

Lensing of the CMB: Non-Gaussian aspects

Matias Zaldarriaga*

Institute for Advanced Studies, School of Natural Sciences, Princeton, New Jersey 08540

(Received 6 January 2000; published 18 August 2000)

We study the generation of cosmic microwave background (CMB) anisotropies by gravitational lensing on small angular scales. We show these fluctuations are not Gaussian. We prove that the power spectrum of the tail of the CMB anisotropies on small angular scales directly gives the power spectrum of the deflection angle. We show that the generated power on small scales is correlated with the large scale gradient. The cross correlation between the large scale gradient and the small scale power can be used to test the hypothesis that the extra power is indeed generated by lensing. We compute the three and four point functions of the temperature in the small angle limit. We relate the non-Gaussian aspects presented in this paper as well as those in our previous studies of the lensing effects on large scales to the three and four point functions. We interpret the statistics proposed in terms of different configurations of the four point function and show how they relate to the statistic that maximizes the S/N .

PACS number(s): 98.80.Es, 95.85.Bh, 98.35.Ce, 98.70.Vc

I. INTRODUCTION

The anisotropies in the cosmic microwave background (CMB) are thought to contain detailed information about the underlying cosmological model. In conventional models the anisotropies on most angular scales were created at the last scattering surface, at a redshift of $z \sim 1000$. At these early times the evolution of perturbations can be calculated accurately with linear theory. The calculation of theoretical predictions is almost straightforward, thus detailed observations of the microwave sky can, at least in principle, greatly constrain the cosmological model. We expect to be able to measure many of the parameters of the cosmological model with percent accuracy [1].

There are several physical processes that imprint anisotropies on the CMB after decoupling. Some of them will degrade our ability to learn about cosmology, such as foreground emission from our galaxy. Others will allow us to constrain processes that happen after decoupling and help us understand the evolution of our universe. For example the reionization of hydrogen by the ultraviolet light from the first generation of objects leaves a distinct mark in the polarization of the CMB [2], Sunyaev-Zeldovich emission from hot gas along the line of sight creates temperature anisotropies and the large scale structure (LSS) of the universe deflects the CMB photons, lensing the anisotropies [3–6,9].

When studying the lensing effect produced by the large scale structure of the universe we are trying to detect lensing produced by random mass fluctuations, the LSS, on a random background image, the CMB. The characteristics of the lensing effect depends on the relative size of the coherence lengths of these two random fields. In [5,6] we studied the limit of a rapidly fluctuating CMB background being lensed by a slowly varying mass distribution. This is the appropriate limit to recover the power spectrum of the projected mass density on scales much larger than the coherence length of the CMB, $\xi \sim 0.15^\circ$. In this paper we study the opposite

limit, the generation of power on scales much smaller than ξ .

The lensing effect is expected to be the dominant nonprimordial contribution to the CMB anisotropies on small scales ($l \sim 3000$). It has been shown [7] that an accurate determination of the power generated by lensing can help break some of the parameter degeneracies in the CMB. Interferometric observations of the anisotropies such as those that will be carried out by the Cosmic Background Imager (CBI) [8] are designed to make measurements at these angular scales.

To be able to use the observed power on small scales to break the degeneracies in the parameters one must be sure that one is observing the lensing signal. We will show that the small scale power generated by gravitational lensing has a very definite signature, it is correlated with the large scale gradient. Regions of the sky where the large scale gradient is larger will have more small scale power. The physical effect can be understood easily in the case of a cluster of galaxies lensing a smooth CMB gradient.

We will also compute the general three and four point function of the temperature field induced by lensing and show that both the statistic discussed in this paper and those proposed in [6] are particular subsets of the possible configurations of the four point function. We will construct the statistic that maximizes the signal to noise ratio.

II. GENERATION OF POWER ON SMALL SCALES

The measured temperature field $T(\boldsymbol{\theta})$ can be expressed in terms of the unlensed CMB field at the last scattering surface $\tilde{T}(\boldsymbol{\theta})$ and the deflection angle of the CMB photons $\delta\boldsymbol{\theta}$:

$$T(\boldsymbol{\theta}) = \tilde{T}(\boldsymbol{\theta} + \delta\boldsymbol{\theta}) \\ \approx \tilde{T}(\boldsymbol{\theta}) + \delta\boldsymbol{\theta} \cdot \nabla \tilde{T}(\boldsymbol{\theta}) + \frac{1}{2} \delta\theta_i \delta\theta_j \partial_{ij} \tilde{T}(\boldsymbol{\theta}). \quad (1)$$

In this paper we are interested in the effect of modes of the deflection angle of spatial wavelength much smaller than that of the unlensed CMB.

*Electronic address: matiasz@ias.edu

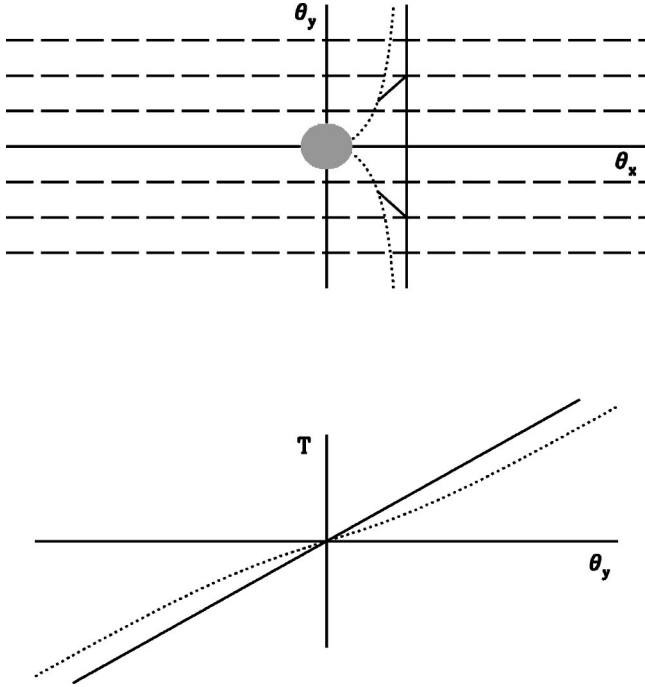


FIG. 1. In the upper panel we show a cluster lensing a background gradient. The bottom panel shows the temperature measured for a fixed θ_x as a function of θ_y , in the presence and absence of the cluster. Points with $\theta_y > 0$ get deflected to a smaller θ_y in the lens plane and thus for a positive gradient they will have a lower temperature in the lensed example than in the unlensed one. The opposite is true if $\theta_y < 0$.

A. A toy example

To understand the physics we will first consider the lensing induced by a cluster of galaxies. We will discuss a very simplified example in this paper, the reader interested in the detectability of the effect for real cluster should look at Ref. [12]. In most cases a cluster will subtend a few arcminutes on the sky. Over such a scale the primary anisotropies are expected to have negligible power so we will treat the unlensed temperature field as a pure gradient. Without loss of generality we will take the gradient to be along the y axis with an amplitude \tilde{T}_{y0} . The observed temperature becomes

$$T(\boldsymbol{\theta}) = \tilde{T}_{y0}(\theta_y + \delta\theta_y). \quad (2)$$

For a spherically symmetric cluster at the origin the deflection is of the form

$$\delta\boldsymbol{\theta} = -\delta\theta \frac{\boldsymbol{\theta}}{\theta}. \quad (3)$$

For a singular isothermal sphere $\delta\theta = 4\pi(\sigma_v/c)^2 D_{LS}/D_{OS}$, with σ_v being the cluster velocity dispersion, D_{LS} is the distance from the lens to the source, and D_{OS} is that between the observer and the source. In an Einstein DeSitter universe $D_{LS}/D_{OS} \approx 1/\sqrt{1+z_L}$, with z_L being the redshift of the cluster and where we have assumed that z_L is much smaller than the redshift of recombination. A singular isothermal profile will be a good approximation for a cluster up

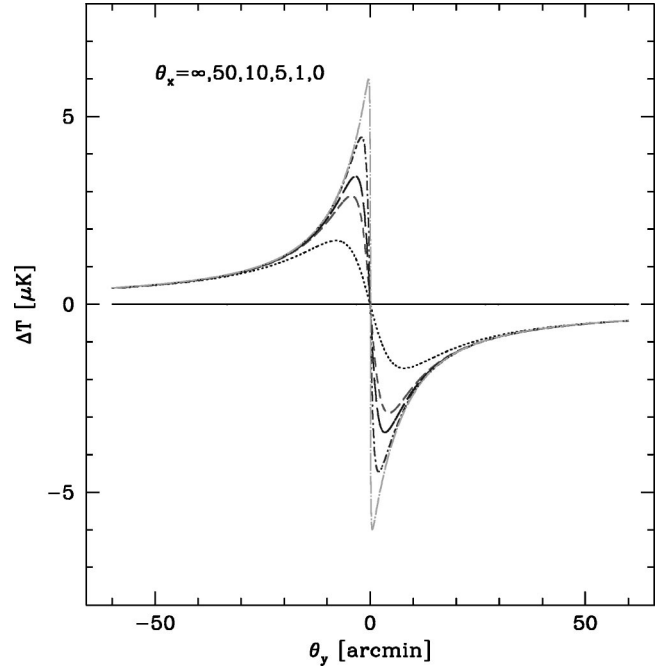


FIG. 2. Temperature profile of a CMB gradient lensed by a cluster. We took $T_{y0}\delta\theta = 13 \mu\text{K}$. The gradient part has been subtracted out for clarity. The cluster profile was cut off at 4 arcmin.

to some maximum radius after which the deflection will fall as $1/\theta$. In the central part of the cluster there might be a core. The typical value for $\delta\theta$ is a fraction of an arcminute.

