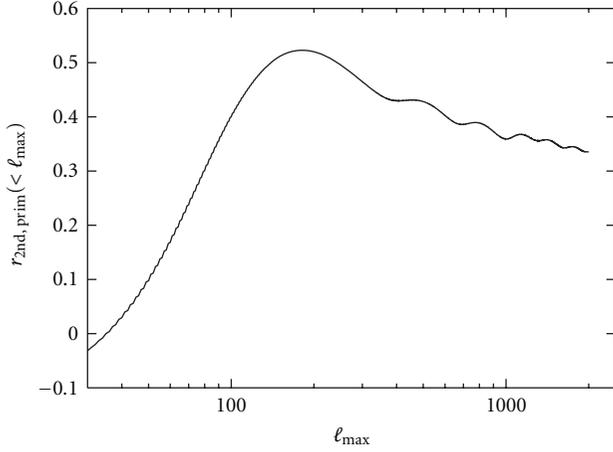


FIGURE 4: The cross-correlation coefficient between the second-order bispectrum from the products of the first-order terms and the local primordial bispectrum.

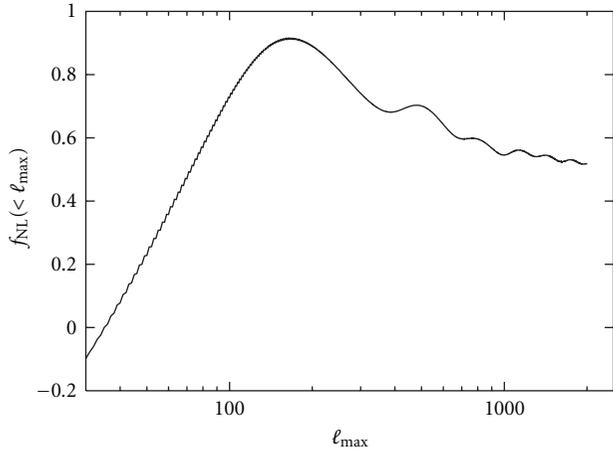


FIGURE 5: Contamination of the local primordial bispectrum as measured by  $f_{\text{NL}}^{\text{con}}$  (359).

## 11. Conclusions

In this paper we have addressed a basic question in cosmology: how a primordial NG propagates into an observable like the CMB anisotropy. Answering this question is fundamental as it will help us in getting some knowledge about the way the primordial cosmological perturbation was generated at the very early stages of the evolution of the universe. In the first sections, we have shown how to set the initial conditions at second-order for the (gauge-invariant) CMB anisotropy when some source of primordial NG is present. This was more or less straightforward because on large angular scales it is basically gravity which dictates the nonlinear dynamics. On small angular scales, the computation of the second-order effects in the CMB anisotropy

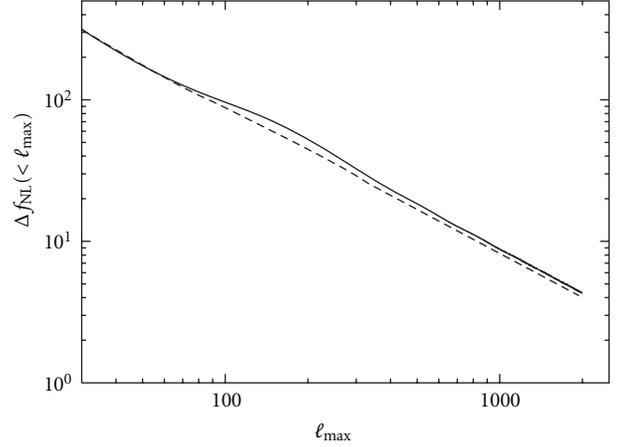


FIGURE 6: Projected uncertainty of  $f_{\text{NL}}$  with (dashed) and without (solid) the second-order bispectrum marginalized over.

is far more difficult, as so many sources of nonlinearities appear. In this paper, we have focussed on the study of the second-order effects appearing at the recombination era when the CMB anisotropy is left imprinted. We have shown how to derive the equations which allow to evaluate CMB anisotropies, by computing the Boltzmann equations describing the evolution of the baryon-photon fluid up to second order. This allows to follow the time evolution of CMB anisotropies (up to second order) on all scales, from the early epoch, when the cosmological perturbations were generated, to the present time, through the recombination era. Through some analytical and simplified example, we have also shown how to estimate the contamination of the recombination secondary effects onto the detection of primordial NG. More refined numerical results confirm our estimates and are also reported here. It goes without saying that this line of research should be pursued until the level of accuracy in the theoretical prediction is reached and is comparable to the one provided by the current or future satellite experiments.

## Appendices

### A. Energy-Momentum Tensor

*A.1. Energy-Momentum Tensor for Photons.* The energy-momentum tensor for photons is defined as

$$T_{\gamma\nu}^{\mu} = \frac{2}{\sqrt{-g}} \int \frac{d^3P}{(2\pi)^3} \frac{P^{\mu}P_{\nu}}{P^0} f, \quad (\text{A.1})$$

where  $g$  is the determinant of the metric (1) and  $f$  is the distribution function. We thus obtain

TABLE 1

Symbol	Definition	Equation
$\Phi, \Psi$	Gravitational potentials in Poisson gauge	(1)
$\omega_i$	2nd-order vector perturbation in Poisson gauge	(1)
$\chi_{ij}$	2nd-order tensor perturbation in Poisson gauge	(1)
$\eta$	Conformal time	(1)
$f$	Photon distribution function	(64)
$g$	Distribution function for massive particles	(84) & (156)
$f^{(i)}$	$i$ th order perturbation of the photon distribution function	(79)
$f_{\ell m}^{(i)}$	Moments of the photon distribution function	(98)
$C(\mathbf{p})$	Collision term	(83) & (87)
$p$	Magnitude of photon momentum ( $\mathbf{p} = pn^i$ )	(59)
$n^i$	Propagation direction	(62)
$\Delta^{(1)}(x^i, n^i, \eta)$	First-order fractional energy photon fluctuations	(128)
$\Delta^{(2)}(x^i, n^i, \eta)$	Second-order fractional energy photon fluctuations	(132)
$n_e$	Electron number density	(166)
$\delta_e(\delta_b)$	Electron (baryon) density perturbation	(94)
$\mathbf{k}$	Wavenumber	(142)
$v_m$	Baryon velocity perturbation	(145) & (146)
$v_{\text{CDM}}^{(2)i}$	Cold dark matter velocity	(201)
$v_\gamma^{(2)i}$	Second-order photon velocity	(187)
$S_{\ell m}$	Temperature source term	(143)
$\tau$	Optical depth	(140)
$\bar{\rho}_\gamma(\bar{\rho}_b)$	Background photon (baryon) energy density	(198)
$B_{\ell_1, \ell_2, \ell_3}$	CMB angular-averaged bispectrum	(353)
$F_{ij}$	Fisher matrix for the amplitudes of the bispectra	(354)
$f_{\text{NL}}^{\text{con}}$	contamination to primordial non-Gaussianity	(359)

$$T_{\gamma 0}^0 = -\bar{\rho}_\gamma \left( 1 + \Delta_{00}^{(1)} + \frac{\Delta_{00}^{(2)}}{2} \right),$$

$$T_{\gamma 0}^i = -\frac{4}{3} e^{\Psi+\Phi} \bar{\rho}_\gamma \left( v_\gamma^{(1)i} + \frac{1}{2} v_\gamma^{(2)i} + \Delta_{00}^{(1)} v_\gamma^{(1)i} \right) + \frac{1}{3} \bar{\rho}_\gamma e^{\Psi-\Phi} \omega^i$$
(A.2)

$$T_{\gamma j}^i = \bar{\rho}_\gamma \left( \Pi_{\gamma j}^i + \frac{1}{3} \delta_j^i \left( 1 + \Delta_{00}^{(1)} + \frac{\Delta_{00}^{(2)}}{2} \right) \right),$$

where  $\bar{\rho}_\gamma$  is the background energy density of photons and

$$\Pi_{\gamma}^{ij} = \int \frac{d\Omega}{4\pi} \left( n^i n^j - \frac{1}{3} \delta^{ij} \right) \left( \Delta^{(1)} + \frac{\Delta^{(2)}}{2} \right),$$
(A.3)

is the quadrupole moment of the photons.

**A.2. Energy-Momentum Tensor for Massive Particles.** The energy-momentum tensor for massive particles of mass  $m$ , number density  $n$  and degrees of freedom  $g_d$

$$T_{m\nu}^\mu = \frac{g_d}{\sqrt{-g}} \int \frac{d^3Q}{(2\pi)^3} \frac{Q^\mu Q_\nu}{Q^0} g_m,$$
(A.4)

where  $g_m$  is the distribution function. We obtain

$$T_{m0}^0 = -\rho_m = -\bar{\rho}_m \left( 1 + \delta_m^{(1)} + \frac{1}{2} \delta_m^{(2)} \right),$$

$$T_{m0}^i = -e^{\Psi+\Phi} \rho_m v_m^i$$

$$= -e^{\Phi+\Psi} \bar{\rho}_m \left( v_m^{(1)i} + \frac{1}{2} v_m^{(2)i} + \delta_m^{(1)} v_m^{(1)i} \right)$$
(A.5)

$$T_{mj}^i = \rho_m \left( \delta_j^i \frac{T_m}{m} + v_m^i v_{mj} \right) = \bar{\rho}_m \left( \delta_j^i \frac{T_m}{m} + v_m^{(1)i} v_{mj}^{(1)} \right),$$

where  $\bar{\rho}_m$  is the background energy density of massive particles and we have included the equilibrium temperature  $T_m$ .

## B. Solutions of Einstein's Equations in Various Eras

**B.1. Matter-Dominated Era.** During the phase in which the CDM is dominating the energy density of the Universe,  $a \sim \eta^2$  and we may use (34) to obtain an equation for the gravitational potential at first order in perturbation theory

(for which  $\Phi^{(1)} = \Psi^{(1)}$ )

$$\Phi^{(1)''} + 3\mathcal{H}\Phi^{(1)'} = 0, \quad (\text{B.1})$$

which has two solutions  $\Phi_+^{(1)} = \text{constant}$  and  $\Phi_-^{(1)} = \mathcal{H}/a^2$ . At the same order of perturbation theory, the CDM velocity can be read off from (30)

$$v^{(1)i} = -\frac{2}{3\mathcal{H}}\partial^i\Phi^{(1)}. \quad (\text{B.2})$$

The matter density contrast  $\delta^{(1)}$  satisfies the first-order continuity equation

$$\delta^{(1)'} = -\frac{\partial v^{(1)i}}{\partial x^i} = -\frac{2}{3\mathcal{H}}\nabla^2\Phi^{(1)}. \quad (\text{B.3})$$

Going to Fourier space, this implies that

$$\delta_k^{(1)} = \delta_k^{(1)}(0) + \frac{k^2\eta^2}{6}\Phi_k^{(1)}, \quad (\text{B.4})$$

where  $\delta_k^{(1)}(0)$  is the initial condition in the matter-dominated period.

At second order, using (34) and (32), and the fact that the first-order gravitational potential is constant, we find an equation for the gravitational potential at second-order  $\Psi^{(2)}$

$$\begin{aligned} \Psi^{(2)''} + 3\mathcal{H}\Psi^{(2)'} &= S_m, \\ S_m &= -\partial_k\Phi^{(1)}\partial^k\Phi^{(1)} + N \\ &= -\partial_k\Phi^{(1)}\partial^k\Phi^{(1)} + \frac{10}{3}\frac{\partial_i\partial^j}{\nabla^2}(\partial_i\Phi^{(1)}\partial_j\Phi^{(1)}), \end{aligned} \quad (\text{B.5})$$

whose solution is

$$\begin{aligned} \Psi^{(2)} &= \Psi_m^{(2)}(0) \\ &+ \int_0^\eta d\eta' \frac{\Phi_+^{(1)}(\eta)\Phi_-^{(1)}(\eta') - \Phi_-^{(1)}(\eta)\Phi_+^{(1)}(\eta')}{W(\eta')} S_m(\eta') \\ &= \Psi_m^{(2)}(0) \\ &- \frac{1}{14} \left( \partial_k\Phi^{(1)}\partial^k\Phi^{(1)} - \frac{10}{3}\frac{\partial_i\partial^j}{\nabla^2}(\partial_i\Phi^{(1)}\partial_j\Phi^{(1)}) \right) \eta^2, \end{aligned} \quad (\text{B.6})$$

with  $W(\eta) = W_0/a^3$  ( $a_0 = 1$ ) the Wronskian obtained from the corresponding homogeneous equation. In (B.6),  $\Psi_m^{(2)}(0)$  represents the initial condition (taken conventionally at  $\eta \rightarrow 0$ ) deep in the matter-dominated phase.

From (35), we may compute the vector perturbation in the metric

$$-\frac{1}{2}\nabla^2\omega^i = 3\mathcal{H}^2\frac{1}{\nabla^2}\partial_j(\partial^i\delta^{(1)}v^{(1)j} - \partial^j\delta^{(1)}v^{(1)i}), \quad (\text{B.7})$$

where we have made use of the fact that the vector part of the CDM velocity satisfies the relation  $(\delta_j^i - (\partial^i\partial_j/\nabla^2))v^{(2)i} = -\omega^i$ .

**B.2. Radiation-Dominated Era.** We consider a universe dominated by photons and massless neutrinos. The energy-momentum tensor for massless neutrinos has the same form as that for photons. During the phase in which radiation is dominating the energy density of the Universe,  $a \sim \eta$  and we may combine (29) and (34) to obtain an equation for the gravitational potential  $\Psi^{(1)}$  at first order in perturbation theory

$$\begin{aligned} \Psi^{(1)''} + 4\mathcal{H}\Psi^{(1)'} - \frac{1}{3}\nabla^2\Psi^{(1)} &= \mathcal{H}Q^{(1)'} + \frac{1}{3}\nabla^2Q^{(1)}, \\ \nabla^2Q^{(1)} &= \frac{9}{2}\mathcal{H}^2\frac{\partial_i\partial^j}{\nabla^2}\Pi_{Tj}^{(1)i}, \end{aligned} \quad (\text{B.8})$$

where the total anisotropic stress tensor is

$$\Pi_{Tj}^i = \frac{\bar{\rho}_\gamma}{\bar{\rho}_T}\Pi_{\gamma j}^i + \frac{\bar{\rho}_\nu}{\bar{\rho}_T}\Pi_{\nu j}^i. \quad (\text{B.9})$$

We may safely neglect the quadrupole and solve (B.8) setting  $u_\pm = \Phi_\pm^{(1)}\eta$ . Then, (B.8), in Fourier space, becomes

$$u'' + \frac{2}{\eta}u' + \left(\frac{k^2}{3} - \frac{2}{\eta^2}\right)u = 0. \quad (\text{B.10})$$

This equation has as independent solutions  $u_+ = j_1(k\eta/\sqrt{3})$ , the spherical Bessel function of order 1, and  $u_- = n_1(k\eta/\sqrt{3})$ , the spherical Neumann function of order 1. The latter blows up as  $\eta$  gets small and we discard it on the basis of initial conditions. The final solution is therefore

$$\begin{aligned} \Phi_k^{(1)} &= 3\Phi^{(1)}(0) \\ &\times \frac{\sin(k\eta/\sqrt{3}) - (k\eta/\sqrt{3})\cos(k\eta/\sqrt{3})}{(k\eta/\sqrt{3})^3} \end{aligned} \quad (\text{B.11})$$

where  $\Phi^{(1)}(0)$  represents the initial condition deep in the radiation era.

At the same order in perturbation theory, the radiation velocity can be read off from (30)

$$v_\gamma^{(1)i} = -\frac{1}{2\mathcal{H}^2}\frac{(a\partial^i\Phi^{(1)})}{a}. \quad (\text{B.12})$$

At second order, combining (29), (34), we find

$$\Psi^{(2)'} + 4\mathcal{H}\Psi^{(2)'} - \frac{1}{3}\nabla^2\Psi^{(2)} = S_y, \quad (\text{B.13})$$

$$\begin{aligned} S_y &= 4\left(\Psi^{(1)'}\right)^2 + 2\Phi^{(1)'}\Psi^{(1)'} + \frac{4}{3}\left(\Phi^{(1)} + \Psi^{(1)}\right)\nabla^2\Psi^{(1)} \\ &\quad - \frac{2}{3}\left(\partial_k\Phi^{(1)}\partial^k\Phi^{(1)} + \partial_k\Psi^{(1)}\partial^k\Psi^{(1)} - \partial_k\Phi^{(1)}\partial^k\Psi^{(1)}\right) \\ &\quad + \mathcal{H}Q^{(2)'} + \frac{1}{3}\nabla^2Q^{(2)} + \frac{4}{3}\left(\Phi^{(1)} + \Psi^{(1)}\right)\nabla^2Q^{(1)}, \end{aligned} \quad (\text{B.14})$$

$$\begin{aligned} &\frac{1}{2}\nabla^2Q^{(2)} \\ &= -\partial_k\Phi^{(1)}\partial^k\Psi^{(1)} - \frac{1}{2}\left(\partial_k\Phi^{(1)}\partial^k\Phi^{(1)} - \partial_k\Psi^{(1)}\partial^k\Psi^{(1)}\right) \\ &\quad + 3\frac{\partial_i\partial^j}{\nabla^2}\left[\partial^i\Phi^{(1)}\partial_j\Psi^{(1)} + \frac{1}{2}\left(\partial^i\Phi^{(1)}\partial_j\Phi^{(1)} - \partial^i\Psi^{(1)}\partial_j\Psi^{(1)}\right)\right] \\ &\quad + \frac{9}{2}\mathcal{H}^2\frac{\partial_i\partial^j}{\nabla^2}\frac{\Pi_{Tj}^{(2)i}}{2} - 9\mathcal{H}^2\frac{\partial_i\partial^j}{\nabla^2}\left(\Psi^{(1)}\Pi_{Tj}^{(1)i}\right), \end{aligned} \quad (\text{B.15})$$

whose solution is

$$\begin{aligned} \Psi^{(2)} &= \Psi_{\text{hom.}}^{(2)} \\ &\quad + \int_0^\eta d\eta' \frac{\Phi_+^{(1)}(\eta)\Phi_-^{(1)}(\eta') - \Phi_-^{(1)}(\eta)\Phi_+^{(1)}(\eta')}{W(\eta')} S_y(\eta'), \end{aligned} \quad (\text{B.16})$$

where  $W(\eta) = (a(0)/a)^4$  is the Wronskian, and  $\Psi_{\text{hom.}}^{(2)}$  is the solution of the homogeneous equation.

The equation of motion for the vector metric perturbations reads

$$\begin{aligned} &-\frac{1}{2}\nabla^2\omega^i + 4\mathcal{H}^2\omega^i \\ &= \left(\delta_j^i - \frac{\partial^i\partial_j}{\nabla^2}\right) \\ &\quad \times \left[ 2\Psi^{(1)'}\partial^j\Phi^{(1)} + \mathcal{H}^2\frac{\bar{\rho}_\gamma + \bar{p}_\gamma}{\bar{\rho}_T}\omega^j \right. \\ &\quad - 2\mathcal{H}^2\left(\frac{\bar{p}_\gamma}{\bar{\rho}_T}v_\gamma^{(2)j} + \frac{\bar{p}_\nu}{\bar{\rho}_T}v_\nu^{(2)i} + 2\frac{\bar{p}_\gamma}{\bar{\rho}_T}\Delta_{00}^{(1)\gamma}v_\gamma^{(1)j} \right. \\ &\quad + 2\frac{\bar{p}_\nu}{\bar{\rho}_T}\Delta_{00}^{(1)\nu}v_\nu^{(1)j} + 2\left(\Phi^{(1)} - \Psi^{(1)}\right)\frac{\bar{p}_\gamma}{\bar{\rho}_T}v_\gamma^{(1)j} \\ &\quad \left. \left. + 2\left(\Phi^{(1)} - \Psi^{(1)}\right)\frac{\bar{p}_\nu}{\bar{\rho}_T}v_\nu^{(1)j}\right) \right], \end{aligned} \quad (\text{B.17})$$

where  $\bar{\rho}_T$  is the total background energy density. The Einstein equations for a universe filled by CDM and a relativistic component can be found in [47].

## C. Linear Solution of the Boltzmann Equations

In this section, we will solve the Boltzmann equations at first order in perturbation theory. The interested reader will find the extension of these formulae to second order in [47]. The first two moments of the photon Boltzmann equation are obtained by integrating (129) over  $d\Omega_n/4\pi$  and  $d\Omega_n n^i/4\pi$ , respectively, and they lead to the density and velocity continuity equations

$$\Delta_{00}^{(1)'} + \frac{4}{3}\partial_i v_\gamma^{(1)i} - 4\Psi^{(1)'} = 0, \quad (\text{C.1})$$

$$v_\gamma^{(1)i'} + \frac{3}{4}\partial_j \Pi_\gamma^{(1)ji} + \frac{1}{4}\Delta_{00}^{(1),i} + \Phi^{(1),i} = -\tau' \left( v^{(1)i} - v_\gamma^{(1)i} \right), \quad (\text{C.2})$$

where  $\Pi^{ij}$  is the photon quadrupole moment, defined in (195).

Let us recall here that  $\delta_\gamma^{(1)} = \Delta_{00}^{(1)} = \int d\Omega \Delta^{(1)}/4\pi$  and that the photon velocity is given by (186).

The two equations above are complemented by the momentum continuity equation for baryons, which can be conveniently written as

$$v^{(1)i} = v_\gamma^{(1)i} + \frac{R}{\tau'} \left[ v^{(1)i'} + \mathcal{H}v^{(1)i} + \Phi^{(1),i} \right], \quad (\text{C.3})$$

where we have introduced the baryon-photon ratio  $R \equiv 3\rho_b/(4\rho_\gamma)$ .

Equation (C.3) is in a form ready for a consistent expansion in the small quantity  $\tau^{-1}$  which can be performed in the tight-coupling limit. By first taking  $v^{(1)i} = v_\gamma^{(1)i}$  at zero order and then using this relation in the left-hand side of (C.3) one obtains

$$v^{(1)i} - v_\gamma^{(1)i} = \frac{R}{\tau'} \left[ v_\gamma^{(1)i'} + \mathcal{H}v_\gamma^{(1)i} + \Phi^{(1),i} \right]. \quad (\text{C.4})$$

Such an expression for the difference of velocities can be used in (C.2) to give the evolution equation for the photon velocity in the limit of tight coupling

$$v_\gamma^{(1)i'} + \mathcal{H}\frac{R}{1+R}v_\gamma^{(1)i} + \frac{1}{4}\frac{\Delta_{00}^{(1),i}}{1+R} + \Phi^{(1),i} = 0. \quad (\text{C.5})$$

Notice that in (C.5) we are neglecting the quadrupole of the photon distribution  $\Pi^{(1)ij}$  (and all the higher moments) since it is well known that at linear order such moment(s) are suppressed in the tight-coupling limit by (successive powers of)  $1/\tau$  with respect to the first two moments, the photon energy density and velocity. Equations (C.1) and (C.5) are the master equations which govern the photon-baryon fluid acoustic oscillations before the epoch of recombination when photons and baryons are tightly coupled by Compton scattering.

