Primordial non-Gaussianity and the CMB bispectrum

J. R. Fergusson and E. P. S. Shellard

Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge,

Wilberforce Road, Cambridge CB3 0WA, United Kingdom

(Received 29 July 2007; published 26 October 2007)

We present a new formalism, together with efficient numerical methods, to directly calculate the cosmic microwave background (CMB) bispectrum today from a given primordial bispectrum using the full linear radiation transfer functions. Unlike previous analyses which have assumed simple separable *Ansätze* for the bispectrum, this work applies to a primordial bispectrum of almost arbitrary functional form, for which there may have been both horizon-crossing and superhorizon contributions. We employ adaptive methods on a hierarchical triangular grid and we establish their accuracy by direct comparison with an exact analytic solution, valid on large angular scales. We demonstrate that we can calculate the full CMB bispectrum to greater than 1% precision out to multipoles l < 1800 on reasonable computational time scales. We plot the bispectrum for both the superhorizon ("local") and horizon-crossing ("equilateral") asymptotic limits, illustrating its oscillatory nature which is analogous to the CMB power spectrum.

DOI: 10.1103/PhysRevD.76.083523

PACS numbers: 98.80.Es, 98.80.Cq

I. INTRODUCTION

Measurements of the cosmic microwave background (CMB) radiation are now strongly constraining the viability of a variety of cosmological theories for the formation of large-scale structure. While the simplest "vanilla" models of single-field inflation appear sufficient to explain current data [1], forthcoming experiments will test this concordance further. One area in which we can expect significant improvement will be in estimates of non-Gaussianity. Single-field inflation predicts that variations in the CMB should follow simple Gaussian statistics to high accuracy. On the other hand, more natural inflation scenarios which involve multiple fields and more complex potentials can, in principle, produce large non-Gaussian signatures that may become measurable in future experiments (see, for example, Refs. [2-6]). Of course, the CMB power spectrum is not sensitive to these non-Gaussianities so we must look to other higher order correlators. The bispectrum is the three-point correlator of the multipoles and we note that statistics based on it have been shown to be optimal for testing non-Gaussianity for a variety of inflationary scenarios [7,8].

The aim of this paper is to provide a framework for evolving the primordial bispectrum (the bispectrum at the end of inflation) until it leaves an observable imprint in the CMB today. Current theory dictates this can be done by solving a four-dimensional integral over the primordial bispectrum multiplied by six highly oscillatory functions. In practice this is not possible to evaluate numerically. In Sec. II we detail the current most popular approach [9] which assumes a simple separable form for the bispectrum, so that the integral separates and becomes tractable. These methods are good for obtaining quick estimates but their applicability is limited by their ability to fit a suitable analytic approximation for the shape of the primordial bispectrum. In Sec. III, we detail a different more general approach. As the primordial bispectrum is the most important ingredient in the calculation, we place no restrictions on it other than an overall separable scale dependence. From here we reduce the four-dimensional integral to a two-dimensional integral over the primordial bispectrum multiplied by two one-dimensional integrals, one model dependent and one purely geometric.

In Sec. IV, we investigate analytic solutions for the bispectrum. There are two asymptotic categories for the shape function-"local" and "equilateral"-depending on whether the non-Gaussianity is created primarily on superhorizon scales or at horizon crossing, respectively. In the large-angle approximation and for the simplest form of the local bispectrum, there exists an exact solution which can be used to test the accuracy of our numerical methods. While there is no similar solution for the equilateral scenario, we argue that our methods should be more precise in this case. We have developed code that will accurately evolve the bispectrum from the end of inflation until today, evaluating its angular counterpart in the CMB. Note that our primary focus here is on the effect of primordial non-Gaussianity on the CMB, so it suffices to use linear radiation transfer functions for this purpose. In Sec. V, we detail the inner workings of the code and point out the key innovations required to compute the integral quickly and accurately. In Sec. VI, we test against the local analytic solution and demonstrate an accuracy of less than 1% for multipole values less than l < 550 (using the idealized large-angle transfer function). We also study the bispectrum for the equilateral case and find it has better convergence with an error of less than 1% for l values, l < l1300. With the correct radiation transfer functions used at small angles, we find that calculations of the full CMB bispectrum are numerically tractable at large l and that the errors remain below 1% out to multipoles l < 1800. We show that growing errors for even larger l are linked directly to the resolution and the asymptotic cutoffs em-

J. R. FERGUSSON AND E. P. S. SHELLARD

ployed. If the parameters governing these are tightened, arbitrary accuracy can be achieved though at some computational cost. We provide plots of the CMB bispectrum for the full radiation transfer function for both the local and equilateral cases, showing oscillatory behavior analogous to the angular power spectrum. Having established the accuracy and efficiency of these methods, we detail future directions for this work in Sec. VII.

II. BISPECTRUM

We will quickly review the calculation for the bispectrum today from a given primordial bispectrum, this section loosely follows the derivation in [10]. The temperature fluctuations in the CMB can be decomposed into spherical harmonics, Y_{lm}

$$\frac{\Delta T}{T}(\hat{\mathbf{n}}) = \sum_{lm} a_{lm} Y_{lm}(\hat{\mathbf{n}}).$$
(1)

We can invert this expression to find the a_{lm} 's using the spherical harmonic orthogonality condition,

$$a_{lm} = \int d\hat{\mathbf{n}} \frac{\Delta T}{T}(\hat{\mathbf{n}}) Y_{lm}^*(\hat{\mathbf{n}}).$$
(2)

The temperature anisotropy can be decomposed into Legendre polynomials $P_l(\mu)$ in k space,

$$\frac{\Delta T}{T}(\hat{\mathbf{n}}) = \int \frac{d^3k}{(2\pi)^3} \sum_{l=0}^{\infty} (-i)^l (2l+1) \Psi(\mathbf{k}) \Delta_l(k) P_l(\hat{\mathbf{k}} \cdot \hat{\mathbf{n}}),$$
(3)

where $\Psi(\mathbf{k})$ is the primordial gravitational-potential perturbation and $\Delta_l(k)$ is the radiation transfer function. We replace the Legendre polynomials using the spherical harmonic addition theorem,

$$P_l(\hat{\mathbf{k}}\cdot\hat{\mathbf{n}}) = \frac{4\pi}{2l+1} \sum_{m=-l}^{l} Y_{lm}(\hat{\mathbf{k}}) Y_{lm}^*(\hat{\mathbf{n}}).$$
(4)

Making these substitutions, and using the orthogonality of the spherical harmonics, we have

$$a_{lm} = \int \frac{d^{3}k}{(2\pi)^{3}} d^{3}n \sum_{l'} (-i)^{l'} (2l'+1) \Psi(k) \Delta_{l'}(k) P_{l'}(\mu) Y_{lm}^{*}(\mathbf{n})$$

$$= \int \frac{d^{3}k}{(2\pi)^{3}} 4\pi \sum_{l'} (-i)^{l'} \Psi(k) \Delta_{l'}(k) \sum_{m'} Y_{l'm'}^{*}(\hat{\mathbf{k}})$$

$$\times \int d^{3}n Y_{l'm'}(\mathbf{n}) Y_{lm}^{*}(\mathbf{n})$$

$$= 4\pi (-i)^{l} \int \frac{d^{3}k}{(2\pi)^{3}} \Psi(k) \Delta_{l}(k) Y_{lm}^{*}(\hat{\mathbf{k}}).$$
(5)

The bispectrum is defined as the three-point correlator of the a_{lm} 's,

$$B_{l_{1}l_{2}l_{3}}^{m_{1}m_{2}m_{3}} = \langle a_{l_{1}m_{1}}a_{l_{2}m_{2}}a_{l_{3}m_{3}} \rangle$$

$$= (4\pi)^{3}(-i)^{l_{1}+l_{2}+l_{3}} \int \frac{d^{3}\mathbf{k}_{1}}{(2\pi)^{3}} \frac{d^{3}\mathbf{k}_{2}}{(2\pi)^{3}} \frac{d^{3}\mathbf{k}_{3}}{(2\pi)^{3}} \\ \times \langle \Psi(\mathbf{k}_{1})\Psi(\mathbf{k}_{2})\Psi(\mathbf{k}_{3}) \rangle \Delta_{l_{1}}(k_{1})\Delta_{l_{2}}(k_{2})\Delta_{l_{3}}(k_{3}) \\ \times Y_{l_{1}m_{1}}^{*}(\hat{\mathbf{k}}_{1})Y_{l_{2}m_{2}}^{*}(\hat{\mathbf{k}}_{2})Y_{l_{3}m_{3}}^{*}(\hat{\mathbf{k}}_{3}),$$
(6)

where $k_1 = |\mathbf{k}_1|$, $k_2 = |\mathbf{k}_2|$, and $k_3 = |\mathbf{k}_3|$. The primordial bispectrum is defined as

$$\langle \Psi(\mathbf{k}_1)\Psi(\mathbf{k}_2)\Psi(\mathbf{k}_3)\rangle = (2\pi)^3 F(k_1, k_2, k_3)\delta(\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3),$$
(8)

where $F(k_1, k_2, k_3)$ is the primordial bispectrum shape. We use the integral form of the delta function,

$$\delta(\mathbf{k}) = \frac{1}{(2\pi)^3} \int e^{i\mathbf{k}\cdot\mathbf{x}} d^3x, \qquad (9)$$

