

Likelihood Analysis of CMB Temperature and Polarization Power Spectra

Article (Published Version)

Hamimeche, Samira and Lewis, Antony (2008) Likelihood Analysis of CMB Temperature and Polarization Power Spectra. *Physical Review D*, 77 (10). ISSN 1050-2947

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/29262/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Likelihood analysis of CMB temperature and polarization power spectra

Samira Hamimeche* and Antony Lewis⁺

Institute of Astronomy, Madingley Road, Cambridge, CB3 0HA, United Kingdom

(Received 8 January 2008; published 28 May 2008)

Microwave background temperature and polarization observations are a powerful way to constrain cosmological parameters if the likelihood function can be calculated accurately. The temperature and polarization fields are correlated, partial-sky coverage correlates power spectrum estimators at different l , and the likelihood function for a theory spectrum given a set of observed estimators is non-Gaussian. An accurate analysis must model all these properties. Most existing likelihood approximations are good enough for a temperature-only analysis, however they cannot reliably handle temperature-polarization correlations. We give a new general approximation applicable for correlated Gaussian fields observed on part of the sky. The approximation models the non-Gaussian form exactly in the ideal full-sky limit and is fast to evaluate using a precomputed covariance matrix and set of power spectrum estimators. We show with simulations that it is good enough to obtain correct results at $l \gtrsim 30$ where an exact calculation becomes impossible. We also show that some Gaussian approximations give reliable parameter constraints even though they do not capture the shape of the likelihood function at each l accurately. Finally we test the approximations on simulations with realistically anisotropic noise and asymmetric foreground mask.

DOI: [10.1103/PhysRevD.77.103013](https://doi.org/10.1103/PhysRevD.77.103013)

PACS numbers: 98.70.Vc

I. INTRODUCTION

The cosmic microwave background (CMB) anisotropies are a powerful cosmological probe as they depend simply on the primordial inhomogeneities, content and geometry of the Universe. If the perturbations are Gaussian, the full-sky power spectra of the CMB anisotropies and their polarization contain all of the cosmological information. Information in the polarization power spectra can help to break degeneracies that are present if only temperature information is used, and also helps to reduce cosmic variance uncertainty. Parameter constraints can therefore be significantly improved by using polarization information even if the data is significantly noisier than the temperature.

An accurate joint likelihood analysis of the CMB temperature and polarization data is crucial to estimate cosmological parameters reliably. In principle this is straightforward at linear order if the primordial perturbations are Gaussian as the distribution can be calculated exactly. However calculating the likelihood exactly from partial-sky data with anisotropic noise is computationally prohibitive except at low l because large matrices need to be inverted. Most analyses therefore rely on approximations to the likelihood function at high l , using only the information in a set of estimators for the power spectra and a covariance estimated (or calibrated) from simulations. An alternative approach not considered further here would be to use the Gibbs sampling approach of Ref. [1,2], though this has serious convergence problems of its own [3].

If the likelihood of the theory power spectrum C_l as a function of the measured estimators \hat{C}_l were Gaussian, the likelihood could be calculated straightforwardly from the measured \hat{C}_l . However the distribution is non-Gaussian because for a given temperature power spectrum C_l , the \hat{C}_l , a sum of squares of Gaussian harmonic coefficients, have a (reduced) χ -squared distribution. At large l the distribution does tend to Gaussian by the central limit theorem; for example, the mean and maximum likelihood values of \hat{C}_l converge as $1/l$. However the precision with which we can hope to measure the cosmological parameters also improves at $1/l_{\max}$, so the relative bias due to the non-Gaussianity is potentially independent of l . On all scales the distribution must be modeled carefully to get unbiased cosmological parameter constraints.

The importance of the non-Gaussianity of the temperature likelihood function at low l is well known, and there are several well-established likelihood approximations to model it [4–7]. Current polarization data only contributes interesting information at low l where an exact likelihood can be used [8–10], however in the future the small-scale polarization signal will be less noise dominated and contain useful information. On small scales the likelihood function cannot be computed exactly in reasonable time, and the likelihood function is significantly more complicated than for the temperature because the temperature and polarization fields are correlated. The only existing attempt to model the polarized likelihood function at high l , Ref. [7], relies on variable transformations that are not guaranteed to be well defined, and is untested in practice. We give a new general well-defined likelihood approximation that can be used with partial-sky Gaussian polarized CMB data, or any other set of correlated Gaussian fields

*samira@ast.cam.ac.uk

⁺<http://cosmologist.info>

observed on part of the sky. It is exact in the full-sky limit, and can easily be calibrated from simulations. We also discuss under what circumstances a Gaussian likelihood approximation is reliable.

The layout of the paper is as follows. In Sec. II we present a brief overview of the exact full-sky likelihood function for isotropic noise, and discuss the accuracy required in general for unbiased parameter estimation. We start Sec. III with a review of various temperature likelihood approximations available in the literature, discuss the accuracy of the various Gaussian approximations, and move on to derive a new general likelihood approximation [Eq. (47)] that is exact in the full-sky limit. In Sec. IV we test the approximations by comparing with the exact likelihood function for azimuthal sky cuts and consistency with the binned likelihood. Finally in Sec. V we check the approximations with realistically anisotropic noise and demonstrate consistent parameter estimation from simple Planck-like simulations. Some mathematical and analysis details are described in the appendices: Appendix A gives identities relating expressions with symmetric matrices to expressions with a vector of components; Appendix B calculates the non-Gaussian correction to the full-sky effective chi-squared; Appendix C gives results for the likelihood function when using cross-power spectrum estimators from different maps; Appendix D reviews the basic pseudo- C_l estimator and exact likelihood formalism and Appendix E slightly generalizes previous hybrid pseudo- C_l estimators for anisotropic noise and gives details of our Planck-like test simulations.

We assume Gaussianity and statistical isotropy of the fields, and focus on the idealized case of pure CMB observations without the complications of foregrounds, point sources, nonlinear effects, anisotropic beams, and other observational artefacts. Generalizing our work to more realistic situations will be crucial for application to real data. If the fluctuations turn out to be significantly non-Gaussian or anisotropic a more complicated analysis may also be required.

II. EXACT FULL-SKY LIKELIHOOD FUNCTION

Observations on the full sky can be decomposed into spherical harmonics Y_{lm} , for example, the temperature at position Ω can be written

$$T(\Omega) = \sum_{lm} a_{lm}^T Y_{lm}(\Omega). \quad (1)$$

The polarization field can be expanded analogously in terms of E and B harmonics with opposite parity, see e.g. Ref. [11]. If the CMB field is Gaussian, as expected in linear theory, the corresponding harmonic components a_{lm}^T , a_{lm}^E and a_{lm}^B are Gaussian variables with zero mean. The CMB power spectrum C_l^{XY} determines the variance, which is independent of m if we assume statistical isotropy, so that

$$\langle |a_{lm}^T|^2 \rangle = C_l^{TT} \quad \langle |a_{lm}^E|^2 \rangle = C_l^{EE} \quad \langle |a_{lm}^B|^2 \rangle = C_l^{BB}. \quad (2)$$

The temperature and E -polarization fields are expected to be correlated, so there is an additional correlation power spectrum $\langle |a_{lm}^T a_{lm}^E| \rangle = C_l^{TE}$, but for a parity-invariant ensemble the B -polarization is expected to be uncorrelated to the other fields and the other cross-correlation power spectra are zero.

Since we only observe one sky, we cannot measure the power spectra directly, but instead form the rotationally invariant estimators, \hat{C}_l^{XY} , for full-sky CMB maps given by

$$\hat{C}_l^{XY} \equiv \frac{1}{2l+1} \sum_{m=-l}^l a_{lm}^X a_{lm}^Y. \quad (3)$$

The expectation values of these estimators are the true power spectra, $\langle \hat{C}_l^{XY} \rangle = C_l^{XY}$.

To keep things general we consider n (correlated) Gaussian fields, and define an n -dimensional vector \mathbf{a}_{lm} of the harmonic coefficients at each l and m . In the case of the CMB $\mathbf{a}_{lm} = (a_{lm}^T, a_{lm}^E, a_{lm}^B)^T$. The covariance matrix at each l is defined as

$$\mathbf{C}_l \equiv \langle \mathbf{a}_{lm} \mathbf{a}_{lm}^\dagger \rangle, \quad (4)$$

and the equivalent estimator is

$$\hat{\mathbf{C}}_l \equiv \frac{1}{2l+1} \sum_m \mathbf{a}_{lm} \mathbf{a}_{lm}^\dagger. \quad (5)$$

Since the \mathbf{a}_{lm} are assumed to be Gaussian and statistically isotropic, they have independent distributions (for $|m| \geq 0$) and the probability of a set of \mathbf{a}_{lm} at a given l is given by

$$\begin{aligned} -2 \ln(P(\{\mathbf{a}_{lm}\} | C_l)) &= \sum_{m=-l}^l [\mathbf{a}_{lm}^\dagger \mathbf{C}_l^{-1} \mathbf{a}_{lm} + \ln |2\pi \mathbf{C}_l|] \\ &= (2l+1) (\text{Tr}[\hat{\mathbf{C}}_l \mathbf{C}_l^{-1}] + \ln |C_l|) \\ &\quad + \text{const}. \end{aligned} \quad (6)$$

The fact that this likelihood for C_l depends only on the \hat{C}_l^{XY} (components of the matrix $\hat{\mathbf{C}}_l$) shows that on the full sky the CMB data can losslessly be compressed to a set of power spectrum estimators that contain all the relevant information about the posterior distribution. In other words $\hat{\mathbf{C}}_l$ is a sufficient statistic for the likelihood. Integrating out all the $\{\mathbf{a}_{lm}\}$ with the same $\hat{\mathbf{C}}_l$ (or normalizing with respect to $\hat{\mathbf{C}}_l$) gives a Wishart distribution¹ for $\hat{\mathbf{C}}_l$ (for a thorough review see Ref. [12]):

$$P(\hat{\mathbf{C}}_l | C_l) \propto \frac{|\hat{\mathbf{C}}_l|^{(2l-n)/2}}{|C_l|^{(2l+1)/2}} e^{-(2l+1)\text{Tr}(\hat{\mathbf{C}}_l C_l^{-1})/2}. \quad (7)$$

The likelihood function for C_l given the observed $\hat{\mathbf{C}}_l$ is

¹Technically $(2l+1)\hat{\mathbf{C}}_l \sim W_n(2l+1, C_l)$.

$\mathcal{L}(C_l|\hat{C}_l) \propto P(\hat{C}_l|C_l)$, an inverted Wishart distribution. It is straightforward to show that the likelihood has a maximum when $C_l = \hat{C}_l$, so \hat{C}_l is the maximum likelihood estimator. When $n = 1$, for example, when only the temperature is considered, the Wishart distribution reduces to

$$\begin{aligned}
 -2 \ln P(\hat{C}_l|C_l) &= (2l + 1) \\
 &\times \left[\hat{C}_l/C_l + \ln(C_l) - \frac{2l - 1}{2l + 1} \ln(\hat{C}_l) \right] \\
 &+ \text{const.} \quad (8)
 \end{aligned}$$

Considered as a function of \hat{C}_l this is a (reduced) χ -squared distribution with $2l + 1$ degrees of freedom; it has mean $\langle \hat{C}_l \rangle = C_l$, but maximum at $\hat{C}_l = C_l(2l - 1)/(2l + 1)$. This skewness is also apparent in the likelihood distribution $\mathcal{L}(C_l|\hat{C}_l) \propto P(\hat{C}_l|C_l)$, which peaks at $C_l = \hat{C}_l$ but has mean value $\hat{C}_l(2l + 1)/(2l - 3)$. The mean value of C_l calculated from the estimators should be above the \hat{C}_l , which is why using a quadratic approximation symmetric in C_l (with mean at $C_l = \hat{C}_l$) potentially biases results by $\mathcal{O}(1/l)$ at each l .

For n correlated Gaussian fields, there are in general $n(n + 1)/2$ distinct power spectra $[C_l]_{ij} = \langle a_{lm}^{(i)*} a_{lm}^{(j)} \rangle$, and on the full sky their estimators have covariance given by

$$\text{cov}([\hat{C}_l]_{ij}, [\hat{C}_l]_{pq}) = \frac{1}{2l + 1} ([C_l]_{ip}[C_l]_{jq} + [C_l]_{iq}[C_l]_{jp}). \quad (9)$$

It is sometimes convenient to work with vectors rather than matrices, so that $\mathbf{X}_l \equiv \text{vecp}(C_l)$ is a vector of the $n(n + 1)/2$ distinct elements of C_l , and similarly for the estimators. The corresponding covariance matrix is $\mathbf{M}_l \equiv \langle (\hat{\mathbf{X}}_l - \mathbf{X}_l)(\hat{\mathbf{X}}_l - \mathbf{X}_l)^T \rangle$. For symmetric matrices \mathbf{A} and \mathbf{B} a useful and somewhat unobvious identity is (see Appendix A)

$$\text{vecp}(\mathbf{A})^T \mathbf{M}_l^{-1} \text{vecp}(\mathbf{B}) = \frac{2l + 1}{2} \text{Tr}[\mathbf{A} \mathbf{C}_l^{-1} \mathbf{B} \mathbf{C}_l^{-1}], \quad (10)$$

which can be used to relate results involving C_l to results involving \mathbf{X}_l . In particular by writing $\text{Tr}[\hat{C}_l C_l^{-1}] = \text{Tr}[C_l C_l^{-1} \hat{C}_l C_l^{-1}]$ we can write the Wishart distribution in terms of the covariance $\mathbf{M}_l = \mathbf{M}_l(\mathbf{X}_l)$ as

$$\begin{aligned}
 -2 \log P(\hat{\mathbf{X}}_l|\mathbf{X}_l) &= 2\hat{\mathbf{X}}_l^T \mathbf{M}_l^{-1} \mathbf{X}_l + \frac{2l + 1}{n + 1} \log|\mathbf{M}_l| \\
 &- \frac{2l - 1}{n + 1} \log|\hat{\mathbf{M}}_l| + \text{const} \quad (11) \\
 &= 2(\hat{\mathbf{X}}_l - \mathbf{X}_l)^T \mathbf{M}_l^{-1} \mathbf{X}_l \\
 &+ \frac{2l + 1}{n + 1} \log|\mathbf{M}_l| - \frac{2l - 1}{n + 1} \log|\hat{\mathbf{M}}_l| \\
 &+ \text{const}, \quad (12)
 \end{aligned}$$

where we used $\log|\mathbf{M}_l| = (n + 1) \log|C_l| + \text{const}$ and $\hat{\mathbf{M}}_l = \mathbf{M}_l(\hat{\mathbf{X}}_l)$.

We now briefly review the standard Bayesian argument to link the function $P(d|\alpha)$ for the data d given parameters α , to the posterior $P(\alpha|d)$, the distribution of the parameters given the data. Bayes's theorem states that the posterior probability of α given the data is

$$P(\alpha|d) = \frac{P(d|\alpha)P(\alpha)}{P(d)} \propto \mathcal{L}(\alpha|d)P(\alpha), \quad (13)$$

where the prior $P(\alpha)$ gives information we already know about the models. In the case of linear CMB power spectra, the C_l can be computed essentially exactly from a set of parameters using standard Boltzmann codes. The probability distribution function of a set of parameters given observed data $\{\hat{C}_l\} \equiv d$ is therefore given on the noise-free full sky by

$$\begin{aligned}
 P(\alpha|\{\hat{C}_l\}) &\propto \mathcal{L}(\{C_l(\alpha)|\{\hat{C}_l\})P(\alpha) \\
 &= \prod_l \mathcal{L}(C_l(\alpha)|\hat{C}_l)P(\alpha). \quad (14)
 \end{aligned}$$

Since the prior depends on the models under consideration, in this paper we analyze the methods for estimating the likelihood $\mathcal{L}(\{C_l\}|\{\hat{C}_l\})$, which is the required input to cosmological parameter estimation codes such as COSMOMC [13]. When analyzing the likelihood function it is often convenient to normalize so that $\ln \mathcal{L} = 0$ when $C_l = \hat{C}_l$, i.e. to use

$$\begin{aligned}
 -2 \ln \mathcal{L}(\{C_l\}|\{\hat{C}_l\}) &= \sum_l (2l + 1) \{ \text{Tr}[\hat{C}_l C_l^{-1}] \\
 &- \ln|\hat{C}_l C_l^{-1}| - n \}. \quad (15)
 \end{aligned}$$

The expected value for this log likelihood is about $n(n + 1)/2$ per l , corresponding to the $n(n + 1)/2$ distinct components of C_l . For a more detailed analysis and discussion of ‘‘chi-squared’’ goodness of fit see Appendix B.

If there are multiple maps, for example, from different frequencies and detectors, cross-map \hat{C}_l estimators can be used to avoid noise bias. If a set of cross estimators is used the exact full-sky likelihood function is somewhat different from the above, as discussed in Appendix C. However in the limit of many maps the distribution becomes Wishart. In the limit in which there are enough maps that the information loss from using only cross estimators is small, the approximations developed in this paper should therefore also be applicable.

When the underlying fields are non-Gaussian, the analysis in this paper does not apply directly. However in many cases it is likely to be a good approximation to use the same likelihood approximations but with the covariance replaced with its full non-Gaussian version including 4-point terms. Non-Gaussianity associated with mode coupling (e.g. from nonlinear evolution) can also change the effective number of modes at a given scale. For example the B -mode CMB polarization power spectrum is generated by lensing of an E field by a relatively small number of lensing

convergence modes. This leads to strong correlations between l , and a drastically reduced number of modes compared to l_{\max}^2 expected for Gaussian fields. Reference [6] shows that using a likelihood approximation designed for analyzing Gaussian fields, but allowing for the full covariance from the non-Gaussianity, can give acceptable results. They also demonstrate the importance of modeling the non-Gaussianity of the likelihood function accurately when analyzing fields that depend on a small number of underlying modes.

