Hunting for primordial non-Gaussianity in the cosmic microwave background

Eiichiro Komatsu
Texas Cosmology Center and Department of Astronomy, The University of Texas, Austin, TX 78712, USA
E-mail: komatsu@astro.as.utexas.edu

Received 1 April 2010, in final form 24 April 2010
Published 27 May 2010
Online at stacks.iop.org/CQG/27/124010

Abstract
Since the first limit on the (local) primordial non-Gaussianity parameter, \( f_{NL} \), was obtained from the Cosmic Background Explorer (COBE) data in 2002, observations of the cosmic microwave background (CMB) have been playing a central role in constraining the amplitudes of various forms of non-Gaussianity in primordial fluctuations. The current 68% limit from the 7-year data of the Wilkinson Microwave Anisotropy Probe (WMAP) is \( f_{NL} = 32 \pm 21 \), and the Planck satellite is expected to reduce the uncertainty by a factor of 4 in a few years from now. If \( f_{NL} \gg 1 \) is found by Planck with high statistical significance, all single-field models of inflation would be ruled out. Moreover, if the Planck satellite finds \( f_{NL} \sim 30 \), then it would be able to test a broad class of multi-field models using the 4-point function (trispectrum) test of \( \tau_{NL} \geq \left( \frac{6 f_{NL}}{5} \right)^2 \). In this paper, we review the methods (optimal estimator), results (WMAP 7-year) and challenges (secondary anisotropy, second-order effect and foreground) of measuring primordial non-Gaussianity from the CMB data, present a science case for the trispectrum and conclude with future prospects.

PACS numbers: 98.80.−k, 98.80.Es
(Some figures in this article are in colour only in the electronic version)

1. Introduction
The physics of the very early, primordial universe is best probed by measurements of statistical properties of primordial fluctuations. The primordial fluctuations are the seeds for the temperature and polarization anisotropies of the cosmic microwave background (CMB) and the large-scale structure of the universe that we observe today. Therefore, both the CMB and the large-scale structure are excellent probes of the primordial fluctuations. In this paper, we shall focus on the CMB. See the article by Desjacques and Seljak in this issue [1] for the corresponding review on the large-scale structure as a probe of the primordial fluctuations.
This paper reviews a recent progress on using the CMB as a probe of a particular statistical aspect of primordial fluctuations called non-Gaussianity. Reviews on this subject were written in 2001 [2] and 2004 [3]. The former review would be most useful for those who are new to this subject.

In this paper, we focus on the new discoveries that have been made since 2004. Particularly notable ones include the following.

1. It has been proven that all inflation models (not just simple ones [4, 5]) based upon a single scalar field would be ruled out regardless of the details of models [6], if the primordial non-Gaussianity parameter called $f_{NL}$ (more precisely, the ‘local type’ $f_{NL}$ as described later) is found to be much greater than unity.

2. The optimal method for extracting $f_{NL}$ from the CMB data has been developed [7–10] and implemented [11]. The latest limit on the local-type $f_{NL}$ from the WMAP 7-year temperature data is $f_{NL} = 32 \pm 21$ (68% CL) [12].

3. The most serious contamination of the local-type $f_{NL}$ due to the secondary CMB anisotropy, the coupling between the integrated Sachs–Wolfe (ISW) effect and the weak gravitational lensing has been identified [9, 13–17]. However, note that the astrophysical contamination such as the galactic foreground emission and radio point sources may still be the most serious contaminant of $f_{NL}$. These effects would pose a serious analysis challenge to measuring $f_{NL}$ from the Planck data.

4. The importance of distinguishing different triangle configurations of the 3-point function of the CMB was realized [18, 19] and has been fully appreciated. It has been shown by many researchers that different configurations probe distinctly different aspects of the physics of the primordial universe. The list of possibilities is long, and a terribly incomplete list of references on recent work (since $\sim 2004$) is as follows: [20] on a general analysis of various shapes; [21–26] on the local shape ($k_3 \ll k_1 \approx k_2$); [18, 27–29] on the equilateral shape ($k_1 \approx k_2 \approx k_3$); [30, 31] on the flattened (or folded) shape ($k_1 \approx 2k_2 \approx 2k_3$); [32, 33] on the orthogonal shape (which is nearly orthogonal to both local and equilateral shapes); [34–38] on combinations of different shapes and [39, 40] on oscillating bispectra. Also see references therein.

5. The connected 4-point function of primordial fluctuations has been shown to be an equally powerful probe of the physics of the primordial universe. In particular, a combination of the 3- and 4-point functions may allow us to further distinguish different scenarios. Many papers have been written on this subject over the last few years: [41–44] on single-field models; [36, 45–70] on multi-field models and [71, 72] on isocurvature perturbations. CMB data are expected to provide useful limits on the parameters of the ‘local-form trispectrum’, $\tau_{NL}$ and $g_{NL}$ [73, 74]. Preliminary limits on these parameters have been obtained from the WMAP data [75, 76].

The number of researchers working on primordial non-Gaussianity has increased dramatically: Science White Paper on non-Gaussianity submitted to Decadal Survey Astro2010 was co-signed by 61 scientists [77].

### 2. Gaussian versus non-Gaussian CMB anisotropy

#### 2.1. What do we mean by ‘Gaussianity’?

What do we mean by ‘Gaussian fluctuations’? Let us consider the distribution of temperature anisotropy of the CMB that we observe on the sky, $\Delta T(\hat{n})$. The temperature anisotropy is

$$
\Delta T(\hat{n}) = \frac{T(\hat{n}) - T_0}{T_0}
$$

where $T_0$ is the mean temperature of the CMB. If the fluctuations are Gaussian, the distribution of $\Delta T(\hat{n})$ is a normal (Gaussian) distribution with mean zero and some variance.

However, if the fluctuations are not Gaussian, the distribution of $\Delta T(\hat{n})$ will depart from a normal distribution. For example, if the fluctuations are non-Gaussian, there may be higher-order moments such as skewness and kurtosis. These higher-order moments provide a way to quantify the departure from Gaussianity.

The most commonly used measure of non-Gaussianity is the non-Gaussianity parameter $f_{NL}$.

$$
f_{NL} = \frac{\langle (\Delta T)^3 \rangle}{\langle (\Delta T)^2 \rangle^{3/2}}
$$

where $\langle \cdot \rangle$ denotes the statistical average over the sky.

In practice, $f_{NL}$ is often expressed in terms of other parameters, such as the local-type non-Gaussianity parameter $f_{NL}^{\text{local}}$.

$$
f_{NL}^{\text{local}} = f_{NL} \left( \frac{k_3}{k_1} \right)^2
$$

where $k_1$, $k_2$, and $k_3$ are the wave numbers of the local, equilateral, and flattened shapes, respectively.

The value of $f_{NL}$ can be determined from the CMB data. The current limit on $f_{NL}$ from the Planck data is $f_{NL} = 32 \pm 21$ (68% CL) [12].
Gaussian when its probability density function (PDF) is given by

\[
P(\Delta T) = \frac{1}{(2\pi)^{N_{\text{pix}}/2}|\xi|^{1/2}} \exp \left[ -\frac{1}{2} \sum_{ij} \Delta T_i (\xi^{-1})_{ij} \Delta T_j \right],
\]

where \(\Delta T_i \equiv T(\hat{n}_i)\), \(\xi_{ij} \equiv \langle \Delta T_i \Delta T_j \rangle\) is the covariance matrix (or the 2-point correlation function) of the temperature anisotropy, \(|\xi|\) is the determinant of the covariance matrix and \(N_{\text{pix}}\) is the number of pixels on the sky.

We often work in harmonic space by expanding \(\Delta T\) using spherical harmonics: \(\Delta T(\hat{n}) = \sum_{lm} a_{lm} Y_{lm}(\hat{n})\). The PDF for \(a_{lm}\) is given by

\[
P(a) = \frac{1}{(2\pi)^{N_{\text{harm}}/2}|C|^{1/2}} \exp \left[ -\frac{1}{2} \sum_{lm} \sum_{l'm'} a_{lm}^* (C^{-1})_{lm,l'm'} a_{l'm'} \right],
\]

where \(C_{lm,l'm'} \equiv \langle a_{lm}^* a_{l'm'} \rangle\), and \(N_{\text{harm}}\) is the number of \(l\) and \(m\). When \(a_{lm}\) is statistically homogeneous and isotropic (which is not always the case because of, e.g., non-uniform noise), one finds \(C_{lm,l'm'} = C \delta_{ll'} \delta_{mm'}\), and thus the PDF simplifies to

\[
P(a) = \prod_{lm} \frac{e^{-|a_{lm}|^2/(2C_l)}}{\sqrt{2\pi C_l}}.
\]

Here, \(C_l\) is the angular power spectrum. The latest determination of \(C_l\) of the CMB temperature anisotropy is shown in figure 1.

The important property of a Gaussian distribution is that the PDF is fully specified by the covariance matrix. In other words, the covariance matrix contains all the information on statistical properties of Gaussian fluctuations. When the PDF is given by equation (3), the power spectrum \(C_l\) contains all the information on \(a_{lm}\). This is not true for non-Gaussian fluctuations, for which one needs information on higher-order correlation functions.

Let us conclude this subsection by noting that a non-zero deviation of the covariance matrix from the diagonal form, \(\Delta C_{lm,l'm'} \equiv C_{lm,l'm'} - C \delta_{ll'} \delta_{mm'}\), does not imply non-Gaussianity: the PDF can be a Gaussian with a non-diagonal covariance matrix as given in equation (2).
A non-zero $\Delta C_{lm,l'}$ may arise in cosmological models that violate statistical isotropy. Such models may yield anisotropic Gaussian fluctuations; thus, one must distinguish between non-Gaussianity and a violation of statistical isotropy.