and we replace the exponential with spherical harmonics using the Rayleigh expansion,

$$e^{i\mathbf{k}_{1}\cdot\mathbf{x}} = 4\pi \sum_{l} i^{l} j_{l}(k_{1}x) \sum_{m} Y_{lm}(\hat{\mathbf{k}}_{1}) Y_{lm}^{*}(\hat{\mathbf{x}}), \qquad (10)$$

where $x = |\mathbf{x}|$ and the unit vector $\hat{\mathbf{x}} = \mathbf{x}/x$. Substituting, then using the orthogonality of the spherical harmonics and the Gaunt integral,

$$\begin{aligned} \mathcal{G}_{m_{1}m_{2}m_{3}}^{l_{1}l_{2}l_{3}} &= \int d\Omega Y_{l_{1}m_{1}}Y_{l_{2}m_{2}}Y_{l_{3}m_{3}} \\ &= \sqrt{\frac{(2l_{1}+1)(2l_{2}+1)(2l_{3}+1)}{4\pi}} \begin{pmatrix} l_{1} & l_{2} & l_{3} \\ 0 & 0 & 0 \end{pmatrix} \\ &\times \begin{pmatrix} l_{1} & l_{2} & l_{3} \\ m_{1} & m_{2} & m_{3} \end{pmatrix}, \end{aligned}$$
(11)

where

$$\begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}$$

is the Wigner 3j symbol. We find the bispectrum becomes

$$B_{l_{1}l_{2}l_{3}}^{m_{1}m_{2}m_{3}} = \left(\frac{2}{\pi}\right)^{3} G_{m_{1}m_{2}m_{3}}^{l_{1}l_{2}l_{3}} \int dx dk_{1} dk_{2} dk_{3} (xk_{1}k_{2}k_{3})^{2} \\ \times F(k_{1}, k_{2}, k_{3}) \Delta_{l_{1}}(k_{1}) \Delta_{l_{2}}(k_{2}) \Delta_{l_{3}}(k_{3}) j_{l_{1}}(k_{1}x) \\ \times j_{l_{2}}(k_{2}x) j_{l_{3}}(k_{3}x).$$
(12)

It is worth noting that the Wigner 3j symbol places some restrictions on the l values. First, no individual l can be greater than the sum of the remaining two (the triangle condition) and, second, that the sum of all three must be even. As we assume statistical isotropy it is more common to work with the angle averaged bispectrum,

$$B_{l_1 l_2 l_3} = \sum_{m_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} B_{l_1 l_2 l_3}^{m_1 m_2 m_3}.$$
 (13)

Observing that

$$\sum_{n_1 m_2 m_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix}^2 = 1,$$
(14)

we have

$$B_{l_1 l_2 l_3} = \left(\frac{2}{\pi}\right)^3 \sqrt{\frac{(2l_1+1)(2l_2+1)(2l_3+1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\times \int dx dk_1 dk_2 dk_3 (xk_1 k_2 k_3)^2 F(k_1, k_2, k_3)$$
$$\times \Delta_{l_1}(k_1) \Delta_{l_2}(k_2) \Delta_{l_3}(k_3) j_{l_1}(k_1 x) j_{l_2}(k_2 x) j_{l_3}(k_3 x).$$
(15)

Also noticing that the Gaunt integral represents a purely geometrical factor which is independent of the primordial bispectrum, we define

$$B_{l_1 l_2 l_3}^{m_1 m_2 m_3} = \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} b_{l_1 l_2 l_3}, \tag{16}$$

where $b_{l_1 l_2 l_3}$ is the reduced bispectrum,

$$b_{l_1 l_2 l_3} = \left(\frac{2}{\pi}\right)^3 \int dx dk_1 dk_2 dk_3 (xk_1 k_2 k_3)^2 F(k_1, k_2, k_3) \\ \times \Delta_{l_1}(k_1) \Delta_{l_2}(k_2) \Delta_{l_3}(k_3) j_{l_1}(k_1 x) j_{l_2}(k_2 x) j_{l_3}(k_3 x).$$
(17)

This is the quantity that we will be calculating.

To proceed in general we must solve a four-dimensional integral of the shape function multiplied by six highly oscillatory functions, Δ_l and j_l , which is not feasible in practice. Current approaches overcome this by assuming a simple separable form for the shape function. The most common is that developed in Ref. [9] where the primary assumption is that the primordial gravitational-potential perturbation can be approximated by

$$\Psi(x) \approx \Psi_L(x) + f_{NL}(\Psi_L^2(x) - \langle \Psi_L^2(x) \rangle), \quad (18)$$

with $\Psi_L(x)$ the linear Gaussian part and f_{NL} a constant. Following this through gives a shape function of

$$F(k_1, k_2, k_3) = 2f_{NL}(P^{\Psi}(k_1)P^{\Psi}(k_2) + P^{\Psi}(k_2)P^{\Psi}(k_3) + P^{\Psi}(k_3)P^{\Psi}(k_1)).$$
(19)

Substituting this into Eq. (17) we can now separate the integral into

$$\int x^2 dx b_{l_1}^L(x) b_{l_2}^L(x) b_{l_3}^{NL}(x) + \text{ perms}$$
(20)

with

$$b_l^L(x) = \frac{2}{\pi} \int k^2 dk P^{\Psi}(k) \Delta_l(k) j_l(kx)$$
(21)

$$b_l^{NL}(x) = \frac{2}{\pi} f_{NL} \int k^2 dk \Delta_l(k) j_l(kx), \qquad (22)$$

that is, it becomes products of one-dimensional integrals which are manageable numerically.

Searches for the bispectrum in the CMB have consequently focused on establishing limits on f_{NL} via purposebuilt statistics rather than calculation of the bispectrum directly. The WMAP team has constrained non-Gaussianity [1] by narrowing their analysis to the local Ansatz (18), estimating the limits to be $-54 < f_{NL} < 114$. (This analysis has been repeated in Ref. [8] with an improved estimator to obtain the limits $-36 < f_{NL} < 100.$) Now primordial non-Gaussianity can be broadly classified into either of two asymptotic regimes, local or equilateral [11]. The local case describes models in which the non-Gaussianity is produced outside the Hubble radius ("superhorizon") with examples including the curvaton model [2], preheating [12,13], and multiple field inflation models [3]. The equilateral case describes the case in which the primary contribution to non-Gaussianity arises just as the perturbations leave the Hubble radius ("horizon crossing"), with examples including Dirac-Born-Infeld (DBI) model [5] and ghost inflation [6]. Following the "local" WMAP approach described above, there has also been a parallel analysis for the equilateral case [14]. The key element shared by this approach is an analytic approximation for the shape function that allows the separation of the bispectrum integral into the product of three onedimensional integrals over k_i inside an integral over x as in (20). They then proceed by generating a statistic sensitive to their Ansatz to establish limits on a generalization of the non-Gaussianity parameter f_{NL} . They find considerably weaker constraints on f_{NL} in the equilateral case, $-256 < f_{NL} < 332$ [8], but argue that this represents approximately the same "level of non-Gaussianity." Other recent work in this area includes Ref. [15] where, again using the separable local Ansatz (20), the particular scale dependence of f_{NL} generated by postinflationary gravitational evolution is incorporated. We note also for the local assumption (18) in Ref. [16], map-making methods are developed to create individual realizations incorporating the bispectrum and higher-order correlators.

In this paper we have adopted a different more general approach which does not rely on separable *Ansätze* giving rise to the triple decomposition (20). Rather than generating approximations to the bispectrum to allow for easy analytic calculation, instead we have written code that will directly solve the integral for any shape function assuming only the separation of an overall scale dependence. We anticipate that this analysis will encompass all viable inflationary models.

III. NEW PARAMETRIZATION

By looking at the expression for the reduced bispectrum (17), we can rearrange the terms as follows:

$$\left(\frac{2}{\pi}\right)^{3} \int dk_{1} dk_{2} dk_{3} (k_{1} k_{2} k_{3})^{2} F(k_{1}, k_{2}, k_{3}) \Delta_{l_{1}}(k_{1}) \Delta_{l_{2}}(k_{2}) \times \Delta_{l_{3}}(k_{3}) \left(\int x^{2} dx j_{l_{1}}(k_{1} x) j_{l_{2}}(k_{2} x) j_{l_{3}}(k_{3} x)\right).$$
(23)

The integral in the brackets arises as a direct result of the delta function and analytic solutions have been found, [17–19]. These solutions demonstrate that the three lengths k_1 , k_2 , k_3 must be such as to be able to form a triangle, as expected from the delta function, which restricts us to the region of k space highlighted in Fig. 1. The analytic solutions also state that the value on the boundary is half that of a point immediately interior. It is natural to reparametrize this region into triangular slices [20],

$$k_{1} = ka = k(1 - \beta) \qquad k_{2} = kb = \frac{1}{2}k(1 + \alpha + \beta)$$
$$k_{3} = kc = \frac{1}{2}k(1 - \alpha + \beta), \qquad (24)$$

where *a*, *b*, *c* are shorthand for the corresponding expressions involving α , β . They are not independent of each other as a + b + c = 2. Here, the overall wave number scale is given by the semiperimeter, $k \equiv \frac{1}{2}(k_1 + k_2 + k_3)$. The surface k = const defines a plane with normal (1, 1, 1) at a distance $\frac{2}{\sqrt{3}}k$ from the origin and α , β parametrize our position on that plane. The new parameters have the following domains $0 \le k < \infty$, $0 \le \beta \le 1$, and $-(1 - \beta) \le \alpha \le 1 - \beta$. This parametrization has volume element $dk_1 dk_2 dk_3 = k^2 dk d\alpha d\beta$ (representing a minor improvement over the original triangular parametrization given in [20]). Making these substitutions we have