Required accuracy

To assess how accurately we need to be able to model the likelihood we need to know how biases on the posterior C_l translate into constraints on parameters. The simplest case is instructive: consider estimating an amplitude parameter A , where $C_l = AC_{fl}$ for some fiducial fixed spectrum C_{fl} . For zero noise and a range of l with $l_{\min} \leq l \leq l_{\max}$, we have

$$-2 \ln \mathcal{L}(A|\{\hat{C}_l\}) = \sum_l (2l+1) \left\{ \frac{1}{A} \text{Tr}[\hat{C}_l C_{fl}^{-1}] - \ln|\hat{C}_l C_{fl}^{-1}| + n \log A - n \right\}, \quad (16)$$

and the maximum likelihood value is

$$\hat{A} = \frac{\sum_l (2l+1) \text{Tr}[\hat{C}_l C_{fl}^{-1}]}{n \sum_l (2l+1)}. \quad (17)$$

If C_{fl} is the underlying true model then $\langle \hat{A} \rangle = 1$ and the Fisher variance is

$$\begin{aligned} \sigma_A^2 &\equiv - \left\langle \frac{d^2}{dA^2} \ln \mathcal{L}(A|\{\hat{C}_l\}) \right\rangle^{-1} \Big|_{A=1} = \frac{2}{n \sum_l (2l+1)} \\ &= \frac{2}{n((l_{\max}+1)^2 - l_{\min}^2)} \sim \frac{2}{nl_{\max}^2} \end{aligned} \quad (18)$$

for a range of l satisfying $l_{\max} \gg l_{\min}$. We therefore need any biases to give $\Delta \hat{A} \ll \sqrt{(2/n)}/l_{\max}$ in order for the bias on \hat{A} to be small compared to its error bar. If we have an l -dependent bias δC_l , the bias on \hat{A} from Eq. (17) is small compared to its error if

$$|\langle \delta \hat{A} \rangle| = \frac{|\sum_l (2l+1) \text{Tr}(C_l^{-1} \delta C_l)|}{n \sum_l (2l+1)} \ll \frac{\sqrt{2/n}}{l_{\max}}. \quad (19)$$

The tolerated bias scales as $1/(2l+1)$, so this criterion will be satisfied for $l_{\max} \gg l_{\min}$ for any systematic error with

$$\frac{1}{n} |\text{Tr}(C_l^{-1} \delta C_l)| \ll \sqrt{\frac{2}{n}} \frac{1}{2l+1} \sim \sqrt{\frac{1}{2n}} \frac{1}{l}. \quad (20)$$

For a multiplicative bias $\delta C_l = B_l C_l$ this criterion is $|B_l| \ll (2n)^{-1/2}/l$. Alternatively if B_l is a constant the requirement is $|B_l| \ll \sqrt{(2/n)}/l_{\max}$. We shall loosely refer to $1/(l\sqrt{n})$ as the ‘‘systematic error,’’ and require biases to be much smaller than this, which is appropriate for nearly full-sky observations. For a realistic experiment with effective sky coverage f_{sky} the bias can be $\sim f_{\text{sky}}^{-1/2}$ times larger.

In the presence of noise the situation is more complicated. For one field with $C_l \rightarrow C_l + N_l$, using the Gaussian approximation we require

$$\left| \sum_l B_l \frac{(2l+1)C_l^2}{(C_l + N_l)^2} \right| \ll \sqrt{2 \sum_l \frac{(2l+1)C_l^2}{(C_l + N_l)^2}}. \quad (21)$$

The bias should be smaller than the systematic error $\sim 1/l$ where $N_l \ll C_l$, but there is greater tolerance where the noise is important.

III. LIKELIHOOD APPROXIMATIONS

A. Single-field likelihood approximations

To approximate the likelihood on the cut sky, the usual approach when analyzing the CMB temperature is to develop a form for the log likelihood that is quadratic in some function of the C_l , and hence can easily be generalized to the cut sky using an estimate of the C_l covariance matrix. Here we summarize some common approximations in their full-sky form.

At large l , Eq. (8) is approximated by a symmetric Gaussian distribution where the variance is determined by the estimators themselves [4]:

$$-2 \ln \mathcal{L}_S(C_l|\hat{C}_l) = \frac{2l+1}{2} \left[\frac{\hat{C}_l - C_l}{\hat{C}_l} \right]^2. \quad (22)$$

This approximation is well known to produce a poor fitting to the true likelihood function at low l [4]; being symmetric it biases posterior C_l low compared to the true likelihood function. Approximating the exact likelihood of Eq. (8) with a second order expansion in $\hat{C}_l/C_l - 1$ gives the same form but with \hat{C}_l replaced by C_l in the denominator:

$$-2 \ln \mathcal{L}_Q(C_l|\hat{C}_l) = \frac{2l+1}{2} \left[\frac{\hat{C}_l - C_l}{C_l} \right]^2. \quad (23)$$

This distribution is closer to the true likelihood, being skewed in the right direction, however it is still a poor approximation in general, this time biasing the posterior C_l high. It is often somewhat misleadingly referred to as the ‘‘Gaussian approximation,’’ even though it does not have the determinant term required for $P(\hat{C}_l|C_l)$ to be a normalized Gaussian distribution.² Another possibility is

²For this reason we denote it \mathcal{L}_Q —for a quadratic approximation—rather than \mathcal{L}_G used by some other authors.

$$-2 \ln \mathcal{L}_f(C_l | \hat{C}_l) = \frac{2l+1}{2} \left[\frac{\hat{C}_l - C_l}{C_{fl}} \right]^2, \quad (24)$$

where C_{fl} is some fixed fiducial model assumed to be smooth and close to the model C_l under consideration. This is more interesting as although the shape of the likelihood is wrong at any given l , as we shall see when summed over a range of l it can give results consistent with the exact likelihood function. It is equivalent to a Gaussian approximation since the determinant term is a constant when using a fixed fiducial model. Adding a C_l -dependent determinant term to the quadratic approximation can also produce valid results; we refer to this as Gaussian_D, given by

$$-2 \ln \mathcal{L}_D(C_l | \hat{C}_l) = \frac{2l+1}{2} \left[\frac{\hat{C}_l - C_l}{C_l} \right]^2 + \ln |C_l|. \quad (25)$$

See Sec. III B for more details of this approximation.

Beyond these quadratic/Gaussian approximations, other approximations that have been used include the log-normal (LN) distribution where the log likelihood is quadratic in the log of the power [4]

$$-2 \ln \mathcal{L}_{LN}(C_l | \hat{C}_l) = \frac{2l+1}{2} \left[\ln \left(\frac{\hat{C}_l}{C_l} \right) \right]^2. \quad (26)$$

This distribution is also somewhat biased [5,6]: it only matches the exact full-sky result to second order in $\hat{C}_l/C_l - 1$.

A weighted combination of the quadratic and the log-normal distributions can be a more accurate approximation to the exact likelihood, being correct to third order in $\hat{C}_l/C_l - 1$. This approximation was adopted in the analysis of the one, three and five-year WMAP data at high l [5]:

$$\ln \mathcal{L}_{\text{WMAP}}(C_l | \hat{C}_l) = \frac{1}{3} \ln \mathcal{L}_Q(C_l | \hat{C}_l) + \frac{2}{3} \ln \mathcal{L}_{LN}(C_l | \hat{C}_l). \quad (27)$$

Reference [6] suggests even better approximations of the form

$$\begin{aligned} -2 \ln \mathcal{L}(C_l | \hat{C}_l) &\approx (2l+1) \frac{9}{2} \left(\frac{2l+\alpha}{2l+1} \right)^{1/3} \\ &\times \left[\left(\frac{\hat{C}_l}{C_l} \right)^{1/3} - \left(\frac{2l+\alpha}{2l+1} \right)^{1/3} \right]^2 \\ &+ (1-\alpha) \ln C_l, \end{aligned} \quad (28)$$

where α is one (referred to as “ $-1/3$ ” approximation) or minus one (referred to as “ $1/3$ ” approximation). The value $\alpha = 1/3$ corresponds to taking the distribution of $\hat{C}_l^{1/3}$ to be Gaussian. These approximations are correct to third order in $\hat{C}_l/C_l - 1$, and also very nearly correct to fourth order.

B. Gaussian approximation for correlated fields

For a model C_{fl} with corresponding full-sky $\hat{\mathbf{X}}_l$ covariance \mathbf{M}_{fl} , a Gaussian approximation to the likelihood function is given by

$$\begin{aligned} -2 \ln \mathcal{L}_f(C_l | \hat{C}_l) &= (\mathbf{X}_l - \hat{\mathbf{X}}_l)^T \mathbf{M}_{fl}^{-1} (\mathbf{X}_l - \hat{\mathbf{X}}_l) \\ &+ \log |\mathbf{M}_{fl}| \\ &= \frac{2l+1}{2} \\ &\times \text{Tr}[(C_l - \hat{C}_l) C_{fl}^{-1} (C_l - \hat{C}_l) C_{fl}^{-1}] \\ &+ (n+1) \log |C_{fl}|. \end{aligned} \quad (29)$$

In the second line we used Eq. (10). If C_{fl} is fixed (independent of C_l) the determinant factors can be dropped, giving the generalization of the approximation for one field given in Eq. (24). It is worth studying this approximation more carefully as it turns out to be very good for smooth models even if the shape of the likelihood function at each l is not accurate. To see this, consider how the total likelihood varies with a parameter θ ,

$$\begin{aligned} -2 \frac{\partial \ln \mathcal{L}_f(\theta | \hat{C}_l)}{\partial \theta} &= \sum_l (2l+1) \\ &\times \text{Tr} \left[\frac{\partial C_l}{\partial \theta} C_{fl}^{-1} (C_l - \hat{C}_l) C_{fl}^{-1} \right], \end{aligned} \quad (31)$$

and compare with the equivalent result for the exact likelihood function

$$-2 \frac{\partial \ln \mathcal{L}(\theta | \hat{C}_l)}{\partial \theta} = \sum_l (2l+1) \text{Tr} \left[\frac{\partial C_l}{\partial \theta} C_l^{-1} (C_l - \hat{C}_l) C_l^{-1} \right]. \quad (32)$$

This will be zero for the maximum likelihood value $\hat{\theta}$, and if $C_{fl} \propto C_l(\hat{\theta})$ then $\hat{\theta}$ will also maximize the approximate likelihood function \mathcal{L}_f . In other words the approximation returns the exact best-fit value as long as the fiducial model is proportional to the best-fit model. If the true model and the fiducial model are both smooth functions of l , this will often be approximately true locally, even if it is not strictly true everywhere. An error in the normalization of C_{fl} would effect the error bar on $\hat{\theta}$. However since we can easily choose a fiducial model with fractional difference $< \mathcal{O}(1/\sqrt{l})$, this would only be a small fractional error on the error. The numerical values of the log likelihoods typically differ by $\mathcal{O}(\ln(l_{\text{max}}))$ (assuming the fiducial model is accurate to $\mathcal{O}(1/l)$; cf. discussion in Appendix B), but \mathcal{L}_f is otherwise generally a good approximation for smooth models.

Note that the above comments only apply to the Gaussian approximation using a fixed fiducial model. If instead we make the covariance \mathbf{M}_l a function of C_l the best-fit model would differ from the exact result due to additional terms in the derivative from the change in the

covariance with parameters. However the Gaussian approximation is still quite accurate, and unbiased in an average sense. To see this first consider the simple case of estimating an amplitude parameter A , where the exact result for the best-fit value was given in Eq. (17), or in terms of \mathbf{X}_l by

$$\hat{A} = 1 + \frac{2\sum_l \Delta \mathbf{X}_l \mathbf{M}_l^{-1} \mathbf{X}_l}{n\sum_l (2l+1)}, \quad (33)$$

where $\Delta \mathbf{X}_l \equiv \hat{\mathbf{X}}_l - \mathbf{X}_l$. Using the Gaussian approximation with $\mathbf{M}_{fl} = \mathbf{M}_l(\mathbf{X}_l)$ and expanding we instead get the best-fit value

$$\begin{aligned} \hat{A}' = \hat{A} + \frac{\sum_l [\Delta \mathbf{X}_l \mathbf{M}_l^{-1} \Delta \mathbf{X}_l - n(n+1)/2]}{n\sum_l (2l+1)/2} \\ - \left[\frac{\sum_l \mathbf{X}_l \mathbf{M}_l^{-1} \Delta \mathbf{X}_l}{n\sum_l (2l+1)/2} \right]^2 + \mathcal{O}(\Delta_l^{-1/2} l^{-3/2}), \end{aligned} \quad (34)$$

where Δ_l is the size of the range of l under consideration (assuming $l \gg 1$). The second term has expectation value zero in the true model, and typical variation of order $\mathcal{O}(\Delta_l^{-1/2} l^{-1})$. The third term is of order $\mathcal{O}(\Delta_l^{-1} l^{-1})$. So in almost all realizations with $\Delta_l \gg 1$, $l \gg 1$ we have $\hat{A}' = \hat{A} + \mathcal{O}(\Delta_l^{-1/2} l^{-1})$. The Gaussian approximation is therefore almost certainly good to within the required error of $\mathcal{O}(1/l)$ as long as $\Delta_l \gg 1$. However unless Δ_l is large it will not be much better than required: local features are likely to be more problematic than the overall amplitude (determined from $\Delta_l = l_{\max}$). More generally we can consider the expectation of the log likelihood

$$\begin{aligned} -2\langle \ln \mathcal{L}_f(\{\mathbf{X}_l\}|\{\hat{\mathbf{X}}_l\}) \rangle_t = \sum_l \{ (\mathbf{X}_l - \mathbf{X}_l^{(t)})^T \mathbf{M}_{fl}^{-1} (\mathbf{X}_l - \mathbf{X}_l^{(t)}) \\ + \text{Tr}[\mathbf{M}_{fl}^{-1} \mathbf{M}_l^{(t)}] + \log |\mathbf{M}_{fl}| \}, \end{aligned} \quad (35)$$

compared to the exact result

$$-2\langle \ln \mathcal{L}(\{\mathbf{X}_l\}|\{\hat{\mathbf{X}}_l\}) \rangle_t = \sum_l (2l+1) \{ \text{Tr}[\mathbf{C}_l^{(t)} \mathbf{C}_l^{-1}] + \log |\mathbf{C}_l| \}. \quad (36)$$

The exact mean log likelihood has a maximum at the true model, when $\mathbf{X}_l = \mathbf{X}_l^{(t)}$. This is however also true of the Gaussian approximation, both when \mathbf{M}_{fl} is for a fixed fiducial model, and also when we allow it to vary with parameters $\mathbf{M}_{fl} = \mathbf{M}_l(\mathbf{X}_l)$. To the extent that \mathbf{C}_l are constant in l , so that summing over l effectively averages the log likelihood, we therefore expect the Gaussian approximations to be nearly unbiased.

In the case when $\mathbf{M}_{fl} = \mathbf{M}_l(\mathbf{X}_l)$ the reliability of the Gaussian approximation depends critically on the inclu-

sion of the determinant term. For example dropping the determinant, the mean approximate log likelihood for A where $\mathbf{X}_l = A\mathbf{X}_l^{(t)}$ is

$$-2\langle \ln \mathcal{L}_Q(A|\{\hat{\mathbf{X}}_l\}) \rangle_t = \sum_l \left\{ \frac{(2l+1)n}{2} \frac{(1-A)^2}{A^2} + \frac{n(n+1)}{2A^2} \right\}. \quad (37)$$

For large l_{\max} the maximum is at $\hat{A} \sim 1 + (n+1)/l_{\max}$ rather than 1, so we expect A to be biased high by the order of the expected error, confirming that \mathcal{L}_Q is not a good approximation to the likelihood. If a fixed fiducial model is used then the determinant does not affect the likelihood, and we have

$$\begin{aligned} -2\langle \ln \mathcal{L}_f(A|\{\hat{\mathbf{X}}_l\}) \rangle_t = \sum_l \{ (1-A)^2 \mathbf{X}_l^{(t)T} \mathbf{M}_{fl}^{-1} \mathbf{X}_l^{(t)} \\ + \text{Tr}[\mathbf{M}_{fl}^{-1} \mathbf{M}_l^{(t)}] \} \end{aligned} \quad (38)$$

$$\propto (1-A)^2 + \text{const}, \quad (39)$$

which has a minimum in agreement with the exact likelihood function ($\hat{A} = 1$) regardless of the choice of fiducial model (though the variance of A would be wrong by the order of the fractional error in the fiducial model).

The case where $\mathbf{M}_{fl} = \mathbf{M}_l(\hat{\mathbf{X}}_l)$ is harder to analyze, but it is not a good approximation because the covariance is then correlated with the $\hat{\mathbf{C}}_l$ (so the contribution of high-fluctuating $\hat{\mathbf{C}}_l$ is down-weighted by larger covariance there).

C. Noise, binning and the Gaussian approximation

In the presence of isotropic uncorrelated noise n_{lm} with known power spectrum N_l , the observed field $a_{lm} + n_{lm}$ is just another Gaussian field with power spectrum $C_l + N_l$. The likelihood functions are then exactly the same as without noise, where C_l and $\hat{\mathbf{C}}_l$ are replaced with their values including noise.

Consider a toy problem where we wish to constrain the amplitude of the power spectrum A over some range of scales over which the power spectrum is flat. If there are n_m Gaussian modes, and we estimate the power spectrum in n_b equal bins, each bin will have $\nu \equiv n_m/n_b$ modes. If each mode has independent Gaussian noise with known variance N , each $\hat{\mathbf{C}}_b$ estimator then has a χ^2 distribution with ν degrees of freedom and mean $A + N$. The posterior mean of A will differ systematically from its maximum likelihood $\hat{\mathbf{C}}_b - N$ by $\sim (A + N)/\nu$, which we can take as an estimate of the bias obtained in each bin by using a Gaussian approximation. Using all the bins we can constrain A to within an error of $\sim (A + N)/\sqrt{n_m}$. The criterion for the bias to be much smaller than the error bar is then $n_b \ll \sqrt{n_m}$. Perhaps surprisingly this is independent of the noise: when this inequality is violated a Gaussian approxi-

mation would be biased for a given bin, even if the signal is noise dominated. Of course if the bin width is increased so that the signal to noise in each bin remains constant, then the Gaussian approximation for the binned estimates does improve as the noise increases.

In the case of observations of the CMB over a fraction f_{sky} of the sky, with useful signal at $l_{\text{min}} \leq l \leq l_{\text{max}}$, the number of modes is $n_m \sim f_{\text{sky}}(l_{\text{max}}^2 - l_{\text{min}}^2)$, so for the Gaussian approximation to be good for each bin we need the number of bins $n_b \ll f_{\text{sky}}^{1/2} l_{\text{max}}$ (assuming $l_{\text{max}}^2 \gg l_{\text{min}}^2$). This is violated by the natural full-sky binning into l_{max} bins, one at each l (which has optimal l -resolution), regardless of how large l_{max} is. For partial-sky observations with bin width $\Delta_l^{(b)}$ in l , you would need $\Delta_l^{(b)} \gg f_{\text{sky}}^{-1/2}$ for the Gaussian approximation to be reliable. However often we do not actually need each bin to be individually unbiased, so this criterion can in practice be relaxed.