2.2. What do we mean by ‘non-Gaussianity’?

What do we mean by ‘non-Gaussian fluctuations’? Any deviation from a Gaussian distribution (such as equations (1) or (2)) is called non-Gaussianity. When fluctuations in the CMB are non-Gaussian, one cannot generally write down its PDF, unless one considers certain models (e.g. inflation). Nevertheless, when non-Gaussianity is weak, one may expand the PDF around a Gaussian distribution [81] and obtain

$$P(a) = \left[ 1 - \frac{1}{6} \sum_{\text{all}l,m} \left( a_{l_1m_1} a_{l_2m_2} a_{l_3m_3} \right) \frac{\partial}{\partial a_{l_1m_1}} \frac{\partial}{\partial a_{l_2m_2}} \frac{\partial}{\partial a_{l_3m_3}} \right] \times \frac{e^{-\frac{1}{2} \sum_{lm} \sum_{l'm'} a_{lm}^*(n^{C^{-1}})_{lm,l'm'} a_{l'm'}}}{(2\pi)^{N_{\text{harm}}/2} |C|^{1/2}}. \quad (4)$$

Here, the expansion is truncated at the 3-point function (bispectrum) of $a_{lm}$, and thus we have assumed that the connected 4-point and higher-order correlation functions are negligible compared to the power spectrum and bispectrum. (This condition is not always satisfied.) By evaluating the above derivatives, one obtains

$$P(a) = \frac{1}{(2\pi)^{N_{\text{harm}}/2} |C|^{1/2}} \exp \left[ -\frac{1}{2} \sum_{lm} \sum_{l'm'} a_{lm}^*(n^{C^{-1}})_{lm,l'm'} a_{l'm'} \right] \times \left\{ 1 + \frac{1}{6} \sum_{\text{all}l,m} \left( a_{l_1m_1} a_{l_2m_2} a_{l_3m_3} \right) \left[ (n^{C^{-1}})_{l_1m_1} (n^{C^{-1}})_{l_2m_2} (n^{C^{-1}})_{l_3m_3} \right] - 3 (n^{C^{-1}})_{l_1m_1}(n^{C^{-1}})_{l_2m_2}(n^{C^{-1}})_{l_3m_3} \right\}. \quad (5)$$

This formula is useful, as it tells us how to estimate the angular bispectrum, $\langle a_{l_1m_1} a_{l_2m_2} a_{l_3m_3} \rangle$, optimally from a given data by maximizing this PDF. In practice, we usually parametrize the bispectrum using a few parameters (e.g. $f_{NL}$) and estimate those parameters from the data by maximizing the PDF with respect to the parameters.

In the limit that the contribution of the connected 4-point function (trispectrum) to the PDF is negligible compared to those of the power spectrum and bispectrum, equation (5) contains all the information on non-Gaussian fluctuations characterized by the covariance matrix, $C_{l_1m_1,l_2m_2} = \langle a_{l_1m_1} a_{l_2m_2} \rangle$, and the angular bispectrum, $\langle a_{l_1m_1} a_{l_2m_2} a_{l_3m_3} \rangle$. This approach can be extended straightforwardly to the trispectrum if necessary.

3. Extracting $f_{NL}$ from the CMB data

3.1. General formula

We have not defined what we mean by ‘$f_{NL}$’. For the moment, let us loosely define it as the amplitude of a certain shape of the angular bispectrum:

$$\langle a_{l_1m_1} a_{l_2m_2} a_{l_3m_3} \rangle = G_{l_1l_2l_3}^{m_1m_2m_3} \sum_i f_{NL}^{(i)} h_{l_1l_2l_3}^{(i)}, \quad (6)$$

1 Babich [82] derived this formula for $C_{lm,l'm'} = C_{\delta l'm'}$. 

4
where the function, \( b_{l_1l_2l_3}^{(i)} \), is called the ‘reduced angular bispectrum’ \([83]\) and defines the shape of the angular bispectrum for a given model denoted by an index \( i \) (which may refer to, e.g., ‘local,’ ‘equilateral,’ ‘orthogonal,’ etc), and \( G_{l_1l_2l_3}^{m_1m_2m_3} \) is the so-called Gaunt integral, defined by

\[
G_{l_1l_2l_3}^{m_1m_2m_3} \equiv \int d^2\hat{n} Y_{l_1m_1}(\hat{n}) Y_{l_2m_2}(\hat{n}) Y_{l_3m_3}(\hat{n}).
\]  

(7)

The physical role of the Gaunt integral is to assure that \( \Delta \) is valid for general forms of \( \Delta \). Here, the summation over \( m \) can be done using the fast Fourier transform (FFT) as \( Y_{l,m}(\theta, \phi) \propto e^{i m \phi} \). This technique is used by the HEALPix package \([85]\), and thus one may use HEALPix to do this summation. To compute \( d_l \), one may use Monte Carlo simulations. Namely, as \( C_{l_1l_2l_3} = \langle a_{l_1}^* a_{l_2} a_{l_3} \rangle \), we have the exact relation between \( d_l \) and \( e_l \): \( d_l(\hat{n}) = \langle e_l(\hat{n}) e_l(\hat{n}) \rangle_{\text{MC}} \). One can evaluate the ensemble average using the Monte Carlo simulation of the CMB and the instrumental noise. Let us denote this operation by \( d_l(\hat{n}) \). The final formula for \( S_i \) is

\[
S_i = \frac{1}{6} \int d^2\hat{n} \sum_{l_1l_2l_3} b_{l_1l_2l_3}^{(i)} [e_{l_1l_2l_3}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,1l_2}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,2l_3}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,3l_1}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,1l_3}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,1l_3}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,1l_3}(\hat{n})],
\]  

(11)

where

\[
e_{l}(\hat{n}) \equiv \sum_{m} (C_{l}^{-1})_{l,m} Y_{l,m}(\hat{n}),
\]  

(12)

\[
d_{l}(\hat{n}) \equiv \sum_{m m'} (C_{l}^{-1})_{l,m} Y_{l,m}(\hat{n}) Y_{l,m'}(\hat{n}).
\]  

(13)

Here, the summation over \( m \) can be done using the fast Fourier transform (FFT) as \( Y_{l,m}(\theta, \phi) \propto e^{i m \phi} \). This technique is used by the HEALPix package \([85]\), and thus one may use HEALPix to do this summation. To compute \( d_l(\hat{n}) \), one may use Monte Carlo simulations. Namely, as \( C_{l_1l_2l_3} = \langle a_{l_1}^* a_{l_2} a_{l_3} \rangle \), we have the exact relation between \( d_l(\hat{n}) \) and \( e_l \): \( d_l(\hat{n}) = \langle e_l(\hat{n}) e_l(\hat{n}) \rangle_{\text{MC}} \). One can evaluate the ensemble average using the Monte Carlo simulation of the CMB and the instrumental noise. Let us denote this operation by \( d_l(\hat{n}) \). The final formula for \( S_i \) is

\[
S_i = \frac{1}{6} \int d^2\hat{n} b_{l_1l_2l_3}^{(i)} [e_{l_1l_2l_3}(\hat{n}) e_{l_1l_2l_3}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,1l_2}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,2l_3}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,3l_1}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,1l_3}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,1l_3}(\hat{n}) - 3d_{l_1l_2l_3}(\hat{n}) e_{l,1l_3}(\hat{n})],
\]  

(14)

which is valid for general forms of \( b_{l_1l_2l_3}^{(i)} \). Note that the integral, \( \int d^2\hat{n} \), must be done over the full sky, even in the presence of the mask: the information on the mask is included in
the calculation of the Fisher matrix, $F_{ij}$. The only assumptions that we have made so far are as follows: (1) each angular bispectrum component has only one free parameter, i.e. the amplitude, and (2) non-Gaussianity (if any) is weak, and the PDF of $a_{lm}$ is given by equation (5).

Finally, the explicit form of the Fisher matrix is given by

$$F_{ij} = \frac{f_{\text{sky}}}{6} \sum_{\ell_1 \ell_2 \ell_3} \sum_{m_1 m_2 m_3} g_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3} b_{\ell_1 \ell_2 \ell_3}^{(i)} \times (C^{-1})_{\ell_1 \ell_2 \ell_3} (C^{-1})_{\ell_2 \ell_3 \ell_3} (C^{-1})_{\ell_3 \ell_3 \ell_3} b_{\ell_1 \ell_2 \ell_3}^{(j)} g_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3},$$

(15)

where $f_{\text{sky}}$ is the fraction of the sky outside of the mask. When the covariance matrix is diagonal, the expression simplifies to

$$F_{ij} = \frac{f_{\text{sky}}}{6} \sum_{\ell_1 \ell_2 \ell_3} I_{\ell_1 \ell_2 \ell_3} b_{\ell_1 \ell_2 \ell_3}^{(i)} b_{\ell_1 \ell_2 \ell_3}^{(j)} C_{\ell_1} C_{\ell_2} C_{\ell_3},$$

(16)

where

$$I_{\ell_1 \ell_2 \ell_3} = \frac{(g_{\ell_1 \ell_2 \ell_3}^{m_1 m_2 m_3})^2}{4\pi} \left( \frac{2l_1 + 1}{2l_2 + 1} \right) \left( \frac{2l_2 + 1}{2l_3 + 1} \right) \left( \begin{array}{ccc} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{array} \right)^2.$$

(17)

Equation (16) may also be written as

$$F_{ij} = f_{\text{sky}} \sum_{\ell_1 \leq \ell_2 \leq \ell_3} I_{\ell_1 \ell_2 \ell_3} b_{\ell_1 \ell_2 \ell_3}^{(i)} b_{\ell_1 \ell_2 \ell_3}^{(j)} C_{\ell_1} C_{\ell_2} C_{\ell_3} \Delta_{\ell_1 \ell_2 \ell_3},$$

(18)

where $\Delta_{\ell_1 \ell_2 \ell_3} = 1, 2$ and 6 when all of $\ell_i$’s are different, two of $\ell_i$’s are the same and all of $\ell_i$’s are the same, respectively.

### 3.2. Poisson bispectrum

As a warm up, let us consider the simplest example: point sources randomly distributed over the sky. As already mentioned, this is a contamination of the primordial non-Gaussianity parameters, and thus an accurate measurement of this component is quite important, especially for Planck as well as for the South Pole Telescope (SPT) and Atacama Cosmology Telescope (ACT), which are working at high frequencies ($\nu > 100$ GHz) where star-forming galaxies dominate $a_{lm}$ at $l > 1000$.

For the Poisson distribution, the reduced bispectrum is independent of multipoles, $b_{\ell_1 \ell_2 \ell_3}^{\text{src}} = 1$, in the absence of window functions, and is given by

$$b_{\ell_1 \ell_2 \ell_3}^{\text{src}} = w_{\ell_1} w_{\ell_2} w_{\ell_3},$$

(19)

in the presence of window functions. Here, $w_{\ell}$ is an experimental window function (a product of the beam transfer function and the pixel window function). Let us then use $b_{\ell}$ (instead of $f_{\text{NL}}$, because this component has nothing to do with primordial fluctuations) to denote the amplitude of the Poisson bispectrum.

From the data, we measure $S_{\text{src}}$ given by

$$S_{\text{src}} = \frac{1}{6} \int d^2\hat{\nu} [E^2(\hat{\nu}) - 3E(\hat{\nu})E(\hat{\nu})]\tilde{MC}.,$$

(20)

where a map $E(\hat{\nu})$ is defined by [86]

$$E(\hat{\nu}) = \sum_{\ell} w_\ell \ell_\ell(\hat{\nu}) = \sum_{\ell m} w_\ell (C^{-1}a)_{\ell m} Y_{\ell m}(\hat{\nu}),$$

(21)

where $e_\ell(\hat{\nu})$ is given by equation (12). The $E(\hat{\nu})$ map is a Wiener-filtered map of point sources randomly distributed on the sky.
3.3. Primordial bispectra

(Most of this subsection is adopted from section 6.1 of [12].) During the period of cosmic inflation [88–93], quantum fluctuations were generated and became the seeds for the cosmic structures that we observe today [89, 94–97]. See [98–103] for reviews.