The effect of lensing can be understood by looking at Fig. 1. We focus on the temperature as a function of θ_y for a fixed θ_x . In the absence of lensing we would observe the gradient. On the other hand the cluster will deflect the light rays so that for $\theta_y > 0$ the rays are coming from a lower value of θ_y in the last scattering surface. If the gradient is positive this implies that for $\theta_y > 0$ in the presence of the cluster we would observe a lower temperature than what would be observed if the cluster was not there. The opposite is true for $\theta_y < 0$. Far away from the cluster the lensed temperature should coincide again with the gradient. Thus the cluster creates a wiggle on top of the large scale gradient. This effect is shown in Fig. 2. The size of this wiggle is $T_{y0}\delta\theta$, thus it is proportional to the size of the gradient and the deflection angle. For a cluster we can use the determined deflection to infer the mass and some information about the cluster profile.

Note that the proposed signal is directly sensitive to the deflection angle and not the shear. This is the only method that has this property. It is usually argued that we can never measure the deflection angle in a lensing system because we do not know the original position of the background image. In this case we avoid this argument because we do know what the background image is, it is a gradient which we can measure on larger scales than the cluster. The other important point is that the effect is proportional to the gradient T_{y0} ; this is the signature we will use in what follows to identify the lensing effect by the large scale structure in the limit we are studying.

B. Large scale structure

For simplicity we will work in the small angle approximation so that we can expand the temperature field in Fourier modes instead of spherical harmonics. The Fourier components are defined as

$$T(\mathbf{l}) = \int d^2\theta e^{i\mathbf{l}\cdot\theta} T(\theta),$$

$$\delta\theta(\mathbf{l}) = \int d^2\theta e^{i\mathbf{l}\cdot\theta} \delta\theta(\theta). \quad (4)$$

We assume both the CMB and the deflection angle are Gaussian random fields characterized by their power spectra:

$$\langle \tilde{T}(\mathbf{l}_1) \tilde{T}(\mathbf{l}_2) \rangle = (2\pi)^2 \delta^D(\mathbf{l}_{12}) C_{l_1}^{\tilde{T}\tilde{T}},$$

$$\langle \delta\theta(\mathbf{l}_1) \cdot \delta\theta(\mathbf{l}_2) \rangle = (2\pi)^2 \delta^D(\mathbf{l}_{12}) C_{l_1}^{\delta\delta}, \quad (5)$$

where $\mathbf{l}_{12} = \mathbf{l}_1 + \mathbf{l}_2$, $C_{l_1}^{\tilde{T}\tilde{T}}$ is the power spectrum of the primary CMB anisotropies and $C_{l_1}^{\delta\delta}$ is the power spectrum of the deflection angle.

To leading order in the deflection angle the Fourier components of the lensed CMB field are

$$T(\mathbf{l}) = \tilde{T}(\mathbf{l}) + \int \frac{d^2\mathbf{l}'}{(2\pi)^2} \delta\theta(\mathbf{l}') \cdot \nabla \tilde{T}(\mathbf{l} - \mathbf{l}'). \quad (6)$$

We want to calculate the power spectrum of the lensed CMB field (C_l^{TT}) on very small scales, scales much smaller than the coherence length of the unlensed CMB gradient. We start by considering a small patch of the sky of size L , with L small enough that the gradient of the unlensed field can be considered constant so that the unlensed CMB map can be approximated linearly. On these scales, the small scale variations of the deflection angle generate additional power in the lensed CMB field. We define $\tilde{T}_x(\theta) = \tilde{T}_{x0}$ and $\tilde{T}_y(\theta) = \tilde{T}_{y0}$ so that

$$T(\theta) = \tilde{T}_{x0}[\theta_x + \delta\theta_x(\theta)] + \tilde{T}_{y0}[\theta_y + \delta\theta_y(\theta)]. \quad (7)$$

For the Fourier modes at large l we get

$$T(\mathbf{l}) = \tilde{T}_{x0} \delta\theta_x(\mathbf{l}) + \tilde{T}_{y0} \delta\theta_y(\mathbf{l}). \quad (8)$$

Thus the power spectra of the lensed temperature is given by

$$\langle T(\mathbf{l}_1) T(\mathbf{l}_2) \rangle = (2\pi)^2 \delta^D(\mathbf{l}_{12}) \sigma_S C_{l_1}^{\delta\delta} / 2, \quad (9)$$

with $\sigma_S = \langle \tilde{T}_{x0}^2 + \tilde{T}_{y0}^2 \rangle$. For the standard cold dark matter model (SCDM) $\sigma_S \approx 2 \times 10^9 \mu\text{K rad}^{-1} \approx 13 \mu\text{K arcmin}^{-1}$. Note that to obtain Eq. (9) we have replaced \tilde{T}_{x0}^2 and \tilde{T}_{y0}^2 by their averages. This implies that we are measuring the power spectrum over a large enough area of the sky that there are many patches over which the gradient is averaged. Equation (9) is a very interesting result on its own, it shows that the power in the tail of the CMB anisotropies directly gives the power spectrum of the deflection angle. This result can also

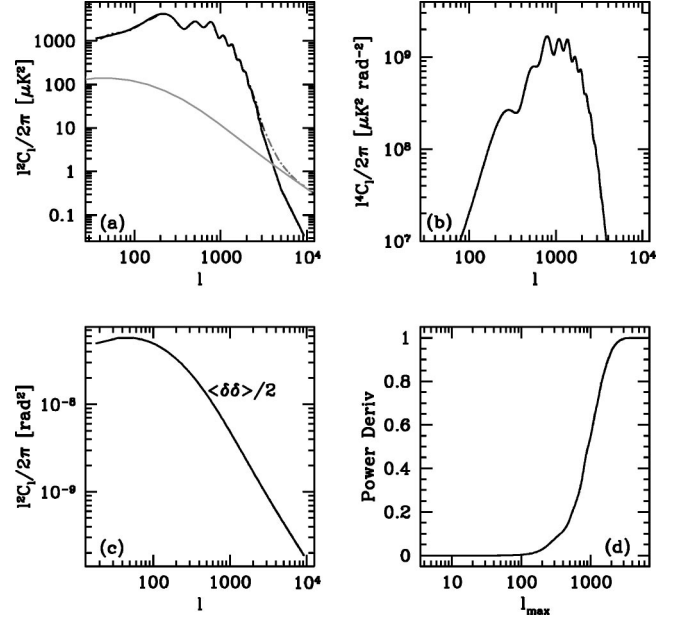


FIG. 3. The panel (a) shows the lensed and unlensed temperature power spectra together with $\sigma_S C_l^{\delta\delta}/2$. Panel (b) the power spectra of the derivative of the unlensed CMB field, (c) the spectra of $\delta\theta$ and (d) the cumulative power in the CMB derivative.

be derived by taking the appropriate limit in the full expression for the lensed CMB spectra as derived for example in [9].

In Fig. 3 we show the various power spectra so as to gain some insight on the orders of magnitude. In panel (a) we show the lensed and unlensed CMB spectra together with $\sigma_S C_l^{\delta\delta}/2$, the limiting value in the damping tail. Several conditions must be met for Eq. (9) to be valid. A very important fact is that the power spectra of both the deflection angle and the CMB derivative fall with l . This means that when we look at the power generated by lensing in a high l mode of the temperature, it could be dominated by a nearby l of the temperature derivative multiplied by a low l deflection angle mode rather than by the power coming from a small l temperature derivative mode multiplied by a high l deflection mode. Equation (9) is only valid in the latter case. We attain this limit because the power spectra of the deflection angle falls more slowly with l than that of the CMB derivative. The power in the deflection angle falls like a power law [Fig. 3(c)], while that in the CMB derivative drops exponentially [Fig. 3(b)]. As a consequence at a high enough l it is always easier to produce power by multiplying low frequency temperature modes by high frequency deflection modes. At $l > 4000$ this effect dominates, at lower l both effects compete. Another condition is that we should apply Eq. (9) for l values large enough that most of the power in the unlensed derivative (contributing to σ_S) comes from a lower l . Figure 3(d) shows that for SCDM most of the power in the derivative comes from scales $l < 2000$.