Binning different l 's together makes the distribution more Gaussian, so binning full-sky C_l into bands of width $\Delta_l^{(b)} \gg 1$ would allow any of the quadratic likelihood approximations to be used with very small bias at each bin. For basic vanilla models it is straightforward to assess the impact of binning on parameter constraints: we generated a toy full-sky simulation at Planck sensitivity [14], generated samples of the posterior parameter values from the exact likelihood function using COSMOMC [15], and then importance sampled using the exact likelihood function on binned values of the C_l (keeping the $l < 30$ spectrum unbinned where in realistic cases the likelihood could also be calculated exactly). Using the quadratic approximation \mathcal{L}_Q in this case (with $\Delta_l^{(b)} = 1$) biases parameters like the spectral index by around 1-sigma compared to the exact result; however using \mathcal{L}_f with a sensible fiducial model produces unbiased constraints (see previous subsection). Binning with a width $\Delta_l^{(b)} = 50$ degrades parameter error bars by only $\leq 10\%$ for basic models; this would be sufficient to make the bias a tiny fraction ($\sim 1/\Delta_l^{(b)}$) of the error bar on each bin. Bins of $\Delta_l \sim 10$ would likely be wide enough to render the error from a quadratic likelihood approximation small relative to other systematic errors. The cost of doing this is that some l -resolution of the acoustic peak structure is lost, and any nonstandard models with features that vary over a few l could not be analyzed reliably (for example see Ref. [16]).

As we shall show, modeling the non-Gaussian distribution accurately is straightforward, and in any case a Gaussian approximation is often adequate, so for full-sky observations there is no need to degrade the data by binning. Note that binning may however be useful for other reasons, for example, to increase the accuracy with which the covariance can be estimated from a fixed number of simulations, or to improve the optimality of the cut-sky C_l estimator. Since almost all theoretical power spectra are very smooth in l , binning is likely to lose little information

as long as the bins are narrow compared to the width of any features.

D. Partial-sky likelihood function

When observations are obtained over part of the sky, or part of the sky is obscured by foregrounds or there is anisotropic noise, the maximum likelihood estimators \hat{C}_l can no longer be measured directly. The CMB is still expected to be Gaussian however, so in principle there is an exact pixel-based likelihood function of the form

$$\mathcal{L}(\{C_l\}|\mathbf{p}) \propto \frac{e^{-\mathbf{p}^T \mathbf{C}_p^{-1} \mathbf{p}/2}}{|\mathbf{C}_p|^{1/2}}, \quad (40)$$

where \mathbf{p} is a vector of pixel values and \mathbf{C}_p is the pixel-pixel covariance (a function of $\{C_l\}$). Equivalently the CMB fields can be expanded in a set of modes that are orthogonal and complete over the observed sky, and the likelihood in terms of these mode coefficients will also be Gaussian [8,17,18]. Neither likelihood function can be expressed solely in terms of a set of maximum likelihood power spectrum estimators, so an optimal analysis does not allow radical compression. The problem with using the exact likelihood function is that the number of pixels goes like l_{max}^2 , so the Cholesky decomposition required to calculate $\mathbf{C}_p^{-1} \mathbf{p}$ will scale like l_{max}^6 , which is prohibitive for l_{max} larger than a few hundred and slow for $l \geq 30$. Gibbs sampling methods avoid doing large matrix inversions, but still have exponential convergence problems if an exact analysis is attempted for general C_l . A sensible strategy is therefore to use an exact likelihood only at low l where it is numerically feasible, and to use an approximate analysis at higher l [9,19,20]. The most obvious way to do this is to compress the high- l data into a set of cut-sky power spectrum estimators, and then find an approximate likelihood function that is a function only of these estimators. There is some evidence that doing this is close to optimal, and it has the advantage of being fast. This means that numerous practical complications can be accounted for simply by adding additional terms to the covariance matrix estimated from simulations.

There are various possible estimators for the cut-sky power spectrum that can be used, varying from maximum likelihood to a variety of quadratic estimators. At high l quadratic estimators can be close to the maximum likelihood and we focus here on the widely used pseudo- C_l methods [19,21–27] that are in many cases equivalent to methods based on correlation functions [28,29]. In principle the statistical distribution of these estimators could be calculated exactly [22], but only at prohibitive numerical cost in general. We therefore look for a fast likelihood approximation that is a function only of the set of cut-sky estimators $\{\hat{C}_l\}$, an estimate of their covariance (e.g. from simulations or calculated), and knowledge of the noise contribution $\{N_l\}$. One of the aims of this work is to

quantify whether such a likelihood approximation is good enough to obtain reliable and nearly optimal parameter constraints. As our guide for modeling the non-Gaussian shape of the likelihood function we will use the known form in the full-sky limit; we aim for our approximation to be exact when the $\{\hat{C}_l\}$ are calculated on the full sky with isotropic noise.

E. New likelihood approximation for correlated fields

We now derive a new likelihood approximation that can be used with \hat{C}_l estimators calculated from correlated Gaussian fields. It is exact on the full sky, and should give reasonable results even for nonstandard models that are not necessarily very smooth functions of l . The approximation involves a fiducial model so that the covariance can easily be precomputed. However errors in the fiducial model are automatically corrected, in that the result remains exact on the full sky however wrong the fiducial model is. We assume that the matrix of estimators \hat{C}_l is positive definite, which may break down for some estimators at low l .

Given the observed estimators \hat{C}_l for the covariance of n Gaussian fields, the full-sky likelihood function can be written

$$\begin{aligned}
 -2 \log \mathcal{L}(C_l | \hat{C}_l) &= (2l + 1) \\
 &\times \{ \text{Tr}[\hat{C}_l C_l^{-1}] - \log |C_l^{-1} \hat{C}_l| - n \} \quad (41) \\
 &= (2l + 1) \{ \text{Tr}[C_l^{-1/2} \hat{C}_l C_l^{-1/2}] \\
 &\quad - \log |C_l^{-1/2} \hat{C}_l C_l^{-1/2}| - n \} \quad (42) \\
 &= (2l + 1) \sum_i [D_{l,ii} - \log(D_{l,ii}) - 1]. \quad (43)
 \end{aligned}$$

The symmetric form is defined using the Hermitian square root and $C_l^{-1/2} \hat{C}_l C_l^{-1/2} = U_l D_l U_l^T$ for orthogonal U_l and diagonal D_l . In the presence of instrumental noise the C_l and \hat{C}_l should include the noise variance.

To generalize to the cut sky we want to make this look quadratic, so we write

$$\begin{aligned}
 -2 \log \mathcal{L}(C_l | \hat{C}_l) &= \frac{2l + 1}{2} \sum_i [g(D_{l,ii})]^2 \\
 &= \frac{2l + 1}{2} \text{Tr}[g(D_l)^2] \quad (44)
 \end{aligned}$$

where

$$g(x) \equiv \text{sign}(x - 1) \sqrt{2(x - \ln(x) - 1)},$$

and $[g(D_l)]_{ij} = g(D_{l,ii}) \delta_{ij}$. Although the sign of the function is irrelevant for consistency with the exact full-sky result, this choice ensures consistency with the Gaussian approximation and that $g(x)$ is a smooth function at $x = 1$. We now want to relate this quadratic form to a version that

is quadratic in the matrix elements. To do this we use Eq. (10) in the form

$$\frac{2l + 1}{2} \text{Tr}[(C_{fl}^{-1/2} C_{gl} C_{fl}^{-1/2})^2] = \mathbf{X}_{gl}^T \mathbf{M}_{fl}^{-1} \mathbf{X}_{gl}, \quad (45)$$

where $\mathbf{X}_{gl} \equiv \text{vecp}(C_{gl})$ [dimension $n(n + 1)/2$] is the vector of distinct elements of C_{gl} , and \mathbf{M}_{fl} is the covariance of $\hat{\mathbf{X}}$ evaluated for $C_l = C_{fl}$. We therefore write the exact result of Eq. (44) as

$$\begin{aligned}
 -2 \log \mathcal{L}(C_l | \hat{C}_l) &= \frac{2l + 1}{2} \text{Tr}[(C_{fl}^{-1/2} C_{gl} C_{fl}^{-1/2})^2] \\
 &= \mathbf{X}_{gl}^T \mathbf{M}_{fl}^{-1} \mathbf{X}_{gl}, \quad (46)
 \end{aligned}$$

where $C_{gl} \equiv C_{fl}^{1/2} U_l g(D_l) U_l^T C_{fl}^{1/2}$ for some fiducial model C_{fl} . This can then be generalized to our final cut-sky approximation where the estimators at different l may be correlated:

$$\begin{aligned}
 -2 \log \mathcal{L}(\{C_l\} | \{\hat{C}_l\}) &\approx \mathbf{X}_g^T \mathbf{M}_f^{-1} \mathbf{X}_g \\
 &= \sum_{ll'} [\mathbf{X}_g]_l^T [\mathbf{M}_f^{-1}]_{ll'} [\mathbf{X}_g]_{l'}. \quad (47)
 \end{aligned}$$

Here \mathbf{M}_f is the fiducial model covariance block matrix with $n(n + 1)/2 \times n(n + 1)/2$ blocks labeled by l and l' , and \mathbf{X}_g is a $(l_{\max} - l_{\min} + 1)n(n + 1)/2$ -row block vector:

$$[\mathbf{M}_f]_{ll'} = \langle (\hat{\mathbf{X}}_l - \mathbf{X}_l)(\hat{\mathbf{X}}_{l'} - \mathbf{X}_{l'})^T \rangle_f \quad (48)$$

$$[\mathbf{X}_g]_l = \text{vecp}(C_{fl}^{1/2} g[C_l^{-1/2} \hat{C}_l C_l^{-1/2}] C_{fl}^{1/2}), \quad (49)$$

where the matrix function g applied to a symmetric positive-definite matrix is defined by application of g to its eigenvalues. On the full sky with isotropic noise $[\mathbf{M}_f]_{ll'} = \delta_{ll'} \mathbf{M}_{fl}$ and the approximation is exact. It is fast to evaluate because \mathbf{M}_f^{-1} is independent of C_l and hence can be precomputed. Remaining diagonalizations on the small matrices at each l are fast. In principle the fiducial model C_{fl} could also be chosen to be equal to \hat{C}_l or C_l , but for most purposes using a fixed smooth theoretical fiducial spectrum that is a good fit to the data is likely to be most convenient. For a general correlation structure \mathbf{M}_f has $[(l_{\max} - l_{\min} + 1)n(n + 1)/2]^2$ elements (but is symmetric). Remember that here C_l and \hat{C}_l include the noise contribution, so for a pure-theory (zero-noise) C_l^{th} the approximation requires an (effective) noise N_l at each l , a covariance matrix, and the set of estimators $\{\hat{C}_l\}$.

If C_l is block diagonal, as in the case of CMB polarization with B modes, the exact full-sky likelihood is separable in the blocks. On the cut sky the estimators for the blocks may however be correlated; in particular, a sky cut will correlate E - and B -mode polarization estimators. The approximation can be applied with full $[(l_{\max} - l_{\min} + 1)n(n + 1)/2]$ vectors, or the approximation can be applied to a truncated vector including only terms in each block.

For example we could use $\mathbf{X}_l = [C_l^{TT}, C_l^{TE}, C_l^{EE}, C_l^{BB}]^T$, with covariance allowing for correlations between E and B power spectra, but ignoring any potential information in components like \hat{C}_l^{TB} (the full-sky likelihood is independent of \hat{C}_l^{TB} , but this may not be the case when there are couplings between T , E and B). If the smaller vector is used the transformation to \mathbf{X}_g can be calculated for each block separately.

For a single Gaussian field the approximation is simply

$$-2 \log \mathcal{L}(\{C_l\}|\{\hat{C}_l\}) \approx \sum_{l'} [g(\hat{C}_l/C_l) C_{f_l}] [M_f^{-1}]_{l'l'} \times [C_{f_{l'}} g(\hat{C}_{l'}/C_{l'})]. \quad (50)$$

1. Generalization

On the full sky, and in some generalizations, the distribution of the estimators \hat{C}_l scales approximately with C_l , so that $P(\hat{C}_l|C_l)d\hat{C}_l = S_l(\hat{C}_l/C_l)(d\hat{C}_l)/C_l$ for some function $S(x)$. The full-sky likelihood function considered above is of this form. In general $S(x)$ can differ from the full-sky form, and could be estimated approximately from simulations using a given fiducial C_l . The likelihood function is then given by $\mathcal{L}(C_l|\hat{C}_l) \propto S_l(\hat{C}_l/C_l)/C_l$. We can then use the same likelihood approximations as above, where for each l

$$g(x) = \text{sign}(x - x_m) \sqrt{2\sigma^2 \log \left[\frac{x_m S_l(x_m)}{x S_l(x)} \right]}, \quad (51)$$

x_m is the value of x that maximizes $x S_l(x)$, and $\sigma^2 = \text{var}(x)$ [on the full sky $x_m = 1$, $\sigma^2 = 2/(2l+1)$]. With multiple fields a similar argument applies as long as the likelihood function can be written in terms of $C_l^{-1/2} \hat{C}_l C_l^{-1/2}$. The function $S_l(x)$ can then be estimated from the distribution of the diagonal elements of $C_l^{-1/2} \hat{C}_l C_l^{-1/2}$ at fixed C_l .

The exact distribution of single-field pseudo- C_l 's is discussed in Ref. [22] for azimuthally symmetric sky cuts. Even in this simple case with no noise the marginalized distribution at each l is of a different functional form from the full-sky result, similarly for the corresponding \hat{C}_l estimators. Using pseudo- C_l estimators with our approximation using $g(x) = \text{sign}(x-1)\sqrt{2(x-\ln(x)-1)}$ amounts to approximating the marginalized distribution of the \hat{C}_l as χ^2 with ν_l degrees of freedom, where $\nu_l = 2C_l^2/\text{var}(\hat{C}_l)$. At high l and for small cuts with uniform weighting outside the cut $\nu_l \sim (2l+1)f_{\text{sky}}^2$ [30]; for binned estimators that are nearly uncorrelated, $\nu_l \sim (2l+1)\Delta_l f_{\text{sky}}$ [23,31].

2. Gaussian approximation

The Gaussian approximations of Sec. III B generalize straightforwardly to a $(l_{\text{max}} - l_{\text{min}} + 1)n(n+1)/2$ -vector of cut-sky estimators $\hat{\mathbf{X}}$ with a covariance matrix \mathbf{M}_f ,

$$-2 \log \mathcal{L}_f(\{\mathbf{X}_l\}|\{\hat{\mathbf{X}}_l\}) = (\hat{\mathbf{X}} - \mathbf{X})^T \mathbf{M}_f^{-1} (\hat{\mathbf{X}} - \mathbf{X}) + \log |\mathbf{M}_f|. \quad (52)$$

Note that even with no correlations between l this cannot be written as a matrix variate normal distribution in the form of Eq. (30) because a general \mathbf{M} has many more degrees of freedom than the exact full-sky matrix where \mathbf{M}_l [a symmetric $n(n+1)/2 \times n(n+1)/2$ matrix] can be expressed in terms of the smaller matrix \mathbf{C}_l (an $n \times n$ symmetric matrix). From the discussion in Sec. III B we expect the Gaussian approximations to be accurate for $l_{\text{max}} \gg 1$ in almost all cases where parameter variations produce changes that are smooth in l .

IV. TESTING THE LIKELIHOOD APPROXIMATIONS

For accurate parameter estimation we need to be able to constrain the theory C_l accurately as a function of l given the estimators \hat{C}_l . On the full sky the likelihood approximations can easily be compared to the exact likelihood function. We fit an amplitude parameter A where $\mathcal{L}(A|\{\hat{C}_l\}) = \mathcal{L}(\{C_l = AC_l^{\text{in}}\}|\{\hat{C}_l\})$, over some range of l using some fiducial model C_l^{in} . The \hat{C}_l are simulated using C_l^{in} , so that on average the best-fit value of A is $A = 1$. Since in almost all models the theory power spectra C_l are smooth functions of l , and we wish to check that off-diagonal correlations are being accounted for correctly, we chose to fit over a range $\Delta l = 10$ in l . This was done for $l = (l_{\text{min}} \rightarrow l_{\text{max}} = l_{\text{min}} + \Delta l - 1)$, i.e., bins of size Δl with l_{min} and l_{max} being the lower and upper values of l in each bin, respectively, as a function of l_{min} .

Using a standard search routine (golden section search in Numerical Recipes), we searched for the best-fit value of A , \hat{A} . In other words, for the exact likelihood and each approximation, we numerically extracted the amplitude that would maximize the likelihood. We then estimate the variance of this estimated maximum likelihood value of A compared to the true maximum likelihood in that realization, $\langle (\hat{A}_i - \hat{A}_{\text{Exact}})^2 \rangle_{\text{simulations}}$. This gives a measure of any error introduced by the approximation. Note that since we are using a range of $\Delta l = 10$ in l , the best-fit value of A depends on the likelihood approximation at each l value, and, in particular, probes the full range of deviations of \hat{C}_l from C_l expected from cosmic variance.

To quantify whether an approximation is good enough, we consider how well we need to know the amplitude of the C_l as a function of l to get unbiased results on an amplitude parameter. We consider the noise-free case.

The cosmic variance error on a single l , is $\sqrt{\frac{2C_l^2}{(2l+1)}}$. Since we are averaging over a range $\Delta l = 10$, the cosmic variance error we can obtain on A from a single bin will be reduced by a factor of Δl , hence a fractional error of $\sim \sqrt{\frac{2}{(2l_{\text{min}}+1)\Delta l}}$ from one band. However as discussed in

Sec. for unbiased results from the full spectrum we need a fractional average systematic error on the C_l much smaller than $\Delta C_l/C_l \ll n^{-1/2}/l$. We therefore require likelihood approximations that give values that are unbiased to better than the systematic error.

A. Full-sky tests

On the full sky with isotropic noise estimators at different l are uncorrelated: the likelihood function is $\mathcal{L}(A|\{\hat{C}_l\}) = \prod_{l=l_{\min}}^{l_{\max}} \mathcal{L}(C_l = AC_l^{in}|\hat{C}_l)$, where \mathcal{L} can take the form of the exact likelihood or any of the approximations described in Sec. III A.