In fact, one can combine these two equations to get a single second-order differential equation for the photon energy density perturbations  $\Delta_{00}^{(1)}$ . Deriving (C.1) with respect to conformal time and using (C.5) to replace  $\partial_i v_\gamma^{(1)i}$  yields

$$\begin{aligned} & \left( \Delta_{00}^{(1)''} - 4\Psi^{(1)''} \right) + \mathcal{H} \frac{R}{1+R} \left( \Delta_{00}^{(1)'} - 4\Psi^{(1)'} \right) \\ & - c_s^2 \nabla^2 \left( \Delta_{00}^{(1)} - 4\Psi^{(1)} \right) = \frac{4}{3} \nabla^2 \left( \Phi^{(1)} + \frac{\Psi^{(1)}}{1+R} \right), \end{aligned} \quad (\text{C.6})$$

where  $c_s = 1/\sqrt{3(1+R)}$  is the speed of sound of the photon-baryon fluid. Indeed, in order to solve (C.6) one needs to know the evolution of the gravitational potentials. We will come back later to the discussion of the solution of (C.6).

A useful relation is obtained by considering the continuity equation for the baryon density perturbation. By perturbing at first order (176), we obtain

$$\delta_b^{(1)'} + v_{,i}^i - 3\Psi^{(1)'} = 0. \quad (\text{C.7})$$

Subtracting (C.7) from (C.1) brings

$$\Delta_{00}^{(1)'} - \frac{4}{3} \delta_b^{(1)'} + \frac{4}{3} (v_\gamma^{(1)i} - v^{(1)i})_{,i} = 0, \quad (\text{C.8})$$

which implies that at lowest order in the tight-coupling approximation

$$\Delta_{00}^{(1)} = \frac{4}{3} \delta_b^{(1)}, \quad (\text{C.9})$$

for adiabatic perturbations.

*C.1. Linear Solutions in the Limit of Tight Coupling.* In this section, we briefly recall how to obtain at linear order the solutions of the Boltzmann equations (C.6). These correspond to the acoustic oscillations of the photon-baryon fluid for modes which are within the horizon at the time of recombination. It is well known that, in the variable  $(\Delta_{00}^{(1)} - 4\Psi^{(1)})$ , the solution can be written as [73, 74]

$$\begin{aligned} & [1+R(\eta)]^{1/4} (\Delta_{00}^{(1)} - 4\Psi^{(1)}) \\ & = A \cos[kr_s(\eta)] + B \sin[kr_s(\eta)] \\ & - 4 \frac{k}{\sqrt{3}} \int_0^\eta d\eta' [1+R(\eta')]^{3/4} \left( \Phi^{(1)}(\eta') + \frac{\Psi^{(1)}(\eta')}{1+R(\eta')} \right) \\ & \times \sin[k(r_s(\eta) - r_s(\eta'))], \end{aligned} \quad (\text{C.10})$$

where the sound horizon is given by  $r_s(\eta) = \int_0^\eta d\eta' c_s(\eta')$ , with  $R = 3\rho_b/(4\rho_\gamma)$ . The constants  $A$  and  $B$  in (C.10) are fixed by the choice of initial conditions.

In order to give an analytical, we will use some simplifications following [70, 99]. First, for simplicity, we are going to neglect the ratio  $R$  wherever it appears, *except* in the arguments of the varying cosines and sines, where we will treat  $R = R_*$  as a constant evaluated at the time of recombination. In this way, we keep track of a damping of the photon velocity amplitude with respect to the case  $R = 0$  which prevents the acoustic peaks in the power-spectrum to disappear. Treating  $R$  as a constant is justified by the fact that for modes within the horizon the time scale of the oscillations is much shorter than the time scale on which  $R$  varies. If  $R$  is a constant the sound speed is just a constant  $c_s = 1/\sqrt{3(1+R_*)}$ , and the sound horizon is simply  $r_s(\eta) = c_s \eta$ .

Second, we are going to solve for the evolutions of the perturbations in two well distinguished limiting regimes. One regime is for those perturbations which enter the Hubble radius when matter is the dominant component, that is at times much bigger than the equality epoch, with  $k \ll k_{\text{eq}} \sim \eta_{\text{eq}}^{-1}$ , where  $k_{\text{eq}}$  is the wavenumber of the Hubble radius at the equality epoch. The other regime is for those perturbations with much smaller wavelengths which enter the Hubble radius when the Universe is still radiation dominated, that is perturbations with wavenumbers  $k \gg k_{\text{eq}} \sim \eta_{\text{eq}}^{-1}$ . In fact, we are interested in perturbation modes which are within the horizon by the time of recombination  $\eta_*$ . Therefore, we will further suppose that  $\eta_* \gg \eta_{\text{eq}}$  in order to study such modes in the first regime. Even though  $\eta_* \gg \eta_{\text{eq}}$  is not the real case, it allows to obtain some analytical expressions.

Before solving for these two regimes let us fix our initial conditions, which are taken on large scales deep in the radiation dominated era (for  $\eta \rightarrow 0$ ). During this epoch, for adiabatic perturbations, the gravitational potentials remain constant on large scales (we are neglecting anisotropic stresses so that  $\Phi^{(1)} \simeq \Psi^{(1)}$ ) and from the  $(0-0)$ -component of Einstein equations

$$\Phi^{(1)}(0) = -\frac{1}{2} \Delta_{00}^{(1)}(0). \quad (\text{C.11})$$

On the other hand, from the energy continuity equation (C.1) on large scales

$$\Delta_{00}^{(1)} - 4\Psi^{(1)} = \text{const.}; \quad (\text{C.12})$$

from (C.11) the constant on the right-hand side of (C.12) is fixed to be  $-6\Phi^{(1)}(0)$ ; thus we find  $B = 0$  and  $A = -6\Phi^{(1)}(0)$ .

With our simplifications, (C.10) then reads

$$\begin{aligned} \Delta_{00}^{(1)} - 4\Psi^{(1)} & = -6\Phi^{(1)}(0) \cos(\omega_0 \eta) \\ & - \frac{8k}{\sqrt{3}} \int_0^\eta d\eta' \Phi^{(1)}(\eta') \sin[\omega_0(\eta - \eta')], \end{aligned} \quad (\text{C.13})$$

where  $\omega_0 = kc_s$ .

C.2. *Perturbation Modes with  $k \gg k_{\text{eq}}$ .* This regime corresponds to perturbation modes which enter the Hubble radius when the universe is matter dominated at times  $\eta \gg \eta_{\text{eq}}$ . During matter domination, the gravitational potential remains constant (both on superhorizon and subhorizon scales), as one can see for example from (B.1), and its value is fixed to  $\Phi^{(1)}(k, \eta) = (9/10)\Phi^{(1)}(0)$ , where  $\Phi^{(1)}(0)$  corresponds to the gravitational potential on large scales during the radiation dominated epoch. Since we are interested in the photon anisotropies around the time of recombination, when matter is dominating, we can perform the integral appearing in (C.10) by taking the gravitational potential equal to its value during matter domination so that it is easily computed

$$\begin{aligned} 2 \int_0^\eta d\eta' \Phi^{(1)}(\eta') \sin[\omega_0(\eta - \eta')] \\ = \frac{18}{10} \frac{\Phi^{(1)}(0)}{\omega_0} (1 - \cos(\omega_0\eta)). \end{aligned} \quad (\text{C.14})$$

Thus, (C.13) gives

$$\Delta_{00}^{(1)} - 4\Psi^{(1)} = \frac{6}{5}\Phi^{(1)}(0) \cos(\omega_0\eta) - \frac{36}{5}\Phi^{(1)}(0). \quad (\text{C.15})$$

The baryon-photon fluid velocity can then be obtained as  $\partial_i v_\gamma^{(1)i} = -3(\Delta_{00}^{(1)} - 4\Psi^{(1)})'/4$  from (C.1). In Fourier space

$$ik_i v_\gamma^{(1)i} = \frac{9}{10}\Phi^{(1)}(0) \sin(\omega_0\eta) \omega_0, \quad (\text{C.16})$$

where, going to Fourier space,  $\partial_i v_\gamma^{(1)i} \rightarrow ik_i v_\gamma^{(1)i}(\mathbf{k})$  and

$$v_\gamma^{(1)i} = -i \frac{k^i}{k} \frac{9}{10}\Phi^{(1)}(0) \sin(\omega_0\eta) c_s, \quad (\text{C.17})$$

since the linear velocity is irrotational.

Notice that under the approximations that  $R = \text{const.}$  and  $\Phi^{(1)} = \Psi^{(1)} = \text{const.}$ , it is easy to find a more accurate solution to (C.6) which better accounts for the presence of the baryons (giving rise to the so called baryon-drag effect, as clearly explained in [73, 74, 95])

$$\frac{\Delta_{00}^{(1)}}{4} = \left[ \frac{\Delta_{00}^{(1)}}{4}(0) + (1+R)\Phi^{(1)} \right] \cos(kr_s) - (1+R)\Phi^{(1)}. \quad (\text{C.18})$$

C.3. *Perturbation Modes with  $k \gg k_{\text{eq}}$ .* This regime corresponds to perturbation modes which enter the Hubble radius when the universe is still radiation dominated at times  $\eta \ll \eta_{\text{eq}}$ . In this case an approximate analytical solution for the evolution of the perturbations can be obtained by considering the gravitational potential for a pure radiation dominated epoch, given by (B.11). For the integral in (C.13), we thus find

$$\int_0^\eta \Phi^{(1)}(\eta') \sin[\omega_0(\eta - \eta')] = -\frac{3}{2\omega_0} \cos(\omega_0\eta), \quad (\text{C.19})$$

where we have kept only the dominant contribution oscillating in time, while neglecting terms which decay in time. The solution (C.13) becomes

$$\Delta_{00}^{(1)} - 4\Psi^{(1)} = 6\Phi^{(1)}(0) \cos(\omega_0\eta), \quad (\text{C.20})$$

and the velocity is given by

$$v_\gamma^{(1)i} = -i \frac{k^i}{k} \frac{9}{2}\Phi^{(1)}(0) \sin(\omega_0\eta) c_s. \quad (\text{C.21})$$

Notice that the solutions (C.20)-(C.21) are actually correct only when radiation dominates. Indeed, between the epoch of equality and recombination, matter starts to dominate. Full account of such a period is given for example, in Section 7.3 of [70], while its consequences for the CMB anisotropy evolution can be found for example, in [100]. We refer to [47], where an alternative way to solve for the acoustic oscillations in this regime is displayed which turns out to be useful for the corresponding computation at second order in Section 7.5.

## Acknowledgments

A. Riotto is on leave of absence from INFN, Sezione di Padova. This research has been partially supported by the ASI contract I/016/07/0 ‘‘COFIS’’, the ASI/INAF Agreement I/072/09/0 for the Planck LFI Activity of Phase E2. A. Riotto acknowledges support by the EU Marie Curie Network UniverseNet (HPRNCT2006035863). The authors would like to thank Michele Liguori for stimulating discussions, and Daisuke Nitta and Eiichiro Komatsu for past and ongoing enjoyable collaborations.

## References

- [1] D. H. Lyth and A. Riotto, ‘‘Particle physics models of inflation and the cosmological density perturbation,’’ *Physics Report*, vol. 314, no. 1-2, pp. 1–146, 1999.
- [2] J. Dunkley, E. Komatsu, M. R. Nolte et al., ‘‘Five-year wilkinson microwave anisotropy probe observations: likelihoods and parameters from the WMAP data,’’ *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 306–329, 2009.
- [3] <http://planck.esa.int/>.
- [4] D. Baumann, M. G. Jackson, P. Adshead et al., ‘‘Probing inflation with CMB polarization,’’ in *CMB Polarization Workshop: Theory and Foregrounds*, vol. 1141 of *AIP Conference Proceedings*, pp. 10–120, 2009.
- [5] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, ‘‘Non-Gaussianity from inflation: theory and observations,’’ *Physics Reports*, vol. 402, no. 3-4, pp. 103–266, 2004.
- [6] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, ‘‘Gauge-invariant second-order perturbations and non-Gaussianity from inflation,’’ *Nuclear Physics B*, vol. 667, no. 1-2, pp. 119–148, 2003.
- [7] J. Maldacena, ‘‘Non-Gaussian features of primordial fluctuations in single field inflationary models,’’ *Journal of High Energy Physics*, vol. 2003, no. 5, article 013, 2003.
- [8] N. Bartolo, S. Matarrese, and A. Riotto, ‘‘Non-Gaussianity from inflation,’’ *Physical Review D*, vol. 65, no. 10, Article ID 103505, 2002.

- [9] F. Bernardeau and J.-P. Uzan, “Non-Gaussianity in multifield inflation,” *Physical Review D*, vol. 66, no. 10, Article ID 103506, 2002.
- [10] D. H. Lyth, C. Ungarelli, and D. Wands, “Primordial density perturbation in the curvaton scenario,” *Physical Review D*, vol. 67, no. 2, Article ID 023503, 2003.
- [11] T. Hamazaki and H. Kodama, “Evolution of cosmological perturbations during reheating,” *Progress of Theoretical Physics*, vol. 96, no. 6, pp. 1123–1145, 1996.
- [12] G. Dvali, A. Gruzinov, and M. Zaldarriaga, “New mechanism for generating density perturbations from inflation,” *Physical Review D*, vol. 69, no. 2, Article ID 023505, 2004.
- [13] L. Kofman, “Probing string theory with modulated cosmological fluctuations,” <http://arxiv.org/abs/astro-ph/0303614>.
- [14] G. Dvali, A. Gruzinov, and M. Zaldarriaga, “Cosmological perturbations from inhomogeneous reheating, freeze-out, and mass domination,” *Physical Review D*, vol. 69, no. 8, Article ID 083505, 2004.
- [15] C. W. Bauer, M. L. Graesser, and M. P. Salem, “Fluctuating annihilation cross sections and the generation of density perturbations,” *Physical Review D*, vol. 72, no. 2, Article ID 023512, 7 pages, 2005.
- [16] D. H. Lyth, “Generating the curvature perturbation at the end of inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 11, article 006, pp. 111–120, 2005.
- [17] M. P. Salem, “Generation of density perturbations at the end of inflation,” *Physical Review D*, vol. 72, no. 12, Article ID 123516, 7 pages, 2005.
- [18] D. H. Lyth and A. Riotto, “Generating the curvature perturbation at the end of inflation in string theory,” *Physical Review Letters*, vol. 97, no. 12, Article ID 121301, 2006.
- [19] M. Bastero-Gil, V. Di Clemente, and S. F. King, “Preheating curvature perturbations with a coupled curvaton,” *Physical Review D*, vol. 70, no. 2, Article ID 023501, 2004.
- [20] E. W. Kolb, A. Riotto, and A. Vallinotto, “Curvature perturbations from broken symmetries,” *Physical Review D*, vol. 71, no. 4, Article ID 043513, 10 pages, 2005.
- [21] C. T. Byrnes and D. Wands, “Scale-invariant perturbations from chaotic inflation,” *Physical Review D*, vol. 73, no. 6, Article ID 063509, 8 pages, 2006.
- [22] T. Matsuda, “Topological curvatons,” *Physical Review D*, vol. 72, no. 12, Article ID 123508, 8 pages, 2005.
- [23] M. Alishahiha, E. Silverstein, and D. Tong, “DBI in the sky: non-Gaussianity from inflation with a speed limit,” *Physical Review D*, vol. 70, no. 12, Article ID 123505, 2004.
- [24] R. Holman and A. J. Tolley, “Enhanced non-Gaussianity from excited initial states,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 5, article 001, 2008.
- [25] D. Babich, P. Creminelli, and M. Zaldarriaga, “The shape of non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, no. 8, article 009, pp. 199–217, 2004.
- [26] J. R. Fergusson and E. P. S. Shellard, “Shape of primordial non-Gaussianity and the CMB bispectrum,” *Physical Review D*, vol. 80, no. 4, Article ID 043510, 2009.
- [27] W. Hu, “Angular trispectrum of the cosmic microwave background,” *Physical Review D*, vol. 64, no. 8, Article ID 083005, 2001.
- [28] T. Okamoto and W. Hu, “Angular trispectra of CMB temperature and polarization,” *Physical Review D*, vol. 66, no. 6, Article ID 063008, 2002.
- [29] G. De Troia, P. A. R. Ade, J. J. Bock et al., “The trispectrum of the cosmic microwave background on subdegree angular scales: an analysis of the BOOMERanG data,” *Monthly Notices of the Royal Astronomical Society*, vol. 343, no. 1, pp. 284–292, 2003.
- [30] J. Lesgourgues, M. Liguori, S. Matarrese, and A. Riotto, “CMB lensing extraction and primordial non-Gaussianity,” *Physical Review D*, vol. 71, no. 10, Article ID 103514, 2005.
- [31] N. Bartolo, S. Matarrese, and A. Riotto, “The full second-order radiation transfer function for large-scale CMB anisotropies,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 010, 2006.
- [32] T. Pyne and S. M. Carroll, “Higher-order gravitational perturbations of the cosmic microwave background,” *Physical Review D*, vol. 53, no. 6, pp. 2920–2929, 1996.
- [33] N. Bartolo, S. Matarrese, and A. Riotto, “Enhancement of non-Gaussianity after inflation,” *Journal of High Energy Physics*, vol. 2004, no. 4, article 006, 2004.
- [34] N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant temperature anisotropies and primordial non-Gaussianity,” *Physical Review Letters*, vol. 93, no. 23, Article ID 231301, 2004.
- [35] K. Tomita, “Relativistic second-order perturbations of nonzero- $\Lambda$  flat cosmological models and CMB anisotropies,” *Physical Review D*, vol. 71, no. 8, Article ID 083504, 11 pages, 2005.
- [36] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity of large-scale cosmic microwave background anisotropies beyond perturbation theory,” *Journal of Cosmology and Astroparticle Physics*, vol. 2005, no. 8, pp. 177–196, 2005.
- [37] L. Boubekour, P. Creminelli, G. D’Amico, J. Norea, and F. Vernizzi, “Sachs-Wolfe at second order: the CMB bispectrum on large angular scales,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 029, 2009.
- [38] G. D’Amico, N. Bartolo, S. Matarrese, and A. Riotto, “CMB temperature anisotropies from third-order gravitational perturbations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 1, article 005, 2008.
- [39] W. Hu and S. Dodelson, “Cosmic microwave background anisotropies,” *Annual Review of Astronomy and Astrophysics*, vol. 40, pp. 171–216, 2002.
- [40] U. Seljak and M. Zaldarriaga, “Direct signature of an evolving gravitational potential from the cosmic microwave background,” *Physical Review D*, vol. 60, no. 4, Article ID 043504, 4 pages, 1999.
- [41] D. M. Goldberg and D. N. Spergel, “Microwave background bispectrum. II. A probe of the low redshift universe,” *Physical Review D*, vol. 59, no. 10, Article ID 103002, 6 pages, 1999.
- [42] D. M. Goldberg and D. N. Spergel, “Microwave background bispectrum. I. Basic formalism,” *Physical Review D*, vol. 59, no. 10, Article ID 103002, 8 pages, 1999.
- [43] E. Komatsu and D. N. Spergel, “Acoustic signatures in the primary microwave background bispectrum,” *Physical Review D*, vol. 63, no. 6, Article ID 063002, 2001.
- [44] K. M. Smith and M. Zaldarriaga, “Algorithms for bispectra: forecasting, optimal analysis, and simulation,” <http://arxiv.org/abs/astro-ph/0612571>.
- [45] P. Creminelli and M. Zaldarriaga, “CMB 3-point functions generated by nonlinearities at recombination,” *Physical Review D*, vol. 70, no. 8, Article ID 083532, 2004.
- [46] N. Bartolo, S. Matarrese, and A. Riotto, “Cosmic microwave background anisotropies at second order: I,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 6, article 024, 2006.