III. CORRELATION BETWEEN SMALL AND LARGE SCALES

In the previous section we have shown that the power in the CMB anisotropies at large enough l measures the power

spectrum of the deflection angle $\delta\theta$. If we want to use the small scale power to measure the fluctuations in the deflection angle we need a way to determine that the power we observe is indeed produced by lensing and is not at the last scattering surface (i.e., it is due to a different underlying primary anisotropies spectrum) or that it is not produced by another secondary effect. To be able to test this we need to consider higher order statistics, we need to go beyond the power spectrum.

The crucial point we are going to use is that the power produced by lensing comes from large scale primary temperature modes promoted to higher l by the small scale deflection of the photons. We will attempt to test the hypothesis that the power we observe is generated by lensing by looking at the correlation between the large scale temperature derivative and the small scale power. If we continue with the simplification that the gradient of the unlensed field is constant, when we high pass filter the temperature map we obtain

$$\begin{aligned} T_H(\boldsymbol{\theta}) &= \int_{l_{h1}}^{l_{h2}} \frac{d^2 l}{(2\pi)^2} e^{-i\boldsymbol{l}\cdot\boldsymbol{\theta}} T(\boldsymbol{l}) \\ &\approx \tilde{T}_x \delta\theta_x(\boldsymbol{\theta}) + \tilde{T}_y \delta\theta_y(\boldsymbol{\theta}), \end{aligned} \quad (10)$$

where l_{h1} and l_{h2} define a ring in l space and we have abused notation and called $\delta\theta_i(\boldsymbol{\theta})$ the high passed filtered deflection angle. We will study the behavior of the square of T_H properly smoothed:

$$\mathcal{H}(\boldsymbol{\theta}) = \int d^2\boldsymbol{\theta}' W_H(\boldsymbol{\theta}-\boldsymbol{\theta}') T_H^2(\boldsymbol{\theta}'). \quad (11)$$

If we smooth the map over a scale bigger than that of the variations in the deflection angle but over which \tilde{T}_x and \tilde{T}_y remain constant, we can replace $\delta\theta_x^2$ and $\delta\theta_y^2$ by their averages. It follows that the smoothed high passed filtered map is approximately

$$\mathcal{H}(\boldsymbol{\theta}) \approx \frac{\langle \delta\theta^2 \rangle}{2} (\tilde{T}_x^2 + \tilde{T}_y^2)(\boldsymbol{\theta}). \quad (12)$$

Only the modes with wave vectors $l_{h1} < l < l_{h2}$ contribute to the average $\langle \delta\theta^2 \rangle$ in Eq. (12).

We now look at the anisotropies on larger scales by applying a low pass filter to the temperature field,

$$T_L(\boldsymbol{\theta}) = \int_0^{l_i} \frac{d^2 l}{(2\pi)^2} e^{-i\boldsymbol{l}\cdot\boldsymbol{\theta}} T(\boldsymbol{l}). \quad (13)$$

We compute the square of the gradient,

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \|\nabla T_L\|^2 \\ &\approx (\tilde{T}_x^2 + \tilde{T}_y^2)(\boldsymbol{\theta}). \end{aligned} \quad (14)$$

Equations (12) and (14) show the maps of \mathcal{L} and \mathcal{H} trace each other,

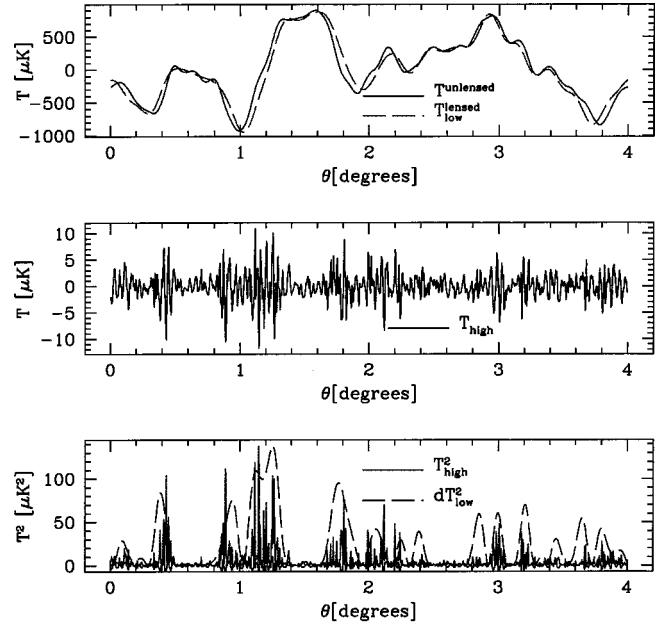


FIG. 4. Example of the generation of power by gravitational lensing in one dimension. The upper panel shows the unlensed temperature and the result of filtering the lensed T . The middle panel shows the high l power generated by lensing. In the bottom we show the square of the small scale power and the square of the large scale gradient (arbitrarily scaled).

$$\mathcal{H}(\boldsymbol{\theta}) \approx \frac{\langle \delta\theta^2 \rangle}{2} \mathcal{L}(\boldsymbol{\theta}). \quad (15)$$

We have a very simple test to establish that the small scale power is generated by lensing, the small scale power has to be correlated with the large scale derivative. Instead, if the small scale power is Gaussian power already present at the last scattering surface, different Fourier modes are uncorrelated. In the Gaussian case if we construct \mathcal{L} and \mathcal{H} with different modes ($l_{h1} > l_i$) there will be no correlation between the two maps. In practice instead of smoothing the high passed temperature map we work in Fourier space and correlate the low l modes of $T_H^2(\boldsymbol{\theta})$ with those of the $\mathcal{L}(\boldsymbol{\theta})$.

Equations (12) and (14) were obtained under simplifying assumptions, more general expressions can also be derived but we only show this limiting case to make the physics more transparent. Comparison with simulations show that these expressions are good approximations in the regime of interest. The procedure to simulate a lensed CMB map was discussed in detail in Ref. [6]. We generate realizations of CMB and projected mass density κ and use a ray tracing technic to produce a lensed CMB map.

We will explain the effect we are studying by first presenting a toy example in 1 dimension. We will consider a 2 dimensional universe so that the last scattering surface becomes a line. In the top panel of Fig. 4 we show the unlensed CMB field together with the low passed filtered ($l_i = 2000$) lensed anisotropies. It is clear that both curves trace each other, although they are displaced. This is the consequence of the large scale modes of the deflection angle, but remember we are after the power generated by the small scale

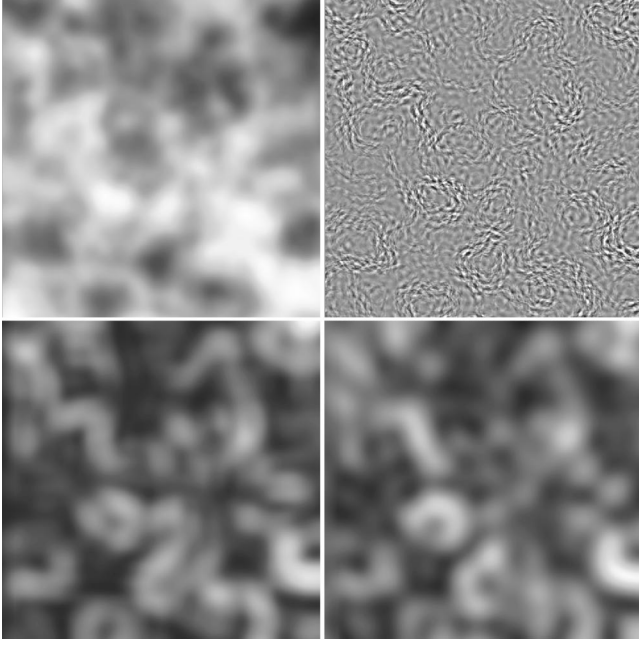


FIG. 5. The upper panel on the left shows the lensed temperature field. The upper right panel shows the high passed temperature. Bottom left has gradient square and bottom right smoothed square of high passed temperature.

modes of $\delta\theta$. In the middle panel we show the high passed ($l > 6000$) lensed CMB, the fluctuations in this panel are the result of lensing. The power is not distributed uniformly, the fluctuations are larger where the derivative of the CMB in the top panel is larger, as Eq. (12) indicates. This is the same physical effect we discussed in the cluster example. On the contrary in regions where the CMB is constant, surface brightness conservation implies that lensing does not create any power. The bottom panel shows the square of the small scale temperature anisotropies and a scaled version of the low l gradient to ease the comparison.

We can now look at the results for a 2 dimensional last scattering surface shown in Fig. 5. The low passed gradient map was constructed using scales $l < 2000$ and the high passed temperature was constructed with modes $l > 6000$. The upper left panel shows the lensed CMB map while the upper right panel has the high passed filtered field. It is clear that the small scale power is not Gaussian and is higher where the large scale derivative is higher. To make this even more apparent the two bottom panels show the \mathcal{L} and \mathcal{H} maps, the correlation is excellent.