In Fig. 1 we show the results for the temperature likelihood approximations on the full sky. We calculate on average over simulations the difference between the posterior amplitudes, $|\langle \hat{A}_i - \hat{A}_{\text{Exact}} \rangle|$ (to probe bias) and the variance $\langle |\hat{A}_i - \hat{A}_{\text{Exact}}|^2 \rangle$ (to probe posterior differences in each realization). We require both quantities to be smaller than $1/l$, where \hat{A}_i is the best-fit value from one of the likelihood approximations given in Sec. III A. As expected, the symmetric Gaussian distribution \mathcal{L}_S shows a very poor fitting as its variance is larger than the systematic error. The

quadratic approximation \mathcal{L}_Q gives results almost identical to the systematic error and hence is not a good enough approximation. The Gaussian $_D$ results are probably good enough, but the WMAP approximation and approximation developed in Ref. [6] are much better. The fiducial Gaussian approximation is exactly unbiased in this simple test and is not shown. Any of these last four approximations should be adequate for temperature parameter estimation, at least assuming cut-sky accuracy with realistic noise follows the full-sky behavior. The new likelihood approximation by construction is also exactly correct in this full-sky case.

B. Cut-sky tests

We now move on to test the approximations on the cut sky. In particular we want to check that any bias on parameter constraints is much smaller than the posterior error, and that the likelihood function has the right shape. To do this we calculate simple pseudo- C_l estimators for azimuthal cuts with isotropic noise where the exact likelihood function can also be computed in reasonable time. Although idealized, realistic cuts are often approximately azimuthal due to the disk shape of the galaxy, and consis-

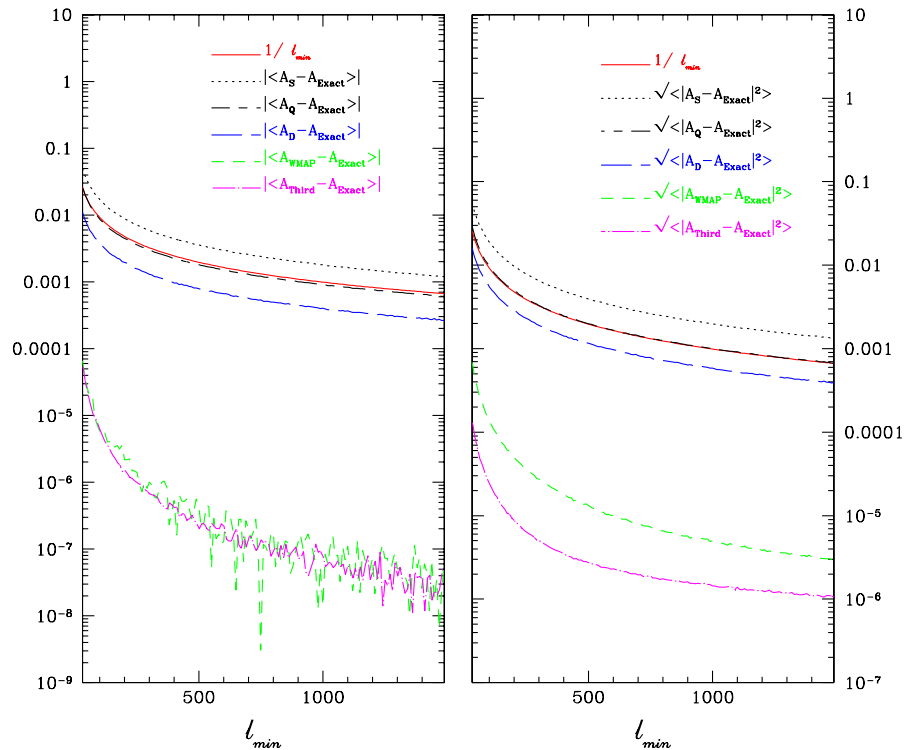


FIG. 1 (color online). The plot compares various likelihood approximations on the full sky for the case of a single field (temperature only) and no noise. The left-hand panel shows the difference between best-fit posterior amplitude of a $\Delta_l = 10$ bin with the likelihood approximations and the exact likelihood over 10 000 simulations where A_{Exact} is the best-fit amplitude of the exact likelihood and A_S , A_Q , A_D and A_{WMAP} are the best-fit amplitudes of the symmetric Gaussian, quadratic, Gaussian $_D$ and WMAP approximations, respectively. The right-hand panel shows the root-mean-square difference. These two quantities are compared to the systematic error tolerance. Only the symmetric Gaussian and the quadratic approximations are clearly not good enough. The fiducial Gaussian and new likelihood approximations are not shown as they are exactly unbiased in this simple test case with a correct fiducial model.

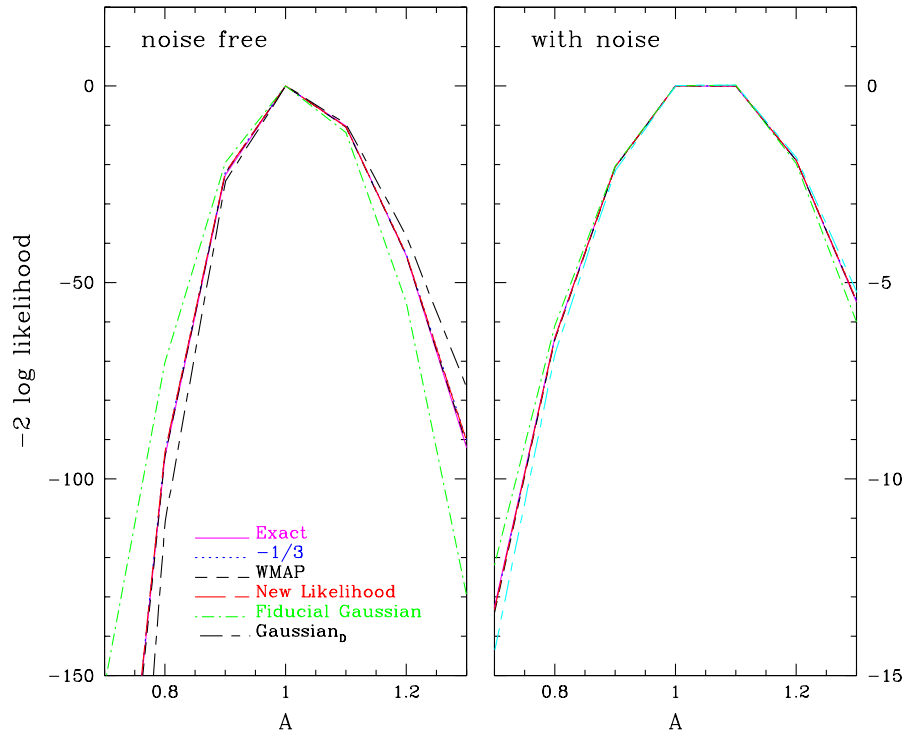


FIG. 2 (color online). Single-field likelihood approximation results for the likelihood as a function of bin amplitude, A . The plot compares the likelihood approximations to the exact likelihood for an azimuthal galactic cut with $f_{\text{sky}} = 0.862$, $l_{\text{max}} = 600$ and bin located at $200 \leq l \leq 209$ for one realization.

tency in this simple case is clearly necessary (if not strictly sufficient) to justify the use of a given likelihood approximation. An azimuthal cut introduces most of the qualitative differences in a cut-sky analysis, namely, correlations between different l and not-exactly Wishart distributions of the \hat{C}_l . The detailed derivations of the pseudo- C_l estimators, the covariance matrix and the exact likelihood for correlated fields are reviewed in Appendix D. We test the more general case of anisotropic noise and asymmetric cuts later in Sec. V.

1. Single-field results

The approximations used in the analysis of the temperature power spectra in the cut sky are given below, where C_l are taken to include noise and $[\mathbf{M}^{-1}]_{ll'}$ is the inverse of the covariance matrix $[\mathbf{M} = \mathbf{M}(\mathbf{X}), \hat{\mathbf{M}} = \mathbf{M}(\hat{\mathbf{X}})]$ when only a single field is considered:

$$-2 \ln \mathcal{L}_{\text{WMAP}} = \frac{1}{3} \sum_{ll'} (\hat{C}_l - C_l) [\mathbf{M}^{-1}]_{ll'} (\hat{C}_{l'} - C_{l'}) + \frac{2}{3} \sum_{ll'} \ln \left(\frac{\hat{C}_l}{C_l} \right) C_l [\mathbf{M}^{-1}]_{ll'} C_{l'} \ln \left(\frac{\hat{C}_{l'}}{C_{l'}} \right). \quad (53)$$

$$-2 \ln \mathcal{L}_{-1/3} = 9 \sum_{ll'} (\hat{C}_l^{-1/3} - C_l^{-1/3}) C_l \hat{C}_l^{1/3} [\hat{\mathbf{M}}^{-1}]_{ll'} \hat{C}_{l'}^{1/3} \times C_{l'} (\hat{C}_{l'}^{-1/3} - C_{l'}^{-1/3}), \quad (54)$$

$$-2 \ln \mathcal{L}_D = \sum_{ll'} (\hat{C}_l - C_l) [\mathbf{M}^{-1}]_{ll'} (\hat{C}_{l'} - C_{l'}) + \log |\mathbf{M}|, \quad (55)$$

$$-2 \ln \mathcal{L}_f = \sum_{ll'} (\hat{C}_l - C_l) [\mathbf{M}_f^{-1}]_{ll'} (\hat{C}_{l'} - C_{l'}), \quad (56)$$

where $[\mathbf{M}_f]_{ll'}$ is the covariance of some fiducial model, similar to the one used in new likelihood [see Eq. (50)].

Figure 2 shows the exact likelihood and the approximations presented in this subsection as a function of the posterior amplitude for a bin in one simulation. We consider both cases of noise-free and noisy power spectra. The approximations compare well to the exact result in both cases, though the results for the Gaussian approximations are not the right shape far away from the peak. Simulations were performed for azimuthal cuts with $f_{\text{sky}} = 0.862$.³ We have also fixed \hat{C}_l at $l \leq 30$ to $\hat{C}_l = C_l$ to prevent occasional negative values in the simulations.

2. Correlated-field results

To obtain unbiased results on an amplitude parameter from n noise-free correlated fields we need the systematic fractional bias on the amplitude to be $\ll 1/l\sqrt{n}$. With more than one field there is of course a lot more freedom than simply a change in amplitude. Nonetheless it is a useful

³That is a galactic cut of 20° .

first test as many important parameters, such as those governing the primordial power spectrum, affect the C_l essentially through an l -dependent scaling. If there is an apparent systematic error δC_l in the C_l spectrum, the criterion for an unbiased amplitude is $\text{Tr}[C_l^{-1} \delta C_l]/n \ll 1/l\sqrt{n}$. Since in practice polarization observations are likely to be noise dominated compared to the temperature for the near future, any approximation that satisfies this criterion will be more than adequate. We should however also test for accuracy of the likelihood to other changes in the spectrum, for example, the degree of cross correlation, as an amplitude scaling is a very special (if relevant) case.

We first test the approximate likelihood function compared to the exact result [see Eq. (34) for exact likelihood function used]; the result is shown in Fig. 3. The new likelihood approximation compares quite well with the exact likelihood, though it is slightly broader due to the loss of information from compressing the data into a set of pseudo- C_l power spectrum estimators \hat{C}_l . The fiducial Gaussian approximation shows significant deviations from the shape of the exact likelihood far from the peak.

For a quick analysis, the tests in the rest of this section were performed for spin-0 T - and E -mode only, i.e. the E -polarization was simulated as a scalar field similar to temperature so that E/B mixing may be ignored (but T - E

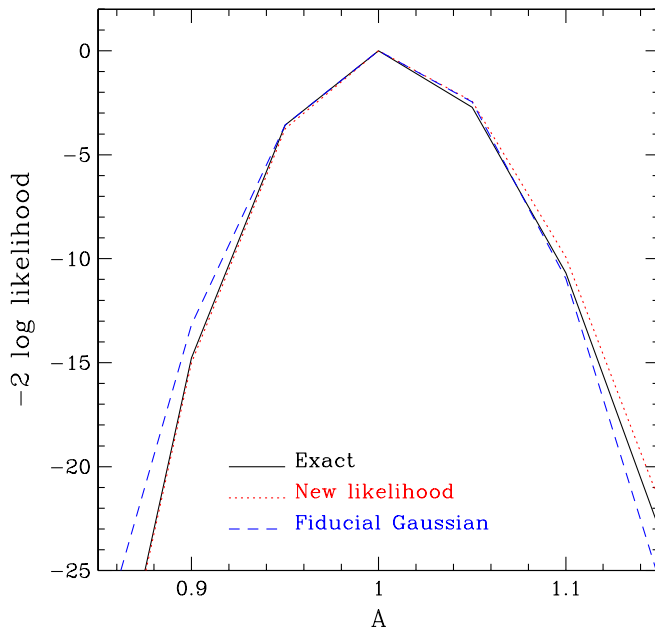


FIG. 3 (color online). The likelihood as a function of bin amplitude, A , for the temperature and polarization fields in one realization. The black (solid) line is the exact likelihood, the red (dotted) line is the new likelihood and the blue (dashed) line is the fiducial-Gaussian distribution. Unlike the fiducial-Gaussian distribution which only agrees well around the peak, the new likelihood captures the shape of the exact one well. We used an azimuthal cut with $f_{\text{sky}} = 0.862$, $l_{\text{max}} = 500$ and bin at $150 \leq l \leq 159$. Noise is isotropic and uncorrelated and the E and B modes noise is twice the T noise.

correlations are correctly accounted for). For all simulations, we also fix \hat{C}_l at $l \leq 30$ to $\hat{C}_l = C_l$ to avoid negative estimators and use a bin size of $\Delta_l = 10$.

The first consistency check is that on average over simulations $|\langle \hat{A} \rangle - 1| \ll 1/l\sqrt{n}$: this is sufficient to check that there is no significant bias in the posterior amplitude. We ran simulations for an azimuthal cut with $f_{\text{sky}} = 0.826$ with the results shown in Fig. 4. The new likelihood and the fiducial-Gaussian approximations appear to be unbiased.

We can also check the consistency of the likelihood function by comparing the binned and unbinned likelihood: as discussed in Sec. III C the likelihood function for bins with $\Delta_l^{(b)} \gg 1$ should be accurately Gaussian. For a smooth power spectrum binning can be performed with very little loss of information, and so the likelihood $P(\{C_b\}|\{\hat{C}_b\})$ can be calculated essentially exactly in the Gaussian approximation. We can check that this is consistent with the likelihood approximation evaluated using each l ; if it is, then we are using the information in the \hat{C}_l essentially optimally, at least when the spectrum is very smooth (even if compressing the sky into a set of \hat{C}_l estimators is not optimal). Similar to the full-sky single-field analysis, we calculate on average over simulations the difference between the posterior amplitudes, $|\langle \hat{A}_a - \hat{A}_b \rangle|$

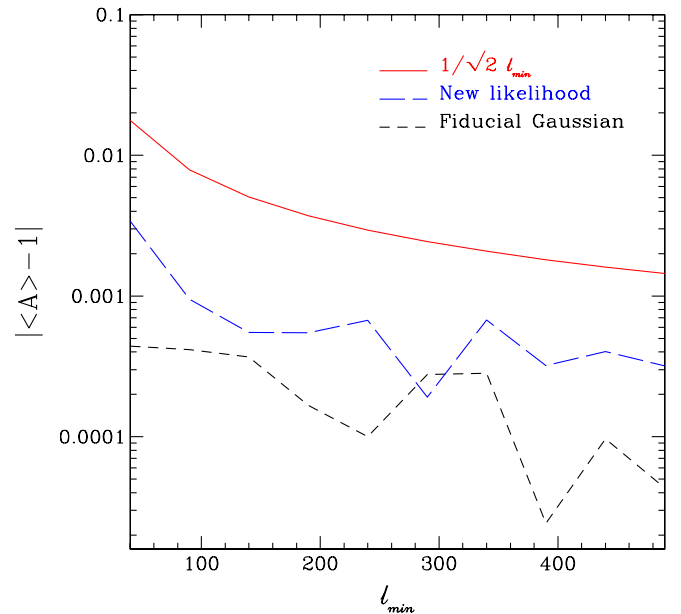


FIG. 4 (color online). The difference between the average of the posterior amplitude and the true input model compared to the systematic error [red (solid) line]. The blue (long-dashed) and the black (dashed) lines represent the differences for the new likelihood and fiducial Gaussian, respectively. The curves clearly do not show any significant bias in the posterior amplitudes. The averages were taken over 5000 simulations (realizations) for $l_{\text{max}} = 800$. Simulations were performed for spin-0 T - and E -mode only and for azimuthal cuts with $f_{\text{sky}} = 0.862$ and a bin size set to 10.

and the variance which is the square of the difference, $\langle |\hat{A}_a - \hat{A}_b|^2 \rangle$. We again require both quantities to satisfy the criterion set earlier, i.e. $|\langle \hat{A}_a - \hat{A}_b \rangle|, \langle |\hat{A}_a - \hat{A}_b|^2 \rangle^{1/2} \ll 1/l\sqrt{n}$. Figure 5 compares fiducial Gaussian, binned fiducial Gaussian, new likelihood, binned new likelihood and Gaussian_D. The plot clearly demonstrates that these approximations would produce the same results and are good enough to be used in analyzing CMB data. Figure 6 shows the comparison between the \mathcal{L}_S approximation (Gaussian with variance given by \hat{C}_l), binned \mathcal{L}_S Gaussian, new likelihood and binned new likelihood. This shows that \mathcal{L}_S is strongly biased when used with unbinned estimators, but when the data is binned it can produce consistent results as expected.

The Gaussian approximation with varying covariance, Gaussian_D, is significantly slower to compute than the other approximations. It is compared to the fiducial-model Gaussian in Fig. 7 for a small number of simulations. Since the fiducial-Gaussian result is unbiased this shows that Gaussian_D is also unbiased to good enough (though not excellent) accuracy in this case.

V. PARAMETER ESTIMATION TESTS WITH ANISOTROPIC NOISE

So far we have been using azimuthally symmetric cuts and assuming that the noise is isotropic. Isotropic noise is particularly simple case because the variance of the \hat{C}_l estimators scales as $\propto (C_l + N_l)^2$ to a good approximation. When the noise is anisotropic, as in realistic observations, this is no longer the case in general and it is important to test the likelihood approximations in this more realistic situation. For example, using a fiducial model covariance in our approximation of Eq. (47) was motivated in the case where everything is a function only of $(C_l + N_l)$. In general it may be necessary to instead evaluate the covariance for each theoretical model to correctly account for the more complicated scaling of the covariance with the signal. This could be done, for example, by rescaling a sum of covariance matrices calculated for noise-only, signal-only and signal plus noise realizations in some fiducial model. Although perfectly tractable, we shall see that in the case of Planck the simpler fiducial model approximation appears to be adequate.