Inflation predicts that the statistical distribution of primordial fluctuations is nearly a Gaussian distribution with random phases. Measuring deviations from a Gaussian distribution, i.e. non-Gaussian correlations in primordial fluctuations, is a powerful test of inflation, as how precisely the distribution is (non-)Gaussian depends on the detailed physics of inflation. See [3, 77] for reviews.

The observed angular bispectrum is related to the three-dimensional bispectrum of primordial curvature perturbations, \( \langle \zeta_{k_1} \zeta_{k_2} \zeta_{k_3} \rangle = (2\pi)^3 \delta^D(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3) B_{\zeta}(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3) \). In the linear order, the primordial curvature perturbation is related to Bardeen’s curvature perturbation [104] in the matter-dominated era, \( \Phi \), by \( \zeta = \frac{5}{3} \Phi \) [105]. The CMB temperature anisotropy in the Sachs–Wolfe limit [106] is given by \( \Delta T / T = -\frac{1}{5} \Phi = -\frac{1}{5} \zeta \). We write the bispectrum of \( \Phi \) as

\[
\langle \Phi(k_1) \Phi(k_2) \Phi(k_3) \rangle = (2\pi)^3 \delta^D(k_1 + k_2 + k_3) F(k_1, k_2, k_3).
\]  

There is a useful way of visualizing the shape dependence of the bispectrum. We can study the structure of the bispectrum by plotting the magnitude of \( F(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2 \) as a function of \( k_2/k_1 \) and \( k_3/k_1 \) for a given \( k_1 \), with a condition that \( k_1 \geq k_2 \geq k_3 \) is satisfied. In order to classify various shapes of the triangles, let us use the following names: squeezed \((k_1 \approx k_2 \gg k_3)\), elongated \((k_1 = k_2 + k_3)\), folded \((k_1 = 2k_2 = 2k_3)\), isosceles \((k_2 = k_3)\) and equilateral \((k_1 = k_2 = k_3)\). See \((a)\)–\((e)\) of figure 2 for the visual representations of these triangles.

We shall explore three different shapes of the primordial bispectrum: ‘local’ ‘equilateral’ and ‘orthogonal’. They are defined as follows.

\[
(a) \quad \text{squeezed triangle} \quad (k_1 \approx k_2 \gg k_3) \\
(b) \quad \text{elongated triangle} \quad (k_1 = k_2 + k_3) \\
(c) \quad \text{folded triangle} \quad (k_1 = 2k_2 = 2k_3) \\
(d) \quad \text{isosceles triangle} \quad (k_2 = k_3) \\
(e) \quad \text{equilateral triangle} \quad (k_1 = k_2 = k_3)
\]
Figure 3. Shapes of the primordial bispectra. Each panel shows the normalized amplitude of $F(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2$ as a function of $k_2/k_1$ and $k_3/k_1$ for a given $k_1$, with a condition that $k_3 \leq k_2 \leq k_1$ is satisfied. As the primordial bispectra shown here are (nearly) scale invariant, the shapes look similar regardless of the values of $k_1$. The amplitude is normalized such that it is unity at the point where $F(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2$ takes on the maximum value. Top left: the local form given in equation (23), which peaks at the squeezed configuration. Note that the most squeezed configuration shown here has $k_1 = k_2 = 100k_3$. Top right: the orthogonal form given in equation (27), which has a positive peak at the equilateral configuration, and a negative valley along the elongated configurations. Bottom left: the equilateral form given in equation (26), which peaks at the equilateral configuration. Note that all of these shapes are nearly orthogonal to each other.

1. Local form. The local form bispectrum is given by \[ F_{\text{local}}(k_1, k_2, k_3) = 2 f_{\text{local}} f_{NL} \left[ P_{\Phi} (k_1) P_{\Phi} (k_2) + P_{\Phi} (k_2) P_{\Phi} (k_3) + P_{\Phi} (k_3) P_{\Phi} (k_1) \right] \]
\[ = 2 A^2 f_{\text{local}} \left[ \frac{1}{k_1^{d-2} k_2^{d-2}} + (2 \text{ perm.}) \right], \quad (23) \]
where $P_{\Phi} = A / k^{d-2}$ is the power spectrum of $\Phi$ with a normalization factor $A$. This form is called the local form, as this bispectrum can arise from the curvature perturbation in the form of $\Phi = \Phi_L + f_{\text{local}} \Phi_L^2$, where both sides are evaluated at the same location in space ($\Phi_L$ is a linear Gaussian fluctuation). The local form, $F_{\text{local}}(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2$, peaks at the so-called squeezed triangle for which $k_3 \ll k_2 \approx k_1$ [19]. See the top-left panel of figure 3. In this limit, we obtain
\[ F_{\text{local}}(k_1, k_1, k_3 \to 0) = 4 f_{\text{local}} f_{NL} P_{\Phi} (k_1) P_{\Phi} (k_3). \quad (24) \]
How large is $f_{\text{local}} f_{NL}$ from inflation? The earlier calculations showed that $f_{\text{local}} f_{NL}$ from single-field slow-roll inflation would be of the order of the slow-roll parameter, $\epsilon \sim 10^{-2}$ [107, 115, 116]. More recently, Maldacena [4] and Acquaviva et al [5] have found that

\[ \text{However, } \Phi = \Phi_L + f_{\text{local}} \Phi_L^2 \text{ is not the only way to produce this type of bispectrum. One can also produce this form from multi-scalar field inflation models where scalar field fluctuations are nearly scale invariant [23]; multi-scalar models called 'curvaton' scenarios [21, 109]; multi-field models in which one field modulates the decay rate of inflaton field [22, 110, 111]; multi-field models in which a violent production of particles and nonlinear reheating, called 'preheating,' occur due to parametric resonances [25, 112–114]; models in which the universe contracts first and then bounces [24].} \]
the coefficient of $P_\Phi(k_1)P_\Phi(k_3)$ from the simplest single-field slow-roll inflation with
the canonical kinetic term in the squeezed limit is given by

$$F_{\text{local}}(k_1, k_1, k_3 \rightarrow 0) = \frac{5}{3}(1 - n_s)P_\Phi(k_1)P_\Phi(k_3). \ (25)$$

Comparing this result with the form predicted by the $f_{\text{NL}}^\text{local}$ model, one obtains $f_{\text{NL}}^\text{local} = (5/12)(1 - n_s)$, which gives $f_{\text{NL}}^\text{local} = 0.015$ for $n_s = 0.963$.

2. Equilateral form. The equilateral form bispectrum is given by [8]

$$F_{\text{equil}}(k_1, k_2, k_3) = 6A^2f_{\text{NL}}\left\{ -\frac{1}{k_1^{4-n_s}k_2^{4-n_s}} - \frac{1}{k_2^{4-n_s}k_3^{4-n_s}} - \frac{1}{k_3^{4-n_s}k_1^{4-n_s}} \right\}.$$  \ (26)

This function approximates the bispectrum forms that arise from a class of inflation models in which scalar fields have non-canonical kinetic terms. One example is the so-called Dirac–Born–Infeld (DBI) inflation [27, 117], which gives $f_{\text{NL}}^\text{equil} \propto -1/c_s^2$ in the limit of $c_s \ll 1$, where $c_s$ is the effective sound speed at which scalar field fluctuations propagate relative to the speed of light. There are various other models that can produce $f_{\text{NL}}^\text{equil}$ [20, 28, 29, 118, 119]. The equilateral form, $F_{\text{equil}}(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2$, peaks at the equilateral configuration for which $k_1 = k_2 = k_3$. See the bottom-left panel of figure 3. The local and equilateral forms are nearly orthogonal to each other, which means that both can be measured nearly independently.

3. Orthogonal form. The orthogonal form, which is constructed such that it is nearly orthogonal to both the local and equilateral forms, is given by [32]

$$F_{\text{orthog}}(k_1, k_2, k_3) = 6A^2f_{\text{NL}}\left\{ -\frac{3}{k_1^{4-n_s}k_2^{4-n_s}} - \frac{3}{k_2^{4-n_s}k_3^{4-n_s}} - \frac{3}{k_3^{4-n_s}k_1^{4-n_s}} \right\} - \frac{8}{(k_1k_2k_3)^{2(4-n_s)/3}} + \left[ \frac{3}{k_1^{4-n_s}k_2^{4-n_s}k_3^{4-n_s}} + (5 \text{ perm.}) \right]. \ (27)$$

This form approximates the forms that arise from a linear combination of higher-derivative scalar-field interaction terms, each of which yields forms similar to the equilateral shape. Senatore, Smith and Zaldarriaga [32] found that, using the ‘effective field theory of inflation’ approach [118], a certain linear combination of similarly equilateral shapes can yield a distinct shape which is orthogonal to both the local and equilateral forms. The orthogonal form $F_{\text{orthog}}(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2$ has a positive peak at the equilateral configuration and a negative valley along the elongated configurations. See the top-right panel of figure 3.

Note that these are not the most general forms one can write down, and there are other forms which would probe different aspects of the physics of inflation [20, 30, 34, 38, 70, 120].

Of these forms, the local form bispectrum has special significance. Creminelli and Zaldarriaga [6] showed that not only models with the canonical kinetic term, but all single-inflation models predict the bispectrum in the squeezed limit given by equation (25), regardless of the form of potential, kinetic term, slow-roll or initial vacuum state. Also see [20, 29, 118]. This means that a convincing detection of $f_{\text{NL}}^\text{local}$ would rule out all single-field inflation models.
3.4. Optimal estimator for $f_{NL}^{\text{local}}$

Given the form of $\Phi$, one can calculate the harmonic coefficients of temperature and $E$-mode polarization anisotropies as

$$a_T^{lm} = 4\pi (-i)^l \int \frac{d^3k}{(2\pi)^3} \Phi(k) g_{Tl}(k) Y_{lm}^*(k),$$

(28)

$$a_E^{lm} = 4\pi (-i)^l \sqrt{\frac{(l+2)!}{(l-2)!}} \int \frac{d^3k}{(2\pi)^3} \Phi(k) g_{Pl}(k) Y_{lm}^*(k),$$

(29)

where $g_{Tl}(k)$ and $g_{Pl}(k)$ are the radiation transfer functions of the temperature and polarization anisotropies, respectively, which can be calculated by solving the linearized Boltzmann equations. One may use the publicly available Boltzmann codes such as CMBFAST [121] or CAMB [122] for computing the radiation transfer functions.⁴

From now on, we shall focus on the temperature anisotropy, largely for simplicity. (See [123, 124] for the treatment of polarization in the angular bispectrum.) The limits on $f_{NL}$ expected from Planck are dominated by the temperature information, and thus the polarization information is not expected to yield competitive limits over the next, say, > 5 years.