FIG. 1 (color online). Region of integration for the bispectrum calculation. The radial (red) lines through the origin are $k_1 = k_2$, $k_3 = 0$; $k_2 = k_3$, $k_1 = 0$; $k_3 = k_1$, $k_2 = 0$ and the region of integration is shaded (yellow). This area can be parametrized into slices represented by the triangle (green) formed from the end points of the three radial (red) lines and the distance $\frac{2}{\sqrt{3}}k$ of the center of the triangle from the origin.

$$\left(\frac{2}{\pi}\right)^{3} \int k^{2} dk d\alpha d\beta (abc)^{2} k^{6} F(ak, bk, ck) \Delta_{l_{1}}(ak) \Delta_{l_{2}}(bk)$$
$$\times \Delta_{l_{3}}(ck) \left(\int x^{2} dx j_{l_{1}}(akx) j_{l_{2}}(bkx) j_{l_{3}}(ckx)\right), \qquad (25)$$

and, if we rescale x to absorb k, we have

$$\left(\frac{2}{\pi}\right)^{3} \int d\alpha d\beta (abc)^{2} \\
\times \left(\int \frac{dk}{k} k^{6} F(ak, bk, ck) \Delta_{l_{1}}(ak) \Delta_{l_{2}}(bk) \Delta_{l_{3}}(ck)\right) \\
\times \left(\int x^{2} dx j_{l_{1}}(ax) j_{l_{2}}(bx) j_{l_{3}}(cx)\right).$$
(26)

Given strong observational limits on the scalar tilt we expect the shape function to exhibit near scale invariance [1,11], so we make the substitution

$$F(k_1, k_2, k_3) \approx \frac{k^n}{k^6} F(\alpha, \beta), \qquad (27)$$

where *n* is the bispectrum tilt which is expected to be small. This is the only restriction we place on the shape function, up until this point the discussion has been entirely general. In fact this is slightly over restrictive and we should note that the method continues to work if we replace k^n with a more general overall scale dependence f(k). Hence, we can rewrite (23) as

$$\left(\frac{2}{\pi}\right)^3 \int d\alpha d\beta F^{\rm SI}(\alpha,\beta) I^T(\alpha,\beta) I^G(\alpha,\beta), \quad (28)$$

where

$$I^{T}(\alpha,\beta) \equiv \int \Delta_{l_{1}}(ak) \Delta_{l_{2}}(bk) \Delta_{l_{3}}(ck) k^{n} \frac{dk}{k}$$
(29)

$$I^{G}(\alpha,\beta) \equiv \int j_{l_{1}}(ax)j_{l_{2}}(bx)j_{l_{3}}(cx)x^{2}dx \qquad (30)$$

$$F^{\rm SI}(\alpha,\beta) \equiv (abc)^2 F(\alpha,\beta), \qquad (31)$$

where we have defined the "scale-independent" shape function, F^{SI} . (A similar concept was discussed in Ref. [11], but can be contrasted with the definition (31) in our more symmetric triangular parametrization.)

The end result is that we have broken up the fourdimensional integral for the reduced bispectrum (17) into a two-dimensional integral (28), restricted to an equilateral triangle, over the scale-independent shape multiplied by two one-dimensional integrals (29) and (30). The first $I^{T}(\alpha, \beta)$ is over the transfer functions and the second is purely geometric $I^{G}(\alpha, \beta)$. These two one-dimensional integrands consist entirely of highly oscillatory functions. However, as we shall discuss later, since they are only one dimensional it is feasible to evaluate them accurately numerically, freeing us from the severe difficulties previously encountered.

The two basic asymptotic limits (local and equilateral) for the scale-independent shape can be approximated by the analytic expressions [14]:

PHYSICAL REVIEW D 76, 083523 (2007)



FIG. 2 (color online). F^{SI} plotted on the $\alpha\beta$ triangle for the "local" superhorizon shape (left) and the "equilateral" horizon-crossing shape (right). The equilateral case has been scaled so that the centers of both plots are at the same height.

$$F_{\text{local}}^{\text{SI}}(a, b, c) = \frac{a^3 + b^3 + c^3}{abc}$$
(32)

$$F_{\text{equilateral}}^{\text{SI}}(a, b, c) = \frac{(1-a)(1-b)(1-c)}{abc}.$$
 (33)

We have plotted these two bispectrum shapes in Fig. 2. While much emphasis has been placed on these two limits, we note in passing the caveat that even single-field inflation yields a more complicated non-Gaussian shape [21].

We will take this opportunity to define a general momentum-dependent f_{NL} that will work in both the local and the equilateral cases (as in [22])

$$2f_{NL}(k_1, k_2, k_3) = \frac{F(k_1, k_2, k_3)}{P^{\Psi}(k_1)P^{\Psi}(k_2) + P^{\Psi}(k_2)P^{\Psi}(k_3) + P^{\Psi}(k_3)P^{\Psi}(k_1)}.$$
(34)

This agrees with the definitions given in both [9,14], given their respective assumptions, but it can now be applied across the full k_i parameter space. Further, if we make the same parameter substitutions as those used in obtaining (28), we see that f_{NL} has an overall scale dependence given by $n_{NL} = n + 2(1 - n_s)$ where n_s is the usual scalar spectral index of the power spectrum.

IV. ANALYTIC RESULTS

As well as the analytic approximations for the shape function in the local and the equilateral limits (32), on large angles ($l \ll 200$) the transfer function can be approximated by

$$\Delta_l(k) = \frac{1}{3} j_l(\Delta \eta k), \tag{35}$$

where $\Delta \eta$ is the conformal time period between today η_0 and the surface of last scattering η_{dec} . Substituting these into the bispectrum, the whole integrand can be expressed in closed form and so we can look for analytic solutions. In the local case, we make the replacement

$$F^{\rm SI}(k_1, k_2, k_3) = \frac{k_1^2}{k_2 k_3} + \frac{k_2^2}{k_3 k_1} + \frac{k_3^2}{k_1 k_2}, \qquad (36)$$

and the bispectrum integral separates into three integrals over the k_i 's inside an integral over x,

$$\left(\frac{2}{3\pi}\right)^{3} \int x^{2} dx \left(\int k_{1}^{2} dk_{1} j_{l_{1}}(\Delta \eta k_{1}) j_{l_{1}}(xk_{1})\right) \\ \times \left(\int \frac{dk_{2}}{k_{2}} j_{l_{2}}(\Delta \eta k_{2}) j_{l_{2}}(xk_{2})\right) \\ \times \left(\int \frac{dk_{3}}{k_{3}} j_{l_{3}}(\Delta \eta k_{3}) j_{l_{3}}(xk_{3})\right),$$
(37)

plus permutations. The k_1 integral evaluates to $(\pi/2x^2)\delta(x - \Delta \eta)$ and integrating out the delta function leaves

$$\left(\frac{4}{27\pi^2}\right)\left(\int \frac{dk_2}{k_2} j_{l_2}^2(\Delta \eta k_2)\right)\left(\int \frac{dk_3}{k_3} j_{l_3}^2(\Delta \eta k_3)\right).$$
 (38)

These two integrals evaluate to $(2l(l + 1))^{-1}$ so with permutations we have the analytic result

$$\binom{1}{27\pi^2} \binom{1}{l_1(l_1+1)l_2(l_2+1)} + \frac{1}{l_2(l_2+1)l_3(l_3+1)} + \frac{1}{l_3(l_3+1)l_1(l_1+1)}.$$
(39)

This exact solution (in the large-angle approximation) provides an excellent reference for checking the accuracy of our numerical methods in later sections.

For the equilateral case, we make the replacement

$$F^{SI}(k_1, k_2, k_3) = \frac{1}{8} \frac{(-k_1 + k_2 + k_3)(k_1 - k_2 + k_3)(k_1 + k_2 - k_3)}{k_1 k_2 k_3}$$
(40)

$$= \frac{1}{8} \left(-2 - \frac{k_1^2}{k_2 k_3} + \frac{k_1}{k_2} + \text{ all possible permutations} \right).$$
(41)

When we substitute this into the full integral it again separates into three integrals over the k_i 's inside an integral over x. The middle term is the same as the local case with the exact solution given above. For the remaining two terms we are left to solve integrals like

$$\int dkk^{n} j_{l}(k) j_{l}(xk) = \frac{\pi}{2\sqrt{x}} \int dkk^{n-1} J_{l+1/2}(k) J_{l+1/2}(xk),$$
(42)

for n = 1, 0, -1. The case n = 0 has a nice solution [23] (p. 405),

$$x > 1 \Rightarrow \frac{\pi}{2} \frac{x^{-(l+1)}}{2l+1} : x < 1 \Rightarrow \frac{\pi}{2} \frac{x^l}{2l+1}.$$
 (43)

Unfortunately, however, both n = 1, -1 produce complicated expressions involving hypergeometric functions [23] (p. 401),

$$n = 1:x < 1 \Rightarrow \frac{\sqrt{\pi}\Gamma(l+1)}{2\Gamma(l+\frac{3}{2})} x_2^l F_1\left(\frac{1}{2}, l+1; l+\frac{3}{2}; x^2\right)$$
(44)

$$x > 1 \Rightarrow \frac{\sqrt{\pi}\Gamma(l+1)}{2\Gamma(l+\frac{3}{2})} x_2^{-(l+2)} F_1\left(\frac{1}{2}, l+1; l+\frac{3}{2}; \frac{1}{x^2}\right)$$
(45)

$$n = -1: x < 1 \Rightarrow \frac{\sqrt{\pi}\Gamma(l)}{4\Gamma(l+\frac{3}{2})} x_2^l F_1\left(-\frac{1}{2}, l; l+\frac{3}{2}; x^2\right)$$
(46)

$$x > 1 \Rightarrow \frac{\sqrt{\pi}\Gamma(l)}{4\Gamma(l+\frac{3}{2})} x_2^{-l} F_1\left(-\frac{1}{2}, l; l+\frac{3}{2}; \frac{1}{x^2}\right),$$
(47)

so apparently there is no analogous route to a simple analytic result.