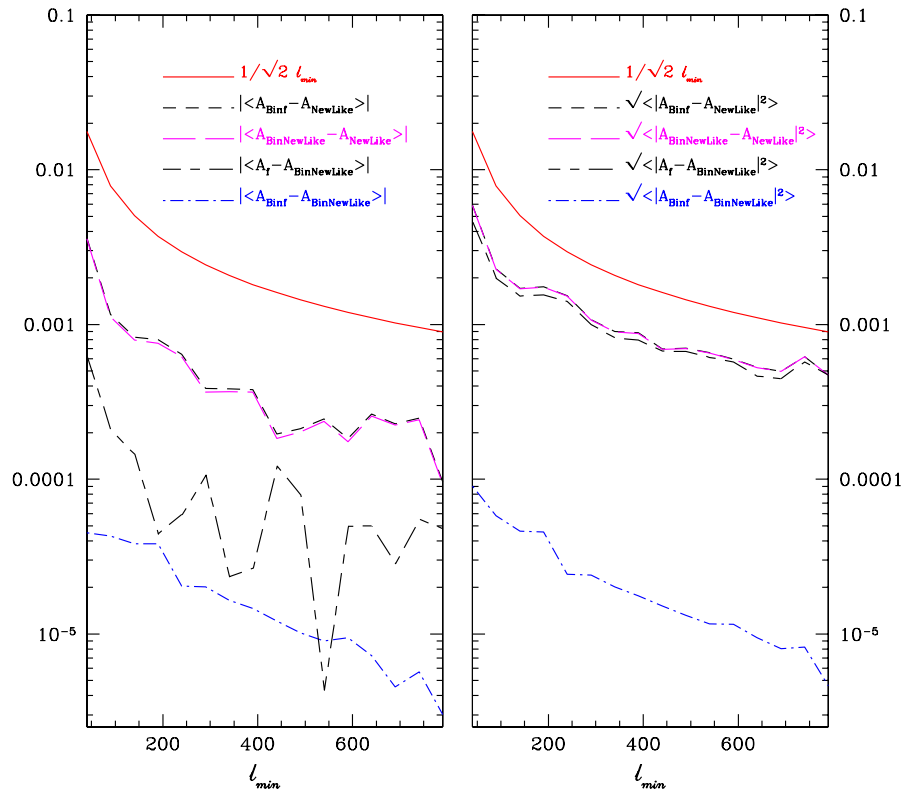


FIG. 5 (color online). Comparison between various binned and unbinned likelihood approximations. The left plot shows the average of the difference between the posterior amplitudes of these likelihoods and the right plot shows the variance, both compared to the systematic error [red (solid) line]. The black (dashed), the cyan (long-dashed), the black (dashed long-dashed), and the blue (dash-dotted) lines represent the comparison between binned fiducial Gaussian and new likelihood, binned new likelihood and new likelihood, fiducial Gaussian and binned new likelihood, and binned fiducial Gaussian and binned new likelihood, respectively. Averages were taken over 200 simulations (realizations) for $l_{\max} = 800$. Simulations were performed as previously mentioned. Results are all consistent to the required accuracy.

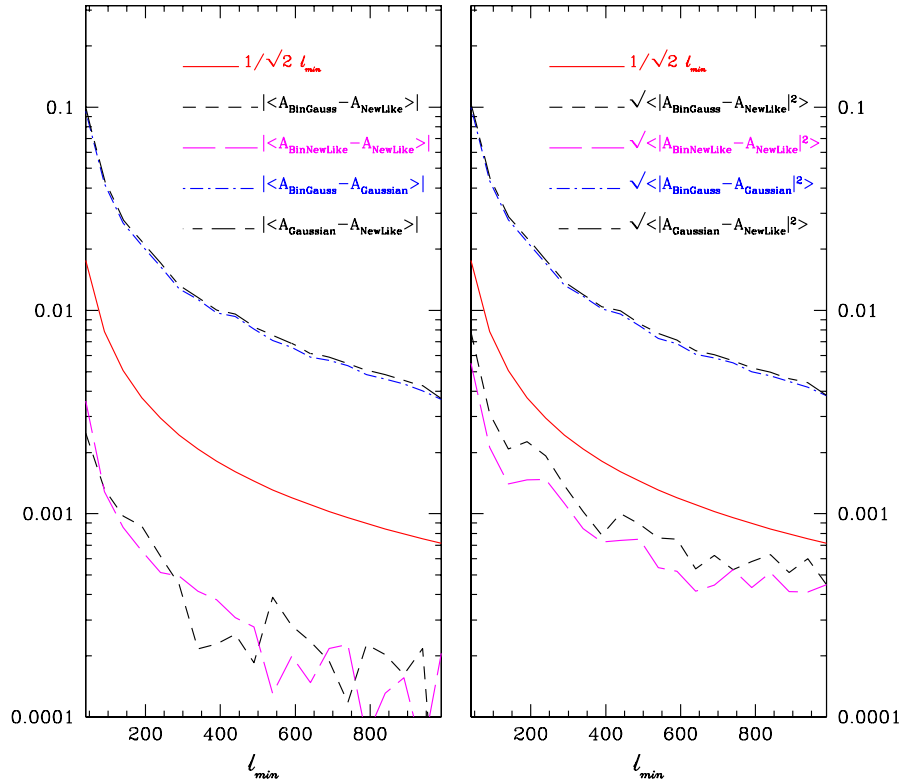


FIG. 6 (color online). Similar comparison as in Fig. 5 but using the symmetric Gaussian approximation \mathcal{L}_S . Unlike the binned and unbinned fiducial Gaussian, the binned and unbinned symmetric Gaussian approximations \mathcal{L}_S show significant bias. Averages are over 100 simulations (realizations) for $l_{max} = 1000$.

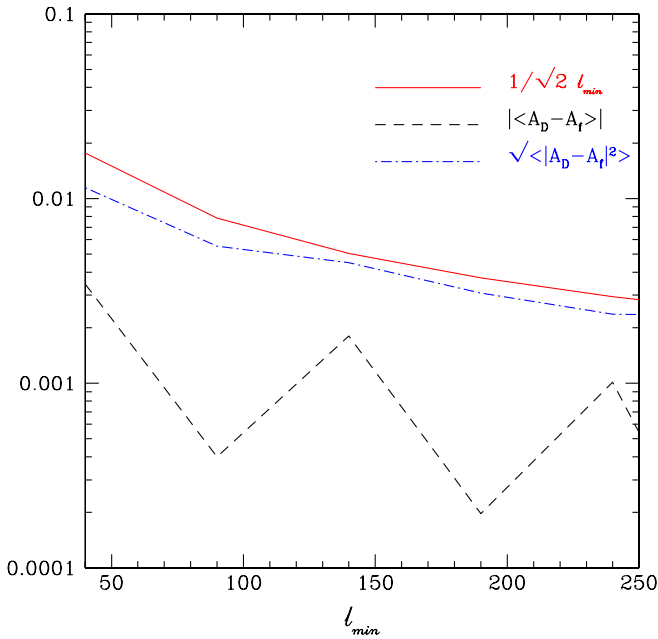


FIG. 7 (color online). A test over 20 simulations to compare the $Gaussian_D$ and $Gaussian_f$ distributions for $l_{max} = 300$ and bin width $\Delta_l = 10$.

We test the likelihood approximations by performing parameter estimation using single-sky maps simulated corresponding to an idealization of the combined Planck 143 GHz channels with 7 arcmin symmetric Gaussian beam [14]. The Planck satellite scanning strategy samples points near the ecliptic poles more densely than near the equator, and so there is a large (~ 100 factor) range of

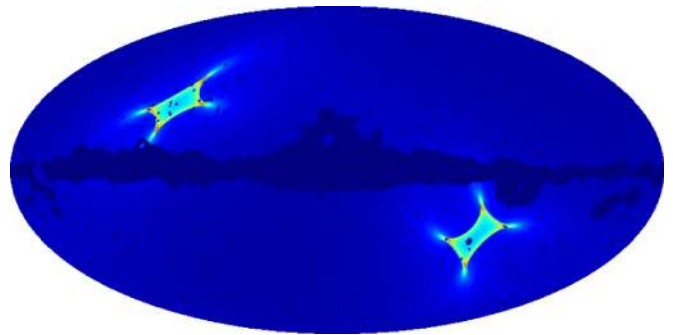


FIG. 8 (color online). Smoothed regularized inverse-noise weight map with WMAP kp2 cut as used by our test Planck-like simulation analysis. Noise is lowest in the cuspy regions around the ecliptic poles. The cut gives zero weight to regions around the galactic plane and numerous point sources. Noise and cut are smoothed with a 7 arcmin-FWHM Gaussian.

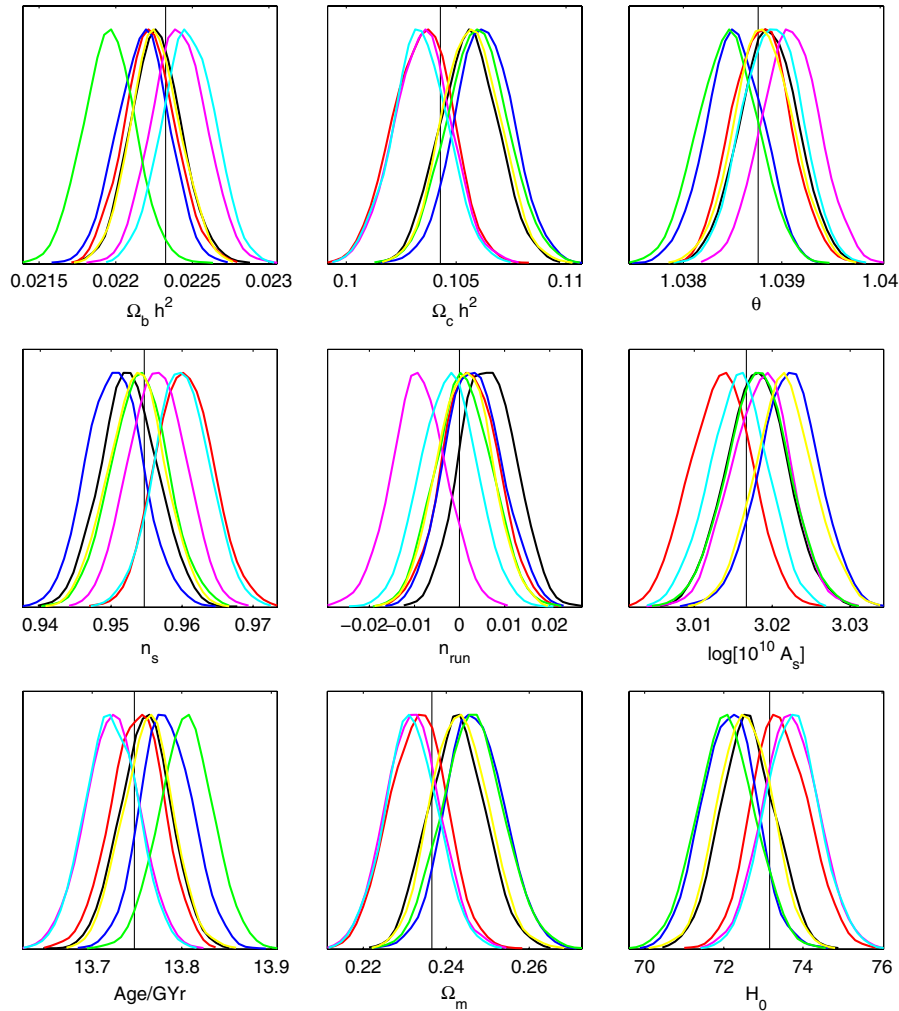


FIG. 9 (color online). Parameter constraints from six idealized Planck-like single map simulations with anisotropic noise as described in the text. The 1-dimensional marginalized posteriors are from using the new likelihood approximation with hybrid pseudo- C_l temperature, E -polarization and cross-correlation estimators at $l > 30$. The optical depth was fixed, and the simulation input parameters are shown with vertical lines. Very similar results are obtained if the noise-dominated B -polarization estimators are included with no tensor modes.

noise values across the sky [32]. In addition we use the “kp2” map [33,34] as a semirealistic sky cut to simulate masking out the galaxy and point sources. Details of our simulation, hybrid pseudo- C_l analysis and covariance model (following Ref. [19]) are given in Appendix E. In the high signal to noise regime the hybrid estimator uses an approximate inverse-noise weighted map with sky cut. As shown in Fig. 8 this is highly anisotropic. This inverse-noise weighted map is combined with a uniform-weighted map to give C_l estimators that are fairly close to optimal on all scales with $l \gtrsim 30$. For our simple test we assume a noise level average equivalent to the number for the 143 GHz channel quoted in the Planck science case [14]. We take the polarization and temperature pixel noise to be uncorrelated and proportional, with the polarization noise a factor of 4 larger than the temperature.

We use the range $30 \leq l \leq 2000$ for test parameter estimation from simulations; the low l likelihood is prob-

lematic because the pseudo- C_l estimators are not guaranteed to be positive definite, and the covariance structure becomes complicated due to E/B mixing effects on the cut sky. It may be possible to obtain reliable results from the \tilde{C}_l directly (without inverting to the unbiased estimators), using maximum likelihood or other more optimal estimators, however at low l the likelihood function can also be calculated essentially exactly in reasonable computational time, so here we focus on the higher l region where an exact analysis is intractable. Investigation of the low l likelihood function for Planck-like noise, how to combine with higher- l approximations, and dealing with real-world complications such as foregrounds is beyond the scope of this paper.⁴

⁴If only temperature is used then the new likelihood approximation works reliably with pseudo- C_l estimators down to $l = 2$ in almost all realizations.

From the simulated \hat{C}_l estimators we calculate the likelihood function of a given theoretical model using a likelihood approximation. This is used in the COSMOMC [13,35] parameter estimation code to sample from the posterior parameter distribution [15]. For our tests we consider a vanilla adiabatic flat Λ -CDM model, with baryon density $\Omega_b h^2$, dark matter density $\Omega_c h^2$, amplitude, spectral index and running of the primordial power spectrum (A_s , n_s and n_{run}), and the parameter θ , 100 times an approximation of the ratio of the sound horizon to the angular diameter distance at recombination. The age, Hubble parameter (H_0 km s $^{-1}$ Mpc $^{-1}$) and matter density relative to critical Ω_m are derived parameters. Since we are only considering the likelihood at $l \geq 30$ we fix the optical depth to reionization; our simulated parameter constraints are therefore tighter than expected from a full realistic analysis.

Figure 9 shows the consistent marginalized parameter constraints obtained when using the new likelihood approximation to analyze a set of sky simulations. Very similar constraints are obtained whether noise-dominated B power spectrum estimators are included or not, at least when there are no tensor modes. The new likelihood approximation seems to work well with realistically anisotropic noise.

Since in reality we will not know *a priori* exactly what fiducial model to choose, it is important that results be robust to choosing a slightly wrong model. Figure 10 compares the results from one simulation using the new likelihood approximation compared to using the fiducial-model Gaussian approximation; the fiducial models have $n_s = 1$ (wrong) and $n_s \sim 0.955$ (true), a difference of many sigma at Planck sensitivity. All the results are broadly consistent, but the fiducial-model Gaussian ap-

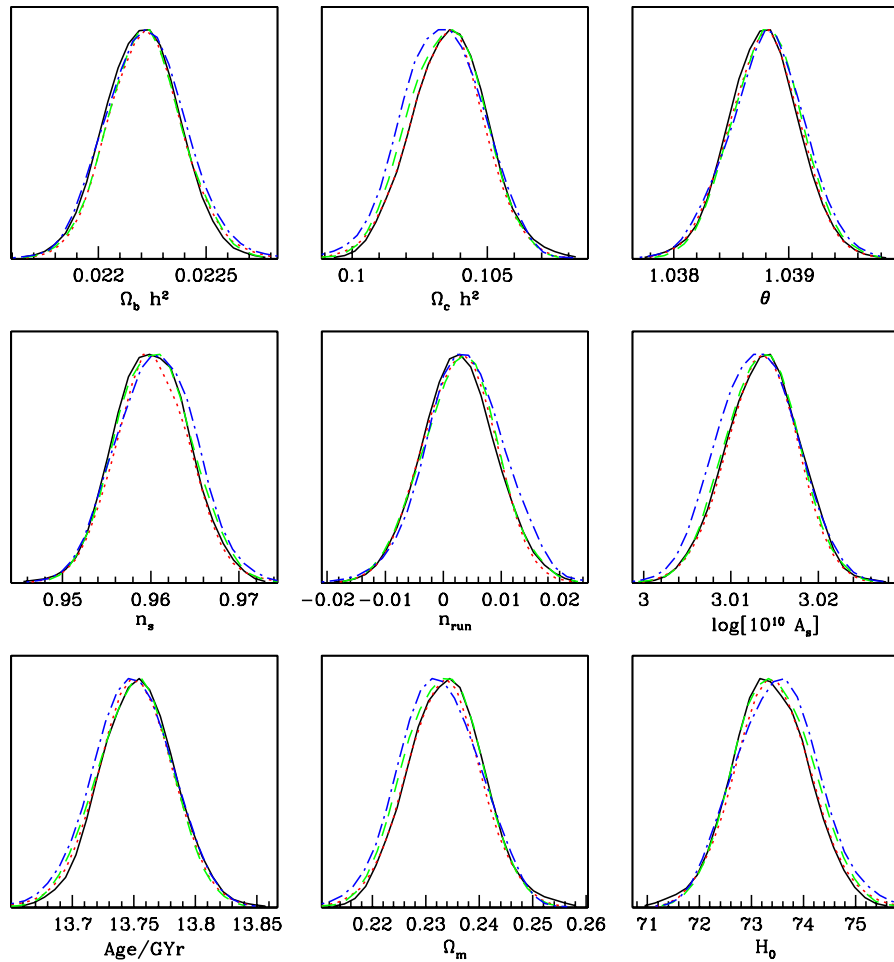


FIG. 10 (color online). Parameter constraints from single idealized Planck-like simulations with anisotropic noise. The 1-dimensional marginalized posteriors are from using the new likelihood and the fiducial-Gaussian approximations, and compare the results obtained when assuming an exactly correct fiducial model or using a wrong $n_s = 1$ model. The red (dotted) line is the new likelihood with the right model, the black (solid) line is new likelihood with the wrong model, which agree very well. The green (dashed) line is the fiducial Gaussian with the right model and the blue (dash-dotted) line is the fiducial Gaussian with the wrong model. The new likelihood results are consistent but the fiducial-Gaussian results are slightly affected by the choice of the model.

proximation shows some dependence on the choice of fiducial model. The new likelihood approximation results are more independent of the choice of fiducial model, and so appear to be more robust as expected.⁵ The values of the goodness-of-fit parameter χ_{eff}^2 (see Appendix B) are also much more stable for the new approximation compared to the fiducial Gaussian; the new likelihood approximation best fits differ by $\Delta\chi_{\text{eff}}^2 \sim 4$, but the fiducial-model Gaussian approximations differ by $\Delta\chi_{\text{eff}}^2 \sim 400$. With a fiducial model chosen to be sensibly closer to the maximum likelihood model both numbers should be significantly smaller.