For the local-form bispectrum given in equation (23), the reduced bispectrum (equation (6)) is given by [83]

$$b_{\text{local}}^{l_1l_2l_3} = 2 \int r^2 dr \left[ \beta_{l_1}(r) \beta_{l_2}(r) \alpha_{l_3}(r) + (2 \text{ perm.}) \right] w_l w_{l_2} w_{l_3},$$

(30)

where

$$\alpha_l(r) = \frac{2}{\pi} \int k^2 dk g_{Tl}(k) j_l(kr),$$

(31)

$$\beta_l(r) = \frac{2}{\pi} \int k^2 dk P_0(k) g_{Tl}(k) j_l(kr).$$

(32)

Using this form in equation (14), one finds $S_{\text{local}}$ as

$$S_{\text{local}} = \int r^2 dr \int d^2\hat{n} [A(\hat{n}, r) B^2(\hat{n}, r) - 2B(\hat{n}) \langle A(\hat{n}, r) B(\hat{n}, r) \rangle_{\text{MC}} - A(\hat{n}, r) \langle B^2(\hat{n}, r) \rangle_{\text{MC}}],$$

(33)

which can be measured from the data. Here, maps $A(\hat{n}, r)$ and $B(\hat{n}, r)$ are defined by [7]

$$A(\hat{n}, r) \equiv \sum_l w_l \alpha_l(r) e_l(\hat{n}) = \sum_l w_l \alpha_l(r) (C^{-1}a)_{lm} Y_{lm}(\hat{n}),$$

(34)

$$B(\hat{n}, r) \equiv \sum_l w_l \beta_l(r) e_l(\hat{n}) = \sum_l w_l \beta_l(r) (C^{-1}a)_{lm} Y_{lm}(\hat{n}),$$

(35)

where $e_l(\hat{n})$ is given by equation (12).

---

⁴ A CMBFAST-based code for computing $g_{Tl}(k)$ and $g_{Pl}(k)$ is available at http://gyudon.as.utexas.edu/komatsu/CRL. A recent version of CAMB has an option to calculate these functions (http://camb.info).
3.5. Optimal estimator for $f_{\text{NL}}^{\text{equil}}$

For the equilateral-form bispectrum given in equation (26), the reduced bispectrum is given by [8]

$$b_{l_1l_2l_3}^{\text{equil}} = -3b_{l_1l_2l_3}^{\text{local}} + 6 \int r^2 dr [\gamma_l(r) \delta_l(r) + (5 \text{ perm.})]$$

where

$$\gamma_l(r) = \frac{2}{\pi} \int k^2 dk P_1(k) \Phi_1(k) g_{Tl}(kr),$$

$$\delta_l(r) = \frac{2}{\pi} \int k^2 dk P_2(k) \Phi_1(k) g_{Tl}(kr).$$

Using this form in equation (14), one finds $S_{\text{equil}}$ as

$$S_{\text{equil}} = -3S_{\text{local}} + 6 \int r^2 dr \int d^2 \hat{n} \left[ \right.$$

$$B(\hat{n}, r) C(\hat{n}, r) D(\hat{n}, r)$$

$$- B(\hat{n})(C(\hat{n}, r) D(\hat{n}, r))_{\text{MC}} - C(\hat{n})(B(\hat{n}, r) D(\hat{n}, r))_{\text{MC}}$$

$$- D(\hat{n})(B(\hat{n}, r) C(\hat{n}, r))_{\text{MC}} - \frac{1}{3} [D^3(\hat{n}, r) - 3D(\hat{n}, r)(D^2(\hat{n}, r))_{\text{MC}}] \left. \right],$$

which can be measured from the data. Here, maps $C(\hat{n}, r)$ and $D(\hat{n}, r)$ are defined by [8]

$$C(\hat{n}, r) \equiv \sum_{lm} w_l \gamma_l(r) e_l(\hat{n}) = \sum_{lm} w_l \gamma_l(r)(C^{-1} a)_{lm} Y_{lm}(\hat{n}),$$

$$D(\hat{n}, r) \equiv \sum_{lm} w_l \delta_l(r) e_l(\hat{n}) = \sum_{lm} w_l \delta_l(r)(C^{-1} a)_{lm} Y_{lm}(\hat{n}).$$

3.6. Optimal estimator for $f_{\text{NL}}^{\text{orthog}}$

For the equilateral-form bispectrum given in equation (27), the reduced bispectrum is given by [32]

$$b_{l_1l_2l_3}^{\text{orthog}} = 3b_{l_1l_2l_3}^{\text{equil}} - 12 \int r^2 dr \delta_l(r) \delta_l(r) \delta_l(r) w_{l_1} w_{l_2} w_{l_3},$$

Using this form in equation (14), one finds $S_{\text{orthog}}$ as

$$S_{\text{orthog}} = 3S_{\text{equil}} - 2 \int r^2 dr \int d^2 \hat{n} [D^3(\hat{n}, r) - 3D(\hat{n}, r)(D^2(\hat{n}, r))_{\text{MC}}],$$

which can be measured from the data.

4. Secondary anisotropy

4.1. General formula for the lensing–secondary coupling

Given the special importance of the local-form bispectrum, we must understand which other (non-primordial) effects might also produce the local form, potentially preventing us from measuring $f_{\text{NL}}^{\text{local}}$.

The local-form bispectrum is generated when the power spectrum of short-wavelength fluctuations is modulated by long-wavelength fluctuations; thus, a mechanism that couples small scales to large scales can potentially generate the local-form bispectrum.
The weak gravitational lensing provides one such mechanism. The local-form bispectrum may then be generated when long- and short-wavelength fluctuations are coupled by the lensing. To see how this might happen, let us write the observed temperature anisotropy in terms of the original (unlensed) contribution from the last scattering surface at \( z = 1090 \), \( \Delta T' \) (where ‘\( P \)’ stands for ‘primary’), the lensing potential \( \phi \) and the secondary anisotropy generated between \( z = 1090 \) and \( z = 0 \), \( \Delta T^\Delta \) (where ‘\( S \)’ stands for ‘secondary’):

\[
\Delta T(\hat{n}) = \Delta T' (\hat{n} + \hat{\phi}) + \Delta T^\Delta (\hat{n}) \\
\approx \Delta T' (\hat{n}) + \left[(\hat{\phi} \cdot (\hat{\phi} \Delta T'))(\hat{n}) + \Delta T^\Delta (\hat{n})\right],
\]

where

\[
\phi(\hat{n}) = -2 \int_0^{r_s} \frac{r - r}{r} \Phi(r, \hat{n}) d\Omega, \tag{45}
\]

and \( r_s \) is the comoving distance out to \( z = 1090 \), and \( \Phi \) is Bardeen’s curvature perturbation, which is related to the usual Newtonian gravitational potential by \( \Phi = -\Phi_{\text{Newton}} \).

Transforming this into harmonic space and computing the reduced bispectrum, one obtains [13]

\[
b_{\ell_1\ell_2\ell_3}^{\text{local}-S} = \frac{1}{2} \left\{ \ell_1(l_1 + 1) - l_2(l_2 + 1) + l_3(l_3 + 1) \right\} \frac{C_{\ell}^P C_{\ell}^S + (5 \text{ perm.})}{2} \sum_l w_l w_{l_1} w_{l_2} w_{l_3}, \tag{46}
\]

where \( C_{\ell}^P \) is the power spectrum of the CMB from the decoupling epoch only (i.e., no ISW or lensing), and \( C_{\ell}^S \equiv \langle \phi_{\ell m}^* a_{\ell m}^S \rangle \) is the lensing–secondary cross-correlation power spectrum.

From this result, one finds that a non-zero bispectrum is generated when the secondary anisotropy traces the large-scale structure (i.e., \( \Phi \)). Various secondary effects have been studied in the literature: the Sunyaev–Zel’dovich effect [13], cosmic reionization [125], point sources [126] and ISW [13]. It has been shown that the last one, the ISW–lensing coupling, is the most dominant contamination of \( f_{\text{NL}} \) [15].

Using equation (46) in equation (14), one finds \( S_{\text{local}-S} \) as

\[
S_{\text{local}-S} = \frac{1}{2} \int d^2 \hat{n} \left\{ P(\hat{n}) \langle \hat{\phi}^2 \rangle Q(\hat{n}) - P(\hat{n}) E(\hat{n}) \langle \hat{\phi} \hat{Q} \rangle(\hat{n}) + \text{(linear terms)} \right\}, \tag{47}
\]

which can be measured from the data. Here, the ‘linear terms’ contain nine terms with \( \langle \rangle_{MC} \), such as \(- P(\hat{n}) \langle [\hat{\phi}^2 E(\hat{n}) Q(\hat{n})]_{MC} \rangle\), etc. The map \( E(\hat{n}) \) is given by equation (21), and the other maps are defined by

\[
P(\hat{n}) = \sum_l w_l C_{\ell}^P e_l(\hat{n}) = \sum_{l m} w_l C_{\ell}^P (C^{-1} a)_{lm} Y_{lm}(\hat{n}), \tag{49}
\]

\[
Q(\hat{n}) = \sum_l w_l C_{\ell}^S e_l(\hat{n}) = \sum_{l m} w_l C_{\ell}^S (C^{-1} a)_{lm} Y_{lm}(\hat{n}). \tag{50}
\]

The maps with \( \hat{\phi}^2 \) are given by \( \hat{\phi}^2 P = - \sum_l l(l + 1) w_l C_{l}^P e_l(\hat{n}) \), etc. The map \( P(\hat{n}) \) is a Wiener-filtered map of the primary temperature anisotropy from \( z = 1090 \).

### 4.2. Lensing–ISW coupling

A change in the curvature perturbation yields a secondary temperature anisotropy via the ISW effect [106]:

\[
\frac{\Delta T_{\text{ISW}}(\hat{n})}{T} = -2 \int_0^{r_s} \frac{\Delta \Phi}{dr} (r, \hat{n}) d\Omega. \tag{51}
\]
where \( r \) is the comoving distance and \( r_* \) is the comoving distance out to \( z = 1090 \). Here, note again \( \Phi = -\Phi_{\text{Newton}} \). The cross-power spectrum of \( \phi \) and the ISW effect is then given by

\[
C_{\phi}^{\Phi,\text{ISW}}(l) = 4 \int_0^{r_*} \frac{dr}{r^3} P_{\phi\Phi} \left( \frac{l}{r^2} \right),
\]

where \( P_{\phi\Phi}(k, r) \) is the cross-power spectrum of \( \Phi \) and \( \Phi' \equiv \partial \Phi / \partial r \), which can be calculated from the power spectrum of \( \Phi \), \( P_{\phi\Phi}(k, r) = \frac{1}{2}[\partial P_{\phi}(k, r) / \partial r] \) [14, 127]. Here, \( P_{\phi}(k, r) \) is not the primordial power spectrum, but it includes the linear transfer function, \( T(k) \), and the growth factor of \( \Phi \), \( g(r) \):

\[
P_{\phi}(k, r) = \frac{A}{k^{4 - n_s}} \left[ T(k)g(r) \right]^2.
\]

Using this, one finds

\[
P_{\phi\Phi}(k, r) = \left( \frac{g'}{g} \right) P_{\phi}(k, r).
\]

Note that \( g(r) \) is normalized such that \( g(r) = 1 \) during the matter-dominated era.