- [47] N. Bartolo, S. Matarrese, and A. Riotto, “CMB anisotropies at second-order II: analytical approach,” *Journal of Cosmology and Astroparticle Physics*, vol. 2007, no. 1, article 019, 2007.
- [48] C. Pitrou, J.-P. Uzan, and F. Bernardeau, “Cosmic microwave background bispectrum on small angular scales,” *Physical Review D*, vol. 78, no. 6, Article ID 063526, 2008.
- [49] C. Pitrou, “The radiative transfer at second order: a full treatment of the Boltzmann equation with polarization,” *Classical and Quantum Gravity*, vol. 26, no. 6, Article ID 065006, 2009.
- [50] C. Pitrou, “The radiative transfer for polarized radiation at second order in cosmological perturbations,” *General Relativity and Gravitation*, vol. 41, no. 11, pp. 2587–2595, 2009.
- [51] S. Matarrese, S. Mollerach, and M. Bruni, “Relativistic second-order perturbations of the Einstein-de Sitter universe,” *Physical Review D*, vol. 58, no. 4, Article ID 043504, 1998.
- [52] D. S. Salopek and J. R. Bond, “Nonlinear evolution of long-wavelength metric fluctuations in inflationary models,” *Physical Review D*, vol. 42, no. 12, pp. 3936–3962, 1990.
- [53] E. W. Kolb, S. Matarrese, A. Notari, and A. Riotto, “Cosmological influence of super-Hubble perturbations,” *Modern Physics Letters A*, vol. 20, no. 35, pp. 2705–2710, 2005.
- [54] K. A. Malik and D. Wands, “Evolution of second-order cosmological perturbations,” *Classical and Quantum Gravity*, vol. 21, no. 11, pp. L65–L71, 2004.
- [55] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity in the curvaton scenario,” *Physical Review D*, vol. 69, no. 4, Article ID 043503, 2004.
- [56] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, “Ghost inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, no. 4, article 001, pp. 1–18, 2004.
- [57] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $f_{NL}^{local}$  from WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article 006, 2009.
- [58] L. Senatore, K. M. Smith, and M. Zaldarriaga, “Non-Gaussianities in single field inflation and their optimal limits from the WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 1, article 028, 2010.
- [59] S. Mollerach and S. Matarrese, “Cosmic microwave background anisotropies from second order gravitational perturbations,” *Physical Review D*, vol. 56, no. 8, pp. 4494–4502, 1997.
- [60] K. N. Ananda, C. Clarkson, and D. Wands, “Cosmological gravitational wave background from primordial density perturbations,” *Physical Review D*, vol. 75, no. 12, Article ID 123518, 2007.
- [61] D. Baumann, P. Steinhardt, K. Takahashi, and K. Ichiki, “Gravitational wave spectrum induced by primordial scalar perturbations,” *Physical Review D*, vol. 76, no. 8, Article ID 084019, 2007.
- [62] A. Mangilli, N. Bartolo, S. Matarrese, and A. Riotto, “Impact of cosmic neutrinos on the gravitational-wave background,” *Physical Review D*, vol. 78, no. 8, Article ID 083517, 2008.
- [63] L. Senatore, S. Tassev, and M. Zaldarriaga, “Cosmological perturbations at second order and recombination perturbed,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 031, 2009.
- [64] S. Dodelson and J. M. Jubas, “Reionization and its imprint on the cosmic microwave background,” *Astrophysical Journal*, vol. 439, no. 2, pp. 503–516, 1995.
- [65] W. Hu, D. Scott, and J. Silk, “Reionization and cosmic microwave background distortions: a complete treatment of second-order Compton scattering,” *Physical Review D*, vol. 49, no. 2, pp. 648–670, 1994.
- [66] C. Pitrou, F. Bernardeau, and J. P. Uzan, “The  $\gamma$ -sky: diffuse spectral distortions of the cosmic microwave background,” <http://arxiv.org/abs/0912.3655>.
- [67] W. Hu, U. Seljak, M. White, and M. Zaldarriaga, “Complete treatment of CMB anisotropies in a FRW universe,” *Physical Review D*, vol. 57, no. 6, pp. 3290–3301, 1998.
- [68] W. Hu and M. White, “CMB anisotropies: total angular momentum method,” *Physical Review D*, vol. 56, no. 2, pp. 596–615, 1997.
- [69] W. Hu, “Reionization revisited: secondary cosmic microwave background anisotropies and polarization,” *Astrophysical Journal*, vol. 529, pp. 12–25, 2000.
- [70] S. Dodelson, *Modern Cosmology*, Academic Press, Amsterdam, The Netherlands, 2003.
- [71] E. Sefusatti, M. Liguori, A. P. S. Yadav, M. G. Jackson, and E. Pajer, “Constraining running non-Gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 12, article 022, 2009.
- [72] E. Komatsu, J. Dunkley, M. R. Nolte et al., “Five-year wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal, Supplement Series*, vol. 180, no. 2, pp. 330–376, 2009.
- [73] W. T. Hu, *Wandering in the background: a CMB explorer*, Ph.D. thesis, University of California at Berkeley, Berkeley, Calif, USA, 1995.
- [74] W. Hu and N. Sugiyama, “Small-scale cosmological perturbations: an analytic approach,” *Astrophysical Journal*, vol. 471, no. 2, pp. 542–570, 1996.
- [75] N. Bartolo, S. Matarrese, and A. Riotto, “Evolution of second-order cosmological perturbations and non-gaussianity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, 2004.
- [76] A. Cooray and W. Hu, “Imprint of reionization on the cosmic microwave background bispectrum,” *Astrophysical Journal*, vol. 534, no. 2, pp. 533–550, 2000.
- [77] S. Donzelli, F. K. Hansen, M. Liguori, and D. Maino, “Impact of the  $1/f$  noise and the asymmetric beam on non-Gaussianity searches with planck,” *Astrophysical Journal*, vol. 706, no. 2, pp. 1226–1240, 2009.
- [78] P. Serra and A. Cooray, “Impact of secondary non-Gaussianities on the search for primordial non-Gaussianity with CMB maps,” *Physical Review D*, vol. 77, no. 10, Article ID 107305, 2008.
- [79] D. Babich and M. Zaldarriaga, “Primordial bispectrum information from CMB polarization,” *Physical Review D*, vol. 70, no. 8, Article ID 083005, 2004.
- [80] D. Hanson, K. M. Smith, A. Challinor, and M. Liguori, “CMB lensing and primordial non-Gaussianity,” *Physical Review D*, vol. 80, no. 8, Article ID 083004, 2009.
- [81] A. Cooray, D. Sarkar, and P. Serra, “Weak lensing of the primary CMB bispectrum,” *Physical Review D*, vol. 77, no. 12, Article ID 123006, 2008.
- [82] A. Mangilli and L. Verde, “Non-Gaussianity and the CMB bispectrum: confusion between primordial and lensing-Rees-Sciama contribution?” *Physical Review D*, vol. 80, no. 12, Article ID 123007, 2009.
- [83] D. Babich and E. Pierpaoli, “Point source contamination in CMB non-Gaussianity analyses,” *Physical Review D*, vol. 77, no. 12, Article ID 123011, 2008.

- [84] A. Cooray, “Large scale pressure fluctuations and the Sunyaev-Zel’dovich effect,” *Physical Review D*, vol. 62, no. 10, Article ID 103506, 15 pages, 2000.
- [85] P. G. Castro, “The bispectrum and the trispectrum of the ostriker and vishniac effect,” *Physical Review D*, vol. 67, Article ID 044039, 2004.
- [86] R. Khatri and B. D. Wandelt, “O-V-S-Z and friends: non-Gaussianity from inhomogeneous reionization,” *Astrophysical Journal*, vol. 711, no. 2, pp. 1310–1315, 2010.
- [87] R. Khatri and B. D. Wandelt, “Crinkles in the last scattering surface: non-Gaussianity from inhomogeneous recombination,” *Physical Review D*, vol. 79, no. 2, Article ID 023501, 2009.
- [88] R. Khatri and B. D. Wandelt, “More on crinkles in the last scattering surface,” *Physical Review D*, vol. 81, no. 10, Article ID 103518, 2010.
- [89] L. Senatore, S. Tassev, and M. Zaldarriaga, “Non-gaussianities from perturbing recombination,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9, article 038, 2009.
- [90] L. Verde and D. N. Spergel, “Dark energy and cosmic microwave background bispectrum,” *Physical Review D*, vol. 65, no. 4, Article ID 043007, 2002.
- [91] N. Aghanim, S. Majumdar, and J. Silk, “Secondary anisotropies of the CMB,” *Reports on Progress in Physics*, vol. 71, no. 6, Article ID 066902, 2008.
- [92] E. Komatsu et al., “Non-Gaussianity as a probe of the physics of the primordial universe and the astrophysics of the low redshift universe,” <http://arxiv.org/abs/0902.4759>.
- [93] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, “Limits on non-Gaussianities from WMAP data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, no. 5, article 004, 2006.
- [94] W. Hu, “Weak lensing of the CMB: a harmonic approach,” *Physical Review D*, vol. 62, no. 4, Article ID 043007, 17 pages, 2000.
- [95] W. Hu and M. White, “The damping tail of cosmic microwave background anisotropies,” *Astrophysical Journal*, vol. 479, no. 2, pp. 568–579, 1997.
- [96] D. Nitta, E. Komatsu, N. Bartolo, S. Matarrese, and A. Riotto, “CMB anisotropies at second order III: bispectrum from products of the first-order perturbations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 014, 2009.
- [97] D. Nitta, N. Bartolo, S. Matarrese, and A. Riotto, to appear.
- [98] M. Liguori, F. K. Hansen, E. Komatsu, S. Matarrese, and A. Riotto, “Testing primordial non-Gaussianity in CMB anisotropies,” *Physical Review D*, vol. 73, no. 4, Article ID 043505, 12 pages, 2006.
- [99] M. Zaldarriaga and D. D. Harari, “Analytic approach to the polarization of the cosmic microwave background in flat and open universes,” *Physical Review D*, vol. 52, no. 6, pp. 3276–3287, 1995.
- [100] V. Mukhanov, ““cMB-Slow” or how to determine cosmological parameters by hand?” *International Journal of Theoretical Physics*, vol. 43, no. 3, pp. 623–668, 2004.

## Review Article

# Non-Gaussianity and Statistical Anisotropy from Vector Field Populated Inflationary Models

**Emanuela Dimastrogiovanni,<sup>1,2</sup> Nicola Bartolo,<sup>1,2</sup> Sabino Matarrese,<sup>1,2</sup> and Antonio Riotto<sup>2,3</sup>**

<sup>1</sup>*Dipartimento di Fisica “G. Galilei”, Università degli Studi di Padova, via Marzolo 8, 35131 Padova, Italy*

<sup>2</sup>*INFN, Sezione di Padova, via Marzolo 8, 35131 Padova, Italy*

<sup>3</sup>*CERN, Theory Division, 1211 Geneva 23, Switzerland*

Correspondence should be addressed to Emanuela Dimastrogiovanni, dimastro@pd.infn.it

Received 22 January 2010; Accepted 12 April 2010

Academic Editor: Eiichiro Komatsu

Copyright © 2010 Emanuela Dimastrogiovanni et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We present a review of vector field models of inflation and, in particular, of the statistical anisotropy and non-Gaussianity predictions of models with  $SU(2)$  vector multiplets. Non-Abelian gauge groups introduce a richer amount of predictions compared to the Abelian ones, mostly because of the presence of vector fields self-interactions. Primordial vector fields can violate isotropy leaving their imprint in the comoving curvature fluctuations  $\zeta$  at late times. We provide the analytic expressions of the correlation functions of  $\zeta$  up to fourth order and an analysis of their amplitudes and shapes. The statistical anisotropy signatures expected in these models are important and, potentially, the anisotropic contributions to the bispectrum and the trispectrum can overcome the isotropic parts.

## 1. Introduction

In the standard cosmological model, at very early times the Universe undergoes a quasi de Sitter exponential expansion driven by a scalar field, the inflaton, with an almost flat potential. The quantum fluctuations of this field are thought to be at the origin of both the Large Scale Structures and the Cosmic Microwave Background (CMB) fluctuations that we are able to observe at the present epoch [1]. CMB measurements indicate that the primordial density fluctuations are of order  $10^{-5}$  have an almost scale-invariant power spectrum and are fairly consistent with Gaussianity and statistical isotropy [2–6]. All of these features find a convincing explanation within the inflationary paradigm. Nevertheless, deviations from the basic single-(scalar)field slow-roll model of inflation are allowed by the experimental data. On one hand, it is then important to search for observational signatures that can help discriminate among all the possible scenarios; on the other hand, it is important to understand what the theoretical predictions are in this respect for the different models.

Non-Gaussianity and statistical anisotropy are two powerful signatures. A random field is defined “Gaussian” if it is entirely described by its two-point function, higher order connected correlators being equal to zero. Primordial non-Gaussianity [7, 8] is theoretically predicted by inflation: it arises from the interactions of the inflaton with gravity and from self-interactions. However, it is observably too small in the single-field slow-roll scenario [9–11]. Alternatives to the latter have been proposed that predict higher levels of non-Gaussianity such as multifield scenarios [12–18], curvaton models [19–25], and models with noncanonical Lagrangians [26–30]. Many efforts have been directed to the study of higher order (three and four-point) cosmological correlators in these models [11, 22, 29, 31–47] and towards improving the prediction for the two-point function, through quantum loop calculations [10, 48–55]. From WMAP, the bounds on the bispectrum amplitude are given by  $-4 < f_{NL}^{loc} < 80$  [56] and by  $-125 < f_{NL}^{equil} < 435$  [57] at 95% CL, respectively in the local and in the equilateral configurations. For the trispectrum, WMAP provides  $-5.6 \times 10^5 < g_{NL} < 6.4 \times 10^5$  [58],  $g_{NL}$  being the “local” trispectrum amplitude from cubic

contributions (see also [59]), whereas from Large-Scale-Structures data  $-3.5 \times 10^5 < g_{NL} < 8.2 \times 10^5$  [60], at 95% CL. Planck [61] and future experiments are expected to set further bounds on primordial non-Gaussianity.

Statistical isotropy has always been considered one of the key features of the CMB fluctuations. The appearance of some “anomalies” [62–65] in the observations though, after numerous and careful data analysis, suggests a possible breaking of this symmetry that might have occurred at some point of the Universe history, possibly at very early times. This encouraged a series of attempts to model this event, preferably by incorporating it in theories of inflation. Let us shortly describe the above mentioned “anomalies”. First of all, the large-scale CMB quadrupole appears to be “too low” and the octupole “too planar”; in addition to that, there seems to exist a preferred direction along which quadrupole and octupole are aligned [3, 62, 66–68]. Also, a “cold spot”, that is, a region of suppressed power, has been observed in the southern Galactic sky [63, 69]. Finally, an indication of asymmetry in the large-scale power spectrum and in higher-order correlation functions between the northern and the southern ecliptic hemispheres was found [64, 65, 70–72]. Possible explanations for these anomalies have been suggested such as improper foreground subtraction, WMAP systematics, statistical flukes; the possibilities of topological or cosmological origins for them have been proposed as well. Moreover, considering a power spectrum anisotropy due to the existence of a preferred spatial direction  $\hat{n}$  and parametrized by a function  $g(k)$  as

$$P(\vec{k}) = P(k) \left( 1 + g(k) (\hat{k} \cdot \hat{n})^2 \right), \quad (1)$$

the five-year WMAP temperature data have been analyzed in order to find out what the magnitude and orientation of such an anisotropy could be. The magnitude has been found to be  $g = 0.29 \pm 0.031$  and the orientation aligned nearly along the ecliptic poles [73]. Similar results have been found in [74], where it is pointed out that the origin of such a signal is compatible with beam asymmetries (uncorrected in the maps) which should therefore be investigated before we can find out what the actual limits on the primordial  $g$  are.

Several fairly recent works have taken the direction of analysing the consequences, in terms of dynamics of the Universe and of cosmological fluctuations, of an anisotropic preinflationary or inflationary era. A cosmic no-hair conjecture exists according to which the presence of a cosmological constant at early times is expected to dilute any form of initial anisotropy [75]. This conjecture has been proven to be true for many (all Bianchi type cosmologies except for the Bianchi type-IX, for which some restrictions are needed to ensure the applicability of the theorem), but not all kinds of metrics and counterexamples exist in the literature [76–78]. Moreover, even in the event isotropization should occur, there is a chance that signatures from anisotropic inflation or from an anisotropic preinflationary era might still be visible today [79–82]. In the same context of searching for models of the early Universe that might produce some anisotropy signatures at late time, new theories have been proposed such as spinor models [83–86], higher p-forms [87–92] and

primordial vector field models (see Section 2 for a quick review).

We are going to focus on statistical anisotropy and non-Gaussianity predictions of primordial vector field models. As mentioned above, there are great expectations that Planck and new experiments will, among other things, shed more light on the level of non-Gaussianity of the CMB fluctuations and on the nature of the unexpected anisotropy features we mentioned (see, e.g., [93]). Models that combine both types of predictions could be more easily testable and, from non-Gaussianity measurement, more stringent statistical anisotropy predictions could be produced or viceversa.

Within vector field models, higher order correlators had been computed in [94–98] and, more recently, in [99, 100] for  $U(1)$  vector fields. We considered  $SU(2)$  vector field models in [101, 102]. Non-Abelian theories offer a richer amount of predictions compared to the Abelian case. Indeed, self-interactions provide extra contributions to the bispectrum and trispectrum of curvature fluctuations that are naturally absent in the Abelian case. We verified that these extra contributions can be equally important in a large subset of the parameter space of the theory and, in some cases, can even become the dominant ones.

This paper is structured as follows: in Section 2 we review some vector field models of inflation; in Section 3 we present the  $SU(2)$  model; in Section 4 we provide the results for the two, three and, four-point functions of the curvature fluctuations; in Section 5 we present the non-Gaussianity amplitudes for the bispectrum and for the trispectrum; in Section 6 we show and discuss their shapes; finally in Section 7 we draw our conclusions.

## 2. Inflation and Primordial Vector Fields

The attempt to explain some of the CMB “anomalous” features as the indication of a break of statistical isotropy is the main reason behind ours and many of the existing inflationary models populated by vector fields, but not the only one. The first one of these models [103] was formulated with the goal of producing inflation by the action of vector fields, without having to invoke the existence of a scalar field. The same motivations inspired the works that followed [104–106]. Lately, models where primordial vector fields can leave an imprint on the CMB have been formulated as an alternative to the basic inflationary scenario, in the search for interesting non-Gaussianity predictions [94–102]. Finally, vector fields models of dark energy have been proposed [107–111]. All this appears to us as a rich bag of motivations for investigating these scenarios.

Before we quickly sketch some of them and list the results so far achieved in this direction, it is important to briefly indicate and explain the main issues and difficulties that these models have been facing. We will also shortly discuss the mechanisms of production of the curvature fluctuations in these models.

Building a model where primordial vector fields can drive inflation and/or produce the observed spectrum of large scale fluctuations requires a more complex Lagrangian than the basic gauge invariant  $\mathcal{L}_{\text{vector}} = -(\sqrt{-g}/4)F^{\mu\nu}F_{\mu\nu}$ . In fact, for a

conformally invariant theory as the one described by  $\mathcal{L}_{\text{vector}}$ , vector fields fluctuations are not excited on superhorizon scales. It is then necessary to modify the Lagrangian. For some of the existing models, these modifications have been done to the expense of destabilizing the theory, by “switching on” unphysical degrees of freedom. This was pointed out in [112–114], where a large variety of vector field models was analyzed in which longitudinal polarization modes exist that are endowed with negative squared masses (the “wrong” signs of the masses are imposed for the theory to satisfy the constraints that allow a suitable background evolution). It turns out that, in a range of interest of the theory, these fields acquire negative total energy, that is, behave like “ghosts”, the presence of which is known to be responsible for an unstable vacuum. A related problem for some of these theories is represented by the existence of instabilities affecting the equations of motion of the ghost fields [112–114].

In the remaining part of this section, we are going to present some of these models together with some recent attempts to overcome their limits.

In all of the models we will consider, primordial vector fields fluctuations end up either being entirely responsible for or only partially contributing to the curvature fluctuations at late times. This can happen through different mechanisms. For example, if the vector fields affect the universe expansion during inflation, its contribution  $\zeta_A$  to the total  $\zeta$  can be derived from combining the definition of the number of e-foldings ( $N = \int H dt$ ) with the Einstein equation ( $H^2 = (8\pi G/3)(\rho_\phi + \rho_A)$ ,  $\rho_A$  being the energy density of the vector field and  $\rho_\phi$  the inflaton energy density) and using the  $\delta N$  expansion of the curvature fluctuation in terms of both the inflaton and the vector fields fluctuations (see Section 4). To lowest order we have [96]

$$\zeta_A = \frac{A_i}{2m_p^2} \delta A_i, \quad (2)$$

where a single vector field has been taken into account for simplicity ( $m_p$  is the reduced Planck mass,  $A$  is the background value of the field and  $\delta A$  its perturbation). When calculating the amplitude of non-Gaussianity in Section 5, we will refer to this case as “vector inflation” for simplicity.

A different fluctuation production process is the curvaton mechanism which was initially formulated for scalar theories but it is also applicable to vectors [96, 115, 116]. Specifically, inflation is driven by a scalar field, whereas the curvaton field(s) (now played by the vectors), has a very small (compared to the Hubble rate) mass during inflation. Towards the end of the inflationary epoch, the Hubble rate value starts decreasing until it equates the vector mass; when this eventually happens, the curvaton begins to oscillate and it will then dissipate its energy into radiation. The curvaton becomes responsible for a fraction of the total curvature fluctuation that is proportional to a parameter ( $r$ ) related to the ratio between the curvaton energy density and the total energy density of the universe at the epoch of the curvaton decay [96]

$$\zeta_A = \frac{r}{3} \frac{\delta \rho_A}{\rho_A}, \quad (3)$$

where  $r \equiv 3\rho_A/(3\rho_A + 4\rho_\phi)$ . Anisotropy bounds on the power spectrum favour small values of  $r$ .

From (2) and (3) we can see that, depending on which one of these two mechanisms of production of the curvature fluctuations is considered, different coefficients will result in the  $\delta N$  expansion (see (2)).

In this section, we will describe both models where inflation is intended to be vector field driven and those models in which, instead, the role of the inflaton is played by a scalar field, whereas the energy of the vector gives a subdominant contribution to the total energy density of the universe during the entire inflationary phase.

**2.1. Self-Coupled Vector Field Models.** A pioneer work on vector field driven inflation was formulated by Ford [103], who considered a single self-coupled field  $A_\mu$  with a Lagrangian

$$L_{\text{vector}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + V(\psi), \quad (4)$$

where  $F_{\mu\nu} \equiv \partial_\mu B_\nu - \partial_\nu B_\mu$  and the potential  $V$  is a function of  $\psi \equiv B_\alpha B^\alpha$ . Different scenarios of expansion are analyzed by the author for different functions  $V$ . The universe expands anisotropically at the end of the inflationary era and this anisotropy either survives until late times or is damped out depending on the shape and the location of the minima of the potential.

The study of perturbations in a similar model was proposed by Dimopoulos in [115] where he showed that for a Lagrangian

$$L_{\text{vector}} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} m^2 B_\mu B^\mu, \quad (5)$$

and for  $m^2 \simeq -2H^2$ , the transverse mode of the vector field is governed by the same equation of motion as a light scalar field in a de Sitter stage. A suitable superhorizon power spectrum of fluctuations could therefore arise. In order to prevent production of large scale anisotropy, in this model the vector field plays the role of the curvaton while inflation is driven by a scalar field.

**2.2. Vector-Field Coupled to Gravity.** The Lagrangian in (5) may be also intended, at least during inflation, as including a nonminimal coupling of the vector field to gravity; indeed the mass term can be rewritten as

$$L_{\text{vector}} \supset \frac{1}{2} (m_0^2 + \xi R) B_\mu B^\mu, \quad (6)$$

where, for the whole duration of the inflationary era, the bare mass  $m_0$  is assumed to be much smaller than the Hubble rate and the Ricci scalar  $R = -6[\ddot{a}/a + (\dot{a}/a)^2]$  can be approximated as  $R \simeq -12H^2$ . For the specific value  $\xi = 1/6$ , (5) is retrieved.

For the Lagrangian just presented, Golovnev et al. [104] proved that the problem of excessive anisotropy production in the case where inflation is driven by vector fields can be avoided if either a triplet of mutually orthogonal or a large number  $N$  of randomly oriented vector fields is considered.

The Lagrangian (6) with  $\xi = 1/6$  was also employed in [116], where inflation is scalar field-driven and a primordial vector field affects large-scale curvature fluctuations and, similarly, in [117], which includes a study of the backreaction of the vector field on the dynamics of expansion, by introducing a Bianchi type-I metric.