Although detailed analysis of secondary signals is beyond the scope of this paper, in Appendix E 4 we show that with Planck noise levels our likelihood approximations also work when applied to lensed CMB fields and the covariance is estimated simply by using the lensed power spectra.

VI. CONCLUSIONS

In this paper we have attempted to find solutions to the problems facing the likelihood analysis of the CMB temperature and polarization estimators on small scales. With realistic data we need to be able to calculate the likelihood accurately from partial-sky observations. Previous attempts have established some excellent approximations to model the non-Gaussianity of the temperature likelihood function. However, no good general approximation has been derived to model the polarized likelihood. At large l computing the likelihood function exactly is computationally prohibitive and the correlation between the temperature and polarization fields makes it more complicated than for the temperature field only. We gave a new general approximation that can account for this correlation and is exact on the full sky. This new approximation is fast to evaluate as it involves a precomputed covariance independent of C_l , and appears to be more than adequate to obtain robust parameter constraints from clean small-scale CMB temperature and polarization data.

In summary, our conclusions regarding the modeling of the likelihood function of power spectrum estimators are:

- (i) In the case of binned power spectra, the number of modes per bin (n_m/n_b) must be much larger than the number of bins (n_b) for non-Gaussian corrections to the likelihood function to be unimportant in all cases; i.e. $n_b \ll \sqrt{n_m}$ is required to ensure that parameter bias is much smaller than the error bar.
- (ii) A Gaussian approximation with fixed fiducial-model covariance gives unbiased results for smooth power spectra at high l , but error bars have some dependence on the choice of the fiducial model.

⁵Some of the difference here is due to changes in the hybrid pseudo- C_l estimators when the fiducial model is changed.

Goodness-of-fit estimators χ_{eff}^2 can be misleading even for small differences between the fiducial and true model.

- (iii) A Gaussian approximation with covariance that varies with parameters can give reliable results at high l for smooth spectra, but only if the determinant term is consistently included; the quadratic approximation without determinant, \mathcal{L}_Q , is biased in general.
- (iv) The new likelihood approximation presented in Sec. III E appears to work well for power spectrum estimators with correlated fields and can give nearly optimal results when applied to good power spectrum estimators. It is fast to evaluate as it relies on a precomputed fiducial-covariance matrix, but is insensitive to small errors in the fiducial model. We recommend it for future work.
- (v) Most likelihood approximations with binned estimators ($n_b \ll \sqrt{n_m}$) can produce consistent results by the central limit theorem; for smooth power spectra consistency of parameter constraints with those from binned power spectra is a good check.

Since the new likelihood approximation is based on estimators and a covariance matrix, it is likely to generalize well to more realistic data where additional uncertainties, non-Gaussianities and correlations can be accounted for via changes to the estimator covariance. It is also likely to produce good results down to low l if positive-definite estimators are used, though this has not been the focus of this paper. Complications such as correlated noise may be well encapsulated in the covariance of a set of maximum likelihood (or similar) estimators, giving a fast alternative to much slower brute-force likelihood calculations. If the approximation is nearly correct, importance sampling techniques could be used to correct the results with a much smaller number of high-accuracy calculations.

We have not touched at all on the complications of foreground modeling, point sources, nonlinear and non-Gaussian anisotropies (e.g. due to Sunyaev-Zel'dovich), beam uncertainties, or a plethora of other real-world complications. Extending our work to account for these will be crucial for the correct interpretation of future data.

ACKNOWLEDGMENTS

S.H. gratefully acknowledges the support of the Algerian Ministry of Higher Education and Scientific Research (MESRS). A.L. acknowledges support from PPARC/STFC and thanks Anthony Challinor, Steven Gratton, George Efstathiou, Dipak Munshi, Mark Ashdown and Carlo Contaldi for discussion. Some of the results in this paper have been derived using the HEALPIX [39] package.

APPENDIX A: USEFUL RESULTS FOR MATRIX VECTORIZATION

In this appendix we review some results from matrix theory relating equations involving matrices to those involving vectors of their components, and establish Eq. (10) in the main text. For further details and references see e.g. Ref. [12].

The elements of a general matrix A can be assigned columnwise into a vector $\text{vec}(A)$. For matrices A and B

$$\text{Tr}[A^T B] = \text{vec}(A)^T \text{vec}(B). \quad (\text{A1})$$

The Kronecker product of an $m \times n$ matrix A with an $p \times q$ matrix B is defined to be the $mp \times nq$ matrix

$$A \otimes B = \begin{pmatrix} A_{11}B & A_{12}B & \dots & A_{1n}B \\ A_{21}B & A_{22}B & \dots & A_{2n}B \\ \vdots & \vdots & & \vdots \\ A_{m1}B & A_{m2}B & \dots & A_{mn}B \end{pmatrix}. \quad (\text{A2})$$

Using this we can write

$$\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B), \quad (\text{A3})$$

and using Eq. (A1) this implies

$$\text{Tr}[A^T DEF] = \text{vec}(A)^T (F^T \otimes D)\text{vec}(E). \quad (\text{A4})$$

For a symmetric $n \times n$ matrix there are only $n(n+1)/2$ distinct elements, and we define $\text{vecp}(A)$ to be the corresponding vector of distinct components of A

$$\text{vecp}(A) = (A_{11}, A_{21}, \dots, A_{n1}, A_{22}, A_{32}, \dots)^T. \quad (\text{A5})$$

The matrix $n^2 \times n(n+1)/2$ matrix B_n is defined so that for a general square matrix A

$$\text{vecp}(A) = B_n^T \text{vec}(A) = B_n^T \text{vec}(A + A^T)/2. \quad (\text{A6})$$

For example, a 2×2 matrix A has $B_n^T \text{vec}(A) = (A_{11}, (A_{12} + A_{21})/2, A_{22})^T$. The pseudoinverse $B_n^+ \equiv (B_n^T B_n)^{-1} B_n^T$ can be used to construct $\text{vec}(A)$ from $\text{vecp}(A)$ when A is symmetric:

$$\text{vec}(A) = (B_n^+)^T \text{vecp}(A). \quad (\text{A7})$$

Applying Eq. (A4) to symmetric matrices A, C, D and E we then have

$$\text{Tr}[ACDE] = \text{vecp}(A)^T B_n^+ (E \otimes C) (B_n^+)^T \text{vecp}(D), \quad (\text{A8})$$

and since $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ (for nonsingular matrices) also

$$\text{Tr}[AC^{-1}DE^{-1}] = \text{vecp}(A)^T [B_n^T (E \otimes C) B_n]^{-1} \text{vecp}(D). \quad (\text{A9})$$

The \hat{C}_l covariance matrix of Eq. (9) is defined by

$$\begin{aligned} M_l &\equiv \langle \text{vecp}(\hat{C}_l - C_l) \text{vecp}(\hat{C}_l - C_l)^T \rangle \\ &= B_n^T \langle \text{vec}(\hat{C}_l - C_l) \text{vec}(\hat{C}_l - C_l)^T \rangle B_n, \end{aligned} \quad (\text{A10})$$

where since $\hat{C}_l \equiv \sum_m \mathbf{a}_{lm} \mathbf{a}_{lm}^\dagger / (2l+1)$ we have

$$\text{vec}(\hat{C}_l) = \frac{1}{2l+1} \sum_m \mathbf{a}_{lm} \otimes \mathbf{a}_{lm}^*. \quad (\text{A11})$$

Using the general results (for appropriately sized matrices) that $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ and $(A \otimes B)^T = A^T \otimes B^T$ gives

$$\begin{aligned} \text{vec}(\hat{C}_l) \text{vec}(\hat{C}_l)^T &= \frac{1}{(2l+1)^2} \sum_{mm'} (\mathbf{a}_{lm} \otimes \mathbf{a}_{lm}^*) (\mathbf{a}_{lm'}^\dagger \otimes \mathbf{a}_{lm'}^T) \\ &= \frac{1}{(2l+1)^2} \sum_{mm'} (\mathbf{a}_{lm} \mathbf{a}_{lm'}^\dagger) \otimes (\mathbf{a}_{lm} \mathbf{a}_{lm'}^\dagger)^*. \end{aligned} \quad (\text{A12})$$

Hence since $\langle \mathbf{a}_{lm} \mathbf{a}_{lm'}^\dagger \rangle = \delta_{mm'} C_l$ we have

$$\langle \text{vec}(\hat{C}_l - C_l) \text{vec}(\hat{C}_l - C_l)^T \rangle = \frac{2}{2l+1} (C_l \otimes C_l), \quad (\text{A13})$$

so that $M_l = 2B_n^T (C_l \otimes C_l) B_n / (2l+1)$. Then from Eq. (A9) we have

$$\text{Tr}[AC_l^{-1}DC_l^{-1}] = \frac{2}{2l+1} \text{vecp}(A)^T M_l^{-1} \text{vecp}(D), \quad (\text{A14})$$

establishing Eq. (10). As a special case

$$\text{vecp}(C_l)^T M_l^{-1} \text{vecp}(C_l) = \frac{(2l+1)n}{2}. \quad (\text{A15})$$

If C has eigenvectors $\{\mathbf{e}_i^c\}$ with eigenvalues $\{\lambda_i^c\}$ then

$$(C \otimes D)(\mathbf{e}_i^c \otimes \mathbf{e}_j^d) = (C\mathbf{e}_i^c) \otimes (D\mathbf{e}_j^d) = \lambda_i^c \lambda_j^d (\mathbf{e}_i^c \otimes \mathbf{e}_j^d), \quad (\text{A16})$$

so the determinant is $|C \otimes D| = \prod_{ij} \lambda_i^c \lambda_j^d = |C|^n |D|^n$. Also for symmetric C , so that $C \otimes C = (C \otimes C) B_n B_n^+$, we have

$$\begin{aligned} &B_n^T (C \otimes C) B_n B_n^+ (\mathbf{e}_i \otimes \mathbf{e}_j) \\ &= [B_n^T (C \otimes C) B_n (B_n^T B_n)^{-1}] B_n^T (\mathbf{e}_i \otimes \mathbf{e}_j) \\ &= \lambda_i \lambda_j B_n^T (\mathbf{e}_i \otimes \mathbf{e}_j). \end{aligned} \quad (\text{A17})$$

So there are $n(n+1)/2$ distinct eigenvectors $B_n^T (\mathbf{e}_i \otimes \mathbf{e}_j)$ of $[B_n^T (C \otimes C) B_n (B_n^T B_n)^{-1}]$, and hence

$$|B_n^T (C \otimes C) B_n (B_n^T B_n)^{-1}| = \prod_i \prod_{j \geq i} \lambda_i \lambda_j = |C|^{n+1}. \quad (\text{A18})$$

The matrix $B_n^T B_n$ is diagonal with n unit entries and $n(n+1)/2 - n = n(n-1)/2$ that are a half, so $|B_n^T B_n| = 2^{-n(n-1)/2}$ and hence

$$|\mathbf{B}_n^T(\mathbf{C} \otimes \mathbf{C})\mathbf{B}_n| = 2^{-n(n-1)/2} |\mathbf{C}_l|^{n+1}. \quad (\text{A19})$$

The covariance matrix therefore has determinant

$$|\mathbf{M}_l| = \frac{2^n}{(2l+1)^{n(n+1)/2}} |\mathbf{C}_l|^{n+1}. \quad (\text{A20})$$

APPENDIX B: FULL-SKY GOODNESS OF FIT

Often people like to quote a chi-squared value as a crude measure of how well the data fit a given model. In the context of the full-sky CMB, where the \mathbf{a}_{lm} are Gaussian, we could define

$$\chi^2 \equiv \sum_l (2l+1) \text{Tr}[\hat{\mathbf{C}}_l \mathbf{C}_l^{-1}] \quad (\text{B1})$$

so that $P(\{\mathbf{a}_{lm}\}|\mathbf{C}) \propto e^{-\chi^2/2}$. This is minimized ($\chi^2 = 0$) when the \mathbf{a}_{lm} take their maximum likelihood values (zero). The mean is $\langle \chi^2 \rangle = \sum_l (2l+1)n$ and variance $\sum_l 2(2l+1)n$.

Alternatively, we could define an ‘‘effective’’ chi-squared, measuring the goodness of fit of the $\{\mathbf{C}_l\}$ to $\{\hat{\mathbf{C}}_l\}$ [5]:

$$\begin{aligned} \chi_{\text{eff}}^2 &\equiv -2 \ln(P(\{\mathbf{C}_l\}|\{\hat{\mathbf{C}}_l\})) \\ &= \sum_l (2l+1) \{ \text{Tr}[\hat{\mathbf{C}}_l \mathbf{C}_l^{-1}] - \log |\hat{\mathbf{C}}_l \mathbf{C}_l^{-1}| - n \} \end{aligned} \quad (\text{B2})$$

(to within a \mathbf{C}_l -independent constant). This is normalized so that if $\mathbf{C}_l = \hat{\mathbf{C}}_l$ then $\chi_{\text{eff}}^2 = 0$. To assess the goodness of fit we could compare χ_{eff}^2 to the value expected if \mathbf{C}_l were the true model. The expectation value under the Wishart distribution can be calculated by performing a Cholesky decomposition into a lower triangular matrix L , where $\mathbf{C}_l^{-1/2} \hat{\mathbf{C}}_l \mathbf{C}_l^{-1/2} = LL^T$, and using the independence of L_{ij} (the off-diagonal elements being Gaussian distributed, the diagonal elements chi-squared) [12]. The result is

$$\langle \chi_{\text{eff}}^2 \rangle = \sum_l (2l+1) \left\{ n \ln(l+1/2) - \sum_{i=1}^n \psi(l+1-i/2) \right\}, \quad (\text{B3})$$

where $\psi(x) \equiv d(\ln \Gamma(x))/dx$. For $l \gg n$ we have

$$\begin{aligned} (2l+1) \left\{ n \ln(l+1/2) - \sum_{i=1}^n \psi(l+1-i/2) \right\} \\ = \frac{n(n+1)}{2} + \frac{1}{12} \frac{n(2n^2+3n-1)}{2l+1} + \mathcal{O}(1/l^2), \end{aligned} \quad (\text{B4})$$

so for a large range of l with $n \ll l_{\min} \leq l \leq l_{\max}$ we have

$$\begin{aligned} \langle \chi_{\text{eff}}^2 \rangle &\approx (l_{\max} - l_{\min} + 1) \frac{n(n+1)}{2} + \frac{1}{24} n(2n^2+3n-1) \\ &\quad \times \ln(l_{\max}/l_{\min}). \end{aligned} \quad (\text{B5})$$

The first term is just what we would expect for a Gaussian distribution in $\hat{\mathbf{X}}_l$, the $n(n+1)/2$ distinct components $\hat{\mathbf{C}}_l$.

The second term is the logarithmic leading-order correction. For $l_{\min} = 30$, $l_{\max} = 2000$ it is ~ 0.7 (for $n = 1$), ~ 4.6 (for $n = 2$) and ~ 13.7 (for $n = 3$). The variance can be calculated similarly, giving

$$\text{var}(\chi_{\text{eff}}^2) = \sum_l (2l+1) \left\{ (2l+1) \sum_{i=1}^n \psi'(l+1-i/2) - 2n \right\} \quad (\text{B6})$$

$$= \sum_l \left\{ n(n+1) + \frac{1}{3} \frac{n(2n^2+3n-1)}{2l+1} + \mathcal{O}(1/l^2) \right\} \quad (\text{B7})$$

$$\approx 2 \langle \chi_{\text{eff}}^2 \rangle + \frac{1}{12} n(2n^2+3n-1) \ln(l_{\max}/l_{\min}), \quad (\text{B8})$$

where the prime denotes the derivative.

Note that even on the full-sky CMB lensing and other secondaries would give a nonzero connected four-point function that would change the variance of the $\hat{\mathbf{C}}_l$ from that calculated here for Gaussian fields.

APPENDIX C: MULTIPLE MAPS

In realistic experiments there are often many maps at different frequencies, from different detectors, and/or from different observation periods. Often the noise on these maps can be taken to be independent to an excellent approximation. Here we consider the very simple case where each map has isotropic noise. If there are two maps $a_{lm}^{(1)}$ and $a_{lm}^{(2)}$, each containing sky signal plus noise, the difference map $a_{lm}^{(1)} - a_{lm}^{(2)}$ will be independent of the signal. With n maps, there are therefore $n-1$ linear combinations that do not depend on the signal, and hence can be integrated out of the likelihood function. The remaining uncorrelated linear combination is the inverse-noise weighted combined map

$$a_{lm}^{(i)} \equiv \frac{\sum_{i=1}^n (N_l^{(i)})^{-1} a_{lm}^{(i)}}{\sum_{i=1}^n (N_l^{(i)})^{-1}}. \quad (\text{C1})$$

A similar argument applies in real space with anisotropic noise. The combined map $\{a_{lm}^{(i)}\}$ is a sufficient statistic for the likelihood function, and the likelihood analysis could therefore be based on C_l estimators from the combined map $a_{lm}^{(i)}$. Alternatively we could consider estimating a set of $\hat{\mathbf{C}}_l^{(ij)}$ from all possible combinations of maps

$$\hat{\mathbf{C}}_l^{(ij)} = \frac{1}{2l+1} \sum_m a_{lm}^{(i)*} a_{lm}^{(j)}. \quad (\text{C2})$$

In the simple case considered above, the optimal linear combination of the $\hat{\mathbf{C}}_l^{ij}$ is $\propto \sum_{ij} (N_l^{(i)})^{-1} (N_l^{(j)})^{-1} \hat{\mathbf{C}}_l^{(ij)}$, and

using this would be equivalent to using the estimator $\hat{C}_l^{(t)}$ from the combined map $a_{lm}^{(t)}$. The likelihood approximations in the main text could be applied directly to realistic pseudo- C_l generalizations of this estimator.