The \mathcal{L} and \mathcal{H} maps only differ by a factor $\langle \delta\theta^2 \rangle / 2$ [Eq. (15)], so their cross correlation defined as $\langle \mathcal{H}(l_1) \mathcal{L}(l_2) \rangle = (2\pi)^2 \delta^D(l_{12}) C_l^{\mathcal{H}\mathcal{L}}$, is simply related to the power spectrum of the \mathcal{L} map ($C_l^{\mathcal{L}\mathcal{L}}$),

$$C_l^{\mathcal{H}\mathcal{L}} \approx \frac{\langle \delta\theta^2 \rangle}{2} C_l^{\mathcal{L}\mathcal{L}},$$

$$C_l^{\mathcal{H}\mathcal{H}} \approx \frac{\langle \delta\theta^2 \rangle^2}{4} C_l^{\mathcal{L}\mathcal{L}}. \quad (16)$$

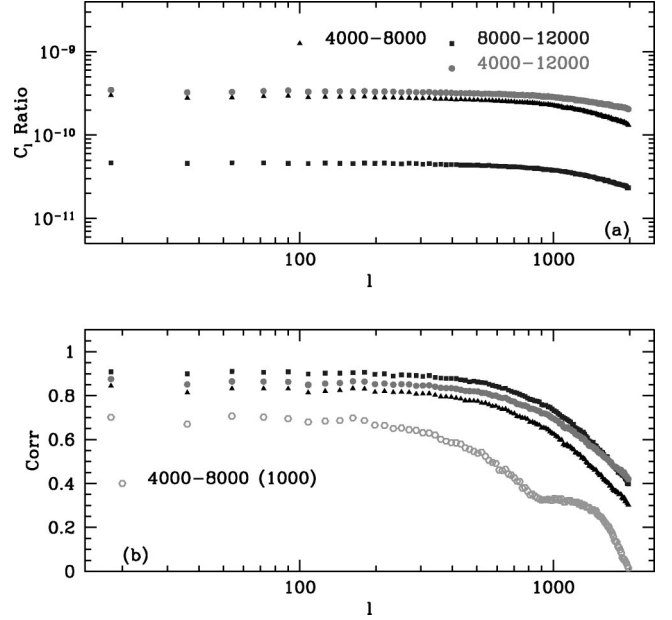


FIG. 6. Panel (a) shows the ratio $C_l^{\mathcal{H}\mathcal{L}}/C_l^{\mathcal{L}\mathcal{L}}$ for several different rings used to construct \mathcal{H} . Panel (b) shows the cross correlation coefficient in l space.

In Fig. 6(a) we show the ratio $C_l^{\mathcal{H}\mathcal{L}}/C_l^{\mathcal{L}\mathcal{L}}$ obtained in simulations. We have used $l < 2000$ to construct \mathcal{L} , and several different rings at higher l for \mathcal{H} . The order of magnitude of the ratio is consistent with Eq. (16), it remains constant over a wide range of l but starts to fall once we approach $l \sim 2000$. The first relation in Eq. (16) is really $C_l^{\mathcal{H}\mathcal{L}} = W_l \langle \delta\theta^2 \rangle C_l^{\mathcal{L}\mathcal{L}} / 2$ where W_l is some window function. The origin of the window can be understood as follows: let us call l_h a high l mode of the deflection and l_1 and l_2 two modes of the derivative. When we compute the gradient square these modes combine to give power at $l_3 = l_1 + l_2$. In the high passed case, we recover this component of the field by looking at the modes with $l_h + l_1$ and $-l_h + l_2$ so that when we multiply them they give the l_3 variations. As l_3 becomes larger at least one of l_1 or l_2 also become larger and because we are selecting only a ring in l space for the high passed map there are higher chances either $l_h + l_1$ or $-l_h + l_2$ will fall outside the ring and we cannot reconstruct l_3 . So as l_3 gets larger the correlation between the two maps fall. This effect explains the window W_l is less important the wider rings for the high passed maps is [compare the high l behavior of the 4000–12000 and 4000–8000 rings in Fig. 6(a)].

In Fig. 6(b) we show the correlation coefficient $C_l \equiv C_l^{\mathcal{H}\mathcal{L}} / \sqrt{(C_l^{\mathcal{H}\mathcal{H}} C_l^{\mathcal{L}\mathcal{L}})}$. Note that the cross correlation for low l is very high, almost one. This proves our claim that on large scales the \mathcal{L} and \mathcal{H} maps trace each other almost perfectly. There are several reasons why the two maps do not correlate exactly. First, although most of the power in the derivative of the CMB comes from $l < 2000$ there is still about 5% additional power coming from higher l . This degrades the cross-correlation because some of the power on the high passed map is coming from these modes of the derivative field. A more extreme case is shown in the figure, where only modes

with $l < 1000$ were used to construct the \mathcal{L} map, the cross-correlation in this case is significantly smaller. The other important effect is that some of the power in the high l map is due to primary anisotropies, these modes are uncorrelated with the low l derivative and thus reduce the cross correlation coefficient. This effect is more important for the 4000–8000 ring than for the 8000–12000 so the cross correlation is smaller for the former. The 4000–12000 rings fall somewhere in the middle of this two cases.

We will calculate $C_l^{\mathcal{L}\mathcal{L}}$, which can be done analytically. If we assume that only a Gaussian component is contributing to the low pass filtered field then it is possible to calculate the power spectra of the gradient squared, $\mathcal{L} = \|\nabla T_L\|^2$. The correlation function in real space for two points separated an angle θ in the \mathbf{x} direction gives

$$\langle \mathcal{L}(0)\mathcal{L}(\theta) \rangle = \langle \mathcal{L}^2 \rangle + 2[C_{xx}^2(\theta) + C_{yy}^2(\theta)]. \quad (17)$$

We have introduced $C_{xx}(\theta) = \langle T_x(0)T_x(\theta) \rangle$ and $C_{yy}(\theta) = \langle T_y(0)T_y(\theta) \rangle$. They are given by

$$\begin{aligned} C_{xx}(\theta) &\equiv \langle \tilde{T}_x(0)\tilde{T}_x(\theta) \rangle_{CMB} \\ &= (2\pi)^{-2} \int^{l_1} d^2\mathbf{l} e^{i\mathbf{l}\cdot\theta \cos\phi_l} l^2 C_l^{\tilde{T}\tilde{T}} \cos^2\phi_l \\ &= \int_0^{l_1} \frac{ldl}{4\pi} l^2 C_l^{\tilde{T}\tilde{T}} [J_0(l\theta) - J_2(l\theta)] \\ &\equiv \frac{1}{2} [C_0(\theta) - C_2(\theta)], \end{aligned}$$

$$\begin{aligned} C_{yy}(\theta) &\equiv \langle \tilde{T}_y(0)\tilde{T}_y(\theta) \rangle_{CMB} \\ &= (2\pi)^{-2} \int^{l_1} d^2\mathbf{l} e^{i\mathbf{l}\cdot\theta \cos\phi_l} l^2 C_l^{\tilde{T}\tilde{T}} \sin^2\phi_l \\ &= \int_0^{l_1} \frac{ldl}{4\pi} l^2 C_l^{\tilde{T}\tilde{T}} [J_0(l\theta) + J_2(l\theta)] \\ &\equiv \frac{1}{2} [C_0(\theta) + C_2(\theta)], \end{aligned}$$

$$\begin{aligned} C_{xy}(\theta) &\equiv \langle \tilde{T}_x(0)\tilde{T}_y(\theta) \rangle_{CMB} \\ &= (2\pi)^{-2} \int^{l_1} d^2\mathbf{l} e^{i\mathbf{l}\cdot\theta \cos\phi_l} l^2 C_l^{\tilde{T}\tilde{T}} \cos\phi_l \sin\phi_l \\ &= 0, \end{aligned} \quad (18)$$

where $C_0(\theta)$ and $C_2(\theta)$ are defined as the integrals over $l^3 C_l^{\tilde{T}\tilde{T}} dl/2\pi$ weighted with $J_0(l\theta)$ and $J_2(l\theta)$, respectively.

The constant part in Eq. (17) only contributes to the $l=0$ mode. The second term is $\sigma_S^2 N^{SS}(\theta)$ in the notation of [6]. The power spectra of the low pass filtered map is that of S in [6], but with only the low l modes included. For $l \neq 0$,

$$C_l^{\mathcal{L}\mathcal{L}} = 2\pi \int \theta d\theta (C_0^2 + C_2^2) J_0(l\theta). \quad (19)$$

This calculation coincides perfectly with the results of simulations. In the presence of detector noise the above expression generalizes to $C_l^{\mathcal{L}\mathcal{L}} \rightarrow C_l^{\mathcal{L}\mathcal{L}} + N_l^{\mathcal{L}\mathcal{L}}$, where $N_l^{\mathcal{L}\mathcal{L}}$ is the detector noise contribution calculated using Eqs. (18) and (19) with the power spectrum of the detector noise.