We have also pursued an alternative approach by deriving series solutions for the integrals required for the bispectrum. These are presented in the Appendix, along with a minor correction to the literature, but their utility is yet to be established since they exhibit slow numerical convergence in key limits.

V. NUMERICAL SOLUTIONS

The main problem with a numerical evaluation of the bispectrum is the intensive sampling required to deal with the extreme oscillatory nature of the integrand. Recall the progress made so far with a 4D integral over six highly oscillatory functions broken down into two 1D integrals (29) and (30) over three highly oscillatory functions each, inside a 2D integral (28) over a fixed triangular domain. This is a significant improvement because it has reduced the overall dimensionality of the problem and limited the worst oscillatory behavior to one dimension. However, we still have to work hard to get the integrals to converge fast enough for the calculation to be feasible. At each point in the $\alpha\beta$ triangle we must solve the two 1D integrals, $I^{T}(\alpha, \beta)$ and $I^{G}(\alpha, \beta)$ with each integrand composed of either Bessel or transfer functions which are nontrivial to calculate. Fortunately, the only part of the integrand that changes as we move about the $\alpha\beta$ triangle (i.e., at fixed l_1, l_2, l_3 are the values of a, b, c [see (24)]. These simply rescale the three functions relative to each other so we do not need to reevaluate them at each point. Instead we can calculate the functions once beforehand, retaining them for future reference. At each point in the $\alpha\beta$ triangle, we then rescale and multiply them together to form the integrand, saving lengthy calculation.

The preliminary calculations of the Bessel and transfer functions are performed for a closely spaced range of lvalues with the results stored in a table, which can then be read in prior to every bispectrum calculation. The transfer functions for all l's were calculated using the line-of-sight code CMB2000 (or else an equivalent such as CMBFAST or CAMB), along with the associated spherical Bessel function at the same l, and stored in separate tables. Next we take the tabulated functions and interpolate using a cubic spline to create a dense grid of values. Bessel and transfer functions for large l start with a long flat section which is approximately zero before the first peak and hence does not contribute to the integral (see Fig. 3). This section is set



FIG. 3. A plot of the spherical Bessel function l = 200. Note how the function essentially vanishes up to $l \approx 180$.

identically to zero with a view to skipping it in the later integration; we define this as the part for which the function magnitude remains 10^{-10} times below its maximum value.

Now consider the two-dimensional integral (28) over the $\alpha\beta$ equilateral triangle at fixed multipoles l_1, l_2, l_3 . Having produced the two tables of Bessel and transfer functions we access the relevant rows for the given multipole set. At each point in the $\alpha\beta$ triangle we then calculate a, b, c and stretch the data appropriately using linear interpolation to calculate the new points. We skip the first section where the values are zero and then calculate the two one-dimensional integrals using the trapezoidal rule. This stretching and linear interpolation will tend to magnify any errors in our original data, so this is mitigated by using a very large number of points $[\mathcal{O}(10^6)]$, even though the integrands use a much smaller selection (say every 20th point). This balances accuracy, using dense initial sampling, with speed, using sparser sampling for the integration. Next we recall that for the geometric integral the value on the boundary is half that of the interior, causing slow convergence of the two-dimensional integral near the discontinuous edges. We avoid this problem by doubling the result of the geometric integration for points actually on the boundary. The important further check that needs to be made is when to terminate the integration. There is a hard boundary when the tabulated data runs out (set by an overall range), but most integrals converge well before this limit. Convergence is determined by comparing the contribution of the last x sections [typically $\mathcal{O}(10\,000)$] to the value of the total integral. If it remains unchanged up to a certain percentage tolerance (typically $\approx 0.0001\%$) then the integration is stopped. However, given the oscillatory nature of the integrand it is important to cover a large region over many periods in order to correctly determine this threshold. This is most important in the corners of the triangle where the data is stretched, along with the period of oscillation, by the largest amount. As we shall note in the next section, these truncation errors in the 1D integrals can be significant at large *l* unless handled carefully.

When attempting to perform the 2D integral (28) over the triangle we have to balance two key considerations. Every point used entails the calculation of both onedimensional integrals so we want to minimize their number to reduce calculation time. Unfortunately, while the majority of the highly oscillatory behavior is confined to the one-dimensional integrals there is still some oscillatory behavior which carries over to the $\alpha\beta$ triangle (see Fig. 4). As the $\alpha\beta$ triangle has small scale structure we need to use a fine mesh to gain accuracy. To achieve this we have developed an adaptive method that first uses a sparse triangular grid to estimate the integral using a 2Dtrapezoidal rule. This then steps through each triangular cell and divides it into four, calculating the change in the local integrand contribution (see Fig. 5). If this change is



FIG. 4 (color online). The equilateral triangle we need to integrate over to obtain the bispectrum for $l_1 = l_2 = l_3 = 20$. Note the small scale structure to the right of the peak. The first few oscillations make a significant contribution and must be sampled intensively for an accurate result. The rest of the oscillations, however, do not because of cancellations. Hence any adaptive algorithm must be able to determine the importance of any structure it finds.

below a tunable accuracy parameter then it uses the original result, but if it is above then it continues the refinement process by dividing each of these triangles into four new triangles and estimating their contribution. This will be repeated until the change in the local result is below an acceptable limit (or the recursion hits a predefined tunable recursion depth). The net effect is that we only have dense grids in areas where the most significant contributions to the integral are made, with the remaining areas only sampled sparsely. For the $l_1 = l_2 = l_3 = 20$ triangle integration, we achieve convergence for a recursion depth of five after which there is little improvement. One must be careful, however, to set the original sampling grid to broadly capture most of the structure or else the refinement algorithm may cease prematurely (see Fig. 6).

We note that these calculations naturally coarse grain the computational work either through the sampling of 1D integrals on the 2D triangular grid or, at a higher level, simply by evaluating the bispectrum at different multipole values. The problem is well suited to parallelization on a large supercomputer or cluster and this has been achieved with the present code using an MPI (message passing



FIG. 5 (color online). Refinement method. We start with the black cell defined by the three corner points (black). The cell is then divided into four by calculating the midpoints of the three sides (red) and the change in area is then $\frac{1}{4} | \sum$ corner points – \sum midpoints|.



FIG. 6. Successive refinement of the grid for integration of the $l_1 = l_2 = l_3 = 20$ triangle, the left is a recursion depth of one proceeding to five on the far right. After a recursion depth of five there is little change in the grid or result, however the error for the integration is still 4.45%. If we look at the diagram we see a blank triangle about halfway between the corner and the center. This region contains important structure but it has been missed as the initial sampling grid was too sparse. By doubling the initial grid the error is reduced to 0.8%.

interface) implementation which dramatically reduces calculation time scales.

VI. RESULTS

As we previously noted, bispectrum multipole values l_1 , l_2 , l_3 are restricted by the same triangle condition as the wave numbers k_1 , k_2 , k_3 . This allows us to propose a similar parametrization to that we used for the k_i 's in (24), that is,

$$l_{1} = \frac{3}{2}l(1-\gamma) \qquad l_{2} = \frac{3}{4}l(1+\delta+\gamma) l_{3} = \frac{3}{4}l(1-\delta+\gamma).$$
(48)

We will use these parameters to graph the bispectrum in two ways, (i) the equal multipole case $l_1 = l_2 = l_3$, i.e., the distance from the origin plotting against *l* (holding $\delta = 0$ and $\gamma = \frac{1}{3}$) and (ii) transverse triangular slices with l = const.

In order to characterize the main sources of systematic error, we begin by comparing to the exact solution in the



FIG. 7 (color online). Plot of the bispectrum for the local case in the large-angle approximation divided by the analytic result. The wiggly line (blue) is with the adaptive code starting with a grid of 200 and allowing two recursions, the smooth line (red) is for a grid of 1600. For the fine 1600 grid the oscillations disappear indicating that this eliminates error from the grid sampling. Note the adaptive method remains within a percent of this line. Both the results are within a percent up to $l \approx 500$ then we notice a downward trend due to the premature truncation of the one-dimensional geometric integral.

local case (39) using the large-angle transfer functions (35). This is a stringent test because the falloff in the radiation transfer functions is much faster at large l in the realistic case; the errors in the test case are more pronounced and help delineate both resolution and truncation errors. We have calculated the equal momentum bispectrum from l = 2 to l = 1800 on a fairly sparse initial triangular grid of side length N = 200 and a maximum recursive depth of only 2, and we find that the numerical result has an error less than 1% for l < 550 and a maximum error of 9% for l = 1800. We also repeated the same calculations with a much finer grid with N = 1600 to find comparable results which are shown in Fig. 7, demonstrating little improvement from the increased resolution. The calculations were performed on the Cosmos supercomputer and took between 16 min and 4 h 55 min CPU time with a median of about 80 min for the adaptive method. The average time to calculate each bispectrum point b_{111} on the finer grid was greater than 11 h.