An alternative is to use only the off-diagonal correlations, where $i \neq j$ [30]. In the simplest case we can define the optimal weighted combination

$$\hat{C}_l^{\text{off}} \equiv \frac{\sum_{ij} (N_l^{(i)})^{-1} (N_l^{(j)})^{-1} \hat{C}_l^{(ij)} (1 - \delta_{ij})}{\sum_{ij} (N_l^{(i)})^{-1} (N_l^{(j)})^{-1} (1 - \delta_{ij})}. \quad (\text{C3})$$

Since $\langle \hat{C}_l^{\text{off}} \rangle = C_l$ the estimator is an unbiased estimator of the C_l regardless of the noise. In some instances it might therefore be more robust than including the diagonal correlations, where an error in the noise model can lead to an immediate bias in the estimator. However this estimator is no longer equivalent to the estimator on the weighted map $a_{lm}^{(t)}$, and has a different distribution. In particular it is not

positive definite. If \hat{C}_l^{off} are to be used for parameter estimation, in principle it may therefore be necessary to use a different likelihood approximation from those designed for analyzing Wishart-like distributions.

To see how different the distribution is we consider the very simplest case of foreground-free full-sky maps where all the maps have identical isotropic noise $N_l^{(i)} = N_l$, and we consider only a single scalar field (no polarization). We can define a n -dimensional vector of $a_{lm}^{(i)}$, \mathbf{a}_{lm} . The estimator is then

$$\hat{C}_l^{\text{off}} = \frac{1}{(2l+1)n(n-1)} \sum_m \mathbf{a}_{lm}^\dagger (\mathbf{e}\mathbf{e}^\dagger - \mathbf{I}) \mathbf{a}_{lm}, \quad (\text{C4})$$

where \mathbf{e} is a vector of ones, $e_i = 1$. The covariance of the \mathbf{a}_{lm} is given by

$$\mathbf{M}_l \equiv \langle \mathbf{a}_{lm} \mathbf{a}_{lm}^\dagger \rangle = C_l \mathbf{e}\mathbf{e}^\dagger + N_l \mathbf{I}. \quad (\text{C5})$$

The distribution of the \hat{C}_l^{off} is then given by

$$P(\hat{C}_l^{\text{off}} | C_l, N_l) = \int d\mathbf{a}_{lm} P(\mathbf{a}_{lm} | \mathbf{M}_l) \delta\left(\hat{C}_l^{\text{off}} - \alpha_{ln} \sum_m \mathbf{a}_{lm}^\dagger (\mathbf{e}\mathbf{e}^\dagger - \mathbf{I}) \mathbf{a}_{lm}\right) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \frac{e^{-ik\hat{C}_l^{\text{off}}}}{|\mathbf{I} - 2ik\alpha_{ln}\mathbf{M}_l(\mathbf{e}\mathbf{e}^\dagger - \mathbf{I})|^{l+1/2}}, \quad (\text{C6})$$

where the last line follows from writing the δ -function as a Fourier transform and $\alpha_{ln}^{-1} \equiv (2l+1)n(n-1)$. Substituting for \mathbf{M}_l and using $|\mathbf{I} + a\mathbf{e}\mathbf{e}^\dagger| = 1 + na$, the characteristic function (Fourier transform of the distribution function) is therefore given by

$$\tilde{P}(k | C_l, N_l) = \frac{1}{[(1 + 2ik\alpha_{ln}N_l)^{n-1} (1 - 2ik(2l+1)^{-1}(C_l + N_l/n))]^{l+1/2}}. \quad (\text{C7})$$

The quantity $C_l + N_l/n \equiv C_l + N_l^{(t)}$ is just the expectation value of $C_l^{(t)}$ from the optimal map. The distribution of \hat{C}_l^{off} is therefore the same as that of the variable $\hat{C}_l^{(t)} - \sum_{j=1}^{n-1} \hat{N}_l^{(t)(j)} / (n-1)$, where $\hat{N}_l^{(t)(j)}$ is the estimator from one of $n-1$ independent realizations of the noise. In the limit of many maps, $n \rightarrow \infty$ keeping the total noise $N_l^{(t)}$ fixed, we have

$$\lim_{n \rightarrow \infty} P(\hat{C}_l^{\text{off}} | C_l, N_l) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \frac{e^{-ik(N_l^{(t)} + \hat{C}_l^{\text{off}})}}{[1 - 2ik(2l+1)^{-1}(C_l + N_l^{(t)})]^{l+1/2}}. \quad (\text{C8})$$

This evaluates to the exact full-sky likelihood for $C_l^{(t)}$, so asymptotically with many maps $\hat{C}_l^{\text{off}} + N_l^{(t)}$ has the same distribution as $C_l^{(t)}$, and hence the likelihood can be approximated using the same approximations.

The distribution of \hat{C}_l^{off} can be calculated analytically for the special case $n = 2$ (as for the marginal distribution of C_l^{TE} [7]), but usually the off-diagonal estimator would be used only when there are several maps. In general the moments and cumulants of the distribution of \hat{C}_l^{off} can be calculated from the characteristic function, since

$$\langle (\hat{C}_l^{\text{off}})^p \rangle = \left[i^{-p} \frac{d^p \tilde{P}(k)}{dk^p} \right]_{k=0} \quad (\text{C9})$$

$$\kappa_p = \left[i^{-p} \frac{d^p \log \tilde{P}(k)}{dk^p} \right]_{k=0}.$$

In particular we have

$$\kappa_1 = \langle \hat{C}_l^{\text{off}} \rangle = C_l \quad (\text{C10})$$

$$\kappa_2 = \langle (\hat{C}_l^{\text{off}} - C_l)^2 \rangle = \frac{2}{2l+1} \left((C_l + N_l^{(t)})^2 + \frac{(N_l^{(t)})^2}{(n-1)} \right) \quad (\text{C11})$$

$$\kappa_3 = \langle (\hat{C}_l^{\text{off}} - C_l)^3 \rangle = \frac{8}{(2l+1)^2} \left((C_l + N_l^{(t)})^3 - \frac{(N_l^{(t)})^3}{(n-1)^2} \right) \quad (\text{C12})$$

$$\kappa_p = \frac{2^{p-1}(p-1)!}{(2l+1)^{p-1}} \left((C_l + N_l^{(i)})^p + (-1)^p \frac{(N_l^{(i)})^p}{(n-1)^{p-1}} \right). \quad (\text{C13})$$

The terms involving $(C_l + N_l^{(i)})$ are the equivalent results for $C_l^{(i)}$. The distribution of \hat{C}_l^{off} is therefore slightly less skewed than for the optimal estimator, but (as expected) with a slightly broader distribution. The third and higher moments will be close to those for $C_l^{(i)}$ if $n \gg 1 + N_l^{(i)}/(N_l^{(i)} + C_l)$. We therefore anticipate that if there are enough maps that this criterion is satisfied, $n \gg 2$, the likelihood approximations presented in this paper should also work well using the estimator $\hat{C}_l^{\text{off}} + N_l^{(i)}$.

Note that even though \hat{C}_l^{off} is unbiased regardless of the noise, the posterior mean of C_l will depend on the noise, and there could therefore be a posterior bias on parameters even if there is no bias directly on the estimators. This bias due to noise error is however suppressed by a factor of $\sim 1/l$ compared the direct bias that would arise from using $C_l^{(i)}$ with an incorrect noise model.

APPENDIX D: CUT-SKY ESTIMATORS, COVARIANCE AND EXACT LIKELIHOOD

1. Calculating the CMB cut-sky estimators

For limited sky coverage the temperature field is observed over only part of the sky. For full-sky observations part of the sky is likely to be dominated by galactic foregrounds, and CMB observations are effectively only available over the region of the sky outside a galactic (and point source) cut. In addition noise properties are generally not uniform across the sky; indeed a cut sky can be thought of a full-sky observation with infinite noise in the cut region. For these reasons it is useful to define a weighted temperature field \tilde{T} given by

$$\tilde{T}(\Omega) \equiv W^T(\Omega)T(\Omega), \quad (\text{D1})$$

where W^T is a weighting function defined over the whole sky that lies in the range 0 to 1. The simplest weighting function is zero in the cut region and one in the region with useful data; however more general window functions can be useful to obtain more optimal estimators. The pseudo-harmonics \tilde{a}_{lm}^T are then defined by the spherical harmonic transform of $\tilde{T}(\Omega)$. They are related to the underlying unweighted full-sky coefficients by

$$\tilde{a}_{lm}^T = \sum_{l'm'} W_{ll'}^{mm'} a_{l'm'}^T, \quad (\text{D2})$$

where the harmonic window function is defined as

$$W_{ll'}^{mm'} = \int d\Omega W^T(\Omega) Y_{l'm'}(\Omega) Y_{lm}^*(\Omega).$$

This can also be expressed as [23]

$$W_{ll'}^{mm'} = \sum_{l''m''} w_{l''m''}^T \left(\frac{(2l+1)(2l'+1)(2l''+1)^{1/2}}{4\pi} \right) (-1)^m \times \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ -m & m' & m'' \end{pmatrix}, \quad (\text{D3})$$

with the spherical harmonic transform coefficient of the window function given by

$$w_{lm}^T = \int W^T(\Omega) Y_{lm}^*(\Omega) d\Omega.$$

Similarly, for the polarization field the cut-sky pseudo-harmonic modes can be expanded as (see for example [18])

$$\tilde{a}_{lm}^E = \sum_{l'm'} ({}_+ W_{ll'}^{mm'} a_{l'm'}^E + i {}_- W_{ll'}^{mm'} a_{l'm'}^B), \quad (\text{D4})$$

$$\tilde{a}_{lm}^B = \sum_{l'm'} ({}_+ W_{ll'}^{mm'} a_{l'm'}^B - i {}_- W_{ll'}^{mm'} a_{l'm'}^E). \quad (\text{D5})$$

Here

$${}_+ W_{ll'}^{mm'} \equiv \frac{1}{2} ({}_2 W_{ll'}^{mm'} + {}_{-2} W_{ll'}^{mm'}), \quad (\text{D6})$$

$${}_- W_{ll'}^{mm'} \equiv \frac{1}{2} ({}_2 W_{ll'}^{mm'} - {}_{-2} W_{ll'}^{mm'}), \quad (\text{D7})$$

with the spin-weighted harmonic window function for spin $s = \pm 2$ given by

$${}_s W_{ll'}^{mm'} = \int d\Omega W_p(\Omega) {}_s Y_{l'm'}(\Omega) {}_s Y_{lm}^*(\Omega), \quad (\text{D8})$$

where ${}_s Y_{lm}(\Omega)$ are the spin-weighted harmonic functions. For azimuthal cuts the coupling matrices are diagonal in m , so $W_{ll'}^{mm'} = \delta_{mm'} W_{ll'}^m$, and they can be calculated quickly using a set of recursion relations [18].

The pseudo- C_l power spectra are defined by

$$\begin{aligned} \tilde{C}_l^{TT} &\equiv \frac{1}{2l+1} \sum_m \tilde{a}_{lm}^T (\tilde{a}_{lm}^T)^* \\ \tilde{C}_l^{TE} &\equiv \frac{1}{2l+1} \sum_m \tilde{a}_{lm}^T (\tilde{a}_{lm}^E)^* \\ \tilde{C}_l^{EE} &\equiv \frac{1}{2l+1} \sum_m \tilde{a}_{lm}^E (\tilde{a}_{lm}^E)^* \\ \tilde{C}_l^{BB} &\equiv \frac{1}{2l+1} \sum_m \tilde{a}_{lm}^B (\tilde{a}_{lm}^B)^*. \end{aligned} \quad (\text{D9})$$

Their expectation values are related to the full-sky power spectra via the relation

$$\begin{pmatrix} \langle \tilde{C}_l^{TT} \rangle \\ \langle \tilde{C}_l^{TE} \rangle \\ \langle \tilde{C}_l^{EE} \rangle \\ \langle \tilde{C}_l^{BB} \rangle \end{pmatrix} = \sum_{l'} \begin{pmatrix} \mathcal{M}_{ll'}^{TT} & 0 & 0 & 0 \\ 0 & \mathcal{M}_{ll'}^{TE} & 0 & 0 \\ 0 & 0 & \mathcal{M}_{ll'}^{EE} & \mathcal{M}_{ll'}^{EB} \\ 0 & 0 & \mathcal{M}_{ll'}^{BE} & \mathcal{M}_{ll'}^{BB} \end{pmatrix} \begin{pmatrix} C_{l'}^{TT} \\ C_{l'}^{TE} \\ C_{l'}^{EE} \\ C_{l'}^{BB} \end{pmatrix}, \quad (\text{D10})$$

where the coupling matrices are [36]

$$\mathcal{M}_{ll'}^{TT} = \frac{1}{2l+1} \sum_{mm'} |W_{ll'}^{mm'}|^2 = (2l'+1) \Xi_{TT}(l, l', \mathcal{W}^{TT}) \quad (\text{D11})$$

$$\begin{aligned} \mathcal{M}_{ll'}^{TE} &= \frac{1}{(2l+1)} \sum_{mm'} |W_{ll'}^{(mm')}(+W_{ll'}^{(mm')})| \\ &= (2l'+1) \Xi_{TE}(l, l', \mathcal{W}^{PT}) \end{aligned} \quad (\text{D12})$$

$$\begin{aligned} \mathcal{M}_{ll'}^{EE} &= \mathcal{M}_{ll'}^{BB} = \frac{1}{(2l+1)} \sum_{mm'} |(W_{ll'}^{(mm')})|^2 \\ &= (2l'+1) \Xi_{EE}(l, l', \mathcal{W}^{PP}) \end{aligned} \quad (\text{D13})$$

$$\begin{aligned} \mathcal{M}_{ll'}^{EB} &= \mathcal{M}_{ll'}^{BE} = \frac{1}{(2l+1)} \sum_{mm'} |(-W_{ll'}^{(mm')})|^2 \\ &= (2l'+1) \Xi_{EB}(l, l', \mathcal{W}^{PP}). \end{aligned} \quad (\text{D14})$$

The window function enters via its power spectrum \mathcal{W}_l^{XY} given by

$$\mathcal{W}_l^{XY} = \frac{1}{2l+1} \sum_m \omega_{lm}^X \omega_{lm}^{Y*}, \quad (\text{D15})$$

and X and Y being either T or P . For isotropic noise tests we only consider $\omega_{lm}^X = \omega_{lm}^Y$. The symmetric Ξ -matrices are defined by

$$\begin{aligned} \Xi_{TT}(l_1, l_2, \tilde{W}) &\equiv \sum_{l_3} \frac{(2l_3+1)}{4\pi} \tilde{W}_{l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2, \\ \Xi_{TE}(l_1, l_2, \tilde{W}) &\equiv \sum_{l_3} \frac{(2l_3+1)}{8\pi} \tilde{W}_{l_3} (1 + (-1)^L) \\ &\quad \times \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l_1 & l_2 & l_3 \\ -2 & 2 & 0 \end{pmatrix}, \\ \Xi_{EE}(l_1, l_2, \tilde{W}) &\equiv \sum_{l_3} \frac{(2l_3+1)}{16\pi} \tilde{W}_{l_3} (1 + (-1)^L)^2 \\ &\quad \times \begin{pmatrix} l_1 & l_2 & l_3 \\ -2 & 2 & 0 \end{pmatrix}^2, \\ \Xi_{EB}(l_1, l_2, \tilde{W}) &\equiv \sum_{l_3} \frac{(2l_3+1)}{16\pi} \tilde{W}_{l_3} (1 - (-1)^L)^2 \\ &\quad \times \begin{pmatrix} l_1 & l_2 & l_3 \\ -2 & 2 & 0 \end{pmatrix}^2, \end{aligned} \quad (\text{D16})$$

for $L = l_1 + l_2 + l_3$. All other coupling matrices are zero.

Provided that the sky cut is small (the usable region is larger than half the sky), the coupling matrix in Eq. (D10) is invertible and pseudo- C_l estimators for the power spectrum are given by (see for example [19,29])

$$\begin{pmatrix} \hat{C}^{TT} \\ \hat{C}^{TE} \\ \hat{C}^{EE} \\ \hat{C}^{BB} \end{pmatrix} = \begin{pmatrix} \mathcal{M}^{TT} & 0 & 0 & 0 \\ 0 & \mathcal{M}^{TE} & 0 & 0 \\ 0 & 0 & \mathcal{M}^{EE} & \mathcal{M}^{EB} \\ 0 & 0 & \mathcal{M}^{BE} & \mathcal{M}^{BB} \end{pmatrix}^{-1} \begin{pmatrix} \tilde{C}^{TT} \\ \tilde{C}^{TE} \\ \tilde{C}^{EE} \\ \tilde{C}^{BB} \end{pmatrix}. \quad (\text{D17})$$

The estimators are unbiased, $\langle \hat{C}_l \rangle = C_l$. When the observed area is small the matrix is not invertible. In this case the C_l can be binned into bands to construct band-power estimates of the power spectrum [23] in an analogous manner. Here we shall focus on nearly full-sky observations such as expected from the Planck satellite where estimates can be obtained for each C_l individually.

Unlike in the full-sky case, the exact cut-sky likelihood function cannot be written purely in terms of a set of pseudo- C_l estimators, so the compression of the observed data to the estimators is not lossless. However it can be a good approximation, and the estimators are convenient because the correlations between the \hat{C}_l induced by the sky cut are accounted for easily.