Let us now consider $C_l^{\mathcal{H}\mathcal{H}}$. If the contribution to the power were Gaussian then the correlation function of \mathcal{H} in real space would be

$$\langle \mathcal{H}(0)\mathcal{H}(\theta) \rangle = \langle \mathcal{H}^2 \rangle + 2C^2(\theta), \quad (20)$$

where

$$\begin{aligned} C(\theta) &\equiv \langle \tilde{T}_H(0)\tilde{T}_H(\theta) \rangle_{CMB} \\ &= (2\pi)^{-2} \int d^2\mathbf{l} e^{i\mathbf{l}\cdot\theta \cos\phi_l} C_l^{\tilde{T}\tilde{T}} \\ &= \int_{l_{h1}}^{l_{h2}} \frac{ldl}{2\pi} C_l^{\tilde{T}\tilde{T}} J_0(l\theta). \end{aligned} \quad (21)$$

The power spectrum becomes

$$C_l^{\mathcal{H}\mathcal{H}} = 4\pi \int \theta d\theta C^2(\theta) J_0(l\theta). \quad (22)$$

This expression only coincides with the results of simulations when either detector noise or intrinsic CMB anisotropies dominate the power in the l range used to construct \mathcal{H} , that is, if the unlensed power is Gaussian. In the idealized detector-noise-free examples we have discussed above, this corresponds to $l_{h2} < 4000$. Fortunately for the power generated by lensing, the power spectra of the \mathcal{H} map is a scaled version of the power spectrum of \mathcal{L} , thus it can be obtained using Eqs. (16) and (21). Again in the presence of noise $C_l^{\mathcal{H}\mathcal{H}} \rightarrow C_l^{\mathcal{H}\mathcal{H}} + N_l^{\mathcal{H}\mathcal{H}}$ with $N_l^{\mathcal{H}\mathcal{H}}$ calculated using Eq. (22) with the detector noise power spectra.

To assess the signal to noise of our lensing signal we need to calculate the variance in the estimator of $C_l^{\mathcal{H}\mathcal{L}}$:

$$\hat{C}_l^{\mathcal{H}\mathcal{L}} = \frac{A_f}{(2\pi)^2} \int \frac{dA_l}{A_l} \mathcal{H}(\mathbf{l}) \mathcal{L}(-\mathbf{l}), \quad (23)$$

where the integral is done over a small area in l space of size A_l centered around l and $A_f = (2\pi)^2/\Omega$ is the area of the fundamental cell in l space. We have denoted Ω the area of sky observed.

If we assume that \mathcal{L} and \mathcal{H} are Gaussian fields, we can calculate the variance as

$$\text{Cov}[(\hat{C}_l^{\mathcal{H}\mathcal{L}})^2] = \frac{A_f}{A} [(C_l^{\mathcal{H}\mathcal{L}})^2 + C_l^{\mathcal{H}\mathcal{H}} C_l^{\mathcal{L}\mathcal{L}}]. \quad (24)$$

The ratio A/A_f is the number of available modes one can use to measure the cross correlation. It can be approximated by $A_l/A_f = f_{sky}(2l+1)$ for a unit width ring. In Fig. 7 we show the variance in the cross correlation measured in simulations normalized to the Gaussian prediction in Eq. (24). The agreement is very good implying that we can use the Gaussian formula to compute the variance.

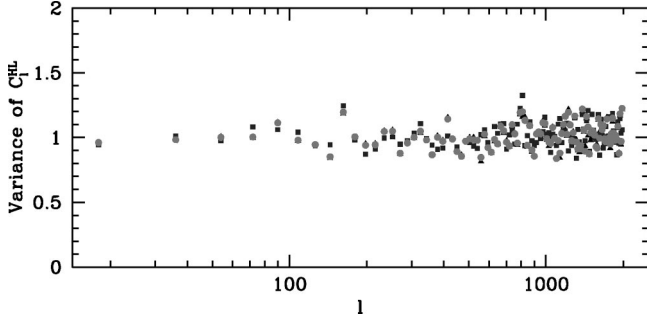


FIG. 7. Ratio of variance of power spectrum measured in the simulation to the predicted result assuming Gaussian fields.

To compute the total S/N we want to combine the signal in all the l modes of the cross correlation that we can measure. To do this we will compute $\hat{X} = \sum_l \alpha_l \hat{C}_l^{\mathcal{H}\mathcal{L}}$ choosing α_l to maximize $S/N = [\langle \hat{X} \rangle^2 / \text{Cov}(\hat{X}^2)]^{1/2}$. It is straightforward to show that $\alpha_l = C_l^{\mathcal{H}\mathcal{L}} / \text{Cov}[(C_l^{\mathcal{H}\mathcal{L}})^2]$ [10]. We then get

$$\begin{aligned} \frac{S}{N} &= \left[\sum_{l < l_1} \frac{(C_l^{\mathcal{H}\mathcal{L}})^2}{\text{Cov}[(C_l^{\mathcal{H}\mathcal{L}})^2]} \right]^{1/2} \\ &= \left[f_{\text{sky}} \sum_{l < l_1} (2l+1) \right. \\ &\quad \left. \times \frac{C_l^2}{[(1 + N_l^{\mathcal{L}\mathcal{L}}/C_l^{\mathcal{L}\mathcal{L}})(1 + N_l^{\mathcal{H}\mathcal{H}}/C_l^{\mathcal{H}\mathcal{H}}) + C_l^2]} \right]^{1/2}. \end{aligned} \quad (25)$$

We have implicitly assumed rings of unit width.

We want to relate Eq. (25) to the S/N with which we can measure the small scale power generated by lensing. It is straightforward to show using Eqs. (19), (22), and the fact that $C_l \sim 1$ that $N_l^{\mathcal{H}\mathcal{H}}/C_l^{\mathcal{H}\mathcal{H}} \approx (w_T^{-1}/\bar{C}^{TT})^2$, here \bar{C}^{TT} is the averaged power in the band $l_{h1} < l < l_{h2}$ and w_T^{-1} is the power spectrum of the detector noise, assumed to be white noise.

We will consider the limit in which w_T^{-1} is small enough that $N_l^{\mathcal{L}\mathcal{L}} \ll C_l^{\mathcal{L}\mathcal{L}}$ but is large enough that $N_l^{\mathcal{H}\mathcal{H}} \gg C_l^{\mathcal{H}\mathcal{H}}$. If the noise were even lower, then the S/N for detecting the cross correlation will be large $S/N \sim N_L^{1/2}$ where N_L is the total number of cross correlations that can be measured, $N_L \approx f_{\text{sky}} l_1^2$. In the limit we are considering,

$$\begin{aligned} \left(\frac{S}{N} \right)^2 &= N_L C_l^2 \left(\frac{\bar{C}_l}{w_T^{-1}} \right)^2 \\ &= 2 \frac{N_L}{N_P} C_l^2 \left(\frac{S}{N} \right)_P^2, \end{aligned} \quad (26)$$

where we have introduced $(S/N)_P^2 = N_P/2(\bar{C}_l/w_T^{-1})^2$, the signal to noise with which we can measure the power in the band $l_{h1} - l_{h2}$. We call N_P the number of modes that we can use to estimate the power, which is related to N_L by

$$\frac{N_L}{N_P} = \frac{l_1^2}{l_{h2}^2 - l_{h1}^2}. \quad (27)$$

Thus the S/N to measure the cross correlation is comparable although always somewhat smaller than that to detect the power directly. For example if we take $l_1 = 2000$, $l_{h1} = 4000$ and $l_{h2} = 6000$ then $S/N \approx \sqrt{2/5} \bar{C}_l (S/N)_P$.

IV. THREE AND FOUR POINT FUNCTIONS IN THE SMALL ANGLE LIMIT

In this paper and in our previous studies [5,6,10] we have investigated several ways of detecting the effect of gravitational lensing on the CMB. As we have explained, this amounts to trying to detect the distortions on the random CMB maps created by the random distributions of the dark matter in the universe. In [10] we used the integrated Sachs-Wolfe (ISW) effect as a tracer of the dark matter distribution and combinations of the CMB derivatives to measure the effect of lensing. The cross correlation of these two effects allowed us to gain information about the time evolution of the gravitational potential. Our method combined the information in particular configurations of the three point function of the temperature. Other studies have used other combinations of the bispectrum to detect the signal [4] and also calculated the contributions to the bispectrum coming from other secondary processes [4,11].

In [6] we used the power spectrum of a quadratic combination of derivatives of the CMB to measure the power spectrum of the projected mass density κ . This method was valid in the limit in which we wanted to recover the long wavelength modes of κ from information in the small scale CMB. This regime is analogous to weak lensing of background galaxies. In essence the different estimates of the power spectrum of κ at different scales were obtained by combining different configurations of the four point function of the lensed temperature.