The largest source of error in the calculation is a tendency to underestimate the integral due to premature truncation of the one-dimensional integrals for large l. There are two reasons for this. First, in the case of l = 1800 the Bessel function is approximately zero until about 1700. Truncating at a value of 10000 we only need to rescale by a < 0.17 for the nonzero section of the integrand to be shifted off the interval, making the whole integral vanish numerically. However, this only affects the outside of the $\alpha\beta$ triangle so it only generates a small error. The second reason is that the intelligent truncation routine that decides if the one-dimensional integral has converged can be too aggressive. This affects all points and is the primary source of numerical error. If we double the integration cutoff and eliminate the intelligent truncation the error in the onedimensional integrals is drastically reduced (see Fig. 8). The cost is a fivefold increase in calculation times for an adaptive N = 400 grid. This is mainly a problem for the geometric integral as this only decays as 1/x requiring late truncation for accuracy. Here, the large-angle transfer integrand decays as $1/x^4$ so it converges more quickly. If we calculate the test case again with an N = 1600 grid, doubling the 1D region of integration, and turning off the



FIG. 8 (color online). Sources of error from the onedimensional integrals. The four lines are all for the same grid of 400 on the triangle. The lowest line (red) is the calculation for the standard conditions, the next lowest line (blue) is when we double the region of integration, the second highest line (yellow) is with the intelligent truncation turned off, and the highest line (green) is when we do both.

truncation routine, then the error is reduced comfortably within 1%. Unfortunately the integrals then take just under 80 h to complete.

Having confirmed the accuracy of the method, in principle, the calculations were then repeated with the full radiation transfer function. Here, with an initial N = 200grid and a recursion depth of 2, the calculations completed on Cosmos in a median time of 31 min. The results for the equal momentum bispectrum are shown in Fig. 9. Its behavior is as expected with the two main peaks appearing at $l \approx 200$ and $l \approx 500$, mirroring those observed in the CMB power spectrum. We have also calculated the bispectrum using a dense N = 1600 grid with the difference also plotted in Fig. 9, confirming the absence of a significant resolution error. On the other hand, we might be concerned about the important truncation error we observed in the large-angle test case discussed above. To check this we ran



FIG. 9 (color online). Plot of the bispectrum in the local case using the full radiation transfer function over the result for the large-angle approximation. Note the three main peaks where we would expect from the plot of the power spectrum. The flat (red) line is the difference from using the dense grid. At 650 it reaches a maximum of only 0.045 which justifies our use of the much faster adaptive method.

long calculations for the N = 1600 grid but without intelligent truncation and while also doubling the cutoff in the geometric one-dimensional integral. The results gave a maximum correction of only 0.004%, indicating that truncation error is also insignificant in the realistic case. The explanation for this improvement in accuracy lies in the much more rapid exponential falloff at large l of the true radiation transfer function, when compared to the largeangle solution (35). Since the transfer function integral multiplies the geometric integral in the final 2D integrand (28), it also suppresses any errors due to the slow convergence of the latter. The comparison in Fig. 9 indicates that even the N = 200 adaptive grid achieves a bispectrum accuracy better than 1% across the full multipole range l < 1800.

In the equilateral case we no longer have an analytic solution with which to compare. However, both $I^{T}(\alpha, \beta)$ and $I^G(\alpha, \beta)$ are strongly peaked in the center of the $\alpha\beta$ triangle (as we can see from Fig. 4). Since both bispectrum shapes are flat in the center, we expect the equilateral case to exhibit approximately the same scaling with l as the local case for b_{III} . To test accuracy we can again take the large-angle approximation, dividing by the local analytic solution to normalize. As anticipated, we find that numerical accuracy is improved over the local test case, because the more oscillatory contributions near the boundaries are further suppressed by the primordial shape. Using an adaptive N = 200 grid (recursive depth 2), we find that this test case has an error of less than 1% for l < 1300 and maximum error of 1.7% for l = 1800 when compared to the calculation on a N = 1600 grid without the intelligent truncation and doubling the cutoff (see Fig. 10).

The calculation for the equilateral case with the full radiation transfer function gives much the same results as for the local case (see Fig. 11). Here we again divide by the



FIG. 10 (color online). Plot of the bispectrum for the equilateral case in the large-angle approximation over the analytic result. The result has been scaled so that the result for l = 20 is 1 as this is the most stable part of the plot. As we have no analytic solution to compare to in the equilateral case we have calculated the bispectrum on a dense N = 1600 grid without the intelligent truncation and doubling the cutoff. By comparing the two plots we see the error in within 1% up until l = 1300, a significant improvement over the previous case.



FIG. 11 (color online). Plot of the bispectrum for both the local case and the equilateral case using the full radiation transfer function over the analytic result. The equilateral case has been multiplied by 24, the ratio between the height of the center point of the local case to the equilateral case. We again see the main peaks where we would expect from the power spectrum calculation but the local case is smaller for low *l*. For large *l* the two one-dimensional integrals act like delta functions picking out the point on the shape functions where k_1 , k_2 , k_3 have the same ratio as l_1 , l_2 , l_3 . As we have scaled the center points to be the same height the bispectrum curves merge for large *l*.

analytic solution scaling the result by the factor 24 so that the centers of the two shape functions are both the same height. The main difference between the two is that the peaks are slightly smaller in the equilateral case for low l(Fig. 11). For large l, the results are almost identical because the larger the l values the more peaked I^T and I^G become when plotted on the $\alpha\beta$ triangle. For large lthey act like a delta function picking out only the shape function value where k_1, k_2, k_3 have the same ratio as $l_1, l_2,$ l_3 . So for large l, the equal l bispectrum is essentially only proportional to the height in the center of the shape function rather than its shape. These bispectrum calculations on Cosmos took an average of approximately 30 min for each l.

It is clear from Figs. 9 and 11 that it would be difficult to distinguish between the local and equilateral cases on the basis of the equal momentum bispectrum. Instead, to achieve this we must look at the transverse triangular slices. For both cases three slices were made at 3l = 850, 3l = 1850, and at 3l = 3650 with points being calculated at l_i 's that are multiples of 50. For the slowly converging test case with a local shape function and the large-angle approximation, we find that the error in the calculation for 3l = 850 is less than 0.6%, for 3l = 1850 less than 3%, and for 3l = 3650 less than 5%. However the error is always negative and its range is small in all cases, indicating that it is primarily systematic and due to the premature truncation issue discussed above. Most importantly we note that the error appears to be a function only of the sum, and not the specific combination, of the l_i 's so there are no new issues in dealing with the slices.

For the local case with the full transfer function with an adaptive N = 200 grid and recursion depth 2 the typical

calculation time was 44 min. In the triangular bispectrum plots (see Fig. 12), we have l_1 constant along the lines parallel to the left bottom edge, where it is at a maximum (400, 900, and 1800, respectively), reducing linearly to a minimum in the top right corner (50 in all three cases). The dependence on l_2 and l_3 is similar with respect to the remaining two edges. As we used the analytic version of the local shape function, which has no initial structure, the diagrams are easily understood from Fig. 9. For 3l = 850, each of the edges corresponds to one $l_i = 400$. As 400 is close to zero in the plot along the equal l direction, the edges are close to zero in the slice. On the other hand, near the center is the triple (250, 300, 300) all of which are large and negative so they produce a broad central trough. In the plot of 3l = 1850 we have a three peaks at the triple (450, 700, 700) (plus two permutations) as 700 is large and negative and 450 is positive. We also have troughs triples (250, 800, 800) (near the corner) and for (500, 500, 850) (near the middle of the edge) as both 250 and 500 are close to maximums and give strong contributions. The plot of 3l = 3650 has many peaks and troughs arising from various combinations of the l_i 's but they are an order of magnitude smaller than the previous results. This is because all triples have at least two $l_i \ge 900$ so the contribution from the remaining l_i is suppressed. We can see a 3D comparison in Fig. 13.

For the equilateral case with an initial N = 200 grid, the bispectrum calculations were completed on Cosmos again in about 44 min and are also plotted in Fig. 12 (right). We see the same basic pattern of peaks and troughs as in the local case. The main difference is that the heights of these features become strongly suppressed towards the edges of the triangle. This is a direct reflection of the difference in the two shape function shapes. The local shape (superhorizon case) magnifies these features near the edge with strong correlations between very different multipoles, whereas for the equilateral case (horizon crossing) they are suppressed.