**2.3. Ackerman-Carroll-Wise (ACW) Model.** A model was proposed in [118, 119] where Lagrange multipliers ( $\lambda$ ) are employed to determine a fixed norm primordial vector field  $B_\mu B^\mu = m^2$

$$L_{\text{vector}} \supset \lambda (B^\mu B_\mu - m^2) - \rho_\Lambda, \quad (7)$$

where  $\rho_\Lambda$  is a vacuum energy. The expansion rate in this scenario is anisotropic: if we orient the  $x$ -axis of the spatial frame along the direction determined by the vector field, we find two different Hubble rates: along the  $x$ -direction it is equal to

$$H_b^2 = \frac{\rho_\Lambda}{m_p^2} \frac{1}{P(\mu)}, \quad (8)$$

and it is given by  $H_a = (1 + c\mu^2)H_b$  along the orthogonal directions;  $\mu \equiv m/m_p$ ,  $P$  is a polynomial function of  $\mu$  and  $c$  is a parameter appearing in the kinetic part of the Lagrangian that we omitted in (7) (see [118, 119] for its complete expression). As expected, an isotropic expansion is recovered if the vev of the vector field is set to zero.

**2.4. Models with Varying Gauge Coupling.** Most of the models mentioned so far successfully solve the problem of attaining a slow-roll regime for the vector-fields without imposing too many restrictions on the parameters of the theory and of avoiding excessive production of anisotropy at late times. None of them though escapes those instabilities related to the negative energy of the longitudinal modes (although a study of the instabilities for fixed-norm field models was done in [120] where some stable cases with non-canonical kinetic terms were found). As discussed in [112–114], in the self-coupled model a ghost appears at small (compared to the horizon) wavelengths; in the non-minimally coupled and in the fixed-norm cases instead the instability concerns the region around horizon crossing.

Models with varying gauge coupling can overcome the problem of instabilities and have recently attracted quite some attention. In [94], the authors consider a model of hybrid inflation [121–124] with the introduction of a massless vector field

$$L \supset \frac{1}{2} (\partial_\mu \phi \partial^\mu \phi + \partial_\mu \chi \partial^\mu \chi) - \frac{1}{4} f^2(\phi) F_{\mu\nu} F^{\mu\nu} + V(\phi, \chi, B_\mu), \quad (9)$$

where  $\phi$  is the inflaton and  $\chi$  is the so-called “waterfall” field. The potential  $V$  is chosen in such a way as to preserve gauge invariance; this way the longitudinal mode disappears and instabilities are avoided.

Similarly, Kanno et al. in [125] (see also [96, 126]) consider a vector field Lagrangian of the type

$$L_{\text{vector}} = -\frac{1}{4} f^2(\phi) F^{\mu\nu} F_{\mu\nu}, \quad (10)$$

but in a basic scalar field-driven inflation model. Very recently, in [127] the linear perturbations in these kind of models have been investigated.

Finally, in [96, 100] varying mass vector field models have been introduced

$$L_{\text{vector}} = -\frac{1}{4} f^2(\phi) F^{\mu\nu} F_{\mu\nu} + \frac{1}{2} m^2 B_\mu B^\mu, \quad (11)$$

where  $f \simeq a^\alpha$  and  $m \simeq a$  ( $a$  is the scale factor and  $\alpha$  is a numerical coefficient). The special cases  $\alpha = 1$  and  $\alpha = -2$  are of special interest. In fact, introducing the fields  $\tilde{A}_\mu$  and  $A_\mu$ , related to one another by  $\tilde{A}_\mu \equiv f B_\mu = a A_\mu$  ( $\tilde{A}_\mu$  and  $A_\mu$  are, resp., the comoving and the physical vectors), it is possible to verify that the physical gauge fields are governed by the same equations of motion as a light scalar field in a de Sitter background. Vector fields in this theory can then generate the observed (almost) scale-invariant primordial power spectrum.

### 3. $SU(2)$ Vector Model: Equations of Motion for the Background and for Linear Perturbations

Let us consider some models where inflation is driven by a scalar field in the presence of an  $SU(2)$  vector multiplet [101, 102]. A fairly general Lagrangian can be the following:

$$S = \int d^4x \sqrt{-g} \left[ \frac{m_p^2 R}{2} - \frac{f^2(\phi)}{4} g^{\mu\alpha} g^{\nu\beta} \sum_{a=1,2,3} F_{\mu\nu}^a F_{\alpha\beta}^a - \frac{M^2}{2} g^{\mu\nu} \sum_{a=1,2,3} B_\mu^a B_\nu^a + L_\phi \right], \quad (12)$$

where  $L_\phi$  is the Lagrangian of the scalar field and  $F_{\mu\nu}^a \equiv \partial_\mu B_\nu^a - \partial_\nu B_\mu^a + g_c \epsilon^{abc} B_\mu^b B_\nu^c$  ( $g_c$  is the  $SU(2)$  gauge coupling). Both  $f$  and the effective mass  $M$  can be viewed as generic functions of time. The fields  $B_\mu^a$  are comoving and related to the physical fields by  $A_\mu^a = (B_0^a, B_i^a/a)$ . The free field operators can be Fourier expanded in their creation and annihilation operators

$$\begin{aligned} & \delta A_i^a(\vec{x}, \eta) \\ &= \int \frac{d^3q}{(2\pi)^3} e^{i\vec{q}\cdot\vec{x}} \\ & \times \sum_{\lambda=L,R,\text{long}} \left[ e_i^\lambda(\hat{q}) a_{\vec{q}}^{a,\lambda} \delta A_\lambda^a(q, \eta) \right. \\ & \quad \left. + e_i^{*\lambda}(-\hat{q}) (a_{-\vec{q}}^{a,\lambda})^\dagger \delta A_\lambda^{*a}(q, \eta) \right], \end{aligned} \quad (13)$$

where the polarization index  $\lambda$  runs over left ( $L$ ), right ( $R$ ) and longitudinal (long) modes and

$$\left[ a_{\vec{k}}^{a,\lambda}, (a_{\vec{k}'}^{a',\lambda'})^\dagger \right] = (2\pi)^3 \delta_{a,a'} \delta_{\lambda,\lambda'} \delta^{(3)}(\vec{k} - \vec{k}'). \quad (14)$$

Here,  $\eta$  the conformal time ( $d\eta = dt/a(t)$ ). Once the functional forms of  $f$  and  $M$  have been specified, the equations of motion for the vector bosons can be written. For the most part, the calculations are quite general in this respect. In fact, the expression of all correlation functions, prior to explicitating the wavefunction for the gauge bosons, apply to any  $SU(2)$  theory with an action as in (12), both for what we will call the ‘‘Abelian’’ and for the ‘‘non-Abelian’’ contributions. In particular, the structure of the interaction Hamiltonian is independent of the functional dependence of  $f$  and  $M$  and determines the general form of and the anisotropy coefficients appearing in the final ‘‘non-Abelian’’ expressions (see Section 4). When it comes to explicitate the wavefunctions, a choice that can help keeping the result as easy to generalize as possible is the following:

$$\delta B^T = -\frac{\sqrt{\pi x}}{2\sqrt{k}} [J_{3/2}(x) + iJ_{-3/2}(x)], \quad (15)$$

for the transverse mode and

$$\delta B^{\parallel} = n(x)\delta B^T, \quad (16)$$

for the longitudinal mode ( $n$  is a unknown function of  $x \equiv -k\eta$ ) [101, 102]. Let us see why. As previously stated, for  $f \simeq a^\alpha$  and with  $\alpha = 0, 1, -2$ , it is possible to verify that the (physical) transverse mode behaves exactly like a light scalar field in a de Sitter background. Considering the solution (15) then takes into account at least these special cases. As to the longitudinal mode, a parametrization was adopted as in (16) in order to keep the analysis more general and given that, because of the instability issues, introducing this degree of freedom into the theory requires special attention. We are going to keep the longitudinal mode ‘‘alive’’ in the calculations we present, by considering a nonzero function  $n(x)$ , and focus on the simplest case of  $f = 1$ . This case is known to be affected by quantum instabilities in the longitudinal mode, anyway we choose  $f = 1$  for the sake of simplicity in our presentation. The results can be easily generalized to gauge invariant models (please refer to [102] for a sample generalization of the calculations to massless  $f \simeq a^{(1,-2)}$  models).

#### 4. Correlation Functions of Curvature Fluctuations: Analytic Expressions

We are now ready to review the computation of the power spectrum, bispectrum, and trispectrum for the curvature fluctuations  $\zeta$  generated during inflation

$$\langle \zeta_{\vec{k}_1} \zeta_{\vec{k}_2} \rangle = (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2) P_\zeta(\vec{k}), \quad (17)$$

$$\begin{aligned} \langle \zeta_{\vec{k}_1} \zeta_{\vec{k}_2} \zeta_{\vec{k}_3} \rangle &= (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3), \\ &\times B_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3), \end{aligned} \quad (18)$$

$$\begin{aligned} \langle \zeta_{\vec{k}_1} \zeta_{\vec{k}_2} \zeta_{\vec{k}_3} \zeta_{\vec{k}_4} \rangle &= (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4), \\ &\times T_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4). \end{aligned} \quad (19)$$

Notice that, on the right-hand side of (17) through (19), we indicated a dependence from the direction of the wavevectors; in models of inflation where isotropy is preserved, the power spectrum and the bispectrum only depend on the moduli of the wave vectors. This will not be the case for the  $SU(2)$  model.

The  $\delta N$  formula [31, 128–130] will be employed

$$\zeta(\vec{x}, t) = N(\vec{x}, t^*, t) - N(t^*, t) \equiv \delta N(\vec{x}, t), \quad (20)$$

which holds if times  $t^*$  and  $t$  are chosen, respectively, on a flat and on a uniform density temporal slices ( $N$  is the number of e-foldings of inflation occurring between these two times). We employ a spatial metric  $g_{ij} = a^2(t)e^{-2\Psi}(e^{\gamma})_{ij}$ , and at linear level the curvature perturbation corresponds to  $\zeta \equiv -\Psi + H\delta u$ , where  $\delta u$  is the fluctuation in the total energy density and  $\Psi$  is a scalar metric fluctuation. In the presence of a single scalar field  $\phi$ , (20) is further expandable as

$$\zeta(\vec{x}, t) = \sum_n \frac{N^{(n)}(t^*, t)}{n!} (\delta\phi(\vec{x}, t^*))^n, \quad (21)$$

where  $N^{(n)}$  is the partial derivative of the e-folding number w.r.t.  $\phi$  on the initial hypersurface  $t^*$ .

If we apply (21) to the inflaton +  $SU(2)$  vector model, we have

$$\begin{aligned} \zeta(\vec{x}, t) &= N_\phi \delta\phi + N_a^\mu \delta A_\mu^a + \frac{1}{2} N_{\phi\phi} (\delta\phi)^2 + \frac{1}{2} N_{ab}^{\mu\nu} \delta A_\mu^a \delta A_\nu^b \\ &+ N_{\phi a}^\mu \delta\phi \delta A_\mu^a + \frac{1}{3!} N_{\phi\phi\phi} (\delta\phi)^3 + \frac{1}{3!} N_{abc}^{\mu\nu\lambda} \delta A_\mu^a \delta A_\nu^b \delta A_\lambda^c \\ &+ \frac{1}{2} N_{\phi\phi a}^\mu (\delta\phi)^2 \delta A_\mu^a + \frac{1}{2} N_{\phi ab}^{\mu\nu} \delta\phi \delta A_\mu^a \delta A_\nu^b \\ &+ \frac{1}{3!} N_{\phi\phi\phi\phi} (\delta\phi)^4 + \frac{1}{3!} N_{abcd}^{\mu\nu\lambda\eta} \delta A_\mu^a \delta A_\nu^b \delta A_\lambda^c \delta A_\eta^d + \dots, \end{aligned} \quad (22)$$

where now

$$N_\phi \equiv \left( \frac{\partial N}{\partial \phi} \right)_{t^*}, \quad N_a^\mu \equiv \left( \frac{\partial N}{\partial A_\mu^a} \right)_{t^*}, \quad N_{\phi a}^\mu \equiv \left( \frac{\partial^2 N}{\partial \phi \partial A_\mu^a} \right)_{t^*} \quad (23)$$

and so on for higher order derivatives.

On a general ground, given the fact that the temporal modes can always be expressed in terms of the longitudinal modes, (22) can be rewritten by retaining only the spatial part for the vector field fluctuations (see also [96]).

Our plan is to show the derivation the correlation functions of  $\zeta$  from the ones of  $\delta\phi$  and  $\delta A_\mu^a$ , after a replacement of the  $\delta N$  expansion (22) in (17) through (19).

The correlation functions can be evaluated using the Schwinger-Keldysh formula [48–52]

$$\begin{aligned} &\langle \Omega | \Theta(t) | \Omega \rangle \\ &= \langle 0 | \left[ \bar{T} \left( e^{i \int_0^t H_I(t') dt'} \right) \right] \Theta_I(t) \left[ T \left( e^{-i \int_0^t H_I(t') dt'} \right) \right] | 0 \rangle, \end{aligned} \quad (24)$$

where, on the left-hand side, the operator  $\Theta$  and the vacuum  $\Omega$  are in the interacting theory whereas, on the right-hand side, all operators are in the so-called “interaction picture”, that is, they can be treated as free fields (the Fourier expansion in (13) thus apply), and  $|0\rangle$  is the free theory vacuum.

When calculating the spectra of  $\zeta$ , the perturbative expansions in (22) and (24) will be carried out to only include tree-level contributions, neglecting higher order “loop” terms, either classical, that is, from the  $\delta N$  series, or of quantum origin, that is, from the Schwinger-Keldysh series. Assuming that the  $SU(2)$  coupling  $g_c$  is “small” and that we are dealing with “small” fluctuations in the fields and given the fact that a slow-roll regime is being assumed, it turns out that it is indeed safe for the two expansions to be truncated at tree-level.

The correlation functions of  $\zeta$  will then result as the sum of scalar, vector and (scalar and vector) mixed contributions. As to the vector part, this will be made up of terms that are merely generated by the  $\delta N$  expansion, that is, they only include the zeroth order of the in-in formula (we call these terms “Abelian”, being them retrievable in the  $U(1)$  case), and by (“non-Abelian”) terms arising from the Schwinger-Keldysh operator expansion beyond zeroth order, that is, from the gauge fields self-interactions.

Let us now discuss the level of generality of the results we will present in the next sections, w.r.t. the choice of a specific Lagrangian.

The expression for the Abelian contributions provided in Sections 4.1 and 4.2.1 apply to any  $SU(2)$  model of gauge interactions with no direct coupling between scalar and vector fields (extra terms would be otherwise needed in (33) and (34)). The next stage in the Abelian contributions computation would be to explicate the derivatives of the e-foldings number and the wavefunctions of the fields: they both depend on the equations of motion of the system, therefore the fixing of a specific model is required at this point. As to the non-Abelian contributions, the results in (42) and (43) are completely general except for assuming, again, that no direct vector-scalar field coupling exists. The structure of (52) and (55) is instead due to the choice of a non-Abelian gauge group. The expressions of the anisotropy coefficients  $I_n$  and  $L_n$  in (52) and (55) depend on the specific non-Abelian gauge group (for  $SU(2)$  one of the  $I_n$  is given in (54)). Finally, the specific expressions of the isotropic functions  $F_n$  (a sample of which is shown in (53)) and  $G_n$  were derived considering the Lagrangian (12) with  $f = 1$  and the eigenfunctions for the vector bosons provided in (15) and (16).

**4.1. The Power Spectrum.** The power spectrum of  $\zeta$  can be straightforwardly derived at tree-level, using the  $\delta N$  expansion (22), from the inflaton and the vector fields power spectra

$$\begin{aligned} P_\zeta(\vec{k}) &= P^{\text{iso}}(k) \left[ 1 + g^{ab} (\hat{k} \cdot \hat{N}_a) (\hat{k} \cdot \hat{N}_b) + i s^{ab} \hat{k} \cdot (\hat{N}_a \times \hat{N}_b) \right], \\ &\quad (25) \end{aligned}$$

which generalizes the results of [96] for the  $U(1)$  case.

The isotropic part of the previous expression has been factorized in

$$P^{\text{iso}}(k) \equiv N_\phi^2 P_\phi(k) + (\vec{N}_c \cdot \vec{N}_d) P_+^{cd}, \quad (26)$$

where we have defined the following combinations

$$P_\pm^{ab} \equiv \left( \frac{1}{2} \right) (P_R^{ab} \pm P_L^{ab}), \quad (27)$$

from the power spectra for the right, left and longitudinal polarization modes

$$\begin{aligned} P_R^{ab} &\equiv \delta_{ab} \delta A_R^a(k, t^*) \delta A_R^{b*}(k, t^*), \\ P_L^{ab} &\equiv \delta_{ab} \delta A_L^a(k, t^*) \delta A_L^{b*}(k, t^*), \\ P_{\text{long}}^{ab} &\equiv \delta_{ab} \delta A_{\text{long}}^a(k, t^*) \delta A_{\text{long}}^{b*}(k, t^*). \end{aligned} \quad (28)$$

The anisotropic parts are weighted by the coefficients

$$\begin{aligned} g^{ab} &\equiv \frac{N^a N^b (P_{\text{long}}^{ab} - P_+^{ab})}{N_\phi^2 P_\phi + (\vec{N}_c \cdot \vec{N}_d) P_+^{cd}}, \\ s^{ab} &\equiv \frac{N^a N^b P_-^{ab}}{N_\phi^2 P_\phi + (\vec{N}_c \cdot \vec{N}_d) P_+^{cd}}, \end{aligned} \quad (29)$$

(where a sum is intended over indices  $c$  and  $d$  but not over  $a$  and  $b$ ). Equation (26) can also be written as

$$P^{\text{iso}}(k) = N_\phi^2 P_\phi \left[ 1 + \beta_{cd} \frac{P_+^{cd}}{P_\phi} \right], \quad (30)$$

after introducing the parameter

$$\beta_{cd} \equiv \frac{\vec{N}_c \cdot \vec{N}_d}{N_\phi^2}. \quad (31)$$

Notice that what when we say “isotropic”, as far as the expression for the power spectrum is concerned, we simply mean “independent” of the direction of the wave vector. In this case instead, the vector bosons introduce three preferred spatial directions: the r.h.s. of (25) depends on their orientation w.r.t. the wave vector.

As expected, the coefficients  $g^{ab}$  and  $s^{ab}$  that weight the anisotropic part of the power spectrum are related to  $\beta_{cd}$ , that is, to the parameters that quantify how much the expansion of the universe is affected by the vector bosons compared to the scalar field.

Assuming no parity violation in the model, we have  $s^{ab} = 0$ ; the parameters  $g^{ab}$  and  $\beta_{ab}$  are instead unconstrained. In the  $U(1)$  case and for parity conserving theories, (25) reduces to [96]

$$P_\zeta(\vec{k}) = P_\zeta^{\text{iso}}(k) \left[ 1 + g(\hat{k} \cdot \hat{n}) \right], \quad (32)$$

where  $\hat{n}$  indicates the preferred spatial direction; also one can check that in this simple case, if  $P_+ \simeq P_\phi$  and  $P_{\text{long}} = k P_+$  ( $k \neq 1$ ), the relation  $g = (k - 1)\beta/(1 + \beta)$  holds, where

$\beta \equiv (N_A/N_\phi)^2$  (the anisotropy coefficient  $g$  is not to be confused with the  $SU(2)$  coupling constant  $g_c$ ). If it is safe to assume  $|g| \ll 1$  (see discussion following (1) and references [73, 74]), a similar upper bound can also be placed on  $\beta$ .

In the case where more than one special directions exists, as in the  $SU(2)$  model, no such analysis on the anisotropy data has been so far carried out, the  $g^a$  parameters cannot then be constrained, unless assuming that the three directions converge into a single one; in that case a constraint could be placed on the sum  $|g| \equiv |\sum_a g^a|$ , where  $a = 1, 2, 3$  and  $P_\zeta(\vec{k}) = P_\zeta^{\text{iso}}(k)[1 + g^a(\hat{k} \cdot \hat{n}_a)]$ .

**4.2. Higher-Order Correlators.** We will present the results for the tree-level contributions to the bispectrum and to the trispectrum of  $\zeta$ .

These can be classified in two categories, that we indicate as “Abelian” and “non-Abelian”. The former are intended as terms that merely arise from the  $\delta N$  expansion and are thus retrievable in the Abelian case; the latter are derived from the linear and quadratic expansions (in terms of the gauge bosons interaction Hamiltonian) of the Schwinger-Keldysh formula and are therefore peculiar to the non-Abelian case.

We are going to provide both types of contributions, in preparation for discussing and comparing their magnitudes later on in Section 5.

**4.2.1. Abelian Contributions.** By plugging the  $\delta N$  expansion (22) in (18) and (19), we have

$$\begin{aligned} B_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3) & \supset \frac{1}{2} N_\phi^2 N_{\phi\phi} [P_\phi(k_1)P_\phi(k_2) + \text{perms.}] \\ & + \frac{1}{2} N_a^i N_b^j N_{cd}^{kl} [\Pi_{ik}^{ac}(\vec{k}_1) \Pi_{jl}^{bd}(\vec{k}_2) + \text{perms.}] \\ & + \frac{1}{2} N_\phi N_a^i N_b^j [P_\phi(k_1) \Pi_{ij}^{ab}(\vec{k}_2) + \text{perms.}] \\ & + N_\phi N_\phi^2 B_\phi(k_1, k_2, k_3), \end{aligned} \quad (33)$$

for the bispectrum (see also [96]) and

$$\begin{aligned} T_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4) & \supset N_\phi^4 T_\phi(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4) \\ & + N_\phi^3 N_{\phi\phi} [P_\phi(k_1) B_\phi(|\vec{k}_1 + \vec{k}_2|, k_3, k_4) + \text{perms.}] \\ & + N_\phi^2 N_a^i N_b^j [P_{ij}^{ab}(\vec{k}_3) B_\phi(k_1, k_2, |\vec{k}_3 + \vec{k}_4|) + \text{perms.}] \\ & + N_\phi^2 N_{\phi\phi}^2 [P_\phi(k_1) P_\phi(k_2) P_\phi(|\vec{k}_1 + \vec{k}_3|) + \text{perms.}] \\ & + N_\phi^3 N_{\phi\phi\phi} [P_\phi(k_1) P_\phi(k_2) P_\phi(k_3) + \text{perms.}] \\ & + N_\phi^2 N_{\phi a}^i N_{\phi b}^j [P_{ij}^{ab}(\vec{k}_1 + \vec{k}_3) P_\phi(k_1) P_\phi(k_2) + \text{perms.}] \\ & + N_a^i N_b^j N_{\phi c}^k N_{\phi d}^l \end{aligned}$$

$$\begin{aligned} & \times [P_{ik}^{ac}(\vec{k}_1) P_{jl}^{bd}(\vec{k}_2) P_\phi(|\vec{k}_1 + \vec{k}_3|) + \text{perms.}] \\ & + N_\phi^2 N_a^i N_{\phi\phi b}^j [P_\phi(k_1) P_\phi(k_2) P_{ij}^{ab}(\vec{k}_3) + \text{perms.}] \\ & + N_\phi N_a^i N_b^j N_{\phi cd}^{kl} [P_{ik}^{ac}(\vec{k}_1) P_{jl}^{bd}(\vec{k}_2) P_\phi(k_3) + \text{perms.}] \\ & + N_{\phi\phi} N_\phi N_{\phi a}^i N_b^j \\ & \times [P_\phi(k_2) P_\phi(|\vec{k}_1 + \vec{k}_2|) P_{ij}^{ab}(\vec{k}_4) + \text{perms.}] \\ & + N_{ab}^{ij} N_c^k N_{\phi d}^l N_\phi [P_{ac}^{ik}(\vec{k}_2) P_{bd}^{jl}(\vec{k}_1 + \vec{k}_2) P_\phi(k_4) + \text{perms.}] \\ & + N_a^i N_b^j N_{cd}^{kl} N_{ef}^{mn} \\ & \times [P_{ik}^{ac}(\vec{k}_1) P_{jm}^{be}(\vec{k}_2) P_{ln}^{df}(\vec{k}_1 + \vec{k}_3) + \text{perms.}] \\ & + N_a^i N_b^j N_c^k N_{def}^{lmn} [P_{il}^{ad}(\vec{k}_1) P_{jm}^{be}(\vec{k}_2) P_{kn}^{cf}(\vec{k}_3) + \text{perms.}], \end{aligned} \quad (34)$$

for the trispectrum (see also [97]).