In the present paper we studied other configurations of the four point function to illustrate the nature of the non-Gaussianities induced by lensing on small scales. The non-Gaussian nature of the generated power manifested itself in the correlations between the large scale gradient and the small scale generated power.

In order to have a unified picture of the different statistics we have proposed it is convenient to study directly the four point function of the temperature field and a three point function which correlates two temperatures and another field X . The field X stands for any field that cross correlates with κ . In our paper [10] $X = T$ but one can imagine doing this correlation with other tracers of the mass, like the fluctuations of the far-infrared background [13].

We define the connected three and four point functions as

$$\begin{aligned} \langle X(l_1)T(l_2)T(l_3) \rangle_c &= (2\pi)^2 \delta^D(l_{123}) T_3(l_1, l_2, l_3), \\ \langle T(l_1)T(l_2)T(l_3)T(l_4) \rangle_c &= (2\pi)^2 \delta^D(l_{1234}) T_4(l_1, l_2, l_3, l_4). \end{aligned} \quad (28)$$

Gravitational lensing produces

$$\begin{aligned}
T_3(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) &= 2C_{l_1}^{\kappa X} \left[C_{l_2}^{\bar{T}\bar{T}} \frac{\mathbf{l}_2 \cdot \mathbf{l}_1}{l_1^2} + C_{l_3}^{\bar{T}\bar{T}} \frac{\mathbf{l}_3 \cdot \mathbf{l}_1}{l_1^2} \right] \\
T_4(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3, \mathbf{l}_4) &= C_{l_1}^{\bar{T}\bar{T}} C_{l_2}^{\bar{T}\bar{T}} \left[\frac{(\mathbf{l}_1 + \mathbf{l}_3) \cdot \mathbf{l}_1 (\mathbf{l}_1 + \mathbf{l}_3) \cdot \mathbf{l}_2}{\|\mathbf{l}_1 + \mathbf{l}_3\|^2} C_{l_3}^{\delta\delta} + \frac{(\mathbf{l}_1 + \mathbf{l}_4) \cdot \mathbf{l}_1 (\mathbf{l}_1 + \mathbf{l}_4) \cdot \mathbf{l}_2}{\|\mathbf{l}_1 + \mathbf{l}_4\|^2} C_{l_4}^{\delta\delta} \right] \\
&+ \text{permutations (5 terms proportional to } C_{l_1}^{\bar{T}\bar{T}} C_{l_3}^{\bar{T}\bar{T}}, C_{l_1}^{\bar{T}\bar{T}} C_{l_4}^{\bar{T}\bar{T}}, C_{l_2}^{\bar{T}\bar{T}} C_{l_3}^{\bar{T}\bar{T}}, C_{l_2}^{\bar{T}\bar{T}} C_{l_4}^{\bar{T}\bar{T}}, C_{l_3}^{\bar{T}\bar{T}} C_{l_4}^{\bar{T}\bar{T}}). \quad (29)
\end{aligned}$$

The unconnected part of the four point function also gets corrections. To make the calculation of these terms fully consistent up to second order in the deflection angle we need to also consider the contributions coming from the second order in the expansion of Eq. (1). The unconnected terms are not relevant for our study so we will not write them down here.

In our previous papers we introduced three variables \mathcal{E} , \mathcal{B} , and \mathcal{S} . We had defined them in terms of derivatives to the temperature field. Equivalently we can write

$$\begin{aligned}
\mathcal{S}(\mathbf{l}) &= \int \frac{d^2 \mathbf{l}_1}{(2\pi)^2} (\mathbf{l} - \mathbf{l}_1) \cdot \mathbf{l}_1 T(\mathbf{l} - \mathbf{l}_1) T(\mathbf{l}_1), \\
\mathcal{Q}(\mathbf{l}) &= \int \frac{d^2 \mathbf{l}_1}{(2\pi)^2} [(l_x - l_{1x}) l_{1x} - (l_y - l_{1y}) l_{1y}] T(\mathbf{l} - \mathbf{l}_1) T(\mathbf{l}_1), \\
\mathcal{U}(\mathbf{l}) &= \int \frac{d^2 \mathbf{l}_1}{(2\pi)^2} [(l_x - l_{1x}) l_{1y} + (l_y - l_{1y}) l_{1x}] T(\mathbf{l} - \mathbf{l}_1) T(\mathbf{l}_1), \\
\mathcal{E}(\mathbf{l}) &= \mathcal{Q}(\mathbf{l}) \cos(2\phi_l) + \mathcal{U}(\mathbf{l}) \sin(2\phi_l), \\
\mathcal{B}(\mathbf{l}) &= -\mathcal{Q}(\mathbf{l}) \sin(2\phi_l) + \mathcal{U}(\mathbf{l}) \cos(2\phi_l). \quad (30)
\end{aligned}$$

When we average over the CMB random field we get

$$\begin{aligned}
\langle \mathcal{S}(\mathbf{l}) \rangle_{CMB} &= [(2\pi)^2 \delta^D(\mathbf{l}) - 2\kappa(\mathbf{l})] \sigma_{\mathcal{S}}, \\
\langle \mathcal{E}(\mathbf{l}) \rangle_{CMB} &= -2\kappa(\mathbf{l}) \sigma_{\mathcal{S}}, \\
\langle \mathcal{B}(\mathbf{l}) \rangle_{CMB} &= 0. \quad (31)
\end{aligned}$$

We have introduced $\sigma_{\mathcal{S}} = \int l dl / 2\pi l^2 C_l^{\bar{T}\bar{T}}$.

To extract all the information in this three point function we combine all possible configurations with a weight β chosen to maximize the signal to noise ratio. We define

$$\hat{Y} = \frac{A_f}{(2\pi)^2} \int \frac{d^2 \mathbf{l}_1}{A_l} \frac{d^2 \mathbf{l}_2}{A_l} \beta(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) X(\mathbf{l}_1) T(\mathbf{l}_2) T(\mathbf{l}_3). \quad (32)$$

For these mildly non-Gaussian maps, the variance can be calculated by only taking the Gaussian part of the temperature, so that

$$\text{Var}(\hat{Y}) = A_f (2\pi)^2 \int \frac{d^2 \mathbf{l}_1}{A_l^2} \frac{d^2 \mathbf{l}_2}{A_l^2} \beta^2(\mathcal{S}) 2C_{l_1}^{\kappa X} C_{l_2}^{\bar{T}\bar{T}} C_{l_3}^{\bar{T}\bar{T}}. \quad (33)$$

The power spectra in Eq. (33) must include the contribution from detector noise. There are additional terms in the variance if the field X and T had some cross correlation before lensing. In practice these terms are unimportant if one is interested in measuring the cross correlation $C_l^{\kappa X}$ at large angular scales (low l), as was the case in our study in [10]. This is so because most the information of lensing is encoded in the high l modes of the temperature, so it is effectively as if the integral over \mathbf{l}_2 in Eq. (32) is done over high l modes while the \mathbf{l}_1 integral only involves low l . Thus the terms that would involve $C_l^{\kappa X}$ are absent because there are no pair of triangles in which X and T are evaluated on the same l .

The weight β that maximizes the S/N is $\beta \propto T_3(\mathbf{l}_1, \mathbf{l}_2, \mathbf{l}_3) / 2C_{l_1}^{\kappa X} C_{l_2}^{\bar{T}\bar{T}} C_{l_3}^{\bar{T}\bar{T}}$. Finally we get

$$\begin{aligned}
\left(\frac{S}{N} \right)^2 &= A_f^{-1} \int d^2 \mathbf{l}_1 \int \frac{d^2 \mathbf{l}_2}{(2\pi)^2} \frac{T_3^2}{2C_{l_1}^{\kappa X} C_{l_2}^{\bar{T}\bar{T}} C_{l_3}^{\bar{T}\bar{T}}} \\
&= A_f^{-1} \int d^2 \mathbf{l}_1 \frac{4(C_{l_1}^{\kappa X})^2}{C_{l_1}^{\kappa X} C_{l_1}^{eff}} \\
\frac{1}{C_{l_1}^{eff}} &\equiv \int \frac{d^2 \mathbf{l}_2}{(2\pi)^2} \left[C_{l_2}^{\bar{T}\bar{T}} \frac{\mathbf{l}_2 \cdot \mathbf{l}_1}{l_1^2} + C_{l_3}^{\bar{T}\bar{T}} \frac{\mathbf{l}_3 \cdot \mathbf{l}_1}{l_1^2} \right]^2 \\
&\times \frac{1}{2C_{l_2}^{\bar{T}\bar{T}}(l_2) C_{l_3}^{\bar{T}\bar{T}}(l_3)}. \quad (34)
\end{aligned}$$

The power spectra in the denominator include the contribution from detector noise (amplified by the beam response $C_{l \rightarrow C_l + B^2 N_l}$). The easiest way to calculate C_l^{eff} is to use a Monte Carlo technique. We used the implementation of the VEGAS algorithm in Ref. [14].