Having proved the method for the two standard analytic cases we can proceed to look at cases where the shape function cannot be separated. A good example is the DBI model [5]. The shape function for this model is identical to than in [24] where higher derivative operators have been added to the Lagrangian. This shape function has the explicit form of

$$F^{SI}(k_1, k_2, k_3) = \frac{1}{k_1 k_2 k_3 (k_1 + k_2 + k_3)^2} \\ \times \left(\sum_i k_i^5 + \sum_{i \neq j} (2k_i^4 k_j - 3k_i^3 k_j^2) \right) \\ + \sum_{i \neq j \neq l} (k_i^3 k_j k_l - 4k_i^2 k_j^2 k_l) \right).$$
(49)

This shape function has been plotted in Fig. 14.



FIG. 12 (color online). Plots of the bispectrum for the local case (on the left) and for the equilateral case (on the right) for l = const slices using the full radiation transfer function. The top row is 3l = 850, the second 3l = 1850, and the third 3l = 3650. The bispectrum has been divided by the analytic result so its features are clearly visible. Lines of constant l_i are those parallel to their respective edge of the triangle. The extremums of the graphs always appear when the l_i triples contain l_i 's which represent extremums of the plot along the *l* direction. Note that the outside of the graphs are suppressed in the equilateral case compared to the local case due to the differing shape functions.

Because of the factor $(k_1 + k_2 + k_3)^2$ this shape cannot be separated as previously, although some progress has been made by using an integral form [25]. Subsequently we have always been forced to calculate the bispectrum for this model by approximating with the equilateral shape.

With this new approach we can now calculate it in full as the nonseparability no longer poses a restriction.

The calculations run in similar times as for the equilateral case. The equal l bispectrum gives almost identical results as the equilateral case as can be seen in Fig. 15 so it



FIG. 13 (color online). Plots of the bispectrum for the local case (on the left) and for the equilateral case (on the right) for l < 1800. Note how in the equilateral case all perturbations off the central axis are suppressed.

may seem that the previous approach is justified. When we plot the transverse slices however, the differences become apparent (see Fig. 16). From these plots we can see that the variation between the two cases is at the 10% level. A 3D comparison can be seen in Fig. 17.

As a final comment, we note that we can exploit the properties of the bispectrum discussed above to speed up its evaluation further. Very substantial improvements could be achieved by tabulating the product of our 1D integrations over the Bessel and transfer functions (29) and (30), rather than keeping tables of the functions themselves. This would be relevant, for example, if the background cosmology was fixed (i.e., the transfer functions are not modified), perhaps for a likelihood analysis marginalizing over inflationary model parameters. In this case, the table of inte-



FIG. 14 (color online). F^{SI} plotted on the $\alpha\beta$ triangle for the "higher derivative" model.

grals would have to cover the 5D parameter space $(\alpha, \beta, l_1, l_2, l_3)$ which may seem unrealistic, but our discussion has shown otherwise. While the $\alpha\beta$ triangles must be sampled to fairly high resolution (e.g., on an N = 800 grid), in contrast, the CMB bispectrum is intrinsically a slowly varying function of the multipoles (period $l \sim 200$, like the CMB power spectrum). Provided the shape function is relatively featureless, we would only require sparse sampling in multipole space (perhaps as few as 33 points on a 2D slice, given the underlying symmetries, even for $l \sim 1000$). The initial tabulation would entail much parallel computation, but subsequently complete bispectra (up to l < 2000) could be accurately calculated in minutes.



FIG. 15 (color online). Plot of the bispectrum for both the higher derivative case and the equilateral case using the full radiation transfer function over the analytic result. As the two shape function shapes are similar in the region surrounding the center of the triangle the results are almost identical. To see the difference between the two shapes we must look to the transverse slices.



FIG. 16 (color online). Plots of the bispectrum for the higher derivative model (on the left) and the difference with the equilateral case (on the right) for l = const slices using the full radiation transfer function. The top row is 3l = 850, the second 3l = 1850, and the third 3l = 3650. The bispectrum has been divided by the analytic result for the local case so its features are clearly visible. Lines of constant l_i are those parallel to their respective edge of the triangle. The extremums of the graphs always appear when the l_i triples contain l_i 's which represent extremums of the plot along the l direction.

VII. ESTIMATION

Having calculated the bispectrum accurately today we then need to turn to the problem of measurement. The bispectrum itself will unfortunately be too small for direct measurement so we will have to rely on estimators. Optimal estimators for the bispectrum are based on the statistic

$$S = \frac{1}{N} \sum_{l_i} \frac{1}{C_{l_1} C_{l_2} C_{l_3}} B_{l_1 l_2 l_3} \sum_{m_i} \begin{pmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \end{pmatrix} \times a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3},$$
(50)



FIG. 17 (color online). Plots of the bispectrum for the DBI case (on the left) and for the equilateral case (on the right) for l < 1800. Note how in the equilateral case the perturbations on the sides are suppressed.

where $B_{l_1 l_2 l_3}$ is the bispectrum predicted from theory and the a_{lm} 's are the measured values. If the normalization factor is

$$N = \sum_{l_i} \frac{B_{l_1 l_2 l_3}^2}{C_{l_1} C_{l_2} C_{l_3}},$$
(51)

then the statistic gives the relative magnitude of the theoretical bispectrum that best fits the data. With the new methods described in the previous section we can now calculate the theoretical bispectrum accurately for an almost arbitrary shape function. Unfortunately however, we are still restricted by difficulties in calculating the observed bispectrum. One of the biggest challenges is to calculate the Wigner 3j symbols. These have no simple analytic form for the general case and are too many too tabulate for large values of *l*. Even for a relatively modest $l_{\text{max}} = 335$ as used for analysis of the WMAP 1-year data requires calculating over 80×10^9 independent 3j symbols. This makes the sum over *m* impractical to compute even if the a_{lm} 's were all known. The separable case overcomes this by using the Gaunt integral to replace the 3j symbol with an integral over three spherical harmonics which are absorbed with the a_{lm} 's into the separable parts allowing for efficient calculation. In the general case this is impossible so we must find another way around the problem.

In [25] a fast method is proposed for calculating the estimator. When the reduced bispectrum can be represented in a separable form,

$$b_{l_1 l_2 l_3} = \frac{1}{6} \sum_{i=1}^{N_{\text{fact}}} (X_{l_1}^{(i)} Y_{l_2}^{(i)} Z_{l_3}^{(i)} + 5 \text{ permutations}).$$
(52)

Then if we define

$$X_{a}^{(i)}(\hat{\mathbf{n}}) = \sum_{lm} X_{l}^{(i)} \frac{a_{lm}}{C_{l}} Y_{lm}(\hat{\mathbf{n}}),$$
(53)

we can write the estimator as

$$S = \frac{1}{N} \sum_{i=1}^{N_{\text{fact}}} \int d\hat{\mathbf{n}} X_a^{(i)}(\hat{\mathbf{n}}) Y_a^{(i)}(\hat{\mathbf{n}}) Z_a^{(i)}(\hat{\mathbf{n}}).$$
(54)

While the general reduced bispectrum has no simple separable form, we note from our earlier plot that it is smooth. This means that we should be able to represent it as a sum of smooth basis functions. As we have seen from previous figures (11, 12, and 16) when we divide the bispectrum by the analytic result we get a surface which is O(1) with slow oscillations. As a result we choose our basis as follows,

$$b_{l_1 l_2 l_3} = \frac{1}{3} \sum_{\alpha \beta \gamma} a_{\alpha \beta \gamma} (X'_{\alpha}(l_1) X'_{\beta}(l_2) X_{\gamma}(l_3) + 2 \text{ permutations}),$$
(55)

where X_{α} are shifted Legendre polynomials,

$$X_{\alpha}(l) = P_{\alpha}\left(\frac{2l - l_{\max}}{l_{\max}}\right) \tag{56}$$

and

$$X'_{\alpha}(l) = \frac{X_{\alpha}(l)}{l(l+1)}.$$
(57)

However, any set of orthogonal polynomials would suffice. Simplifying we have



FIG. 18 (color online). Plot comparing the equal *l* reduced bispectrum and the representation of it in terms of basis functions up to $\alpha = 30$.

$$\begin{pmatrix} l_1(l_1+1)l_2(l_2+1)l_3(l_3+1)\\ \overline{l_1(l_1+1)+l_2(l_2+1)+l_3(l_3+1)} \end{pmatrix} b_{l_1l_2l_3} \\ = \sum_{\alpha\beta\gamma} a_{\alpha\beta\gamma} X_{\alpha}(l_1) X_{\beta}(l_2) X_{\gamma}(l_3).$$
(58)

And $a_{\alpha\beta\gamma}$ can be easily determined,

$$a_{\alpha\beta\gamma} = (2\alpha + 1)(2\beta + 1)(2\gamma + 1) \int \frac{dl_1 dl_2 dl_3}{l_{\max}^3} \\ \times \left(\frac{l_1(l_1 + 1)l_2(l_2 + 1)l_3(l_3 + 1)}{l_1(l_1 + 1) + l_2(l_2 + 1) + l_3(l_3 + 1)}\right) \\ \times b_{l_1 l_2 l_3} X_{\alpha}(l_1) X_{\beta}(l_2) X_{\gamma}(l_3),$$
(59)

using the orthogonality condition,

$$\int_{-1}^{1} dx P_{\alpha}(x) P_{\beta}(x) = \frac{2\delta_{\alpha\beta}}{2\alpha + 1} \Rightarrow \int_{0}^{l_{\max}} \frac{dl}{l_{\max}} X_{\alpha}(l) X_{\beta}(l)$$
$$= \frac{\delta_{\alpha\beta}}{2\alpha + 1}.$$
(60)