With this result, it is easy to see why the lensing–ISW coupling yields the squeezed configuration: on very large scales, where \( T(k) \to 1 \), \( C_{\phi,\text{ISW}}(l) \propto 1/l^3 \). On smaller scales, \( T(k) \) declines with \( k \), and thus \( C_{\phi,\text{ISW}}(l) \) falls faster than \( 1/l^3 \). The lensing coupling includes \( l(l+1)C_{\phi,\text{ISW}}(l) \), which falls faster than \( 1/l \), i.e. the largest power comes from the smallest \( l \).

A recent estimate by Hanson et al [16] showed that the lensing–ISW coupling, if not included in the parameter estimation, would bias \( f_{\text{local}} \) by \( \Delta f_{\text{local}} = 9.3 \). The expected bias for WMAP is \( \Delta f_{\text{local}} = 2.7 \) [12]. One can remove this bias by including the lensing–ISW coupling (or any other lensing–secondary couplings) using the optimal estimator given by equation (48).

5. Second-order effect

5.1. General discussion

So far, we have assumed that one can use equation (28):

\[
a_{lm} = 4\pi (-i)^l \int \frac{d^3k}{(2\pi)^3} \Phi_p(k) g_{T'l} Y^*_l(k),
\]

to convert the primordial curvature perturbation into the temperature anisotropy. (Here, the subscript ‘\( p \)’ stands for ‘primordial,’ by which we mean \( \Phi_p = \frac{1}{2} \zeta \) without the linear transfer function.) However, this equation is valid only for linear theory. As any nonlinear effect can produce non-Gaussianity, one has to study the impacts of various nonlinear effects on the observed non-Gaussianity.

The origin of the linear radiation transfer function is the nonlinear Boltzmann equation:

\[
\frac{\partial \Delta^{(1)}}{\partial \eta} + i k \mu \Delta^{(1)} + \sigma_T n_e \Delta^{(1)} = S^{(1)}(k, \mu, \eta),
\]

where \( n_e \) is the conformal time, \( \mu \equiv \hat{k} \cdot \hat{n} \), \( \Delta^{(1)} \equiv 4[\Delta T(k, \mu, \eta)/T] \) is the perturbation in the photon energy density and \( S^{(1)} \) is the linear source function, which depends on the metric perturbations as well as on the density, velocity, pressure and stress perturbations of matter and radiation in the universe and the photon polarization.

The second-order Boltzmann equation is then similarly written as

\[
\frac{\partial \Delta^{(2)}}{\partial \eta} + i k \mu \Delta^{(2)} + \sigma_T n_e \Delta^{(2)} = S^{(2)}(k, \hat{n}, \eta),
\]

where \( \Delta^{(2)} \equiv 8[\Delta T^{(2)}(k, \hat{n}, \eta)/T] + 12[\Delta T^{(1)}(k, \mu, \eta)/T]^2 \) and \( S^{(2)} \) is the second-order source function. Note that the azimuthal symmetry is lost at the second order, and thus the
perturbations depend on the directions of $k$ and $\hat{n}$ independently. In this case, the second-order $a_{lm}$ is given by [128]

$$a_{lm}^{(2)} = \frac{4\pi}{8} (-i)^f \int \frac{d^3k'}{(2\pi)^3} \int \frac{d^3k''}{(2\pi)^3} \int d^3k'' \delta(k' + k'' - k) \Phi^{(1)}(k') \Phi^{(1)}(k'')$$

$$\times \sum_{l'm'} F_{lm}^m(k', k'') Y_{lm}^* (\hat{k}),$$

(56)

where $F_{lm}^m$ is the second-order radiation transfer function, whose form is determined by the second-order source function, $S^{(2)}$, in the Boltzmann equation.

The shape of the second-order bispectrum, $\{a_{lm}^{(1)} a_{lm}^{(1)} a_{lm}^{(2)}\}$, is determined by the shape of the second-order radiation transfer function. If the second-order radiation transfer function vanishes in the squeezed limit, i.e., $F_{lm}^m(k', k'', k) \to 0$ for $k \to 0$, then the CMB bispectrum would not peak at the squeezed configuration, and thus the resulting $f_{\text{NL}}$ would be small.

The second-order source function is quite complicated [129–139], but it can be divided into two parts$^4$:

1. the terms given by the products of the first-order perturbations, such as $[\Phi^{(1)}]^2$;
2. the terms given by the ‘intrinsically second-order terms’ such as $\Phi^{(2)}$.

The intrinsically second-order terms are sourced by products of the first-order perturbations, and thus it is created by the late-time evolution of cosmological perturbations, whereas the terms in (1) are set by the initial conditions.

The contamination of $f_{\text{NL}}$ due to the terms in (1) is small, $|f_{\text{NL}}| < 1$ [128]. Recently, Pitrou, Bernardeau and Uzan [138] have reported a surprising result that the terms in (2) would give $f_{\text{NL}} \sim 5$ for the Planck data ($l_{\text{max}} = 2000$).

Why surprising? As the intrinsically second-order terms arise as a consequence of the late-time evolution of the cosmological perturbations, they are generated by the causal mechanism, i.e., gravity and hydrodynamics. It is difficult for the causal mechanism to generate the bispectrum in the squeezed configuration, as it requires very long wavelength perturbations to be coupled to short wavelength ones.

5.2. Newtonian calculation

As an example, let us consider the well-known second-order solution for $\Phi^{(2)}$ in the sub-horizon limit, i.e., $k \gg aH$, which is equivalent to taking the non-relativistic (Newtonian) limit. Here, the second-order Bardeen curvature perturbation is defined by $\Phi = \Phi^{(1)} + \frac{1}{2} \Phi^{(2)}$. The explicit solution is [137]$^5$

$$\frac{1}{2} \Phi^{(2)}(k, \eta) = \frac{1}{6} \int \frac{d^3k'}{(2\pi)^3} \int d^3k'' \delta^3(k' + k'' - k) \left(\frac{k''\eta}{k}\right)^2$$

$$\times F_{2}^{(2)}(k', k'') \Phi^{(1)}(k') \Phi^{(1)}(k''),$$

(57)

where the linear perturbation, $\Phi^{(1)}$, on the right-hand side is constant during the matter-dominated era, and the symmetrized function, $F_{2}^{(2)}$, is defined as

$$F_{2}^{(2)}(k_1, k_2) = \frac{3}{7} + \frac{k_1 \cdot k_2}{2k_1k_2} \left(\frac{k_1}{k_2} + \frac{k_2}{k_1}\right) + \frac{2}{7} \left(\frac{k_1 \cdot k_2}{k_1k_2}\right)^2.$$  

(58)

$^4$ This decomposition is not gauge invariant, and thus which terms belong to (1) or (2) depends on the gauge that one chooses. Therefore, one must specify the gauge when making such a decomposition. Our discussion in this section is based on the gauge choice made by Bartolo, Matarrese and Riotto [129, 130] and Pitrou, Uzan and Bernardeau [137, 138], which reduces to the Newtonian gauge at the linear order. This seems a convenient gauge, as the products of the first-order terms only give $|f_{\text{NL}}| < 1$ [128].

$^5$ Note that our $\Phi$ is $(-1)$ times $\Phi$ used in equation (72) of [137].
Note that $F_2^{(s)}$ is related to the function $G$ given in equation (8.9) of [130] as $G(k_1, k_2, k) = -\frac{14}{7} \left(\frac{k_1}{k_2}\right)^2 F_2^{(s)}(k_1, k_2)$ with $k = k_1 + k_2$.

The function $F_2^{(s)}(k_1, k_2)$ vanishes in the squeezed limit, $k_1 = -k_2$, and thus the CMB bispectrum generated from $\Phi^{(2)}$ in the Newtonian limit is not given by the local form. To see this, let us calculate

$$\langle \Phi(k_1, \eta) \Phi(k_2, \eta) \Phi(k_3, \eta) \rangle = (2\pi)^3 \delta_D(k_1 + k_2 + k_3) F_{2nd}(k_1, k_2, k_3, \eta),$$

(59)

where

$$F_{2nd}(k_1, k_2, k_3, \eta) = \frac{\eta^2}{3} \left[ \left(\frac{k_1 k_2}{k_3}\right)^2 F_2^{(s)}(k_1, k_2) P_\Phi(k_1) P_\Phi(k_2) + \text{(2 perm.)} \right],$$

(60)

and $P_\Phi(k) = AT^2(k)/k_2^{4-n_s}$. The shape dependence of $F_{2nd}(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2$ is shown in figure 4 for various values of $k_1$ (because $F_{2nd}$ is not scale invariant). The CMB data are sensitive to $k_1 < k_{max} \sim 0.2 h \text{ Mpc}^{-1}(l_{max}/2000)$. We find that the bispectrum peaks at the equilateral configuration on large scales ($k_1 < 10^{-2} h \text{ Mpc}^{-1}$), and it peaks along the elongated configurations on small scales ($k_1 \sim 0.1 h \text{ Mpc}^{-1}$). It peaks at the squeezed configuration on a very small scale ($k_1 \sim 1 h \text{ Mpc}^{-1}$), but these scales are not accessible by the CMB due to the Silk damping. Note that the most squeezed configuration shown in this figure has $k_1 = k_2 = 100k_3$. The dominant shape changes with scales, as the linear transfer function $T(k)$ declines with $k$, with the small-scale limit given by $T(k) \propto \ln k/k^2$. From these results, we expect the second-order effect in the Newtonian limit to yield only a small contamination of $f_{\text{local}}^{NL}$. 