The above formula can be compared to what we obtain using the \mathcal{S} and \mathcal{E} variables [10]. Equation (7) of [10] reads

$$\left(\frac{S}{N} \right)^2 = \int d^2 \mathbf{l} \frac{4(C_l^{\kappa X})^2}{C_l^{\kappa X}} W^2(l) \left(\frac{1}{N_l^{SS}} + \frac{1}{N_l^{\mathcal{E}\mathcal{E}}} \right), \quad (35)$$

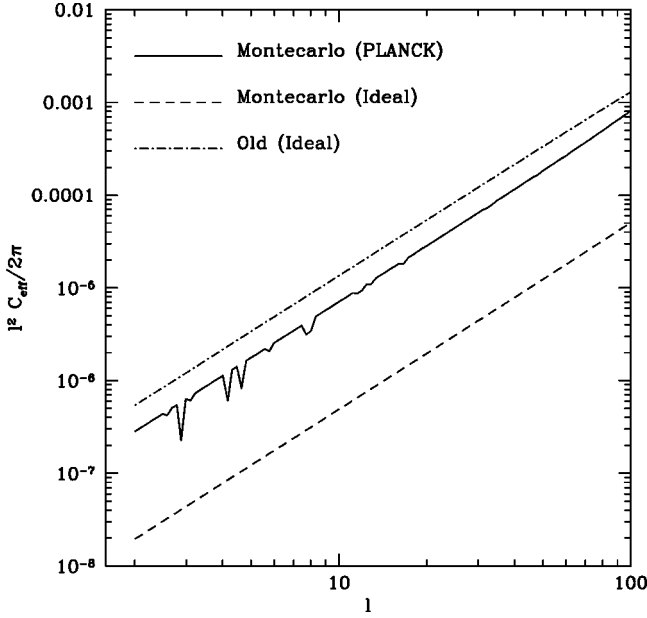


FIG. 8. Comparison between C_l^{eff} and the noise of our old method. We show the results for the Planck satellite and an ideal experiment. For our old method there was no difference in the noise between the Planck satellite and an ideal experiment on these angular scales.

where $W^2(l)$ is a window that encapsulates the effect of beam smearing. For low l N_l^{SS} and $N_l^{\mathcal{E}\mathcal{E}}$ are constant and satisfy, $N_l^{SS} = 2N_l^{\mathcal{E}\mathcal{E}}$.

In Fig. 8 we compare $1/C_l^{eff}$ to $W^2(l)(1/N_l^{SS} + 1/N_l^{\mathcal{E}\mathcal{E}})$. We show the results for two separate examples, an ideal experiment with no noise and infinite resolution and the Planck satellite. We focus on the large scale limit and for the ideal experiment we only consider the information coming from modes with $l < 3000$. There are several salient features of the comparison. Although the difference between the methods is not so large for the Planck satellite it is much larger for the ideal experiment. This can be easily understood. In our previous method the power spectrum of the CMB noise in this limit was

$$N_l^{SS} = (2\pi) \frac{\int l^5 dl (C_l^{\bar{T}\bar{T}})^2}{\left(\int l^3 dl C_l^{\bar{T}\bar{T}}\right)^2}. \quad (36)$$

It is clear from Eq. (36) that once we get into the damping tail where the $C_l^{\bar{T}\bar{T}}$ fall exponentially, N_l^{SS} no longer changes which means that our method does not receive any information from those modes. In contrast, Eq. (34) shows that the amplitude of the $C_l^{\bar{T}\bar{T}}$ cancels in C_l^{eff} as long as the modes had been measured with high S/N . Thus the new method continues to extract information from modes in the damping tail. In most practical cases this is not very important because the detector noise quickly dominates in this regime and then all methods downweight the modes. This explains why the difference between the two methods is not that large for the



FIG. 9. Quadrilaterals corresponding to the two limits discussed in the text. Panel (a) corresponds to configurations relevant for the cross correlation between the large scale gradient and the small scale power. Panel (b) shows the configurations that enter in the calculation of \mathcal{S} and \mathcal{E} .

Planck satellite. There is another relevant consideration. When computing the variance we assumed that the field was only mildly non-Gaussian and that we could take the unlensed temperature power spectrum to calculate it. This is clearly not the case on small scales. As we have shown in previous sections, on small scales the fluctuations become very non-Gaussian as most of the power is generated by lensing. We only consider modes with $l < 3000$ for the calculation of C_l^{eff} to partially take into account this effect.

The four point function

In the first part of this paper we have studied one particular physical limit when we are trying to recover information on the fluctuations of the mass distribution on scales much smaller than the coherence length of the CMB. To recover this limit we have to consider a quadrilateral in which two sides are much smaller than the other two [Fig. 9(a)]. The two small sides correspond to the low pass filtered derivatives while the large l correspond to the high passed filtered one. We consider the case where $l_1, l_2 \ll l_3 \sim l_4$. As we noted before the power spectrum of the primary anisotropies decreases exponentially, while that of the deflection angle is only as a power law. We conclude that of all the terms in Eq. (29) only those explicitly written dominate:

$$T_4(l_1, l_2, l_3, l_4) \approx 2 \frac{l_3 \cdot l_1 l_3 \cdot l_2}{l_3^2} C_{l_3}^{\delta\delta} C_{l_1}^{\bar{T}\bar{T}} C_{l_2}^{\bar{T}\bar{T}}, \quad l_1, l_2 \ll l_3 \sim l_4, \quad (37)$$

where we approximated $C_{l_3}^{\delta\delta} \approx C_{l_4}^{\delta\delta}$.

A different set of quadrilaterals dominate in the calculation of \mathcal{E} and \mathcal{S} . Those variables extract information about the large scale κ fluctuations from small angular scale fluctuations in the CMB. If we focus on modes of the temperature on scales larger than the damping tail $l^2 C_l^{\bar{T}\bar{T}}$ remains approximately constant while the power spectra of the deflection angle falls. Moreover \mathcal{S} and \mathcal{E} are a combination of the derivatives of the CMB and the extra l 's weigh the contribution to smaller scales. The power spectra of \mathcal{S} and \mathcal{E} are dominated by the type of quadrilaterals shown in Fig. 9(b). All the l 's are large but the quadrilaterals are thin. The thin diagonal corresponds to the l of the κ mode being recovered. The terms in Eq. (29) proportional to $C_{l_1}^{\bar{T}\bar{T}} C_{l_2}^{\bar{T}\bar{T}} C_{l_{12}}^{\delta\delta}$ (where l_1 and l_2 represent the length of the sides and l_{12} is the small diagonal) dominate.

To extract all the information in the four point function we can add all the quadrilaterals with an appropriate weight,

$$\hat{Z} = \frac{A_f}{(2\pi)^2} \int \frac{d^2 l_1}{A_l} \frac{d^2 l_2}{A_l} \frac{d^2 l_3}{A_l} \beta(S) T(l_1) T(l_2) T(l_3) T(l_4), \quad (38)$$

where $l_4 = -(l_1 + l_2 + l_3)$ and A_l is the area in l space we are using. For the optimal filter that minimizes S/N one gets $\beta \propto T_4 / C_{l_1}^{\tilde{T}\tilde{T}} C_{l_2}^{\tilde{T}\tilde{T}} C_{l_3}^{\tilde{T}\tilde{T}} C_{l_4}^{\tilde{T}\tilde{T}}$, where we have assumed Gaussianity to compute the variance. For the S/N we get

$$\left(\frac{S}{N}\right)^2 = \frac{A_f^{-1}}{24(2\pi)^4} \int d^2 l_1 d^2 l_2 d^2 l_3 \frac{T_4^2(l_1, l_2, l_3, l_4)}{C_{l_1}^{\tilde{T}\tilde{T}} C_{l_2}^{\tilde{T}\tilde{T}} C_{l_3}^{\tilde{T}\tilde{T}} C_{l_4}^{\tilde{T}\tilde{T}}}, \quad (39)$$

where the power spectra in the denominator should include the contribution from detector noise.

We change integration variables in Eq. (39) and write

$$\begin{aligned} \left(\frac{S}{N}\right)^2 &= A_f^{-1} \int d^2 l_1 \left(\frac{4C_{l_1}^{\kappa\kappa}}{\tilde{C}_{l_1}^{eff}}\right)^2 \\ \left(\frac{1}{\tilde{C}_{l_1}^{eff}}\right)^2 &= \frac{1}{24} \int \frac{d^2 l_2}{(2\pi)^2} \frac{d^2 l_3}{(2\pi)^2} \\ &\quad \times \frac{T_4^2(l_1 - l_2, l_2, l_3, -l_1 - l_3)}{C_{l_1 - l_2}^{\tilde{T}\tilde{T}} C_{l_2}^{\tilde{T}\tilde{T}} C_{l_3}^{\tilde{T}\tilde{T}} C_{l_1 + l_3}^{\tilde{T}\tilde{T}} (4C_{l_1}^{\kappa\kappa})^2}. \end{aligned} \quad (40)$$

This is a useful change of variables because it makes the integral resemble what we had in our old method. In this way the limit $l_1 \rightarrow 0$ corresponds to quadrilaterals that have four large sides but a small diagonal (l_1). Equation (6) of [5] reads

$$\left(\frac{S}{N}\right)^2 = A_f^{-1} \int d^2 l \frac{(4C_l^{\kappa\kappa})^2}{\sigma_{C_l}^2}, \quad (41)$$

with $\sigma_{C_l}^{-2} = \sigma_{C_{SS}}^{-2} + \sigma_{C_{\mathcal{E}\mathcal{E}}}^{-2} + \sigma_{C_{SE}}^{-2}$.