If we define as before

$$\bar{X}_{\alpha}(\hat{\mathbf{n}}) = \sum_{lm} X_{\alpha}(l) \frac{a_{lm}}{C_l} Y_{lm}(\hat{\mathbf{n}}), \qquad (61)$$

and

PHYSICAL REVIEW D 76, 083523 (2007)

$$\bar{X}'_{\alpha}(\hat{\mathbf{n}}) = \sum_{lm} \frac{X_{\alpha}(l)}{l(l+1)} \frac{a_{lm}}{C_l} Y_{lm}(\hat{\mathbf{n}}), \tag{62}$$

then the estimator becomes

$$S = \frac{1}{N} \sum_{\alpha\beta\gamma} a_{\alpha\beta\gamma} M_{\alpha\beta\gamma}.$$
 (63)

Where we have defined

$$M_{\alpha\beta\gamma} = \frac{1}{3} \int d\hat{\mathbf{n}} (\bar{X}'_{\alpha}(\hat{\mathbf{n}}) \bar{X}'_{\beta}(\hat{\mathbf{n}}) \bar{X}_{\gamma}(\hat{\mathbf{n}}) + 2 \text{ permutations}).$$
(64)

This approach has several advantages. First, there is complete separation of theory and measurement. The quantities \bar{X} , \bar{X}' , and $M_{\alpha\beta\gamma}$ only need to be calculated once per map then $M_{\alpha\beta\gamma}$ can simply be stored on disk. For each theory we only need to calculate $a_{\alpha\beta\gamma}$ then perform the sum of the product of it with the waiting $M_{\alpha\beta\gamma}$. This process will be quick as rather than summing over all combinations of l_i we now only need to do so over the range of α , β , γ . This new approach for the estimator makes it possible to produce quick and accurate bispectrum estimates for general models. This decomposition has been tested for $l_{\text{max}} = 400$ for the local model. Using basis functions up to $\alpha = 30$ we managed to reconstruct the reduced bispectrum to a 1% accuracy as shown by the figures (18 and 19).

VIII. FUTURE WORK

With the numerical methods we have developed it is now possible to accurately calculate the CMB bispectrum today from a general shape function. In the stringent test case there is a problem obtaining accuracy without long calculation times because of truncation errors, but arbitrary accuracy can be achieved by altering truncation thresholds and improving resolution. In the real calculation using the full radiation transfer function we have demonstrated that the systematic error sources are much weaker and very



FIG. 19 (color online). Plots of the bispectrum for the local model plotted next to the representation of it in terms of basis functions up to $\alpha = 30$.

accurate results can be obtained rapidly using our adaptive method. We find that the bispectra for the local and equilateral cases are broadly similar in the equal multipole case, with the main difference being that the equilateral case is smaller at low *l*. Where the two cases differ significantly is in the slices $l_1 + l_2 + l_3 = \text{const}$ (and for which we find no significant new sources of error). For the calculation with the full radiation transfer functions, we find that the results are much the same in the center, but with the equilateral case more heavily suppressed where the *l* values differ significantly, as expected. This indicates that the equal multipole bispectrum provides a good measure of the overall size of the non-Gaussianity, while the unequal l bispectrum provides a differentiator between competing models. Obviously, meaningful plots of the CMB bispectrum (like Fig. 12) are not expected to be determined observationally, not least because they will be dominated by noise and convolved with experimental beams. Instead we must use an estimator to determine the fit of theory to observation.

While a parallel implementation of the code runs sufficiently rapidly on a supercomputer for calculations of the full bispectrum, there is still plenty of scope for further improvements. Essentially all of the calculation time is spent evaluating the one-dimensional integrals with the geometric integral taking twice as long as the transfer integral. To reduce overall evaluation times we then have two options. We can improve the integration methods for the one-dimensional integrals or we can reduce the number of points we need to calculate by improving the adaptive algorithm. Improving the cell selection criteria in the adaptive algorithm has largely been a long process of trial and error, and no doubt this can be continued with further success. However, improving the one-dimensional integration method is perhaps the more exciting option. Here, more rapidly convergent numerical methods are possible, but the hope is that for the case of the geometric integral we may be able to switch to an analytic solution should a stable method of evaluation be discovered. Such an advance could reduce overall bispectrum calculation times by a factor of 3.

The decomposition of the bispectrum into Legendre polynomials has also been parallelized but is still hampered by the requirement to use a large number of polynomials. This arises from having to perform the integral over the cube in l space with sides from 0 to l_{max} while the bispectrum is only defined inside the tetrahedron created by the triangle condition on the l's. When we use the code to calculate points outside this tetrahedron we introduce discontinuities which require a large number of polynomials to fit accurately. We are currently looking to edit the code to calculate these points so that the bispectrum remains smooth as we cross the boundary.

The present code has been developed in modular form to link and work together with existing CMB line-of-sight codes for the angular power spectrum (like CMBFAST, CAMB, or CMB2000); it extracts the appropriate radiation transfer functions from these for a given cosmology. In due course, it is our intention to improve the efficiency, portability, and parallelism (as well as the documentation) of this code and to make it publicly available.

The bispectrum remains a highly significant tool for testing for non-Gaussianity in the CMB. Hopefully, future experiments like the Planck satellite will either detect or provide much tighter constraints on this non-Gaussianity, opening a new window on the early Universe. With this code we now have the capability to make accurate predictions for the contributions to the CMB bispectrum which are of primordial origin. For example, we plan to use quantitative bispectra from multifield inflation models [3] and project these forward to make falsifiable CMB predictions. Clearly parallel efforts must also continue in order to pin down and distinguish alternative sources of CMB non-Gaussianity from later times, such as secondorder effects in the radiation transfer functions and nonlinear astrophysical effects.

ACKNOWLEDGMENTS

We are especially grateful to Kendrick Smith for conversations regarding the decomposition of the reduced bispectrum into polynomials and the subsequent method for calculating the estimator. We are also grateful for useful discussions with Paolo Creminelli, Michele Liguori, Gerasimos Rigopoulos, and Bartjan van Tent. Simulations were performed on the COSMOS supercomputer (an Altix 3700) which is funded by PPARC, HEFCE, and SGI. We are very grateful for help from the COSMOS programmer, Victor Travieso, who parallelized the code and improved its efficiency. The CMB2000 code, from which the radiation transfer functions were extracted, was written by Martin Landriau, EPS, Martin Bucher, and Richard Battye. This research was supported by PPARC Grant No. PP/C501676/1, the Sims Fund, Churchill College, the Cambridge Commonwealth trust, the Edward and Isabel Kidson Fund, and the William Georgetti Fund.

APPENDIX: THREE BESSEL FUNCTION INTEGRALS

Higher-order angular correlation functions generically involve integrals over products of Bessel functions. If we allow a general bispectrum but remain in the large-angle approximation then both the 1D integrals in (28), $I^{T}(\alpha, \beta)$ and $I^{G}(\alpha, \beta)$, are geometric and of this form. Assuming n = 0,

$$I^{T}(\alpha, \beta) = \frac{1}{3^{3}} \int j_{l_{1}}(ak) j_{l_{2}}(bk) j_{l_{3}}(ck) \frac{dk}{k}, \qquad (A1)$$

where $\Delta \eta$ has been absorbed via a rescaling of k. This leaves us trying to solve integrals of the form

$$I_{l_1, l_2, l_3}(a, b, c, n) = \int_0^\infty j_{l_1}(ax) j_{l_2}(bx) j_{l_3}(cx) x^n dx \quad (A2)$$

where n = 2, -1.

We already have several solutions for n = 2 given in Refs. [17–19]. For the general case a solution is proposed in Ref. [26] but, unfortunately, their analysis is incomplete. Their method revolves around replacing the spherical PHYSICAL REVIEW D 76, 083523 (2007)

Bessel functions as below

$$j_{l}(x) = \sum_{m=0}^{l} \frac{(l+m)!}{(l-m)!m!} \left(\frac{1}{2x}\right)^{m+1} \times (e^{-ix}(i)^{l-m+1} + e^{ix}(-i)^{l-m+1})$$
(A3)

to rewrite the integral as

$$\sum_{m_{1}=0}^{l_{1}} \sum_{m_{2}=0}^{l_{2}} \sum_{m_{3}=0}^{l_{3}} \left(\prod_{i=1}^{3} \frac{(l_{i}+m_{i})!}{(l_{i}-m_{i})!m_{i}!} \right) \frac{1}{(2a)^{m_{1}+1}(2b)^{m_{2}+1}(2c)^{m_{3}+1}} \int_{0}^{\infty} \frac{(i)^{l_{1}+l_{2}+l_{3}-m_{1}-m_{2}-m_{3}+3}}{x^{m_{1}+m_{2}+m_{3}-n+3}} \\ \times (e^{-iax} + e^{iax}(-1)^{l_{1}-m_{1}+1})(e^{-ibx} + e^{ibx}(-1)^{l_{2}-m_{2}+1})(e^{-icx} + e^{icx}(-1)^{l_{3}-m_{3}+1})dx.$$
(A4)