Let us now provide some definition for the quantities introduced in the previous equations

$$\Pi_{ij}^{ab}(\vec{k}) \equiv T_{ij}^{\text{even}}(\vec{k}) P_+^{ab} + iT_{ij}^{\text{odd}}(\vec{k}) P_-^{ab} + T_{ij}^{\text{long}}(\vec{k}) P_{\text{long}}^{ab}, \quad (35)$$

where

$$\begin{aligned} T_{ij}^{\text{even}}(\vec{k}) & \equiv e_i^L(\hat{k}) e_j^{*L}(\hat{k}) + e_i^R(\hat{k}) e_j^{*R}(\hat{k}), \\ T_{ij}^{\text{odd}}(\vec{k}) & \equiv i[e_i^L(\hat{k}) e_j^{*L}(\hat{k}) - e_i^R(\hat{k}) e_j^{*R}(\hat{k})], \\ T_{ij}^{\text{long}}(\vec{k}) & \equiv e_i^L(\hat{k}) e_j^{*L}(\hat{k}). \end{aligned} \quad (36)$$

The polarization vectors are  $e^L(\hat{k}) \equiv (1/\sqrt{2})(\cos\theta \cos\phi - i \sin\phi, \cos\theta \sin\phi + i \cos\phi, -\sin\theta)$ ,  $e^R(\hat{k}) = e^{*L}(\hat{k})$  and  $e^l(\hat{k}) = \hat{k} = (\sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta)$ , from which we have

$$\begin{aligned} T_{ij}^{\text{even}}(\vec{k}) & = \delta_{ij} - \hat{k}_i \hat{k}_j, \\ T_{ij}^{\text{odd}}(\vec{k}) & = \epsilon_{ijk} \hat{k}_k, \\ T_{ij}^{\text{long}}(\vec{k}) & = \hat{k}_i \hat{k}_j. \end{aligned} \quad (37)$$

The purely scalar terms in (33)-(34) are already known from the literature: in single-field slow-roll inflation  $P_\phi = H_*^2/2k^3$ , where  $H_*$  is the Hubble rate evaluated at horizon exit; the bispectrum and the trispectrum of the scalar field ( $B_\phi$  and  $T_\phi$ ) can be found in [9–11, 38, 131] (they were also reported in [102, (11) and (12)]). As to the mixed (scalar-vector) terms, they can be ignored if one considers a Lagrangian where there is no direct coupling between the inflaton and the gauge bosons and where slow-roll assumptions are introduced for the fields (see [102, Section 4] for a complete discussion on this). Let us then look at the (purely) vector

part. Its anisotropy features can be stressed by rewriting them as follows (see also [96, 97])

$$B_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3) \supset \frac{1}{2} N_a^i N_b^j N_{cd}^{kl} \Pi_{ik}^{ac}(\vec{k}_1) \Pi_{jl}^{bd}(\vec{k}_2) = M_k^c N_{cd}^{kl} M_l^d \quad (38)$$

$$\begin{aligned} T_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4) &\supset N_a^i N_b^j N_{cd}^{kl} N_{ef}^{mn} P_{ik}^{ac}(\vec{k}_1) P_{jm}^{be}(\vec{k}_2) \\ &\times P_{ln}^{df}(\vec{k}_1 + \vec{k}_3) + N_a^i N_b^j N_c^k N_{def}^{lmn} \\ &\times P_{il}^{ad}(\vec{k}_1) P_{jm}^{be}(\vec{k}_2) P_{kn}^{cf}(\vec{k}_3) \\ &= M_i^c L_{ce}^{ij} M_j^e + M_i^f M_j^e M_k^d N_{fed}^{ijk}, \end{aligned} \quad (39)$$

where

$$\begin{aligned} M_k^c(\vec{k}) &\equiv N_a^i P_{ik}^{ac}(\vec{k}) \\ &= P_+^{ac}(k) [\delta_{ik} N_a^i + p^{ac}(k) \hat{k}_k (\hat{k} \cdot \vec{N}_a) + iq^{ac}(k) (\hat{k} \times \vec{N}_a)_k], \\ L_{ce}^{jl}(\vec{k}) &\equiv N_{cd}^j P_{ik}^{df}(\vec{k}) N_{fe}^{kl} \\ &= P_+^{df}(\vec{k}) [\vec{N}_{cd}^j \cdot \vec{N}_{ef}^l + p^{df}(k) (\hat{k} \cdot \vec{N}_{cd}^j) (\hat{k} \cdot \vec{N}_{ef}^l) \\ &\quad + iq^{df}(k) \hat{k} \cdot \vec{N}_{cd}^j \times \vec{N}_{ef}^l]. \end{aligned} \quad (40)$$

In the previous equations, we defined

$$\begin{aligned} p^{ac}(k) &\equiv \frac{P_{\text{long}}^{ac} - P_+^{ac}}{P_+^{ac}}, \\ q^{ac}(k) &\equiv \frac{P_-^{ac}}{P_+^{ac}}, \end{aligned} \quad (41)$$

with  $\vec{N}_a \equiv (N_a^1, N_a^2, N_a^3)$  and  $\vec{N}_{cd}^j \equiv (N_{cd}^{j1}, N_{cd}^{j2}, N_{cd}^{j3})$ .

Notice that, as for the power spectrum (25), also in (38)-(39) the anisotropic parts of the expressions are weighted by coefficients that are proportional either to  $P_-$  or to  $(P_{\text{long}} - P_+)$ . When these two quantities are equal to zero, the (Abelian) bispectrum and trispectrum are therefore isotropized.  $P_- = 0$  in parity conserving theories, like the ones we have been describing. According to the parametrization (16) of the longitudinal mode, we have  $P_{\text{long}} - P_+ = (|n(x)|^2 - 1) P_+$ .

**4.2.2. Non-Abelian Contributions.** We list the non-Abelian terms for the bispectrum

$$B_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3) \supset N_a^i N_b^j N_c^k B_{ijk}^{abc}(\vec{k}_1, \vec{k}_2, \vec{k}_3), \quad (42)$$

and for the trispectrum

$$\begin{aligned} T_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4) &\supset N_a^i N_b^j N_c^k N_d^l T_{ijkl}^{abcd}(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4) \\ &+ N_a^i N_b^j N_\phi N_\phi^k [P_\phi(k_3) B_{ijk}^{abc}(\vec{k}_1, \vec{k}_2, \vec{k}_3 + \vec{k}_4) + \text{perms.}] \\ &+ N_a^i N_b^j N_c^k N_{de}^{lm} \\ &\times [P_{il}^{ad}(\vec{k}_1) B_{jkm}^{bce}(\vec{k}_1 + \vec{k}_2, \vec{k}_3, \vec{k}_4) + \text{perms.}]. \end{aligned} \quad (43)$$

The computation of the vector bosons spectra

$$\langle \delta A_i^a \delta A_j^b \delta A_k^c \rangle = \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) B_{ijk}^{abc}, \quad (44)$$

$$\langle \delta A_i^a \delta A_j^b \delta A_k^c \delta A_l^d \rangle = \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4) T_{ijkl}^{abcd},$$

will be reviewed in this section. This requires the expansion of the in-in formula up to second order in the interaction Hamiltonian

$$\begin{aligned} &\langle \Theta(\eta^*) \rangle \\ &\supset i \left\langle T \left[ \Theta \int_{-\infty}^{\eta^*} d\eta' (H_{\text{int}}^+(\eta') - H_{\text{int}}^-(\eta')) \right] \right\rangle \\ &\quad + \frac{(-i)^2}{2} \left\langle T \left[ \Theta \int_{-\infty}^{\eta^*} d\eta' (H_{\text{int}}^+(\eta') - H_{\text{int}}^-(\eta')) \right. \right. \\ &\quad \left. \left. \times \int_{-\infty}^{\eta^*} d\eta'' (H_{\text{int}}^+(\eta'') - H_{\text{int}}^-(\eta'')) \right] \right\rangle. \end{aligned} \quad (45)$$

The interaction Hamiltonian needs to be expanded up to fourth order in the fields fluctuations, that is,  $H_{\text{int}} = H_{\text{int}}^{(3)} + H_{\text{int}}^{(4)}$ , where

$$\begin{aligned} H_{\text{int}}^{(3)} &= g_c \varepsilon^{abc} g^{ik} g^{jl} (\partial_i \delta B_j^a) \delta B_k^b \delta B_l^c \\ &\quad + g_c^2 \varepsilon^{eab} \varepsilon^{ecd} g^{ik} g^{jl} B_i^a \delta B_j^b \delta B_k^c \delta B_l^d, \end{aligned} \quad (46)$$

$$H_{\text{int}}^{(4)} = g_c^2 \varepsilon^{eab} \varepsilon^{ecd} g^{ij} g^{kl} \delta B_i^a \delta B_k^b \delta B_j^c \delta B_l^d.$$

To tree-level, the relevant diagrams are pictured in Figures 1 and 2. By looking at (46), we can see that there is a bispectrum diagram that is lower in terms of power of the  $SU(2)$  coupling ( $\sim g_c$ ) compared to the trispectrum ( $\sim g_c^2$ ); as a matter of fact, for symmetry reasons that we are going to discuss later in this section,  $g_c^2$  interaction terms are needed to provide a nonzero contributions to the bispectrum.

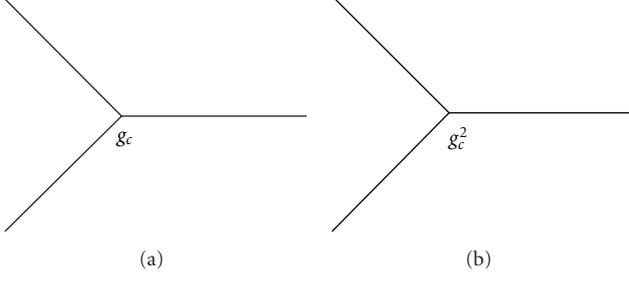


FIGURE 1: Diagrammatic representations of the tree-level contributions to the vector fields bispectrum.

The propagators for “plus” and “minus” fields are

$$\begin{aligned}
\widehat{\delta B_i^{a,+}(\eta') \delta B_j^{b,+}(\eta'')} &= \widetilde{\Pi}_{ij}^{ab}(\eta', \eta'') \Theta(\eta' - \eta'') \\
&\quad + \overline{\Pi}_{ij}^{ab}(\eta', \eta'') \Theta(\eta'' - \eta'), \\
\widehat{\delta B_i^{a,+}(\eta') \delta B_j^{b,-}(\eta'')} &= \overline{\Pi}_{ij}^{ab}(\eta', \eta''), \\
\widehat{\delta B_i^{a,-}(\eta') \delta B_j^{b,+}(\eta'')} &= \widetilde{\Pi}_{ij}^{ab}(\eta', \eta''), \\
\widehat{\delta B_i^{a,-}(\eta') \delta B_j^{b,-}(\eta'')} &= \overline{\Pi}_{ij}^{ab}(\eta', \eta'') \Theta(\eta' - \eta'') \\
&\quad + \widetilde{\Pi}_{ij}^{ab}(\eta', \eta'') \Theta(\eta'' - \eta'),
\end{aligned} \tag{47}$$

or

$$\begin{aligned}
\widetilde{\Pi}_{ij}^{ab}(\vec{k}) &\equiv T_{ij}^{\text{even}}(\hat{k}) \widetilde{P}_+^{ab} + iT_{ij}^{\text{odd}}(\hat{k}) \widetilde{P}_{ij}^{ab} + T_{ij}^{\text{long}}(\hat{k}) \widetilde{P}_{ij}^{ab}, \\
\overline{\Pi}_{ij}^{ab}(\vec{k}) &\equiv T_{ij}^{\text{even}}(\hat{k}) \overline{P}_+^{ab} + iT_{ij}^{\text{odd}}(\hat{k}) \overline{P}_{ij}^{ab} + T_{ij}^{\text{long}}(\hat{k}) \overline{P}_{ij}^{ab},
\end{aligned} \tag{48}$$

in Fourier space. In the previous equations we set  $\widetilde{P}_\pm^{ab} \equiv (1/2)(\widetilde{P}_R^{ab} \pm \widetilde{P}_L^{ab})$ ,  $\widetilde{P}_R^{ab} = \delta_{ab} \delta B_R^{ab}(k, \eta^*) \delta B_R^{*ab}(k, \eta)$  and  $\overline{P}_\pm^{ab} = (\widetilde{P}_\pm^{ab})^*$  (similar definitions apply for  $\widetilde{P}_L^{ab}$  and  $\widetilde{P}_{\text{long}}^{ab}$ ).

We are now ready to show the computation of the following contributions to the bispectrum and trispectrum of  $\zeta$

$$\begin{aligned}
\langle \zeta_{\vec{k}_1}^- \zeta_{\vec{k}_2}^- \zeta_{\vec{k}_3}^- \rangle \\
\supset N_a^i N_b^j N_c^k \langle \delta A_i^a(\vec{k}_1) \delta A_j^b(\vec{k}_2) \delta A_k^c(\vec{k}_3) \rangle,
\end{aligned} \tag{49}$$

$$\begin{aligned}
\langle \zeta_{\vec{k}_1}^- \zeta_{\vec{k}_2}^- \zeta_{\vec{k}_3}^- \zeta_{\vec{k}_4}^- \rangle \\
\supset N_a^i N_b^j N_c^k N_d^l \\
\times \langle \delta A_i^a(\vec{k}_1) \delta A_j^b(\vec{k}_2) \delta A_k^c(\vec{k}_3) \delta A_l^d(\vec{k}_4) \rangle.
\end{aligned} \tag{50}$$

Equation (49) becomes

$$\begin{aligned}
\langle \zeta_{\vec{k}_1}^- \zeta_{\vec{k}_2}^- \zeta_{\vec{k}_3}^- \rangle &\supset N_a^i N_b^j N_c^k \frac{\delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3)}{a^3(\eta^*)} \\
&\times \left[ \int d\eta a^4(\eta) \widetilde{\Pi}_{im}^l(\vec{k}_1) \widetilde{\Pi}_j^l(\vec{k}_2) \widetilde{\Pi}_k^m(\vec{k}_3) \right. \\
&\quad \left. \times \left( g_c \varepsilon^{abc} k_{1l} + g_c^2 \varepsilon^{eda} \varepsilon^{ebc} B_l^d \right) \right] \\
&\quad + \text{perms.} + c.c.
\end{aligned} \tag{51}$$

Even before performing the time integration, one realizes that, because of the antisymmetric properties of the Levi-Civita tensor, the  $\sim g_c$  contribution on the r.h.s. of (51) is equal to zero once the sum over all the possible permutations has been performed. The vector bosons bispectrum is therefore proportional to  $g_c^2$ . The final result from (51) has the following form

$$\begin{aligned}
\langle \zeta_{\vec{k}_1}^- \zeta_{\vec{k}_2}^- \zeta_{\vec{k}_3}^- \rangle &\supset (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3) g_c^2 H_*^2 \\
&\times \sum_n F_n(k_i, \eta^*) I_n(\hat{k}_i \cdot \hat{k}_j, \vec{A}_i \cdot \vec{A}_j, \hat{k}_i \cdot \vec{A}_j),
\end{aligned} \tag{52}$$

where the  $F_n$ 's are isotropic functions of time and of the moduli of the wave vectors ( $i = 1, 2, 3$ ) and the  $I_n$ 's are anisotropic coefficients. The sum in the previous equation is taken over all possible combinations of products of three polarization indices, that is,  $n \in (EEE, EEL, ELE, \dots, lll)$ , where  $E$  stands for “even”,  $l$  for “longitudinal”. The complete expressions for the terms appearing in the sum are quite lengthy (see [101, Section 4.2]). As an example, we report one of these terms

$$\begin{aligned}
F_{lll} &= -n^6(x^*) \frac{1}{24k^6 k_1^2 k_2^2 k_3^2 x^{*2}} \\
&\times [A_{EEE} + (B_{EEE} \cos x^* + C_{EEE} \sin x^*) E_i x^*],
\end{aligned} \tag{53}$$

$$\begin{aligned}
I_{lll} &= \varepsilon^{aa'b'} \varepsilon^{ac'e} \\
&\times \left[ \left( (\hat{k}_1 \cdot \vec{N}^{a'}) (\hat{k}_3 \cdot \vec{N}^{b'}) (\hat{k}_2 \cdot \vec{N}^{c'}) (\hat{k}_1 \cdot \hat{k}_2) \right. \right. \\
&\quad \times (\hat{k}_3 \cdot \hat{A}^e) - (\hat{k}_3 \cdot \vec{N}^{a'}) (\hat{k}_2 \cdot \vec{N}^{b'}) \\
&\quad \times (\hat{k}_1 \cdot \vec{N}^{c'}) (\hat{k}_1 \cdot \hat{k}_2) (\hat{k}_3 \cdot \hat{A}^e) \left. \right. \\
&\quad \left. \left. + (1 \leftrightarrow 3) + (2 \leftrightarrow 3) \right],
\end{aligned} \tag{54}$$

where  $A_{EEE}$ ,  $B_{EEE}$  and  $C_{EEE}$  are functions of  $x^*$  and of the momenta  $k_i \equiv |k_i|$  (they are all reported in [101, Appendix C]),  $E_i$  is the exponential-integral function and  $i \leftrightarrow j$  means “exchange  $\hat{k}_i$  with  $\hat{k}_j$ ”. As we will discuss in more details in Section 6, one of the more interesting features of these models is that the bispectrum and the trispectrum turn out to have an amplitude that is modulated by the preferred directions that break statistical isotropy.

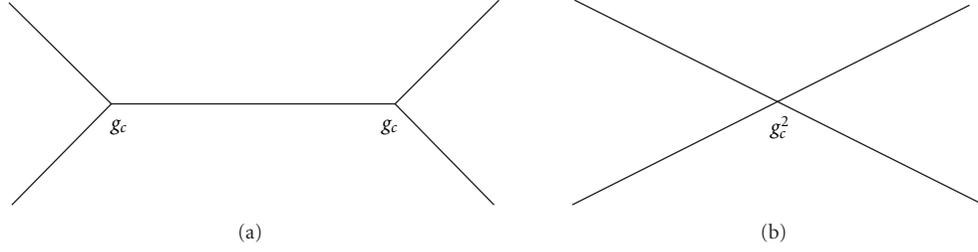


FIGURE 2: Diagrammatic representations of the tree-level contributions to the vector fields trispectrum: vector-exchange (a) and contact-interaction (b) diagrams.

Let us now move to the trispectrum. Again, we count two different kinds of contributions, the first from  $\sim g_c$  and the second from  $\sim g_c^2$  interaction terms, respectively in  $H_{\text{int}}^{(3)}$  and  $H_{\text{int}}^{(4)}$ . The former produce vector-exchange diagrams, the latter are represented by contact-interaction diagrams (see Figure 2). Their analytic expressions are different, but they both have a structure similar to (52)

$$\begin{aligned} & \langle \zeta_{\vec{k}_1} \zeta_{\vec{k}_2} \zeta_{\vec{k}_3} \zeta_{\vec{k}_4} \rangle \\ & \supset (2\pi)^3 \delta^{(3)}(\vec{k}_1 + \vec{k}_2 + \vec{k}_3 + \vec{k}_4) g_c^2 H_*^2 \\ & \quad \times \sum_n G_n(k_i, k_{\hat{1}2}, k_{\hat{1}4}, \eta^*) L_n(\hat{k}_i \cdot \hat{k}_j, \vec{A}_i \cdot \vec{A}_j, \hat{k}_i \cdot \vec{A}_j), \end{aligned} \quad (55)$$

where we define  $k_{\hat{1}2} \equiv |\vec{k}_1 + \vec{k}_2|$  and  $k_{\hat{1}4} \equiv |\vec{k}_1 + \vec{k}_4|$  (see [102, Sections 5.2.1 and 5.2.2] for the explicit expressions of the functions  $G_n$  and  $L_n$ ).

## 5. Amplitude of non-Gaussianity: $f_{NL}$ and $\tau_{NL}$

In this paper, we use the following definitions for the non-Gaussianity amplitudes

$$\begin{aligned} \frac{6}{5} f_{NL} &= \frac{B_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3)}{P_{\text{iso}}(k_1) P_{\text{iso}}(k_2) + \text{perms.}}, \\ \tau_{NL} &= \frac{2T_\zeta(\vec{k}_1, \vec{k}_2, \vec{k}_3, \vec{k}_4)}{P_{\text{iso}}(k_1) P_{\text{iso}}(k_2) P_{\text{iso}}(k_{\hat{1}4}) + 23 \text{ perms.}} \end{aligned} \quad (56)$$

The choice of normalizing the bispectrum and the trispectrum by the isotropic part of the power spectrum, instead of using its complete expression  $P_\zeta$ , is motivated by the fact that the latter would only introduce a correction to the previous equations proportional to the anisotropy parameter  $g$ , which is a small quantity.