In Fig. 10 we compare the results C_l^{eff} in Eq. (40) with σ_{C_l} Eq. (41). The plot is qualitatively similar to Fig. 8 for the three point function. While there are hardly any improvements in our previous method when we go from the Planck satellite to an ideal experiment, there are significant differences for the optimal filter which continues to gather information from the damping tail. For the Planck satellite the situation is different, both the optimal method and our old method obtain a similar amount of information from the data. Even though the optimal method is able to get information from the damping tail in the ideal case, this is unimportant for the Planck satellite because the finite size of the beam makes it impossible. The fact that our previous method seems to have slightly less noise than the optimal method when l is a few hundred is most probably an artifact. The noise in our old method was calculated using a Gaussian approximation which we had seen breaking down slightly in our simulations [6]. As for the three point function we only included modes of the temperature with $l < 3000$ as the

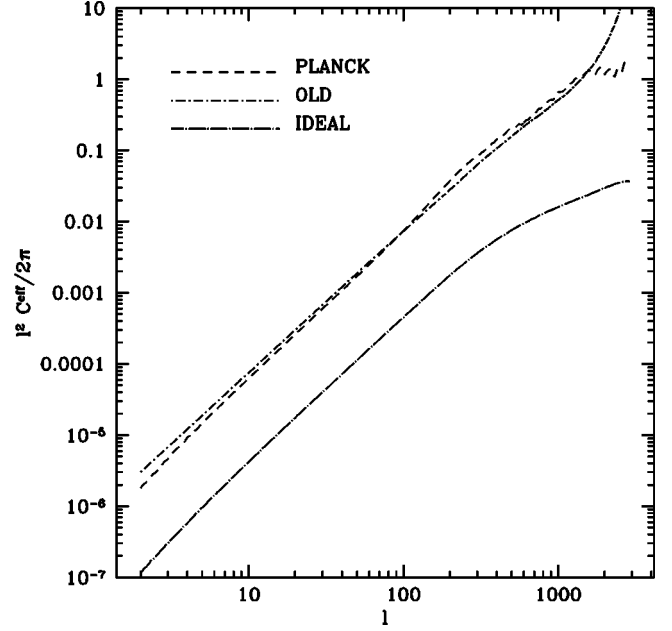


FIG. 10. The result of the Monte Carlo experiment for the four point function together with result from the previous technique.

power generated by lensing is non-Gaussian so our estimate of the variance of \hat{Z} is not valid on smaller scales.

In the limit $l_1 \rightarrow 0$ the four point function becomes approximately,

$$T_4(l_1 - l_2, l_2, l_3, -l_1 - l_3) \approx -C_{l_2}^{\tilde{T}\tilde{T}} C_{l_3}^{\tilde{T}\tilde{T}} l_1^2 C_{l_1}^{\delta\delta}. \quad (42)$$

To obtain Eq. (42) we had to assume that $C_l^{\kappa\kappa}$ is a decreasing function of l and that $C_{l_1 - l_2}^{\tilde{T}\tilde{T}} \approx C_{l_2}^{\tilde{T}\tilde{T}}$ and the equivalent formula for l_3 . Both of these assumptions break down in some range of l 's. For example, $C_l^{\kappa\kappa}$ has a peak at $l \sim 100$ and when l_2 is in the damping tail range, for finite l_1 there might be corrections due to the difference between $C_{l_1 - l_2}^{\tilde{T}\tilde{T}}$ and $C_{l_2}^{\tilde{T}\tilde{T}}$. We can still use this expression as a rough estimate to try to compare how the optimal β compares with the weight used by our previous method. In this limit and for the quadrilaterals relevant for this statistic, the optimal β is equivalent to multiplying each of the temperatures by $(C_l^{\tilde{T}\tilde{T}})^{1/2} / (C_l^{\tilde{T}\tilde{T}} + B_l^2 N_l^{\tilde{T}\tilde{T}})$. Thus for a temperature power spectra that goes as $C_l^{\tilde{T}\tilde{T}} \propto l^{-2}$ and when detector noise is irrelevant, the optimal filter amounts to multiplying the temperatures by l , equivalent to taking derivatives. This is the reason our previous method is not far from optimal in situations where we can neglect detector noise and we are not trying to extract information from the damping tail of the CMB. As we mention when we discussed the three point function, on small enough scales our treatment of the noise breaks down because the power is dominated by the power generated by lensing. The Gaussian approximation for the noise will not be valid.

V. CONCLUSIONS

We have studied the generation of power by gravitational lensing on small angular scales. We have shown that the power spectra of the anisotropies gives a measure of the spectrum of the deflection angle. The generated power is correlated with the size of the large scale gradient.

The generation of power can be understood by studying the lensing of the primary anisotropies by a cluster of galaxies. On the scales of a cluster the CMB can be assumed to be a simple gradient. Lensing generates a wiggle on top of the gradient that can be tens of μK . This signal will be large enough to be detected by a CMB experiment which targeted clusters with sufficient angular resolution, ~ 1 arcmin.

The lensing effect produced by the large scale structure of the universe can be separated from other secondary effects or from intrinsic CMB anisotropies at the last scattering surface by measuring the cross correlation between the map of the large scale gradient and the map of the small scale power. We have shown that this statistic has only a slightly smaller signal to noise than the measurement of the small scale power itself. The power generated by lensing dominates over the intrinsic fluctuations for $l \gtrsim 4000$.

We have calculated the three and four point functions of the lensing field in the small angle limit. The cross correlation between large and small scales as well as the statistics introduced in [6] are particular combinations of the four point function of the temperature field. We have calculated explicitly the dependence of the three and four point functions on the CMB and deflection angle power spectra as well as on the shape of the configuration. It is fair to say that both the statistic introduced here and those used in [6] can be viewed as particular ways of compressing the information in the four point function that take into account the physical intuition coming from our understanding of the lensing effect. The lensing effect predicts a particular dependence of the four point function on the configuration and scale that can be used to separate it from other non-Gaussian signals.

ACKNOWLEDGMENTS

We are grateful to Uros Seljak and Wayne Hu for very useful discussions. M.Z. is supported by NASA through grant HF-01116.01-98A from STScI, operated by AURA, Inc. under NASA contract NAS5-26555.

-
- [1] G. Jungman, M. Kamionkowski, A. Kosowsky, and D. N. Spergel, Phys. Rev. Lett. **76**, 1007 (1996); Phys. Rev. D **54**, 1332 (1996); J. R. Bond, G. Efstathiou, and M. Tegmark, Mon. Not. R. Astron. Soc. **291**, 33 (1997); M. Zaldarriaga, D. N. Spergel, and U. Seljak, Astrophys. J. **488**, 1 (1997); M. Tegmark, D. Eisenstein, W. Hu, and A. de Olivera Costa, *ibid.* **530**, 133 (2000).
 - [2] M. Zaldarriaga, Phys. Rev. D **55**, 1822 (1997).
 - [3] F. Bernardeu, Astron. Astrophys. **432**, 15 (1997); **338**, 767 (1998).
 - [4] D. M. Goldberg and D. N. Spergel, Phys. Rev. D **59**, 103002 (1999).
 - [5] U. Seljak and M. Zaldarriaga, Phys. Rev. Lett. **82**, 2636 (1999).
 - [6] M. Zaldarriaga and U. Seljak, Phys. Rev. D **59**, 123507 (1999).
 - [7] R. B. Metcalf and J. Silk, Astrophys. J. Lett. **492**, L1 (1998).
 - [8] Information on CBI can be found at <http://astro.caltech.edu/~tjp/CBI/abstract.html>
 - [9] U. Seljak, Astrophys. J. **463**, 1 (1996).
 - [10] U. Seljak and M. Zaldarriaga, Phys. Rev. D **60**, 043504 (1999).
 - [11] A. Cooray and W. Hu, astro-ph/9910397.
 - [12] U. Seljak and M. Zaldarriaga, astro-ph/9907254.
 - [13] L. Knox, Z. Haiman, and M. Zaldarriaga (in preparation).
 - [14] W. Press *et al.*, *Numerical Recipes* (Cambridge University Press, Cambridge, England, 1992).