Figure 4. Shapes of the second-order bispectrum due to the second-order curvature perturbations in the Newtonian limit given in equation (60). Each panel shows the normalized amplitude of $F_{2nd}(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2$ as a function of $k_2/k_1$ and $k_3/k_1$ for a given $k_1$, with a condition that $k_1 \leq k_2 \leq k_3$ is satisfied. The amplitude is normalized such that it is unity at the point where $F_{2nd}(k_1, k_2, k_3)(k_2/k_1)^2(k_3/k_1)^2$ takes on the maximum value. Top left: $k_1 = 10^{-3} h \text{ Mpc}^{-1}$. Top right: $k_1 = 10^{-1} h \text{ Mpc}^{-1}$. Bottom left: $k_1 = 1 h \text{ Mpc}^{-1}$. Bottom right: $k_1 = 10 h \text{ Mpc}^{-1}$. The CMB data are sensitive to $k_1 < k_{max} \sim 0.2 h \text{ Mpc}^{-1}(l_{max}/2000)$, where the second-order bispectrum peaks at the equilateral configuration on large scales, and peaks along the elongated configurations on a smaller scale ($k_1 \sim 0.1 h \text{ Mpc}^{-1}$). On a very small scale ($k_1 \sim 1 h \text{ Mpc}^{-1}$), it peaks at the squeezed configuration. Note that the most squeezed configuration shown here has $k_1 = k_2 = 100k_3$. 


The dominant contribution to the second-order temperature anisotropy in the sub-horizon limit is given by the second-order Sachs–Wolfe effect [137]:
\[
\frac{\Delta T^{(2)}}{T}(\hat{n}) = \frac{1}{2} R_s \Phi^{(2)}(r_s, \hat{n}_s),
\]
where \( R_s \equiv 3\rho_b/(4\rho_c) \) is the baryon–photon ratio at the decoupling epoch. (Here, a factor of \(1/2\) comes from our way of defining the second-order temperature anisotropy \( \Delta T = \Delta T^{(1)} + \Delta T^{(2)} \) and the second-order curvature perturbation \( \Phi = \Phi^{(1)} + \frac{1}{2} \Phi^{(2)} \).) This definition follows from [128]. The corresponding second-order \( a_{lm}^{(2)} \) is
\[
a_{lm}^{(2)} = 4\pi (-i)^l \int \frac{d^3 k}{(2\pi)^3} \left[ \frac{1}{2} R_s \Phi^{(2)}(k, \eta_s) \right] j_l(k r_s) Y_{lm}^*(\hat{n}).
\]
Comparing this with equation (56), we identify the term inside the square bracket as the second-order radiation transfer function \( F^{(\text{NL})}_{l,m} \). This result, one can calculate the reduced bispectrum of the Newtonian second-order effect \( b_{\text{NL}}^{\text{2nd}} \). The resulting \( F^{\text{local}}_{l,m} \) is always less than unity regardless of the angular scales ([196]; also see [141, 142]). The calculation was done for \( l \leq 2000 \).

How can we reconcile this result with those found in [138]? The calculations given above (equations (57) and (61)) are valid only in the sub-horizon limits, and thus they are not suitable for calculating the contribution to the squeezed-limit bispectrum, which can correlate super- and sub-horizon fluctuations. Therefore, the difference between these results seems to imply the following.

1. The dominant contamination of \( f^{\text{local}}_{l,m} \) comes from the general relativistic (post-Newtonian) evolution of \( \Phi^{(2)} \) that is not captured by the above Newtonian calculation (equation (57)).
2. The full second-order radiation transfer function beyond the sub-horizon approximation (equation (61)) gives the dominant contribution to \( f^{\text{local}}_{l,m} \).

Perhaps both contributions are important. This is yet to be confirmed; however, if this is true, one should be able to construct a simple template for the second-order bispectrum, and use it to remove the contamination by including its amplitude, \( f^{\text{local}}_{l,m} \), in the fit.

6 The CMB bispectrum from the second-order ISW effect was considered in [140, 141].
7 Incidentally, in the notation of [128] (see their equation (2.29)), \( \mathcal{S}^{(2)}_{\text{ISW}}(k', k'', k, \eta_s) = 2 R_s \left( \frac{\Delta T}{T} \right)^2 F^{(1)}_{2}(k', k'') T(k') T(k'') \).
where \( k_{ij} \equiv |k_i + k_j| \). (In this section, we use \( \Phi \) for the primordial perturbation, i.e. \( \Phi = \Phi_p = \frac{3}{2} \zeta \).) When the curvature perturbation is given by the simplest local form, \( \Phi = \Phi_L + f_{\text{local}}^2 \Phi_L^2 + g_{\text{NL}} \Phi_L^4 \), one finds the above trispectrum with \( \tau_{NL} = (6 f_{\text{NL}}^2 / 5)^2 \) [45]. However, in general, \( \tau_{NL} \) is different from \( (6 f_{\text{NL}}^2 / 5)^2 \).

To see this, let us consider a broad class of multi-field models in which the primordial curvature perturbation, \( \zeta \), is given in terms of the field derivatives of the number of e-folds, \( N = \ln a \), and the perturbation in the \( I \)th scalar field, \( \delta \phi_I \), as
\[
\zeta = \sum_I \frac{\partial N}{\partial \phi_I} \delta \phi_I + \frac{1}{2} \sum_{IJ} \frac{\partial^2 N}{\partial \phi_I \partial \phi_J} \delta \phi_I \delta \phi_J + \cdots ,
\]
(64)
where \( \langle \delta \phi_I \delta \phi_J \rangle = 0 \) for \( I \neq J \). This expansion is known as the ‘\( \delta N \) formalism’ [23, 89, 115, 143, 144]. In this case, \( f_{\text{local}}^2 \) and \( \tau_{NL} \) are given by [23, 47, 145]

\[
6 f_{\text{local}}^2 = \frac{\sum_{IJK} N_{IJ} N_{IK} N_{JK}}{\left[ \sum_j (N_j)^2 \right]^2} ,
\]
(65)
\[
\tau_{NL} = \frac{\sum_{IJK} N_{IJ} N_{IK} N_{JK}}{\left[ \sum_j (N_j)^2 \right]^2} = \frac{\sum I \left( \sum_{J} N_{IJ} N_{J} \right)^2}{\left[ \sum_j (N_j)^2 \right]^2} ,
\]
(66)
where \( N_{IJ} \equiv \partial N / \partial \phi_I \) and \( N_{IJK} \equiv \partial^2 N / \partial \phi_I \partial \phi_J \). Suyama and Yamaguchi [50] showed that the Cauchy–Schwarz inequality implies that the inequality
\[
\tau_{NL} \geq \left( \frac{6 f_{\text{local}}^2}{5} \right)^2
\]
(67)
is satisfied. To derive this result, use the Cauchy–Schwarz inequality
\[
\left( \sum_I a_I^2 \right) \left( \sum_J b_J^2 \right) \geq \left( \sum_I a_I b_I \right)^2 ,
\]
(68)
with
\[
a_I = \frac{\sum_j N_{IJ} N_j}{\left[ \sum_j (N_j)^2 \right]^{3/2}} ,
\]
(69)
\[
b_I = \frac{N_I}{\left[ \sum_j (N_j)^2 \right]^{1/2}} .
\]
(70)
The equality \( \tau_{NL} = (6 f_{\text{local}}^2 / 5)^2 \) is satisfied for the simplest local-form model, \( \zeta = N_1 \delta \phi_1 + (N_{11} / 2) \delta \phi_2 \).

Note, however, that one finds a different relation between \( f_{\text{local}}^2 \) and \( \tau_{NL} \) when the Cauchy–Schwarz inequality becomes trivial, i.e. \( 0 = 0 \). For example, when \( \zeta \) is given by [45]
\[
\zeta = \frac{\partial N}{\partial \phi_1} \delta \phi_1 + \frac{1}{2} \frac{\partial^2 N}{\partial \phi_2^2} \delta \phi_2^2
\]
(71)
and \( \langle \delta \phi_1 \delta \phi_2 \rangle = 0 \), one finds \( \tau_{NL} \sim 10^3 (f_{\text{local}}^2)^{4/3} \) [51]. The Cauchy–Schwarz inequality becomes \( 0 = 0 \) because \( a_I = 0 \) for all \( I \), and thus both the bispectrum and the trispectrum come from the second term in equation (71): the bispectrum is given by the 6-point function of \( \delta \phi_2 \), and the trispectrum is given by the 8-point function of \( \delta \phi_2 \).

In this case, whether \( \tau_{NL} \geq (6 f_{\text{local}}^2 / 5)^2 \) is satisfied depends on the value of \( f_{\text{local}}^2 \). For this particular example, the current limit of \( f_{\text{local}}^2 < 74 \) implies that \( \tau_{NL} \geq (6 f_{\text{local}}^2 / 5)^2 \) is still satisfied parametrically.
The inequality is valid also for the ‘quasi-single field inflation’ of [146]; see equation (7.6) of [70]. (Note that $\tau_{\text{NL}}^{\text{SL}}$ and $\tau_{\text{NL}}^{\text{CI}}$ in [70] correspond to $\tau_{\text{NL}}$ and $g_{\text{NL}}$ in this paper, respectively.) Therefore, if observations indicate $\tau_{\text{NL}} < (6 f_{\text{local}}^{\text{NL}}/5)^2$, a broad class of multi-field models satisfying the above conditions would be ruled out. This property makes the trispectrum a powerful probe of the physics of multi-field models. Note that the simplest local-form limit $\tau_{\text{NL}} = (6 f_{\text{local}}^{\text{NL}}/5)^2$ has no special significance, as this is just one of the many possibilities of multi-field models. (All single-field models predict a non-detectable level of the primordial $f_{\text{local}}^{\text{NL}}$, and thus the observational test using the above relation between $f_{\text{local}}^{\text{NL}}$ and $\tau_{\text{NL}}$ has no relevance to single-field models.)

The expected 95% uncertainties in $\tau_{\text{NL}}$ from the 7-year WMAP data ($l_{\text{max}} \sim 500$) and the 1-year Planck data ($l_{\text{max}} \sim 1500$) are 5000 and 560, respectively [74]. If the Planck finds $f_{\text{local}}^{\text{NL}} \sim 30$, then it would be able to test if the measured $\tau_{\text{NL}}$ would satisfy equation (67). This provides an excellent science case for the trispectrum that would be measured by Planck.

The expected uncertainties in $g_{\text{NL}}$ have not been calculated yet, although we expect them to be much greater than those for $\tau_{\text{NL}}$, as $g_{\text{NL}}$ is the coefficient of the cubic-order term (i.e. $g_{\text{NL}}$ is much more difficult to constrain than $\tau_{\text{NL}}$) [147].

The local-form trispectrum is not the only possibility. Various other inflation models predict distinctly different quadrilateral shape dependence. For some general analyses of shapes, see [36, 43, 59, 70].

Finally, while we do not discuss the large-scale structure of the universe in this paper, the most promising probe of the local-form trispectrum seems to be the bispectrum of galaxies. See [87, 148, 149] for details.

7. Current results

7.1. Bispectrum

(Most of this subsection is adopted from section 6.2 of [12].)