The parameters  $f_{NL}$  and  $\tau_{NL}$  receive contributions both from scalar (“s”) and from vector (“v”) fields

$$f_{NL} = f_{NL}^{(s)} + f_{NL}^{(v)}, \quad (57)$$

$$\tau_{NL} = \tau_{NL}^{(s)} + \tau_{NL}^{(v)}. \quad (58)$$

The latter can again be distinguished into Abelian (A) and non-Abelian (NA)

$$f_{NL}^{(v)} = f_{NL}^{(A)} + f_{NL}^{(NA)}, \quad (59)$$

$$\tau_{NL}^{(v)} = \tau_{NL}^{(A_1)} + \tau_{NL}^{(A_2)} + \tau_{NL}^{(NA_1)} + \tau_{NL}^{(NA_2)}. \quad (60)$$

The contribution  $f_{NL}^{(A)}$  comes from (38),  $f_{NL}^{(NA)}$  from (52),  $\tau_{NL}^{(A_1)}$  and  $\tau_{NL}^{(A_2)}$  from (39), finally  $\tau_{NL}^{(NA_1)}$  from (55) and  $\tau_{NL}^{(NA_2)}$  from the last line of (43).

In order to keep the vector contributions manageable and simple in their structure, all gauge and vector indices will be purposely neglected at this point and so the angular functions appearing in the anisotropy coefficients will be left out of the final amplitude results. This is acceptable considering that these functions will in general introduce numerical corrections of order one. Nevertheless, it is important to keep in mind that the amplitudes also depend on the angular parameters of the theory.

We will now focus on the dependence of  $f_{NL}$  and  $\tau_{NL}$  from the nonangular parameters of the theory and quickly draw a comparison among the different contributions listed in (57) through (60).

The expression of the number of e-foldings depends on the specific model and, in particular, on the mechanism of production of the fluctuations. Two possibilities have been described in Section 2. For vector inflation we have

$$N_a^i = \frac{A_i^a}{2m_p^2}, \quad N_{ab}^{ij} = \frac{\delta_{ab} \delta^{ij}}{2m_p^2}, \quad (61)$$

(see [101, Appendix B] for their derivation). In the vector curvaton model the same quantities become [96, 101]

$$N_a^i = \frac{2}{3} r \frac{A_i^a}{\sum_b |\vec{A}^b|^2}, \quad N_{ab}^{ij} = \frac{1}{3} r \frac{\delta_{ab} \delta^{ij}}{\sum_c |\vec{A}^c|^2}. \quad (62)$$

Neglecting tensor and gauge indices, the expressions above can be simplified as  $N_A \simeq A/m_p^2$  and  $N_{AA} \simeq 1/m_p^2$  in vector inflation,  $N_A \simeq r/A$  and  $N_{AA} \simeq r/A^2$  in the vector curvaton model. Also we have  $N_{AAA} = 0$  in vector inflation and  $N_{AAA} \simeq r/A^3$  in vector curvaton.

We are now ready to provide the final expressions for the amplitudes: in Table 1 we list all the contributions to  $f_{NL}$ ,

TABLE 1: Order of magnitude of  $f_{NL}$  in different scenarios.

	$f_{NL}^s$	$f_{NL}^A$	$f_{NL}^{NA}$
general case	$\frac{1}{(1+\beta)^2} \frac{N_{\phi\phi}}{N_\phi^2}$	$\frac{\beta}{(1+\beta)^2} \frac{N_{AA}}{N_\phi^2}$	$\frac{\beta^2}{(1+\beta)^2} g_c^2 \left(\frac{m}{H}\right)^2$
v.inflation	$\frac{\epsilon_\phi}{(1 + ((A/m_P)\sqrt{\epsilon_\phi})^2)^2}$	$\frac{\epsilon_\phi^2}{(1 + ((A/m_P)\sqrt{\epsilon_\phi})^2)^2} \left(\frac{A}{m_P}\right)^2$	$\frac{\epsilon_\phi^2 g_c^2}{(1 + ((A/m_P)\sqrt{\epsilon_\phi})^2)^2} \left(\frac{A^2}{m_P H}\right)^2$
v.curvaton	$\frac{\epsilon_\phi}{(1 + (Am_P/A_{tot}^2)^2 \epsilon_\phi r^2)^2}$	$\frac{\epsilon_\phi^2 r^3}{(1 + (Am_P/A_{tot}^2)^2 \epsilon_\phi r^2)^2} \left(\frac{Am_P^2}{A_{tot}^3}\right)^2$	$\frac{\epsilon_\phi^2 r^3 g_c^2}{(1 + (Am_P/A_{tot}^2)^2 \epsilon_\phi r^2)^2} \left(\frac{A^2 m_P^2}{A_{tot}^3 H}\right)^2$

TABLE 2: Order of magnitude of the vector contributions to  $\tau_{NL}$  in different scenarios.

	$\tau_{NL}^{NA1}$	$\tau_{NL}^{NA2}$	$\tau_{NL}^{A1}$	$\tau_{NL}^{A2}$
general case	$10^3 \frac{\beta^2 \epsilon g_c^2}{(1+\beta)^3} \left(\frac{m_P}{H}\right)^2$	$10^{-5} \frac{\beta^{3/2} \epsilon^{3/2} g_c^2}{(1+\beta)^3} \left(\frac{A}{H}\right) \left(\frac{m_P}{H}\right) m_P^2 N_{AA}$	$\frac{\beta \epsilon^2}{(1+\beta)^3} m_P^4 N_{AA}^2$	$\frac{\beta^{3/2} \epsilon^{3/2}}{(1+\beta)^3} m_P^3 N_{AAA}$
v.inflation	same as above	$10^{-5} \frac{\beta^{3/2} \epsilon^{3/2} g_c^2}{(1+\beta)^3} \left(\frac{A}{H}\right) \left(\frac{m_P}{H}\right)$	$\frac{\beta \epsilon^2}{(1+\beta)^3}$	0
v.curvaton	same as above	$10^{-5} \frac{r \beta^{3/2} \epsilon^{3/2} g_c^2}{(1+\beta)^3} \left(\frac{A}{H}\right) \left(\frac{m_P}{H}\right) \left(\frac{m_P}{A}\right)^2$	$\frac{r^2 \beta \epsilon^2}{(1+\beta)^3} \left(\frac{m_P}{A}\right)^4$	$\frac{r \beta^{3/2} \epsilon^{3/2}}{(1+\beta)^3} \left(\frac{m_P}{A}\right)^3$

TABLE 3: Order of magnitude of the ratios  $f_{NL}^v/f_{NL}^s$  in different scenarios.

	$f_{NL}^A/f_{NL}^s$	$f_{NL}^{NA}/f_{NL}^s$
general case	$\beta \frac{N_{AA}}{N_{\phi\phi}}$	$\beta^2 g_c^2 \left(\frac{m}{H}\right)^2 \frac{N_\phi^2}{N_{\phi\phi}}$
v.inflation	$\beta$	$\frac{\beta^2 g_c^2}{\epsilon_\phi} \left(\frac{m_P}{H}\right)^2$
v.curvaton	$\beta r \left(\frac{m_P}{A}\right)^2$	$\frac{\beta^2 g_c^2}{\epsilon_\phi r} \left(\frac{A}{H}\right)^2$

Table 2 includes the vector contributions to  $\tau_{NL}$ , the scalar contributions being given by

$$\tau_{NL}^{(s)} = \frac{\epsilon_\phi}{(1+\beta)^3} + \frac{\epsilon_\phi^2}{(1+\beta)^3}. \quad (63)$$

In the expressions appearing in the tables, numerical coefficients of order one have not been reported. Also,  $m$  is by definition equal to  $m_P$  in vector inflation and to  $A/\sqrt{r}$  in the vector curvaton model;  $N_\phi \simeq (m_P \sqrt{\epsilon_\phi})^{-1}$  and  $N_{\phi\phi} \simeq m_P^{-2}$ , with  $\epsilon_\phi \equiv (\dot{\phi}^2)/(2m_P^2 H^2)$ .

The quantities involved in the amplitude expressions are  $g$ ,  $\beta$ ,  $r$ ,  $\epsilon_\phi$ ,  $g_c$ ,  $m_P/H$ ,  $A/m_P$  and  $A/H$ . We already know that  $g$  and  $\beta$  are to be considered smaller than one (see discussion after (32)). Similarly, as mentioned after (3),  $r$  has to remain small at least until inflation ends so as to attain an ‘‘almost isotropic’’ expansion. The slow-roll parameter  $\epsilon_\phi$  and the  $SU(2)$  coupling  $g_c$  are small, respectively, to allow the inflaton to slowly roll down its potential and for perturbation theory to be valid. The ratio  $m_P/H$  is of order  $10^5$  (assuming  $\epsilon_\phi \sim 10^{-1}$ ). Finally,  $A/m_P$  and  $A/H$  have no stringent bounds. A reasonable choice could be to assume that the expectation value of the gauge fields is no larger than the Planck mass, that is,  $A/m_P \leq 1$ . As to the  $A/H$  ratio, different possibilities are allowed, including the one where it is of order one (see [99, Section 6 and Appendix A] for a discussion on this).

Let us now compare the different amplitude contributions. The ratios between scalar and vector contributions are shown in Table 3 for the bispectrum and Table 4 for the trispectrum. We can observe that the dominance of a given contribution w.r.t. another one very much depends on the selected region of parameter space. It turns out that it is allowed for the vector contributions to be larger than the scalar ones and also for the non-Abelian contributions to be larger than the Abelian ones. This is discussed more in details in [99, Section 6]. An interesting point is, for

TABLE 4: Order of magnitude of the ratios  $\tau_{NL}^v/\tau_{NL}^s$  in different scenarios.

	$\tau_{NL}^{NA_1}/\tau_{NL}^s$	$\tau_{NL}^{NA_2}/\tau_{NL}^s$	$\tau_{NL}^{A_1}/\tau_{NL}^s$	$\tau_{NL}^{A_2}/\tau_{NL}^s$
general case	$10^3 \beta^2 g_c^2 \left(\frac{m_P}{H}\right)^2$	$10^{-5} \beta^{3/2} \epsilon^{1/2} g_c^2 \left(\frac{A}{H}\right) \left(\frac{m_P}{H}\right) m_P^2 N_{AA}$	$\beta \epsilon m_P^4 N_{AA}^2$	$\beta^{3/2} \epsilon^{1/2} m_P^3 N_{AAA}$
v.inflation	same as above	$10^{-5} \beta^{3/2} \epsilon^{1/2} g_c^2 \left(\frac{A}{H}\right) \left(\frac{m_P}{H}\right)$	$\beta \epsilon$	0
v.curvaton	same as above	$10^{-5} r \beta^{3/2} \epsilon^{1/2} g_c^2 \left(\frac{A}{H}\right) \left(\frac{m_P}{H}\right) \left(\frac{m_P}{A}\right)^2$	$r^2 \beta \epsilon \left(\frac{m_P}{A}\right)^4$	$r \beta^{3/2} \epsilon^{1/2} \left(\frac{m_P}{A}\right)^3$

TABLE 5: Order of magnitude of the ratios  $\tau_{NL}^v/(f_{NL}^{NA})^2$  in different scenarios.

	$\tau_{NL}^{NA_1}/(f_{NL}^{NA})^2$	$\tau_{NL}^{NA_2}/(f_{NL}^{NA})^2$	$\tau_{NL}^{A_1}/(f_{NL}^{NA})^2$	$\tau_{NL}^{A_2}/(f_{NL}^{NA})^2$
v.i.	$10^9 \frac{\epsilon(1+\beta)}{g_c^2 \beta^2} \left(\frac{H}{m_P}\right)^2$	$10 \frac{\epsilon^{3/2}(1+\beta)}{\beta^{5/2} g_c^2} \left(\frac{A}{H}\right) \left(\frac{H}{m_P}\right)^3$	$10^6 \frac{\epsilon^2(1+\beta)}{\beta^3 g_c^4} \left(\frac{H}{m_P}\right)^4$	0
v.c.	$10^9 \frac{r^2 \epsilon(1+\beta)}{g_c^2 \beta^2} \frac{m_P^2 H^2}{A^2 A^2}$	$10 \frac{r^5 \epsilon^{3/2}(1+\beta)}{\beta^{5/2} g_c^2} \frac{H^3 m_P m_P^2}{A^3 H A^2}$	$10^6 \frac{r^6 \epsilon^2(1+\beta)}{\beta^3 g_c^4} \left(\frac{m_P}{A}\right)^4 \left(\frac{H}{A}\right)^4$	$10^6 \frac{r^3 \epsilon^{3/2}(1+\beta)}{g_c^2 \beta^{5/2}} \frac{m_P^3 H^4}{A^3 A^4}$

instance, the following: ignoring tensor and gauge indices, the ratio  $g_c A/H$ , that appears in many of the Tables entries, is a quantity smaller than one; if we consider the different configurations identified by gauge and vector indices, we realize that this is not always true, in fact the value of this ratio can be  $\gg 1$  in some configurations.

Finally, it is interesting to compare bispectrum and trispectrum amplitudes (see Table 5). Again, it is allowed for the ratios appearing in Table 5 to be either large or small, depending on the specific location within the parameter space of the theory. For instance, the combination of a small bispectrum with a large trispectrum is permitted. The latter is an interesting possibility: if the bispectrum was observably small, we could still hope the information about non-Gaussianity to be accessible thanks to the trispectrum.

Another interesting feature of this model is that the bispectrum and the trispectrum depend on the same set of quantities. If these correlation functions were independently known, that information could then be used to test the theory and place some bounds on its parameters.

## 6. Shape of Non-Gaussianity and Statistical Anisotropy Features

Studying the shape of non-Gaussianity means understanding the features of momentum dependence of the bispectrum and higher order correlators. If they also depend on variables other than momenta, it is important to determine how these other variables affect the profiles for any given momentum set-up. This is the case as far as the bispectrum and the trispectrum of the gauge fields are concerned, given the fact

that they are functions, besides of momenta, also of a large set of angular variables (see (52) and (55)).

*6.1. Momentum Dependence of the Bispectrum and Trispectrum Profiles.* We show the study of the momentum dependence of the  $F_n$  and  $G_n$  functions in (52) and (55) first and then analyze the angular variables dependence of the spectra, once the momenta have been fixed in a given configuration. A natural choice would be to consider the configuration where the correlators are maximized.

The maxima can be easily determined for the bispectrum by plotting the isotropic functions  $F_n$  and  $G_n$  in terms of two of their momenta. These plots are provided in Figure 3, where the variables are  $x_2 \equiv k_2/k_1$  and  $x_3 \equiv k_3/k_1$ . Each one of the plots corresponds to a single isotropic functions of the sum in (52). It is apparent that the maxima are mostly located in the so-called local region, that is, for  $k_1 \sim k_2 \gg k_3$ ; three out of the eight graphs do not have their peaks in this configuration but, at the same time, they show negligible amplitudes compared to the ‘‘local’’ peaked graphs.

The situation is much more complex for the trispectrum, being the number of momentum variables larger than three ( $k_1, k_2, k_3, k_4, k_{\hat{1}2}$  and  $k_{\hat{1}4}$ ). The momentum dependence of the isotropic functions can be studied by selecting different configurations for the tetrahedron made up by the four momentum vectors, in such a way as to narrow the number of independent momentum variables down to two. A list of possible configurations was presented in [43]. We consider two of them, the ‘‘equilateral’’ and the ‘‘specialized planar’’.

In the equilateral configuration the four sides of the tetrahedron have the same length ( $k_1 = k_2 = k_3 = k_4$ ),

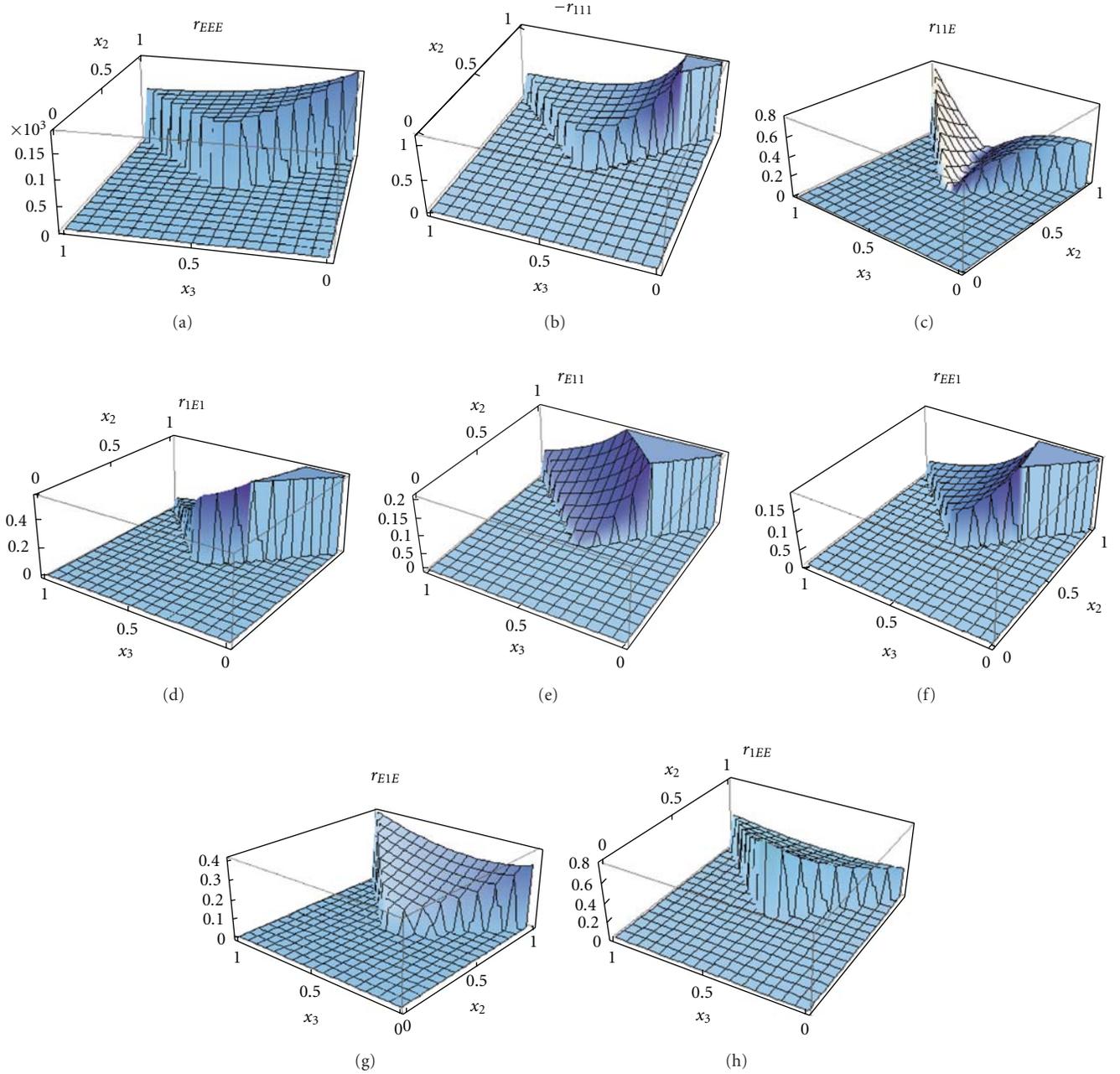


FIGURE 3: Plot of  $r_n \equiv \Theta(x_2 - x_3)\Theta(x_3 - 1 + x_2)x_2^2x_3^2R_n(x_2, x_3)$ , where we define  $R_n = k_1^6 F_n$ . The Heaviside step functions  $\Theta$  help restricting the plot domain to the region  $(x_2, x_3)$  that is allowed for the triangle  $\vec{k}_1 + \vec{k}_2 + \vec{k}_3 = 0$  (in particular, we set  $x_3 < x_2$ ). We also set  $x^* = 1$ .

therefore  $x \equiv k_{\hat{1}2}/k_1$  and  $y \equiv k_{\hat{1}4}/k_1$  can be chosen as variables for the plots. The plots of the isotropic functions of contact interaction and vector exchange contributions are provided in Figure 4. The former (*c.i.*) shows a constant behaviour in this configuration, being independent of  $k_{\hat{1}2}$  and  $k_{\hat{1}4}$ . The latter (*v.e.(I)*, *v.e.(II)* and *v.e.(III)*) diverge as  $k_{\hat{1}i}^{-3}$  ( $i = 1, 2$ , and  $3$ , resp., for the three plots) in the limit of a flat tetrahedron, that is,  $(k_{\hat{1}i}/k_1) \rightarrow 0$ .

In the specialized planar configuration, the tetrahedron is flattened and, in addition to that, three of the six momentum variables are set equal to one another ( $k_1 = k_3 = k_{\hat{1}4}$ ); this

leaves two independent variables, which can be  $x \equiv k_2/k_1$  and  $y \equiv k_3/k_1$ . There is a double degeneracy in this configuration, due to the fact that the quadrangle can have internal angles larger than or smaller/equal to  $\pi$ , as we can see from the plus and minus signs in the expressions for  $k_{\hat{1}2}$  and  $k_{\hat{1}3}$  [43]

$$\begin{aligned} \frac{k_{\hat{1}2}}{k_1} &= \sqrt{1 + \frac{x^2 y^2}{2} \pm \frac{xy}{2} \sqrt{(4-x^2)(4-y^2)}}, \\ \frac{k_{\hat{1}3}}{k_1} &= \sqrt{x^2 + y^2 - \frac{x^2 y^2}{2} \mp \frac{xy}{2} \sqrt{(4-x^2)(4-y^2)}}. \end{aligned} \quad (64)$$

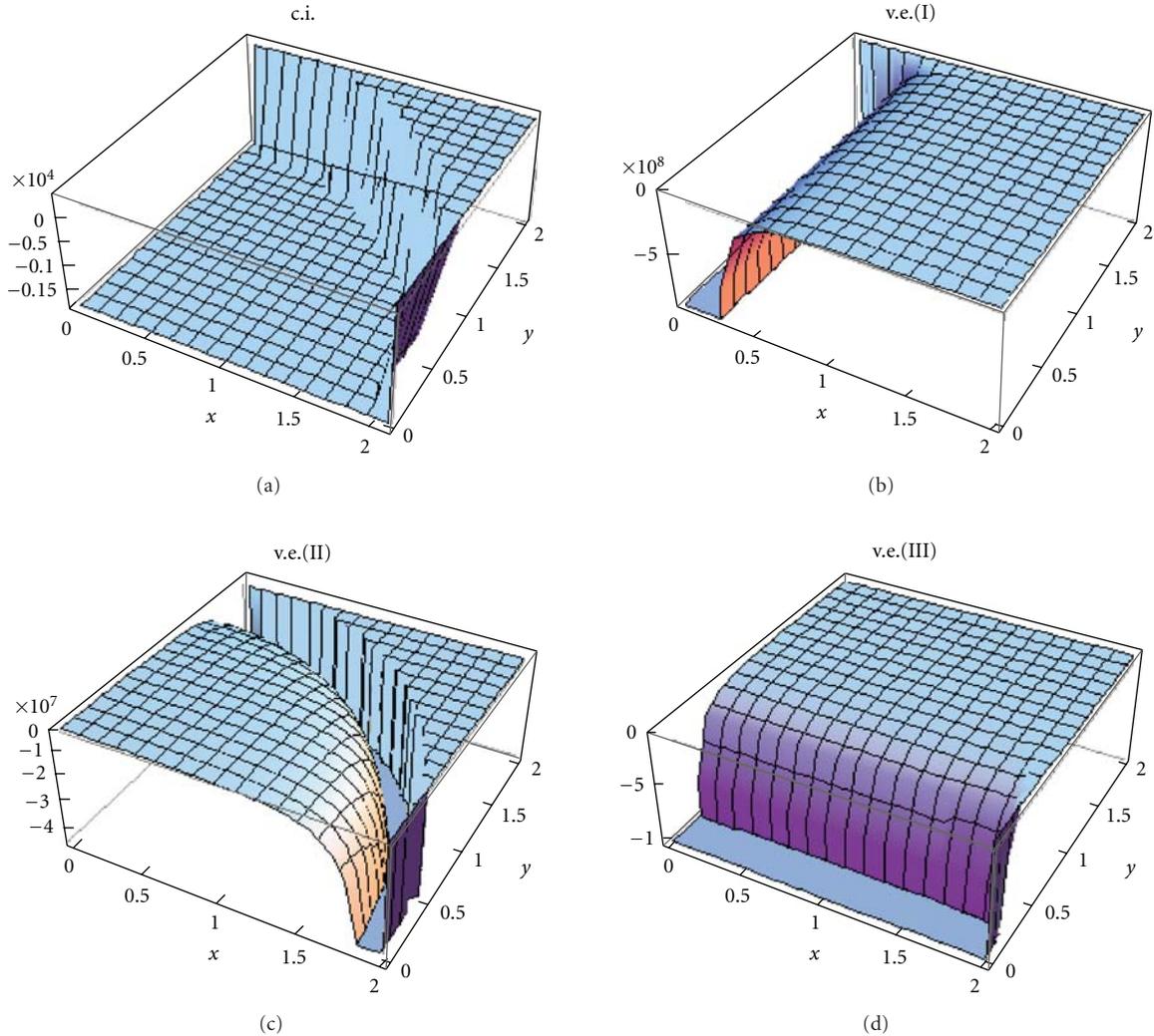


FIGURE 4: Plots of the isotropic functions appearing in the vector fields trispectrum (from (55)): *c.i.* is the contribution from contact-interaction diagrams, *v.e.(I)*, *v.e.(II)*, and *v.e.(III)* are the contributions from the vector-exchange diagrams. The equilateral configuration has been considered in this figure.