The second line of the integrand can be expanded to

$$(i)^{l_{1}+l_{2}+l_{3}-m_{1}-m_{2}-m_{3}+3} [e^{-i(a+b+c)x} + e^{i(a+b+c)x}(-1)^{l_{1}+l_{2}+l_{3}-m_{1}-m_{2}-m_{3}+3} + (-1)^{l_{1}-m_{1}+1}(e^{-i(-a+b+c)x} + e^{i(-a+b+c)x}(-1)^{l_{1}+l_{2}+l_{3}-m_{1}-m_{2}-m_{3}+3}) + (-1)^{l_{2}-m_{2}+1}(e^{-i(a-b+c)x} + e^{i(a-b+c)x}(-1)^{l_{1}+l_{2}+l_{3}-m_{1}-m_{2}-m_{3}+3}) + (-1)^{l_{3}-m_{3}+1}(e^{-i(a+b-c)x} + e^{i(a+b-c)x}(-1)^{l_{1}+l_{2}+l_{3}-m_{1}-m_{2}-m_{3}+3})].$$
(A5)

So we are trying to complete the integral (A4) for eight values of λ ,

$$\int_0^\infty \frac{e^{i\lambda x}}{x^P} dx.$$
 (A6)

This integral is singular but the sum is not. The authors calculate

$$\int_0^\infty \frac{e^{i\lambda x}}{(x+\epsilon)^P} dx,\tag{A7}$$

and assert that the infinite terms cancel explicitly in the sum as $\epsilon \to 0$, which is incorrect. If we instead calculate

$$\int_{\epsilon}^{\infty} \frac{e^{i\lambda x}}{x^{P}} dx,$$
 (A8)

which has solution

$$\frac{(i\lambda)^{p-1}}{(P-1)!}E_1(-i\epsilon\lambda) + e^{i\epsilon\lambda}\sum_{s=0}^{p-2}\frac{(p-2-s)!}{(p-1)!}\frac{(i\lambda)^s}{\epsilon^{p-1-s}},$$
(A9)

where E_1 is the En function with series expansion,

$$E_1(-i\epsilon\lambda) = -\gamma - \ln(\epsilon) - \ln|\lambda| + \frac{i\pi}{2}\operatorname{sgn}(\lambda) + O(\epsilon),$$
(A10)

and γ is the Euler constant. This is identical to that in Ref. [26] result except for the $e^{i\epsilon\lambda}$ factor in front of the sum. The claim that the infinite part canceled led the authors to drop the sum altogether. Here the singular parts of the sum over s cancel in the triple sum over m_i , and as we will set ϵ to zero, we are then only left with in the constant part of the series.

$$e^{i\epsilon\lambda} \sum_{s=0}^{p-2} \frac{(p-2-s)!}{(p-1)!} \frac{(i\lambda)^s}{\epsilon^{p-1-s}} = \sum_{n=0}^{\infty} \frac{(i\epsilon\lambda)^n}{n!} \sum_{s=0}^{p-2} \frac{(p-2-s)!}{(p-1)!} \frac{(i\lambda)^s}{\epsilon^{p-1-s}}.$$
 (A11)

Which gives

$$\sum_{s=0}^{p-2} \frac{(i\epsilon\lambda)^{p-1-s}}{(p-1-s)!} \frac{(p-2-s)!}{(p-1)!} \frac{(i\lambda)^s}{\epsilon^{p-1-s}} = \sum_{s=0}^{p-2} \frac{(i\lambda)^{p-1}}{p-1-s} \frac{1}{(p-1)!} = \frac{(i\lambda)^{p-1}}{(p-1)!} \sum_{s=0}^{p-2} \frac{1}{p-1-s} = \frac{(i\lambda)^{p-1}}{(p-1)!} H_{p-1}, \quad (A12)$$

where $H_a = \sum_{s=1}^{a} \frac{1}{s}$ is the harmonic number. In the triple sum over m_i the parts proportional to γ and $\ln(\epsilon)$ in E_1 cancel exactly. The part of the integral that contributes to the final result is then

$$\frac{(i\lambda)^{P-1}}{(P-1)!} \left(\mathbf{H}_{p-1} - \ln|\lambda| + \frac{i\pi}{2} \operatorname{sgn}(\lambda) \right).$$
(A13)

We can then substitute this into the expression for the integral over the spherical Bessel functions to eventually obtain

$$\begin{aligned} (i)^{l_1+l_2+l_3-n} \sum_{m_1=0}^{l_1} \sum_{m_2=0}^{l_2} \sum_{m_3=0}^{l_3} \frac{1}{(m_1+m_2+m_3-n+2)!} \left(\prod_{i=1}^{3} (-1)^{m_i} \frac{(l_i+m_i)!}{(l_i-m_i)!m_i!} \right) \\ \times \left\{ i(1+(-1)^{l_1+l_2+l_3+n+1}) [F_1(a,b,c)+(-1)^{l_1}F_1(-a,b,c)+(-1)^{l_2}F_1(a,-b,c)+(-1)^{l_3}F_1(a,b,-c)] \right. \\ \left. + \frac{\pi}{2} (1+(-1)^{l_1+l_2+l_3-n}) [F_2(a,b,c)+(-1)^{l_1}F_2(-a,b,c)+(-1)^{l_2}F_2(a,-b,c)+(-1)^{l_3}F_2(a,b,-c)] \right\}, \end{aligned}$$

where

$$F_1(a, b, c) = \frac{(a+b+c)^{m_1+m_2+m_3-n+2}(H_{m_1+m_2+m_3-n+2} - \log|a+b+c|)}{(2a)^{m_1+1}(2b)^{m_2+1}(2c)^{m_3+1}}$$
(A14)

$$F_2(a, b, c) = \frac{(a+b+c)^{m_1+m_2+m_3-n+2}\operatorname{sgn}(a+b+c)}{(2a)^{m_1+1}(2b)^{m_2+1}(2c)^{m_3+1}}.$$
(A15)

These analytic solutions have unfortunately not proved particularly useful in the evaluation of the bispectrum. The solutions that involve large series like above are unstable for large *l*'s and also in the corners of the $\alpha\beta$ triangles, where one of the k_i 's tends to zero. The difficulty with these series for a straightforward numerical implementation is that they generically involve the exact cancellation of many very large terms.

- D. N. Spergel *et al.*, Astrophys. J. Suppl. Ser. **170**, 377 (2007); E. Komatsu *et al.*, Astrophys. J. Suppl. Ser. **148**, 119 (2003).
- [2] D. H. Lyth, C. Ungarelli, and D. Wands, Phys. Rev. D 67, 023503 (2003).
- [3] G.I. Rigopoulos, E.P.S. Shellard, and B.J.W. van Tent, arXiv:astro-ph/0511041 [Phys. Rev. D (to be published)].
- [4] See, for example, F. Bernardeau and J.-P. Uzan, Phys. Rev. D 66, 103506 (2002); F. Vernizzi and D. Wands, J. Cosmol. Astropart. Phys. 05 (2006) 019; I. Zaballa, Y. Rodriguez, and D. H. Lyth, J. Cosmol. Astropart. Phys. 06 (2006) 013.
- [5] M. Alishahiha, E. Silverstein, and D. Tong, Phys. Rev. D 70, 123505 (2004).
- [6] N. Arkani-Hamed, P. Creminelli, S. Mukohyama, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 04 (2004) 001.
- [7] D. Babich, Phys. Rev. D 72, 043003 (2005).
- [8] P. Creminelli, L. Senatore, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 03 (2007) 019.
- [9] E. Komatsu and D. N. Spergel, Phys. Rev. D 63, 063002 (2001).
- [10] L. Wang and M. Kamionkowski, Phys. Rev. D 61, 063504 (2000).
- [11] D. Babich, P. Creminelli, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 08 (2004) 009.
- [12] K. Enquvist, A. Jokinen, A. Mazumdar, T. Multamaki, and

A. Vaihkonen, Phys. Rev. Lett. 94, 161301 (2005).

- [13] A. Jokinen and A. Mazumdar, J. Cosmol. Astropart. Phys. 04 (2006) 003.
- [14] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, J. Cosmol. Astropart. Phys. 05 (2006) 004.
- [15] M. Liguori, F.K. Hansen, E. Komatsu, S. Matarrese, and A. Riotto, Phys. Rev. D 73, 043505 (2006).
- [16] Michele Liguori, Sabino Matarrese, and Lauro Moscardini, Astrophys. J. 597, 57 (2003).
- [17] R. Mehrem, J. T. Londergan, and M. H. Macfarlane, J. Phys. A 24, 1435 (1991).
- [18] C.T. Whelan, J. Phys. B 26, L823 (1993).
- [19] Jian-Xing Chen and Jun-Chen Su, Phys. Rev. D 69, 076003 (2004).
- [20] G. I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, Phys. Rev. D 72, 083507 (2005).
- [21] J. Maldacena, J. High Energy Phys. 05 (2003) 013.
- [22] G.I. Rigopoulos, E. P. S. Shellard, and B. J. W. van Tent, Phys. Rev. D 73, 083522 (2006).
- [23] G.N. Watson, *Theory of Bessel Functions* (Cambridge University Press, Cambridge, England, 1966).
- [24] P. Creminelli, J. Cosmol. Astropart. Phys. 10 (2003) 003.
- [25] K. M. Smith and M. Zaldarriaga, arXiv:astro-ph/0612571.
- [26] A Gervois and H. Navelet, SIAM J. Math. Anal. 20, 1006 (1989).