In 2002, the first limit on $f_{\text{local}}^{\text{NL}}$ was obtained from the COBE 4-year data [150] by Komatsu et al [151], using the angular bispectrum. The limit was improved by an order of magnitude when the WMAP first year data were used to constrain $f_{\text{local}}^{\text{NL}}$ [152]. Since then the limits have improved steadily as WMAP collects more years of data and the bispectrum method for estimating $f_{\text{local}}^{\text{NL}}$ has improved [7, 8, 11, 86, 153–157].

Using the optimal estimators described in section 3, we have constrained the primordial non-Gaussianity parameters as well as the point-source bispectrum using the WMAP 7-year data. The 7-year data and results are described in [12, 78, 158–161].

We use the V- and W-band maps at the HEALPix resolution $N_{\text{side}} = 1024$. As the optimal estimator weights the data optimally at all multipoles, we no longer need to choose the maximum multipole used in the analysis, i.e. we use all the data. We use both the raw maps (before cleaning foreground) and foreground-reduced (clean) maps to quantify the foreground contamination of $f_{\text{NL}}$ parameters. For all cases, we find the best limits on $f_{\text{NL}}$ parameters by combining the V- and W-band maps, and marginalizing over the synchrotron, free–free and dust foreground templates [159]. As for the mask, we always use the KQ75y7 mask [159].

In table 1, we summarize our results.

1. Local form results. The 7-year best estimate of $f_{\text{local}}^{\text{NL}}$ is

$$f_{\text{local}}^{\text{NL}} = 32 \pm 21 \text{(68\%CL)}.$$

The 95\% limit is $-10 < f_{\text{local}}^{\text{NL}} < 74$. When the raw maps are used, we find $f_{\text{local}}^{\text{NL}} = 59 \pm 21 \text{(68\% CL)}$. When the clean maps are used, but foreground templates are
Table 1. Estimates and the corresponding 68% intervals of the primordial non-Gaussianity parameters \((f_{\text{local}}^{\text{NL}}, f_{\text{equil}}^{\text{NL}}, f_{\text{orthog}}^{\text{NL}})\) and the point source bispectrum amplitude, \(b_{\text{src}}\) (in units of \(10^{-5}\mu K^3\text{sr}^2\)), from the WMAP 7-year temperature maps. This table is adopted from [12]

| Band     | Foreground | \(f_{\text{local}}^{\text{NL}}\) | \(f_{\text{equil}}^{\text{NL}}\) | \(f_{\text{orthog}}^{\text{NL}}\) | \(b_{\text{src}}\) |
|----------|------------|---------------------------------|---------------------------------|---------------------------------|-----------------|
| V+W Raw  | 59 ± 21    | 33 ± 140                        | -199 ± 104                      | N/A                             |
| V+W Clean| 42 ± 21    | 29 ± 140                        | -198 ± 104                      | N/A                             |
| V+W Marg. | 32 ± 21    | 26 ± 140                        | -202 ± 104                      | -0.08 ± 0.12                    |
| V Marg.  | 43 ± 24    | 64 ± 150                        | -98 ± 115                       | 0.32 ± 0.23                     |
| W Marg.  | 39 ± 24    | 36 ± 154                        | -257 ± 117                      | -0.13 ± 0.19                    |

not marginalized over, we find \(f_{\text{local}}^{\text{NL}} = 42 ± 21\) (68\% CL). These results (in particular the clean-map versus the foreground marginalized) indicate that the foreground emission makes a difference at the level of \(\Delta f_{\text{local}}^{\text{NL}} \sim 10^8\). We find that the V+W result is lower than the V-band or W-band results. This is possible, as the V+W result contains contributions from the cross-correlations of V and W such as \(\langle \bar{V}\bar{W} \rangle\) and \(\langle \bar{V}\bar{V}\bar{W} \rangle\).

2. **Equilateral form results.** The 7-year best estimate of \(f_{\text{equil}}^{\text{NL}}\) is

   \[
   f_{\text{equil}}^{\text{NL}} = 26 ± 140 \text{ (68\%CL)}.
   \]

   The 95\% limit is \(-214 < f_{\text{equil}}^{\text{NL}} < 266\). For \(f_{\text{equil}}^{\text{NL}}\), the foreground marginalization does not shift the central values very much, \(\Delta f_{\text{equil}}^{\text{NL}} = -3\). This makes sense, as the equilateral bispectrum does not couple small-scale modes to very large-scale modes \(l < 10\), which are sensitive to the foreground emission. On the other hand, the local form bispectrum is dominated by the squeezed triangles, which do couple large- and small-scale modes.

3. **Orthogonal form results.** The 7-year best estimate of \(f_{\text{orthog}}^{\text{NL}}\) is

   \[
   f_{\text{orthog}}^{\text{NL}} = -202 ± 104 \text{ (68\%CL)}.
   \]

   The 95\% limit is \(-410 < f_{\text{orthog}}^{\text{NL}} < 6\). The foreground marginalization has little effect, \(\Delta f_{\text{orthog}}^{\text{NL}} = -4\).

As for the point-source bispectrum, we do not detect \(b_{\text{src}}\) in V, W or V+W. In [86], we estimated that the residual sources could bias \(f_{\text{local}}^{\text{NL}}\) by a small positive amount and applied corrections using Monte Carlo simulations. In this paper, we do not attempt to make such corrections, but we note that sources could give \(\Delta f_{\text{local}}^{\text{NL}} \sim 2\) (note that the simulations used in [86] likely overestimated the effect of sources by a factor of 2). As the estimator has changed from that used in [86], extrapolating the previous results is not trivial. Source corrections to \(f_{\text{equil}}^{\text{NL}}\) and \(f_{\text{orthog}}^{\text{NL}}\) could be larger [86], but we have not estimated the magnitude of the effect for the 7-year data.

As we described in section 4, among various sources of secondary non-Gaussianities which might contaminate measurements of primordial non-Gaussianity (in particular \(f_{\text{local}}^{\text{NL}}\)), a coupling between the ISW effect and the weak gravitational lensing is the most dominant source of confusion for \(f_{\text{local}}^{\text{NL}}\). Calabrese et al [163] used the skewness power spectrum method of [164] to search for this term in the WMAP 5-year data and found a null result.

---

8 The effect of the foreground marginalization depends on an estimator. Using the needlet bispectrum, Cabella et al [162] found \(f_{\text{local}}^{\text{NL}} = 35 ± 42\) and \(38 ± 47\) (68\% CL) with and without the foreground marginalization, respectively.
7.2. Trispectrum

The optimal estimators for the trispectrum have not been implemented, largely because they are computationally demanding. While the first measurements of the angular bispectrum were made from the COBE 4-year data [150] by the authors of [2, 165] in 2001, limits on the physical parameters have not been obtained from the direct trispectrum analysis.

Recently, Smidt et al [76] used the sub-optimal estimator developed in [166] (which becomes optimal in the limit that the instrumental noise is isotropic) and found the 95% CL limits of $-7.4 \times 10^5 < \tau_{NL} < 8.2 \times 10^5$ and $-0.6 \times 10^4 < f_{NL} < 3.3 \times 10^4$. The current limit is consistent with the Suyama–Yamaguchi inequality, $\tau_{NL} \geq \left(6 f_{NL}^{local}/5\right)^2$.

7.3. Other statistical methods

While the optimal estimators for the $f_{NL}$ parameters (with the minimum variance) must be constructed from the PDF as in section 3, there are various other ways of constraining non-Gaussianity. While these other methods are usually sub-optimal, they serve as useful diagnosis tools of the results obtained from the direct bispectrum and trispectrum methods. In some cases, they are easier to implement than the optimal estimators.

A major progress in the topological Gaussianity test using the Minkowski functionals [167–171] since 2004 is the derivation and implementation of the analytical formula for the Minkowski functionals of the CMB [172, 173]. This method has been applied to the WMAP data [174, 175] as well as to the BOOMERanG data [176]. The Planck data are expected to reach the 68% limit of $\Delta f_{NL}^{local} = 20$ [172], which is worse than the limit from the optimal method $\Delta f_{NL}^{local} = 5$ [83]. An advantage of the Minkowski functionals is that the measurements of the Minkowski functionals do not depend on the models, and thus the computational cost is the same for all models. This allows one to obtain limits on various models, for which the optimal estimators are difficult to implement. For example, a limit on the primordial non-Gaussianity in the isocurvature perturbation is currently available only from the Minkowski functionals [175].

Instead of expanding the temperature anisotropy into spherical harmonics, one may choose to expand it using a different basis. One popular basis used in the CMB community is the so-called Spherical Mexican Hat Wavelet (SMHW). See [177, 178] for reviews on this method. A major progress in this method is the realization that the 3-point function of the wavelet coefficients made of large and small smoothing scales is nearly an optimal estimator for the local-form bispectrum: when only the adjacent scales are included, Curto et al [179] found $-8 < f_{NL}^{local} < 111$ (95% CL) from the WMAP 5-year data. When all the scales (including large- and small-scale combinations) are included in the analysis, the limit improved significantly to $-18 < f_{NL}^{local} < 80$ (95% CL) [180], which is similar to the optimal limit from the 5-year data, $-4 < f_{NL}^{local} < 80$ (95% CL) [11]. An advantage of the SMHW is that it retains information on the spatial distribution of the signal. This property can be used to measure $f_{NL}^{local}$ as a function of positions on the sky [180]. In addition, the analytical formula for the SMHW as a function of $f_{NL}^{local}$ has been derived and implemented [197]. See [181, 182] for earlier limits on $f_{NL}^{local}$ from the SMHW.

Another form of spherical wavelets that has been used to constrain $f_{NL}^{local}$ is the spherical needlets [183]. The limits on $f_{NL}^{local}$ are reported in [162, 184, 185]. This method also allows one to look for a spatial variation in $f_{NL}^{local}$, and the results are reported in [186, 187]. For the other types of wavelets considered in the literature, see [188, 189] and references therein.

Many other statistical methods have been proposed and used for constraining $f_{NL}^{local}$ in the literature. An incomplete list of references is as follows: [190, 191] on the real-space 3-point
function: [192] on the integrated bispectrum; [193] on the 2-1 cumulant correlator; [182] on the local curvature; and [75, 194] on the N-point PDF. Also see references therein. While we have not listed the statistical methods that have not been used to constrain the primordial non-Gaussianity parameters yet, there are many other methods proposed in a general context in the literature.

8. Conclusion

Since the last review articles on signatures of primordial non-Gaussianity in the CMB were written in 2001 [2] and 2004 [3], a lot of progress has been made in this field. The current standard lore may be summarized as follows.

1. **Shape and physics.** Different aspects of the physics of the primordial universe appear in different shapes of 3- and 4-point functions.

2. **Importance of local shape.** Of these shapes, the local shapes have special significance: a significant detection of the local-form bispectrum (with \( f_{\text{local}}^{NL} \gg 1 \)) would rule out all single-field inflation models, and the local-form trispectrum can be used to rule out a broad class of multi-field models (if not all multi-field models) by testing \( \tau_{\text{NL}} \geq (6 f_{\text{local}}^{NL} / 5)^2 \).