The two cases are plotted in Figures 5 and 6. Notice that divergences generally occur as  $x, y \rightarrow 0$ , as  $x \rightarrow y$  and  $(x, y) \rightarrow (2, 2)$ .

**6.2. Features and Level of Anisotropy.** Statistical homogeneity and isotropy are considered characterizing features of the CMB fluctuations distribution, if one ignores the issues raised by the “anomalous” detections we presented in the introduction.

Homogeneity of the correlation functions equates translational invariance and hence total momentum conservation, as enforced by the delta functions appearing on the left-hand sides of (17) through (19). This invariance property can then be pictured as the three momentum vectors forming a closed triangle for the bispectrum and the four momenta arranged in a tetrahedron for the trispectrum (see Figure 7).

Statistical isotropy corresponds to invariance w.r.t. rotations in space of the momentum (for the power spectrum) and of the triangle or tetrahedron made up by the momenta,

respectively for the bispectrum and the trispectrum. This symmetry can be broken, as it for example happens in the  $SU(2)$  case, by assuming the existence of preferred spatial directions in the early universe that might be revealed in the CMB observations. When this happens, the correlation functions are expected to be sensitive to the spatial orientation of the wave number or of the momenta triangles and tetrahedrons w.r.t. these special directions. Analytically, the bispectrum and the trispectrum will depend on the angles among the vector bosons and the wave vectors (besides the angles among the gauge bosons themselves), as shown in the coefficients appearing in (52) and (55). This implies that both the amplitude and the shape of bispectrum and trispectrum will be affected by these mutual spatial orientations. The modulation of the shapes by the directions that break statistical anisotropy was discussed with some examples both for the bispectrum and the trispectrum in our previous papers [101, 102]. These examples are here reported in Figures 8 and 9. In Figure 8 we show the plot of the vector

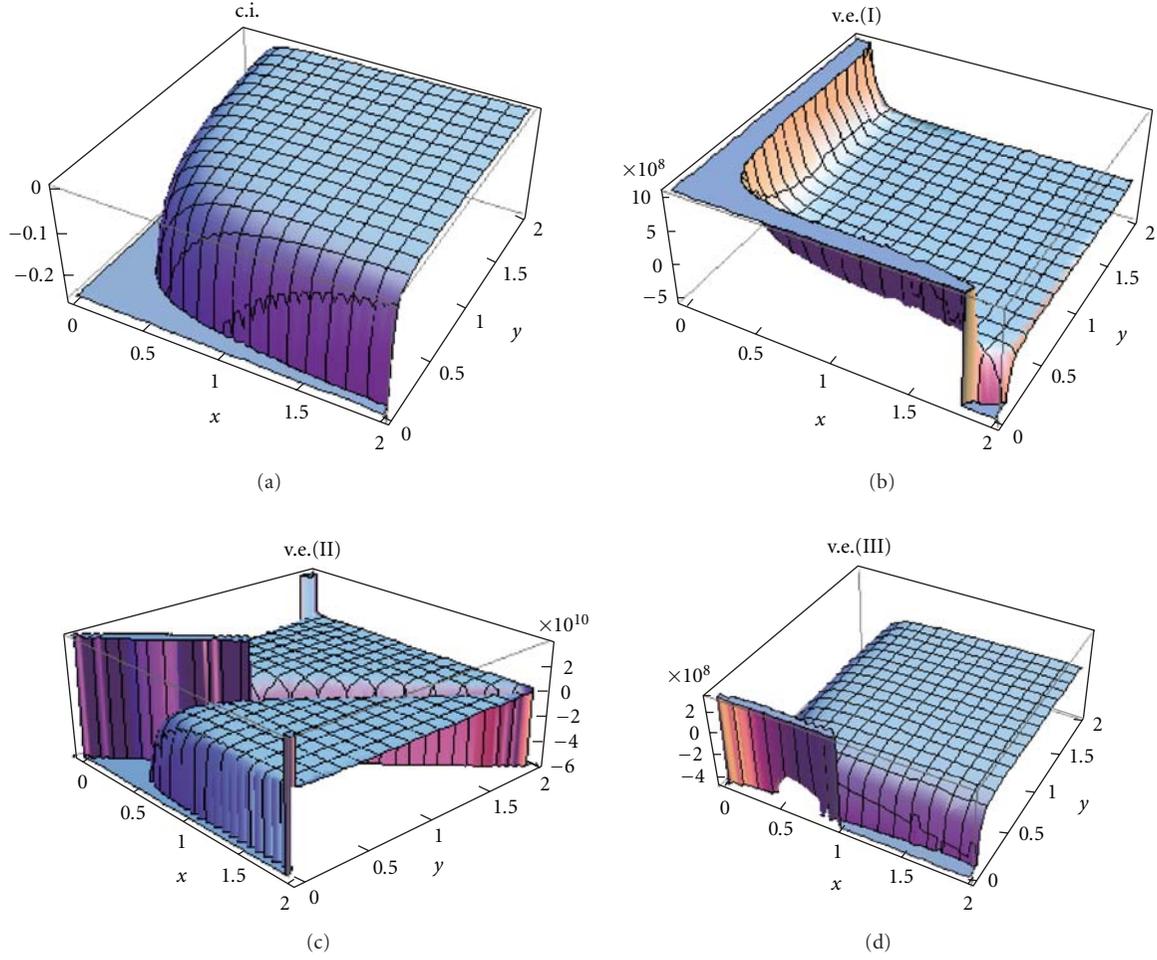


FIGURE 5: Plots of the contact interaction and of the vector-exchange contributions in the specialized planar configuration (plus sign).

contribution to the bispectrum of  $\zeta$ , properly normalized in the configuration

$$\vec{N}_3 = N_A(0, 0, 1), \quad (65)$$

$$\vec{N}_1 = \vec{N}_2 = N_A(\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta),$$

where, the  $(x, y, z)$  coordinate frame is chosen to be  $\hat{k}_3 = \hat{x}$  and  $\hat{k}_1 = \hat{k}_2 = \hat{z}$  and  $\delta$  is the angle between  $\vec{N}_{1,2}$  and  $\hat{k}_3$ .

In Figure 9 we provide a similar plot for the trispectrum, but in a different configuration

$$\begin{aligned} \hat{N}_2 \cdot \hat{k}_i &= 0 \quad (i = 1, \dots, 4) \\ \hat{N}_1 \cdot \hat{k}_1 &= \cos \delta, \quad \hat{N}_1 \cdot \hat{k}_2 = 0, \\ \hat{N}_3 \cdot \hat{k}_2 &= \cos \theta, \quad \hat{N}_3 \cdot \hat{k}_1 = 0. \end{aligned} \quad (66)$$

In both examples, it is assumed for simplicity that the  $\vec{N}_a$  have the same magnitude  $N_A$  for all  $a = 1, 2, 3$ .

Another comment should be added concerning statistical anisotropy in the model. Notice that both the bispectrum and the trispectrum can be written as the sum of a purely

isotropic and an anisotropic parts. The orders of magnitude of these two parts can, for instance, be read from Table 2 for the trispectrum: each one among  $\tau_{NL}^{NA_2}$ ,  $\tau_{NL}^{A_1}$  and  $\tau_{NL}^{A_2}$  provide the order of magnitude of the level of both their isotropic and anisotropic contributions, which are therefore comparable;  $\tau_{NL}^{NA_1}$  instead quantifies a purely anisotropic contribution which, as discussed in Section 5, can be comparable to the other three parts, if not the dominant one. A similar discussion applies to the bispectrum (see  $f_{NL}^A$  and  $f_{NL}^{NA}$  in Table 1). We can then conclude that for the three and for the four point function, there is room in the parameter space of the theory for the anisotropic contributions to be as large as, or even larger than, the isotropic ones.

## 7. Conclusions

Motivated by the interest in models that combine non-Gaussianity and statistical anisotropy predictions for the CMB fluctuations, we have considered models of inflation where primordial vector fields effectively participate in the production of the curvature perturbations  $\zeta$ . More specifically, we have reviewed the computation of the correlation functions up to fourth order, considering an  $SU(2)$  vector

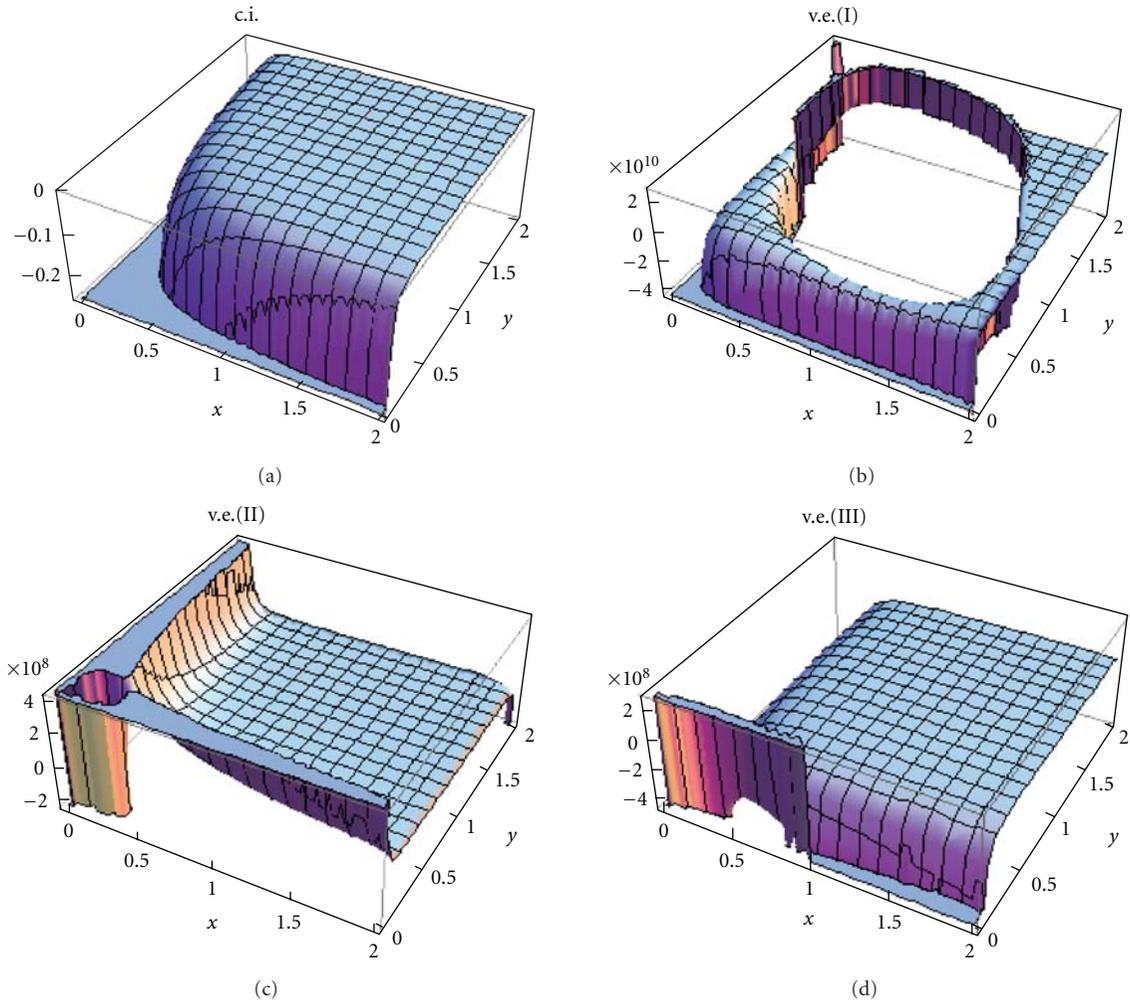


FIGURE 6: Plots of the contact interaction and of the vector-exchange contributions in the specialized planar configuration (minus sign).

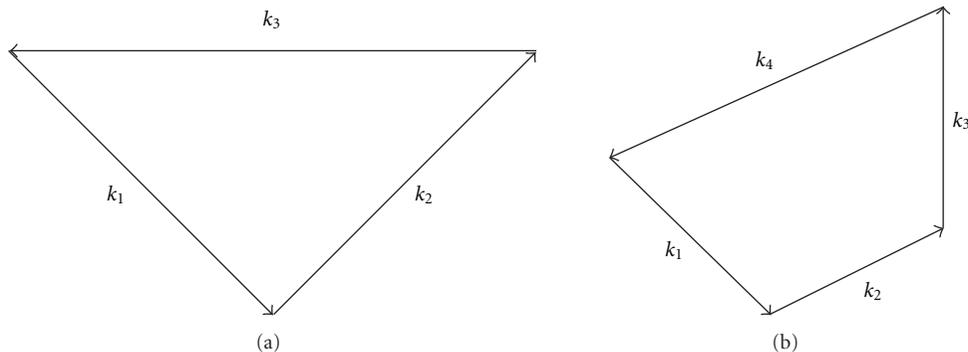


FIGURE 7: Representation of momentum conservation for the bispectrum (the three momenta form a closed triangle) and for the trispectrum (the momenta form a tetrahedron).

multiplet. The  $\delta N$  formalism was employed to express  $\zeta$  in terms of the quantum fluctuations of all the primordial fields. The Schwinger-Keldysh formula was also used in evaluating the correlators.

The correlation functions result as the sum of scalar and vector contributions. The latter are of two kinds, “Abelian”

(i.e., arising from the zeroth order terms in the Schwinger-Keldysh expansion) and “non-Abelian” (i.e., originating from the self-interactions of the vector fields). The bispectrum and the trispectrum final results are presented as a sum of products of isotropic functions of the momenta,  $F_n$  and  $G_n$  in (52) and (55), multiplied by anisotropy coefficients,  $I_n$

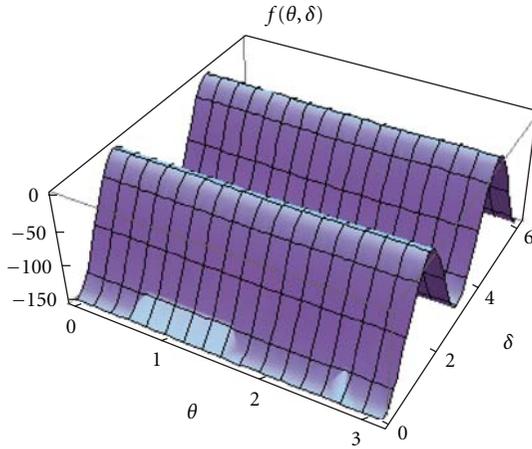


FIGURE 8: Plot of  $f(\theta, \delta) \equiv [(B_{\zeta}(\theta, \delta, x^*, x_2, x_3)x_2^2x_3^2k_1^6)/(g_c^2H^2m^2N_A^4)]$  evaluated at  $(x^* = 1, x_2 = 0.9, x_3 = 0.1)$  in a sample angular configuration. See [99, Appendix D] for its complete analytic expression.

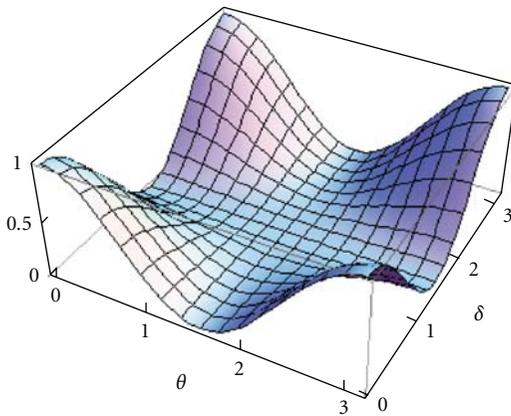


FIGURE 9: Plot of the anisotropic part of the trispectrum from the contribution due to vector-exchange diagrams in a sample angular configuration. See [100, Section 8] for its analytic expression.

and  $L_n$  in (52) and (55), which depend on the angles between the (gauge and wave) vectors.

The amplitude of non-Gaussianity has been presented through the parameters  $f_{NL}$  and  $\tau_{NL}$ ; in particular we have show the dependence of these functions from the nonangular parameters of the theory. We have provided the comparisons among the different (scalar versus vector, Abelian versus non-Abelian) contributions to  $f_{NL}$  and  $\tau_{NL}$ , noticing that any one of them can be the dominant contribution depending on the selected region of parameter space. In particular, we have stressed how the anisotropic contributions to the bispectrum and the trispectrum can overcome the isotropic parts. An interesting feature of these models is that the bispectrum and the trispectrum depend on the same set of parameters and their amplitudes are therefore strictly related to one another.

We have presented the shapes of both the bispectrum and the trispectrum. The isotropic functions appearing in their final expressions had been analyzed separately from their

anisotropy coefficients. The bispectrum isotropic functions had been found to preferably show a local shape. The trispectrum ones had been plotted selecting equilateral and specialized planar configurations. The full expressions (i.e., complete of their anisotropy coefficients) of bispectrum and trispectrum have been presented in specific momenta configuration, in order to provide a hint of the modulation of shapes and amplitudes operated by anisotropy.

We have reviewed old and recent vector field models, indicating both their limits and achievements. We would like to stress that, in our view, the most promising features of these models consists in the possibility of providing both non-Gaussianity and statistical anisotropy predictions that are related to one another because of the fact that they share the same underlying theory. This might, at some point in the future, become a great advantage: measurements of non-Gaussianity could be used to constrain statistical anisotropy or vice versa.

## Acknowledgments

This research has been partially supported by ASI Contract I/016/07/0 “COFIS”. A. Riotto acknowledges support by the EU Marie Curie Network UniverseNet (HPRNCT2006035863).

## References

- [1] D. H. Lyth and A. Riotto, “Particle physics models of inflation and the cosmological density perturbation,” *Physics Report*, vol. 314, no. 1-2, pp. 1–146, 1999.
- [2] G. F. Smoot, C. L. Bennett, A. Kogut, et al., “Structure in the COBE differential microwave radiometer first-year maps,” *Astrophysical Journal*, vol. 396, no. 1, part 2, pp. L1–L5, 1992.
- [3] C. L. Bennett, A. J. Banday, K. M. Górski, et al., “Four-year COBE 1 DMR cosmic microwave background observations: maps and basic results,” *Astrophysical Journal*, vol. 464, no. 1, Part 2, pp. L1–L4, 1996.
- [4] K. M. Górski, A. J. Banday, C. L. Bennett, et al., “Power spectrum of primordial inhomogeneity determined from the four-year COBE 1 DMR sky maps,” *Astrophysical Journal*, vol. 464, no. 1, pp. L11–L15, 1996.
- [5] D. N. Spergel, R. Bean, O. Doré, et al., “Three-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: implications for cosmology,” *Astrophysical Journal, Supplement Series*, vol. 170, no. 2, pp. 377–408, 2007.
- [6] E. Komatsu, J. Dunkley, M. R. Nolta, et al., “Five-year Wilkinson microwave anisotropy probe observations: cosmological interpretation,” *Astrophysical Journal*, vol. 180, no. 2, supplement, pp. 330–376, 2009.
- [7] N. Bartolo, E. Komatsu, S. Matarrese, and A. Riotto, “Non-Gaussianity from inflation: theory and observations,” *Physics Reports*, vol. 402, no. 3-4, pp. 103–266, 2004.
- [8] E. Komatsu, N. Afshordi, and N. Bartolo, “Non-Gaussianity as a probe of the physics of the primordial universe and the astrophysics of the low redshift universe,” preprint <http://arxiv.org/abs/0902.4759>.
- [9] V. Acquaviva, N. Bartolo, S. Matarrese, and A. Riotto, “Gauge-invariant second-order perturbations and non-Gaussianity from inflation,” *Nuclear Physics B*, vol. 667, no. 1-2, pp. 119–148, 2003.