3. **Optimal estimators.** The optimal estimators of the bispectrum and trispectrum can be derived systematically from the expansion of the PDF. The optimal bispectrum estimator has been implemented.

4. **Secondary.** The most serious contamination of \( f_{\text{local}}^{NL} \) is due to the lensing–ISW coupling, which can be removed by using the template given in section 4.

5. **Foreground.** The Galactic foreground contamination is minimal for \( f_{\text{NL}}^{\text{equil}} \) and \( f_{\text{NL}}^{\text{orthog}} \), but it can be as large as \( f_{\text{NL}}^{\text{local}} \sim 10 \) for the local-form bispectrum. This must be carefully studied and eliminated in the Planck data analysis. The random (Poisson) point-source contamination can be removed by using the template given in section 3.2.

Some outstanding issues for the ‘CMB and primordial non-Gaussianity’ include the following.

1. **Second order.** (In Newtonian gauge) the products of the first-order terms and the intrinsically second-order terms in the sub-horizon limit do not contaminate the local-form bispectrum very much \( (\Delta f_{\text{NL}}^{\text{local}} < 1) \). However, would the post-Newtonian effect give \( \Delta f_{\text{NL}}^{\text{local}} \sim 5 \), as found in [138]? If so, we need to construct a template for this effect.

2. **More foreground.** How can we model the non-Poisson (clustered) point source bispectrum? How about the foreground (and secondary) contamination of the primordial trispectrum?

3. **Trispectrum estimators.** How can we implement the optimal trispectrum estimators for both local and non-local shapes?

These issues would become important when the Planck data are analyzed in search of primordial non-Gaussianity. The Planck is expected to reduce the uncertainty in \( f_{\text{local}}^{NL} \) by a factor of 4 compared to the current limit, \( f_{\text{local}}^{NL} = 32 \pm 21 \) (68% CL). If the Planck detected \( f_{\text{local}}^{NL} \sim 30 \), then the trispectrum would provide an important test of multi-field models. In particular, if \( f_{\text{local}}^{NL} \gg 1 \) and \( \tau_{\text{NL}} < (6 f_{\text{local}}^{NL} / 5)^2 \) are found, then all single-field models and many (if not all) multi-field models would be ruled out, and thus the standard paradigm of inflation as the origin of fluctuations would face a serious challenge.

However, do not despair even if the Planck did not detect the primordial bispectrum or trispectrum—while the CMB may end its leading role as a probe of primordial non-Gaussianity (unless the next-generation, comprehensive CMB satellite which can measure...
both the temperature and polarization to the cosmic-variance-limited precision is funded [195]), the large-scale structure of the universe would eventually take over and substantially reduce the uncertainties in the local-form parameters such as $f_{\text{NL}}$, $g_{\text{NL}}$ and $T_{\text{NL}}$ (see [1]).

**Acknowledgments**

We thank the organizers and participants of ‘The Non-Gaussian Universe’ workshop at Yukawa Institute for Theoretical Physics (YITP) for stimulating discussion and presentations, D Jeong for making figures 3 and 4, D Nitta for calculating $f_{\text{local,}}_{\text{NL}}$ from the second-order $/Phi_{1}$ in the Newtonian limit and T Takahashi for valuable comments on section 6. This work is supported in part by an NSF grant PHY-0758153.

**References**

[1] Desjacques V and Seljak U 2010 *Class. Quantum Grav.* 27 124011

[2] Komatsu E 2001 PhD thesis Tohoku University (arXiv:astro-ph/0206039)

[3] Bartolo N, Komatsu E, Matarrese S and Riotto A 2004 *Phys. Rep.* 402 103–266 (arXiv:astro-ph/0406398)

[4] Maldacena J M 2003 *J. High Energy Phys.* JHEP05(2003)013 (arXiv:astro-ph/0210603)

[5] Acquaviva V, Bartolo N, Matarrese S and Riotto A 2003 *Nucl. Phys.* B667 119–48 (arXiv:astro-ph/0209156)

[6] Creminelli P and Zaldarriaga M 2004 *J. Cosmol. Astropart. Phys.* JCAP10(2004)006 (arXiv:astro-ph/0407059)

[7] Komatsu E, Spergel D N and Wandelt B D 2005 *Astrophys. J.* 634 14–9

[8] Creminelli P, Nicolis A, Senatore L, Tegmark M and Zaldarriaga M 2006 *J. Cosmol. Astropart. Phys.* JCAP05(2006)004 (arXiv:astro-ph/0509029)

[9] Smith K M and Zaldarriaga M 2006 arXiv:astro-ph/0612571

[10] Yadav A P S, Komatsu E, Wandelt B D, Liguori M, Hansen F K and Matarrese S 2008 *Astrophys. J.* 678 578–82 (arXiv:0711.4933)

[11] Smith K M, Senatore L and Zaldarriaga M 2009 *J. Cosmol. Astropart. Phys.* JCAP09(2009)006 (arXiv:0901.2572)

[12] Komatsu E et al 2010 *Astrophys. J.* Suppl. (at press) (arXiv:1001.4538)

[13] Goldberg D M and Spergel D N 1999 *Phys. Rev. D* 59 103002

[14] Verde L and Spergel D N 2002 *Phys. Rev. D* 65 043007 (arXiv:astro-ph/0108179)

[15] Serra P and Cooray A 2008 *Phys. Rev. D* 77 103035 (arXiv:0801.3276)

[16] Hanson D, Smith K M, Challinor A and Liguori M 2009 *Phys. Rev. D* 80 043004 (arXiv:0905.4732)

[17] Mangilli A and Verde L 2009 *Phys. Rev. D* 80 123007 (arXiv:0906.2317)

[18] Creminelli P 2003 *J. Cosmol. Astropart. Phys.* JCAP10(2003)003 (arXiv:astro-ph/0306122)

[19] Babich D, Creminelli P and Zaldarriaga M 2004 *J. Cosmol. Astropart. Phys.* JCAP06(2004)009 (arXiv:astro-ph/0407056)

[20] Chen X, Huang M and Kachru S and Shiu G 2007 *J. Cosmol. Astropart. Phys.* JCAP01(2007)002 (arXiv:hep-th/0605045)

[21] Lyth D H, Unal E and Wands D 2003 *Phys. Rev. D* 67 25303

[22] Zaldarriaga M 2004 *Phys. Rev. D* 69 043508 (arXiv:astro-ph/0310606)

[23] Lyth D H and Rodriguez Y 2005 *Phys. Rev. Lett.* 95 071301 (arXiv:astro-ph/0404045)

[24] Lehners J 2008 *Phys. Rev.* D76 083004 (arXiv:0806.1345)

[25] Bond J R, Frolow A V, Huang Z and Kofman L 2009 *Phys. Rev. Lett.* 103 101301 (arXiv:0903.3407)

[26] Chambers A, Nomura S and Rajantie A 2010 *J. Cosmol. Astropart. Phys.* JCAP01(2010)012 (arXiv:0909.4535)

[27] Alishahiha M, Silverstein E and Tong D 2009 *Phys. Rev. D* 79 043516 (arXiv:0904.04084)

[28] Arkani-Hamed N, Creminelli P, Mukohyama S and Zaldarriaga M 2004 *J. Cosmol. Astropart. Phys.* JCAP04(2004)001 (arXiv:hep-th/0312100)

[29] Seery D and Lidsey J E 2005 *J. Cosmol. Astropart. Phys.* JCAP06(2005)003 (arXiv:astro-ph/0503692)

[30] Holman R and Tolley A J 2008 *J. Cosmol. Astropart. Phys.* JCAP08(2008)001 (arXiv:0710.1302)

[31] Meerburg P D, van der Schaar J P and Corasaniti P S 2009 *J. Cosmol. Astropart. Phys.* JCAP05(2009)018 (arXiv:0901.4044)

[32] Senatore L, Smith K M and Zaldarriaga M 2010 *J. Cosmol. Astropart. Phys.* JCAP01(2010)028 (arXiv:0905.3746)
[179] Curto A, Martínez-González E, Mukherjee P, Barreiro R B, Hansen F K, Liguori M and Matarrese S 2009
Mon. Not. R. Astron. Soc. 393 615–22

[180] Curto A, Martínez-González E and Barreiro R B 2009 Astrophys. J. 706 399–403 (arXiv:0902.1523)

[181] Mukherjee P and Wang Y 2004 Astrophys. J. 613 51–60 (arXiv:astro-ph/0402602)

[182] Cabella P et al 2005 Mon. Not. R. Astron. Soc. 358 684–92 (arXiv:astro-ph/0406026)

[183] Marinucci D, Pietrobon D, Balbi A, Baldi P, Cabella P, Kerkyacharian G, Natoli P, Picard D and Vittorio N
2008 Mon. Not. R. Astron. Soc. 383 539–45 (arXiv:0707.0844)

[184] Rudjord O, Groeneboom N E, Eriksen H K, Huey G, Górski K M and Jewell J B 2009 Astrophys. J. 692 1669–77
(arXiv:0809.4624)

[185] Pietrobon D, Cabella P, Balbi A, de Gasperis G and Vittorio N 2009 Mon. Not. R. Astron. Soc. 396 1682–8
(arXiv:0812.2478)

[186] Rudjord O et al 2010 Astrophys. J. 708 1321–5 (arXiv:0906.3232)

[187] Pietrobon D, Cabella P, Balbi A, Crittenden R, de Gasperis G and Vittorio N 2010 Mon. Not. R. Astron.
Soc. 402 L34–8 (arXiv:0905.3702)

[188] Aghanim N, Kunz M, Castro P G and Forni O 2003 Astron. Astrophys. 406 797–816 (arXiv:astro-ph/0301220)

[189] Jin J, Starck J, Donoho D L, Aghanim N and Forni O 2005 J. Appl. Signal Process. 15 2470
(arXiv:astro-ph/0503374)

[190] Gaztañaga E and Wagg J 2003 Phys. Rev. D 68 021302

[191] Chen G and Szapudi I 2005 Astrophys. J. 635 743–9 (arXiv:astro-ph/0508316)

[192] Cabella P, Hansen F K, Liguori M, Marinucci D, Matarrese S, Moscardini L and Vittorio N 2006 Mon. Not.
R. Astron. Soc. 369 819–24 (arXiv:astro-ph/0512112)

[193] Chen G and Szapudi I 2006 Astrophys. J. Lett. 647 L87–90 (arXiv:astro-ph/0606394)

[194] Vielva P and Sanz J L 2009 Mon. Not. R. Astron. Soc. 397 837–48

[195] Baumann D et al (CMBPol Study Team) 2009 AIP Conf. Proc. 1141 10–120 (arXiv:0811.3919)

[196] Nitta D 2010 private communication

[197] Curto A 2009 private communication