- [10] J. Maldacena, “Non-gaussian features of primordial fluctuations in single field inflationary models,” *Journal of High Energy Physics*, vol. 7, no. 5, article 013, pp. 233–264, 2003.
- [11] D. Seery, J. E. Lidsey, and M. S. Sloth, “The inflationary trispectrum,” *Journal of Cosmology and Astroparticle Physics*, no. 1, article 027, 2007.
- [12] A. D. Linde, “Generation of isothermal density perturbations in the inflationary universe,” *Physics Letters B*, vol. 158, no. 5, pp. 375–380, 1985.
- [13] L. A. Kofman, “What initial perturbations may be generated in inflationary cosmological models,” *Physics Letters B*, vol. 173, no. 4, pp. 400–404, 1986.
- [14] D. Polarski and A. A. Starobinsky, “Isocurvature perturbations in multiple inflationary models,” *Physical Review D*, vol. 50, no. 10, pp. 6123–6129, 1994.
- [15] J. García-Bellido and D. Wands, “Metric perturbations in two-field inflation,” *Physical Review D*, vol. 53, no. 10, pp. 5437–5445, 1996.
- [16] V. F. Mukhanov and P. J. Steinhardt, “Density perturbations in multifield inflationary models,” *Physics Letters, Section B*, vol. 422, no. 1–4, pp. 52–60, 1998.
- [17] D. Langlois, “Correlated adiabatic and isocurvature perturbations from double inflation,” *Physical Review D*, vol. 59, no. 12, Article ID 123512, pp. 1–11, 1999.
- [18] C. Gordon, D. Wands, B. A. Bassett, and R. Maartens, “Adiabatic and entropy perturbations from inflation,” *Physical Review D*, vol. 63, no. 2, Article ID 023506, 11 pages, 2001.
- [19] S. Mollerach, “Isocurvature baryon perturbations and inflation,” *Physical Review D*, vol. 42, no. 2, pp. 313–325, 1990.
- [20] K. Enqvist and M. S. Sloth, “Adiabatic CMB perturbations in pre-Big-Bang string cosmology,” *Nuclear Physics B*, vol. 626, no. 1-2, pp. 395–409, 2002.
- [21] D. H. Lyth and D. Wands, “Generating the curvature perturbation without an inflaton,” *Physics Letters, Section B*, vol. 524, no. 1-2, pp. 5–14, 2002.
- [22] D. H. Lyth, C. Ungarelli, and D. Wands, “Primordial density perturbation in the curvaton scenario,” *Physical Review D*, vol. 67, no. 2, Article ID 023503, 2003.
- [23] T. Moroi and T. Takahashi, “Effects of cosmological moduli fields on cosmic microwave background,” *Physics Letters, Section B*, vol. 522, no. 3-4, pp. 215–221, 2001.
- [24] T. Moroi and T. Takahashi, “Erratum: Effects of cosmological moduli fields on cosmic microwave background,” *Physics Letters, Section B*, vol. 539, no. 3-4, p. 303, 2002.
- [25] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity in the curvaton scenario,” *Physical Review D*, vol. 69, no. 4, Article ID 043503, 2004.
- [26] J. Garriga and V. F. Mukhanov, “Perturbations in k-inflation,” *Physics Letters, Section B*, vol. 458, no. 2-3, pp. 219–225, 1999.
- [27] C. Armendáriz-Picón, T. Damour, and V. Mukhanov, “k-inflation,” *Physics Letters, Section B*, vol. 458, no. 2-3, pp. 209–218, 1999.
- [28] M. Alishahiha, E. Silverstein, and D. Tong, “DBI in the sky: non-Gaussianity from inflation with a speed limit,” *Physical Review D*, vol. 70, no. 12, Article ID 123505, 2004.
- [29] X. Chen, M.-X. Huang, S. Kachru, and G. Shiu, “Observational signatures and non-Gaussianities of general single-field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 1, article 002, 2007.
- [30] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, “Ghost inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 4, article 001, pp. 1–18, 2004.
- [31] D. H. Lyth and Y. Rodríguez, “Inflationary prediction for primordial non-gaussianity,” *Physical Review Letters*, vol. 95, no. 12, Article ID 121302, pp. 1–4, 2005.
- [32] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity from inflation,” *Physical Review D*, vol. 65, no. 10, Article ID 103505, 2002.
- [33] F. Vernizzi and D. Wands, “Non-Gaussianities in two-field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 5, article 019, 2006.
- [34] X. Chen and Y. Wang, “Quasi-single field inflation and non-Gaussianities,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 4, article 027, 2010.
- [35] N. Bartolo, S. Matarrese, and A. Riotto, “Non-Gaussianity of large-scale cosmic microwave background anisotropies beyond perturbation theory,” *Journal of Cosmology and Astroparticle Physics*, no. 8, article 010, pp. 177–196, 2005.
- [36] D. Seery and J. E. Lidsey, “Non-Gaussianity from the inflationary trispectrum,” *Journal of Cosmology and Astroparticle Physics*, no. 1, article 008, 2007.
- [37] C. T. Byrnes, M. Sasaki, and D. Wands, “Primordial trispectrum from inflation,” *Physical Review D*, vol. 74, no. 12, Article ID 123519, 2006.
- [38] D. Seery, M. S. Sloth, and F. Vernizzi, “Inflationary trispectrum from graviton exchange,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 3, article 018, 2009.
- [39] M. Sasaki, J. Väliiviita, and D. Wands, “Non-Gaussianity of the primordial perturbation in the curvaton model,” *Physical Review D*, vol. 74, no. 10, Article ID 103003, 2006.
- [40] X. Chen, M. X. Huang, and G. Shiu, “Inflationary trispectrum for models with large non-Gaussianities,” *Physical Review D*, vol. 74, no. 12, Article ID 121301, 5 pages, 2006.
- [41] F. Arroja and K. Koyama, “Non-Gaussianity from the trispectrum in general single field inflation,” *Physical Review D*, vol. 77, no. 8, Article ID 083517, 2008.
- [42] F. Arroja, S. Mizuno, K. Koyama, and T. Tanaka, “Full trispectrum in single field DBI inflation,” *Physical Review D*, vol. 80, no. 4, Article ID 043527, 2009.
- [43] X. Chen, B. Hu, M.-X. Huang, G. Shiu, and Y. Wang, “Large primordial trispectra in general single field inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 8, article 008, 2009.
- [44] X. Gao and B. Hu, “Primordial trispectrum from entropy perturbations in multifield DBI model,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 8, article 012, 2009.
- [45] S. Mizuno, F. Arroja, and K. Koyama, “Full quantum trispectrum in multifield DBI inflation,” *Physical Review D*, vol. 80, no. 8, Article ID 083517, 2009.
- [46] T. Okamoto and W. Hu, “Angular trispectra of CMB temperature and polarization,” *Physical Review D*, vol. 66, no. 6, Article ID 063008, 2002.
- [47] N. Kogo and E. Komatsu, “Angular trispectrum of CMB temperature anisotropy from primordial non-Gaussianity with the full radiation transfer function,” *Physical Review D*, vol. 73, no. 8, Article ID 083007, 2006.
- [48] J. Schwinger, “Brownian motion of a quantum oscillator,” *Journal of Mathematical Physics*, vol. 2, no. 3, pp. 407–432, 1961.
- [49] E. Calzetta and B. L. Hu, “Closed-time-path functional formalism in curved spacetime: application to cosmological back-reaction problems,” *Physical Review D*, vol. 35, no. 2, pp. 495–509, 1987.
- [50] R. D. Jordan, “Effective field equations for expectation values,” *Physical Review D*, vol. 33, no. 2, pp. 444–454, 1986.

- [51] S. Weinberg, “Quantum contributions to cosmological correlations,” *Physical Review D*, vol. 72, no. 4, Article ID 043514, pp. 1–19, 2005.
- [52] S. Weinberg, “Quantum contributions to cosmological correlations. II. Can these corrections become large?” *Physical Review D*, vol. 74, no. 2, Article ID 023508, 2006.
- [53] D. Seery, “One-loop corrections to a scalar field during inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 11, article 025, pp. 1–32, 2007.
- [54] D. Seery, “One-loop corrections to the curvature perturbation from inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 2, article 006, 2008.
- [55] E. Dimastrogiovanni and N. Bartolo, “One-loop graviton corrections to the curvature perturbation from inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 11, article 016, 2008.
- [56] K. M. Smith, L. Senatore, and M. Zaldarriaga, “Optimal limits on  $f_{NL}^{local}$  from WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 9 article 006, 2009.
- [57] L. Senatore, K. M. Smith, and M. Zaldarriaga, “Non-Gaussianities in single field inflation and their optimal limits from the WMAP 5-year data,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 1, article 028, 2010.
- [58] P. Vielva and J. L. Sanz, “Constraints on  $f_{nl}$  and  $g_{nl}$  from the analysis of the N-pdf of the CMB large scale anisotropies,” preprint, <http://arxiv.org/abs/0910.3196>.
- [59] J. Smidt, A. Amblard, A. Cooray, A. Heavens, D. Munshi, and P. Serra, “A measurement of cubic-order primordial non-Gaussianity ( $g_{NL}$ ) with WMAP 5-year data,” preprint, <http://arxiv.org/abs/1001.5026>.
- [60] V. Desjacques and U. Seljak, “Signature of primordial non-Gaussianity of 3 type in the mass function and bias of dark matter haloes,” *Physical Review D*, vol. 81, no. 2, Article ID 023006, 2010.
- [61] <http://planck.esa.int/>.
- [62] A. De Oliveira-Costa, M. Tegmark, M. Zaldarriaga, and A. Hamilton, “Significance of the largest scale CMB fluctuations in WMAP,” *Physical Review D*, vol. 69, no. 6, Article ID 063516, 2004.
- [63] P. Vielva, E. Martínez-González, R. B. Barreiro, J. L. Sanz, and L. Cayón, “Detection of non-gaussianity in the Wilkinson microwave anisotropy probe first-year data using spherical wavelets,” *Astrophysical Journal*, vol. 609, no. 1, pp. 22–34, 2004.
- [64] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Górski, and P. B. Lilje, “Asymmetries in the cosmic microwave background anisotropy field,” *Astrophysical Journal*, vol. 605, no. 1, pp. 14–20, 2004.
- [65] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Górski, and P. B. Lilje, “Erratum: Asymmetries in the cosmic microwave background anisotropy field,” *Astrophysical Journal*, vol. 609, no. 2, pp. 1198–1199, 2004.
- [66] D. N. Spergel, L. Verde, H. V. Peiris, et al., “First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: determination of cosmological parameters,” *Astrophysical Journal, Supplement Series*, vol. 148, no. 1, pp. 175–194, 2003.
- [67] G. Efstathiou, “A maximum likelihood analysis of the low cosmic microwave background multipoles from the Wilkinson Microwave Anisotropy Probe,” *Monthly Notices of the Royal Astronomical Society*, vol. 348, no. 3, pp. 885–896, 2004.
- [68] K. Land and J. Magueijo, “Examination of evidence for a preferred axis in the cosmic radiation anisotropy,” *Physical Review Letters*, vol. 95, no. 7, Article ID 071301, pp. 1–4, 2005.
- [69] M. Cruz, L. Cayón, E. Martínez-González, P. Vielva, and J. Jin, “The non-gaussian cold spot in the 3 year Wilkinson microwave anisotropy probe data,” *Astrophysical Journal*, vol. 655, no. 1, pp. 11–20, 2007.
- [70] F. K. Hansen, P. Cabella, D. Marinucci, and N. Vittorio, “Asymmetries in the local curvature of the wilkinson microwave anisotropy probe data,” *Astrophysical Journal*, vol. 607, no. 2, pp. L67–L70, 2004.
- [71] F. K. Hansen, A. J. Banday, and K. M. Górski, “Testing the cosmological principle of isotropy: local power-spectrum estimates of the WMAP data,” *Monthly Notices of the Royal Astronomical Society*, vol. 354, no. 3, pp. 641–665, 2004.
- [72] D. Pietrobon, P. Cabella, A. Balbi, R. Crittenden, G. de Gasperis, and N. Vittorio, “Needlet bispectrum asymmetries in the WMAP 5-year data,” preprint, <http://arxiv.org/abs/0905.3702>.
- [73] N. E. Groeneboom, L. Ackerman, I. K. Wehus, and H. K. Eriksen, “Bayesian analysis of an anisotropic universe model: systematics and polarization,” preprint, <http://arxiv.org/abs/0905.3702>.
- [74] D. Hanson and A. Lewis, “Estimators for CMB statistical anisotropy,” *Physical Review D*, vol. 80, no. 6, Article ID 063004, 15 pages, 2009.
- [75] R. M. Wald, “Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant,” *Physical Review D*, vol. 28, no. 8, pp. 2118–2120, 1983.
- [76] J. D. Barrow and S. Hervik, “Anisotropically inflating universes,” *Physical Review D*, vol. 73, no. 2, Article ID 023007, 2006.
- [77] J. D. Barrow and S. Hervik, “Evolution of universes in quadratic theories of gravity,” *Physical Review D*, vol. 74, no. 12, Article ID 124017, 2006.
- [78] E. Di Grezia, G. Esposito, A. Funel, G. Mangano, and G. Miele, “Spacetime noncommutativity and antisymmetric tensor dynamics in the early Universe,” *Physical Review D*, vol. 68, no. 10, Article ID 105012, pp. 1–10, 2003.
- [79] T. S. Pereira, C. Pitrou, and J.-P. Uzan, “Theory of cosmological perturbations in an anisotropic universe,” *Journal of Cosmology and Astroparticle Physics*, no. 9, article 006, 2007.
- [80] C. Pitrou, T. S. Pereira, and J. P. Uza, “Predictions from an anisotropic inflationary era,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, article 004, 2008.
- [81] E. Dimastrogiovanni, W. Fischler, and S. Paban, “Perturbation growth in anisotropic cosmologies,” *Journal of High Energy Physics*, vol. 2008, no. 7, article 045, 2008.
- [82] A. E. Gumrukcuoglu, L. Kofman, and M. Peloso, “Gravity waves signatures from anisotropic preinflation,” *Physical Review D*, vol. 78, no. 10, Article ID 103525, 2008.
- [83] C. Armendáriz-Picón and P. B. Greene, “Spinors, inflation, and non-singular cyclic cosmologies,” *General Relativity and Gravitation*, vol. 35, no. 9, pp. 1637–1658, 2003.
- [84] C. G. Böhrer and D. F. Mota, “CMB anisotropies and inflation from non-standard spinors,” *Physics Letters, Section B*, vol. 663, no. 3, pp. 168–171, 2008.
- [85] T. Watanabe, “Dirac-field model of inflation in Einstein-Cartan theory,” <http://arxiv.org/abs/0902.1392>.
- [86] G. De Berredo-Peixoto and E. A. De Freitas, “On the torsion effects of a relativistic spin fluid in early cosmology,” *Classical and Quantum Gravity*, vol. 26, no. 17, Article ID 175015, 2009.

- [87] C. Germani and A. Kehagias, “P-nflation: generating cosmic Inflation with p-forms,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 3, article 028, 2009.
- [88] T. Kobayashi and S. Yokoyama, “Gravitational waves from p-form inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 004, 2009.
- [89] T. S. Koivisto and N. J. Nunes, “Three-form cosmology,” *Physics Letters, Section B*, vol. 685, no. 2-3, pp. 105–109, 2010.
- [90] C. Germani and A. Kehagias, “Scalar perturbations in p-nflation: the 3-form case,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 11, article 005, 2009.
- [91] T. S. Koivisto and N. J. Nunes, “Inflation and dark energy from three-forms,” *Physical Review D*, vol. 80, no. 10, Article ID 103509, 18 pages, 2009.
- [92] T. S. Koivisto, D. F. Mota, and C. Pitrou, “Inflation from N-forms and its stability,” *Journal of High Energy Physics*, vol. 2009, no. 9, article 092, 2009.
- [93] N. Mandolesi, M. Bersanelli, and R.C. Butler, “Planck pre-launch status: the Planck-LFI programme,” preprint, <http://arxiv.org/abs/1001.2657>.
- [94] S. Yokoyama and J. Soda, “Primordial statistical anisotropy generated at the end of inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 005, 2008.
- [95] C.A. Valenzuela-Toledo, Y. Rodríguez, and D. H. Lyth, “Non-Gaussianity at tree and one-loop levels from vector field perturbations,” *Physical Review D*, vol. 80, no. 10, Article ID 103519, 7 pages, 2009.
- [96] K. Dimopoulos, M. Karčiauskas, D. H. Lyth, and Y. Rodríguez, “Statistical anisotropy of the curvature perturbation from vector field perturbations,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 5, article 013, 2009.
- [97] C.A. Valenzuela-Toledo and Y. Rodríguez, “Non-gaussianity from the trispectrum and vector field perturbations,” *Physics Letters, Section B*, vol. 685, no. 2-3, pp. 120–127, 2010.
- [98] M. Karčiauskas, K. Dimopoulos, and D. H. Lyth, “Anisotropic non-Gaussianity from vector field perturbations,” *Physical Review D*, vol. 80, no. 2, Article ID 023509, 2009.
- [99] K. Dimopoulos, M. Karčiauskas, and J. M. Wagstaff, “Vector curvaton with varying kinetic function,” *Physical Review D*, vol. 81, no. 2, Article ID 023522, 2010.
- [100] K. Dimopoulos, M. Karčiauskas, and J. M. Wagstaff, “Vector curvaton without instabilities,” *Physics Letters, Section B*, vol. 683, no. 4-5, pp. 298–301, 2010.
- [101] N. Bartolo, E. Dimastrogiovanni, S. Matarrese, and A. Riotto, “Anisotropic bispectrum of curvature perturbations from primordial non-abelian vector fields,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 10, article 015, 2009.
- [102] N. Bartolo, E. Dimastrogiovanni, S. Matarrese, and A. Riotto, “Anisotropic trispectrum of curvature perturbations induced by primordial non-Abelian vector fields,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 11, article 028, 2009.
- [103] L. H. Ford, “Inflation driven by a vector field,” *Physical Review D*, vol. 40, no. 4, pp. 967–972, 1989.
- [104] A. Golovnev, V. Mukhanov, and V. Vanchurin, “Vector inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 6, article 009, 2008.
- [105] A. Golovnev, V. Mukhanov, and V. Vanchurin, “Gravitational waves in vector inflation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 11, article 018, 2008.
- [106] A. Golovnev and V. Vanchurin, “Cosmological perturbations from vector inflation,” *Physical Review D*, vol. 79, no. 10, Article ID 103524, 2009.
- [107] C. Armendariz-Picon, “Could dark energy be vector-like?” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, article 007, 2004.
- [108] C. G. Böhrer and T. Harko, “Dark energy as a massive vector field,” *European Physical Journal C*, vol. 50, no. 2, pp. 423–429, 2007.
- [109] T. Koivisto and D. F. Mota, “Accelerating cosmologies with an anisotropic equation of state,” *Astrophysical Journal*, vol. 679, no. 1, pp. 1–5, 2008.
- [110] T. Koivisto and D. F. Mota, “Vector field models of inflation and dark energy,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 021, 2008.
- [111] J. B. Jiménez, R. Lazkoz, and A. L. Maroto, “Cosmic vector for dark energy: constraints from supernovae, cosmic microwave background, and baryon acoustic oscillations,” *Physical Review D*, vol. 80, no. 2, Article ID 023004, 2009.
- [112] B. Himmetoglu, C. R. Contaldi, and M. Peloso, “Instability of anisotropic cosmological solutions supported by vector fields,” *Physical Review Letters*, vol. 102, no. 11, Article ID 111301, 2009.
- [113] B. Himmetoglu, C. R. Contaldi, and M. Peloso, “Instability of the Ackerman-Carroll-Wise model, and problems with massive vectors during inflation,” *Physical Review D*, vol. 79, no. 6, Article ID 063517, 19 pages, 2009.
- [114] B. Himmetoglu, C. R. Contaldi, and M. Peloso, “Ghost instabilities of cosmological models with vector fields non-minimally coupled to the curvature,” *Physical Review D*, vol. 80, no. 12, Article ID 123530, 2009.
- [115] K. Dimopoulos, “Can a vector field be responsible for the curvature perturbation in the Universe?” *Physical Review D*, vol. 74, no. 8, Article ID 083502, 2006.
- [116] K. Dimopoulos and M. Karčiauskas, “Non-minimally coupled vector curvaton,” *Journal of High Energy Physics*, vol. 2008, no. 7, article 119, 2008.
- [117] S. Kanno, M. Kimura, J. Soda, and S. Yokoyama, “Anisotropic inflation from vector impurity,” *Journal of Cosmology and Astroparticle Physics*, vol. 2008, no. 8, article 034, 2008.
- [118] L. Ackerman, S. M. Carroll, and M. B. Wise, “Imprints of a primordial preferred direction on the microwave background,” *Physical Review D*, vol. 75, no. 18, Article ID 083502, 7 pages, 2007.
- [119] L. Ackerman, S. M. Carroll, and M. B. Wise, “Erratum: Imprints of a primordial preferred direction on the microwave background,” *Physical Review D*, vol. 80, no. 6, Article ID 069901, 1 pages, 2009.
- [120] S. M. Carroll, T. R. Dulaney, M. I. Gresham, and H. Tam, “Instabilities in the aether,” *Physical Review D*, vol. 79, no. 6, Article ID 065011, 2009.
- [121] D. H. Lyth, “Generating the curvature perturbation at the end of inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 11, article 006, pp. 111–120, 2005.
- [122] L. Alabidi and D. Lyth, “Curvature perturbation from symmetry breaking the end of inflation,” *Journal of Cosmology and Astroparticle Physics*, no. 8, article 006, 2006.
- [123] M. P. Salem, “Generation of density perturbations at the end of inflation,” *Physical Review D*, vol. 72, no. 12, Article ID 123516, pp. 1–7, 2005.
- [124] F. Bernardeau, L. Kofman, and J.-P. Uzan, “Modulated fluctuations from hybrid inflation,” *Physical Review D*, vol. 70, no. 8, Article ID 083004, 2004.
- [125] M.-A. Watanabe, S. Kanno, and J. Soda, “Inflationary universe with anisotropic hair,” *Physical Review Letters*, vol. 102, no. 19, Article ID 191302, 2009.

- [126] B. Himmetoglu, "Spectrum of perturbations in anisotropic inflationary universe with vector hair," *Journal of Cosmology and Astroparticle Physics*, vol. 2010, no. 3, article 023, 2010.
- [127] T. R. Dulaney and M. I. Gresham, "Primordial Power Spectra from Anisotropic Inflation," *Physical Review D*, vol. 81, no. 10, Article ID 103532, 20 pages, 2010.
- [128] A. A. Starobinsky, "Multicomponent de Sitter (inflationary) stages and the generation of perturbations," *JETP Letters*, vol. 42, no. 3, pp. 152–155, 1985, *Pis'ma v ZhETF*, vol. 42, no. 3, 124–127, 1985.
- [129] M. Sasaki and E. D. Stewart, "A general analytic formula for the spectral index of the density perturbations produced during inflation," *Progress of Theoretical Physics*, vol. 95, no. 1, pp. 71–78, 1996.
- [130] D. H. Lyth, K. A. Malik, and M. Sasaki, "A general proof of the conservation of the curvature perturbation," *Journal of Cosmology and Astroparticle Physics*, no. 5, pp. 50–69, 2005.
- [131] D. Seery and J. E. Lidsey, "Primordial non-Gaussianities in single-field inflation," *Journal of Cosmology and Astroparticle Physics*, no. 6, pp. 45–70, 2005.