2. Covariance matrix

The covariance matrix of the \tilde{C}_l^{TT} is given by

$$\begin{aligned} \langle \Delta \tilde{C}_l^{TT} \Delta \tilde{C}_{l'}^{TT} \rangle &= \frac{2}{(2l+1)(2l'+1)} \sum_{mm'} \sum_{l_1 m_1} \sum_{l_2 m_2} C_{l_1}^{TT} C_{l_2}^{TT} W_{ll_1}^{mm_1} \\ &\quad \times (W_{l'l_1}^{m'l_1})^* W_{l'l_2}^{m'l_2} (W_{ll_2}^{mm_2})^*. \end{aligned} \quad (\text{D18})$$

As suggested by Ref. [19], this expression of the \tilde{C}_l^{TT} covariance matrix may be simplified for the case of a narrow galactic cut. In this case, $C_{l_1}^{TT}$ and $C_{l_2}^{TT}$ can be replaced with $C_{l_1}^{TT}$ and $C_{l_2}^{TT}$, respectively, and then by applying the completeness relation for spherical harmonics [37], the temperature \tilde{C}_l 's covariance matrix would be given by

$$\langle \Delta \tilde{C}_l^{TT} \Delta \tilde{C}_{l'}^{TT} \rangle = 2 C_{l_1}^{TT} C_{l_2}^{TT} \Xi_{TT}(l, l', \mathcal{W}^{TT}). \quad (\text{D19})$$

The covariance matrix of the \hat{C}_l estimators is then given by

$$\langle \Delta \hat{C}_l^{TT} \Delta \hat{C}_{l'}^{TT} \rangle = \sum_{l_1 l_2} \mathcal{M}_{ll_1}^{-1} \mathcal{M}_{l'l_2}^{-1} \langle \tilde{C}_{l_1}^{TT} \tilde{C}_{l_2}^{TT} \rangle. \quad (\text{D20})$$

Unfortunately, the other covariances do not simplify as easily since the completeness relation works only for the spherical harmonics with similar spin. For our azimuthal tests we use $W^T(\Omega)$ that takes values 1 or 0 and approximate the pseudocovariances by the following

$$\langle \Delta \tilde{C}_l^{TT} \Delta \tilde{C}_{l'}^{TT} \rangle \approx 2 \frac{C_{l_1}^{TT} C_{l_2}^{TT}}{(2l'+1)} \mathcal{M}_{ll'}^{TT}, \quad (\text{D21})$$

$$\langle \Delta \tilde{C}_l^{TE} \Delta \tilde{C}_{l'}^{TE} \rangle \approx \frac{\sqrt{C_l^T C_{l'}^T C_l^E C_{l'}^E}}{(2l'+1)} \mathcal{M}_{ll'}^{TE} + \frac{C_l^{TE} C_{l'}^{TE}}{(2l'+1)} \mathcal{M}_{ll'}^{TT}, \quad (\text{D22})$$

$$\langle \Delta \tilde{C}_l^{EE} \Delta \tilde{C}_{l'}^{EE} \rangle \approx 2 \frac{C_l^{EE} C_{l'}^{EE}}{(2l'+1)} \mathcal{M}_{ll'}^{EE} + 2 \frac{C_l^{BB} C_{l'}^{BB}}{(2l'+1)} \mathcal{M}_{ll'}^{EB}, \quad (\text{D23})$$

$$\langle \Delta \tilde{C}_l^{BB} \Delta \tilde{C}_{l'}^{BB} \rangle \approx 2 \frac{C_l^{BB} C_{l'}^{BB}}{(2l'+1)} \mathcal{M}_{ll'}^{BB} + 2 \frac{C_l^{EE} C_{l'}^{EE}}{(2l'+1)} \mathcal{M}_{ll'}^{EB}, \quad (\text{D24})$$

$$\langle \Delta \tilde{C}_l^{TT} \Delta \tilde{C}_{l'}^{EE} \rangle \approx 2 \frac{C_l^{TE} C_{l'}^{TE}}{(2l'+1)} \mathcal{M}_{ll'}^{TT}, \quad (\text{D25})$$

$$\langle \Delta \tilde{C}_l^{TT} \Delta \tilde{C}_{l'}^{TE} \rangle \approx \frac{\sqrt{C_l^{TT} C_{l'}^{TT}} (C_l^{TE} + C_{l'}^{TE}) \mathcal{M}_{ll'}^{TT}}{(2l'+1)}, \quad (\text{D26})$$

$$\langle \Delta \tilde{C}_l^{EE} \Delta \tilde{C}_{l'}^{TE} \rangle \approx \frac{\sqrt{C_l^{EE} C_{l'}^{EE}} (C_l^{TE} + C_{l'}^{TE}) \mathcal{M}_{ll'}^{TE}}{(2l'+1)}, \quad (\text{D27})$$

$$\langle \Delta \tilde{C}_l^{EE} \Delta \tilde{C}_{l'}^{BB} \rangle \approx \frac{(\sqrt{C_l^{EE} C_{l'}^{EE}} + \sqrt{C_l^{BB} C_{l'}^{BB}})^2}{2(2l'+1)} \mathcal{M}_{ll'}^{EB}. \quad (\text{D28})$$

Note that in the presence of isotropic noise the C_l here include the noise contribution.

At high l one can approximate $\mathcal{M}_{ll'}^{TE} = \mathcal{M}_{ll'}^{EE} = \mathcal{M}_{ll'}^{BB} = \mathcal{M}_{ll'}^{TT}$, since the spin ± 2 harmonics become close to the spin zero ones. Note our approximations in Eqs. (D26) and (D27) differ from those in Ref. [25]: since the C_l^{TE} can be negative we require consistency with the exact result on the full sky rather than forcing these terms to be positive. Also, note the difference in Eqs. (D23) and (D24) from those in Ref. [25]. More general results applicable with anisotropic noise and general weight function are given in Appendix E. More accurate results accounting for the complications of E/B mixing are given in Ref. [31]; see also Ref. [27]. Note that inaccuracies in the covariance matrix generally only affect the error bars; to this extent accuracy is less crucial than getting the estimators or likelihood function accurate, since there an inaccuracy could introduce biases.

The covariance of the \hat{C}_l estimators can be calculated from the \tilde{C}_l covariance using the relevant coupling matrices.

3. Exact likelihood for temperature and polarization

Although an exact likelihood calculation is prohibitively slow in general, for azimuthal sky cuts the relevant matrices are block-diagonal in m and the calculation is numerically tractable. For the special case of azimuthal cuts we

can therefore test cut-sky likelihood approximations against the exact result.

For each m we can define a vector of pseudoharmonic coefficients

$$\begin{aligned} \tilde{\mathbf{X}} &\equiv \begin{pmatrix} \tilde{a}_{lm}^T \\ \tilde{a}_{lm}^E + i\tilde{a}_{lm}^B \\ \tilde{a}_{lm}^E - i\tilde{a}_{lm}^B \end{pmatrix} \\ &= \begin{pmatrix} W_{ll'}^{(m)} & 0 & 0 \\ 0 & 2W_{ll'}^{(m)} & 0 \\ 0 & 0 & -2W_{ll'}^{(m)} \end{pmatrix} \begin{pmatrix} a_{l'm}^T \\ a_{l'm}^E + ia_{l'm}^B \\ a_{l'm}^E - ia_{l'm}^B \end{pmatrix}, \end{aligned} \quad (\text{D29})$$

which can simply be written as

$$\tilde{\mathbf{X}} = \text{diag}(W_{ll'}^{(m)}, 2W_{ll'}^{(m)}, -2W_{ll'}^{(m)}) \mathbf{X}. \quad (\text{D30})$$

For Gaussian fields $\tilde{\mathbf{X}}$ is just a linear combination of Gaussian harmonics, and hence also Gaussian. However due to the sky cut the coupling matrix is not directly invertible, as the W -matrices will have eigenvalues very close to zero (corresponding to modes localized in the unobserved region). However we can use a singular value decomposition (SVD) to isolate the observable independent modes following Ref. [17,18]. We diagonalize the transformation matrix as $\text{diag}(W_{ll'}^{(m)}, 2W_{ll'}^{(m)}, -2W_{ll'}^{(m)}) = \mathbf{U} \mathbf{D} \mathbf{U}^\dagger$ and define new linear combinations:

$$\mathbf{X}' = \hat{\mathbf{D}}^{-1/2} \hat{\mathbf{U}}^\dagger \tilde{\mathbf{X}} = \hat{\mathbf{D}}^{1/2} \hat{\mathbf{U}}^\dagger \mathbf{X}. \quad (\text{D31})$$

Here $\hat{\mathbf{D}}$ denotes the smaller square matrix obtained from \mathbf{D} by deleting nearly zero rows and columns. $\hat{\mathbf{U}}$ is the corresponding rectangular matrix obtained from \mathbf{U} by deleting the corresponding columns.

The signal correlation is

$$\begin{aligned} \mathbf{S} &= \langle \mathbf{X}' \mathbf{X}'^\dagger \rangle = \hat{\mathbf{D}}^{1/2} \hat{\mathbf{U}}^\dagger \langle \mathbf{X} \mathbf{X}^\dagger \rangle \hat{\mathbf{U}} \hat{\mathbf{D}}^{1/2} \\ &= \hat{\mathbf{D}}^{1/2} \hat{\mathbf{U}}^\dagger \begin{pmatrix} C_l^{TT} & C_l^{TE} & C_l^{TE} \\ C_l^{TE} & C_l^{EE} + C_l^{BB} & C_l^{EE} - C_l^{BB} \\ C_l^{TE} & C_l^{EE} - C_l^{BB} & C_l^{EE} + C_l^{BB} \end{pmatrix} \hat{\mathbf{U}} \hat{\mathbf{D}}^{1/2}. \end{aligned} \quad (\text{D32})$$

If the noise is isotropic and uncorrelated, this frame structure provides a diagonal noise correlation [18]:

$$\begin{aligned} \langle \tilde{\mathbf{X}}_N \tilde{\mathbf{X}}_N^\dagger \rangle &= \sigma_N^2 \text{diag}(W_{ll'}^{(m)}, 2_+ W_{ll'}^{(m)}, 2_- W_{ll'}^{(m)}) \Rightarrow \mathbf{N} \\ &= \langle \mathbf{X}'_N \mathbf{X}'_N^\dagger \rangle = \sigma_N^2 \text{diag}(1, 2, 2), \end{aligned} \quad (\text{D33})$$

where we have considered $\sigma_N^{T2} = \sigma_N^2$ and $\sigma_N^{E2} = \sigma_N^{B2} = 2\sigma_N^2$ for simulation purposes.

Given that the signal and noise are Gaussian, the likelihood function is then given by

$$\mathcal{L}(\{C_l^T, C_l^E, C_l^{TE}, C_l^B\}|\mathbf{X}') \propto \frac{\exp[-\frac{1}{2}\mathbf{X}'^\dagger(\mathbf{S} + \mathbf{N})^{-1}\mathbf{X}']}{|\mathbf{S} + \mathbf{N}|^{1/2}}. \quad (\text{D34})$$

The only approximation is in the choice of cutoff value for the SVD; for nonzero noise the result is insensitive to this choice as long as it is small.

APPENDIX E: ANISOTROPIC NOISE: ESTIMATORS AND TEST SIMULATION

1. Hybrid pseudo- C_l estimators with cross weights

We consider pixelized maps with anisotropic but uncorrelated pixel noise variance σ_s^2 (in this section the C_l do not include noise). We generalize the hybrid pseudo- C_l method of Ref. [26] slightly to include pseudo- C_l estimators from mixed weights, e.g. using a set of pseudo- C_l 's

$$\tilde{C}_l^{XY,ij} \equiv \frac{1}{2l+1} \sum_m \tilde{a}_{lm}^{X,i} \tilde{a}_{lm}^{Y,j*}, \quad (\text{E1})$$

where $\tilde{a}_{lm}^{X,i}$ is defined using weight function w^i . For each X and Y there are therefore $n(n+1)/2$ distinct estimators if $X = Y$, or n^2 if $X \neq Y$, where n is the number of weight functions. For high signal to noise the best weight function should be close to uniform to minimize cosmic variance, for low signal to noise it should be proportional to the inverse noise to minimize the noise [19]. Combining results from two weight functions, one with uniform and one with inverse-noise weighting, is therefore perhaps the most natural choice, especially if the polarization noise is proportional to the temperature noise in each pixel as we assume for our test simulations. Including the cross estimator between maps with different weight functions is particularly useful for estimating C_l^{TE} : since the polarization noise is much larger than the temperature, over a wide range of scales the cross estimator between uniform and inverse-noise weighted maps is much better than using uniform/uniform or inverse noise/inverse noise. Even for the temperature case there is a range of scales in between noise and signal domination where the cross estimator can be useful. Including more than two weighting functions seems to gain very little, so we use just two.

The unbiased \hat{C}_l estimators are constructed using the coupling matrix

$$\hat{C}_l^{XY,ij} = [\mathcal{M}^{XY,ij}]_{ll'}^{-1} \tilde{C}_l^{XY,ij}, \quad (\text{E2})$$

where

$$\begin{aligned} \mathcal{M}_{ll'}^{XY,ij} &= (2l'+1) \Xi_{XY}(l, l', \tilde{W}^{ij}), \\ \tilde{W}^{ij} &\equiv \frac{1}{2l+1} \sum_m w_{lm}^i w_{lm}^j, \end{aligned} \quad (\text{E3})$$

and the coupling matrices are defined in Eq. (D16).

The noise contribution to the pseudo- C_l is given, for uncorrelated pixel noise $(\sigma_s^T)^2$, $(\sigma_s^Q)^2$, $(\sigma_s^U)^2$ and pixel area

Ω_s , by

$$\tilde{N}_l^{TT,ij} = \frac{1}{4\pi} \sum_s (\sigma_s^T)^2 w^i(s) w^j(s) \Omega_s^2, \quad (\text{E4})$$

$$\tilde{N}_l^{EE,ij} = \tilde{N}_l^{BB,ij} = \frac{1}{8\pi} \sum_s [(\sigma_s^Q)^2 + (\sigma_s^U)^2] w^i(s) w^j(s) \Omega_s^2, \quad (\text{E5})$$

with other combinations being zero. We then have $\langle \hat{C}_l^{XY,ij} \rangle = C_l^{XY} + [\mathcal{M}^{XY,ij}]^{-1} \tilde{N}_l^{XY,ij}$.

From multiple pseudo- C_l estimators with different weight functions one can either attempt to apply the likelihood approximations directly to the complete set of estimators, or one can compress into a single hybrid estimator. At low l it is likely to be beneficial to also include more optimal estimators than pseudo- C_l , especially for the polarization [26].

A hybrid pseudo- C_l estimator can be constructed following Ref. [26]: this is defined by constructing the best-fit C_l to the multiple estimators by minimizing the Gaussian approximation to the likelihood using the approximate full covariance. We do this separately for each temperature-polarization spectrum, so that the hybrid estimator is just a linear combination of the individual estimators rather than mixing estimators of different type. Since the polarization noise is higher than for the temperature, we consider cross spectra of the form $C_l^{TE,ij}$ where $i \geq j$, and the weight functions are ordered so that lower i are more optimal in the case of lower noise. We then have the same number of cross-weight spectra for each of the power spectra. Since the hybrid estimators are just linear combinations of the separate estimators, their covariance can easily be calculated from the coupling matrices and full covariance matrix approximations given below. When including C_l^{BB} we impose a uniform weight function at $l < 120$ to minimize E/B mixing effects and ensure that the covariance matrix approximations below remain accurate. This is suboptimal but unbiased; we do not investigate the more difficult problem of optimally constraining the tensor amplitude here.

2. Covariance matrix approximations

Approximations for some components of the covariance matrices for the pseudo- C_l 's were given in Ref. [26] for a general pixel-weighting function $w(s)$ (pixels area Ω_s) and anisotropic but uncorrelated instrumental pixel noise $(\sigma_s^T)^2$ and $(\sigma_s^Q)^2 = (\sigma_s^U)^2$. The approximations essentially make as many assumptions as necessary for the result to simplify to the forms given; the approximations should be reasonably accurate for small cuts at high l (where $sY_{lm} \sim Y_{lm}$) and noise-dominated B -polarization spectra. Here we summarize these results with slight generalization, and extend to include all the terms needed for the full polarized and

correlated estimator covariance. We only consider the case of using pseudo- C_l estimators from single maps of T , Q and U with various weighting; the noise properties of cross spectra between multiple maps with independent noise are a simple generalization.

Assuming the polarization and temperature noise are uncorrelated, the covariance of the pseudo- C_l estimators can be estimated using the approximations (for $l \gg 1$ and significant noise so that E/B mixing effects are small and large f_{sky}):

$$\begin{aligned} \langle \Delta \tilde{C}_l^{TT,ij} \Delta \tilde{C}_l^{TT,pq} \rangle &\approx C_l^{TT} C_l^{TT} [\Xi_{TT}(l, l', \tilde{W}^{(ip)(jq)}) + \Xi_{TT}(l, l', \tilde{W}^{(iq)(jp)})] \\ &\quad + (C_l^T C_l^T)^{1/2} [\Xi_{TT}(l, l', \tilde{W}^{2T(ip)(jq)}) + \Xi_{TT}(l, l', \tilde{W}^{2T(iq)(jp)}) + \Xi_{TT}(l, l', \tilde{W}^{2T(jq)(ip)}) \\ &\quad + \Xi_{TT}(l, l', \tilde{W}^{2T(jp)(iq)})] + \Xi_{TT}(l, l', \tilde{W}^{TT(ip)(jq)}) + \Xi_{TT}(l, l', \tilde{W}^{TT(iq)(jp)}), \end{aligned} \quad (\text{E6})$$

$$\begin{aligned} \langle \Delta \tilde{C}_l^{TE,ij} \Delta \tilde{C}_l^{TE,pq} \rangle &\approx (C_l^{TT} C_l^{TT} C_l^{EE} C_l^{EE})^{1/2} \Xi_{TE}(l, l', \tilde{W}^{(ip)(jq)}) + C_l^{TE} C_l^{TE} \Xi_{TT}(l, l', \tilde{W}^{(iq)(jp)}) + \Xi_{TE}(l, l', \tilde{W}^{TQ(ip)(jq)}) \\ &\quad + (C_l^{TT} C_l^{TT})^{1/2} \Xi_{TE}(l, l', \tilde{W}^{2Q(ip)(jq)}) + (C_l^{EE} C_l^{EE})^{1/2} \Xi_{TT}(l, l', \tilde{W}^{2T(jq)(ip)}), \end{aligned} \quad (\text{E7})$$

$$\begin{aligned} \langle \Delta \tilde{C}_l^{EE,ij} \Delta \tilde{C}_l^{EE,pq} \rangle &\approx C_l^{EE} C_l^{EE} [\Xi_{EE}(l, l', \tilde{W}^{(ip)(jq)}) + \Xi_{EE}(l, l', \tilde{W}^{(iq)(jp)})] + (C_l^{EE} C_l^{EE})^{1/2} [\Xi_{EE}(l, l', \tilde{W}^{2Q(ip)(jq)}) \\ &\quad + \Xi_{EE}(l, l', \tilde{W}^{2Q(iq)(jp)}) + \Xi_{EE}(l, l', \tilde{W}^{2Q(jq)(ip)}) + \Xi_{EE}(l, l', \tilde{W}^{2Q(jp)(iq)})] \\ &\quad + \Xi_{EE}(l, l', \tilde{W}^{QQ(ip)(jq)}) + \Xi_{EE}(l, l', \tilde{W}^{QQ(iq)(jp)}), \end{aligned} \quad (\text{E8})$$

$$\begin{aligned} \langle \Delta \tilde{C}_l^{BB,ij} \Delta \tilde{C}_l^{BB,pq} \rangle &\approx C_l^{BB} C_l^{BB} [\Xi_{EE}(l, l', \tilde{W}^{(ip)(jq)}) + \Xi_{EE}(l, l', \tilde{W}^{(iq)(jp)})] \\ &\quad + (C_l^{BB} C_l^{BB})^{1/2} [\Xi_{EE}(l, l', \tilde{W}^{2Q(ip)(jq)}) + \Xi_{EE}(l, l', \tilde{W}^{2Q(iq)(jp)}) + \Xi_{EE}(l, l', \tilde{W}^{2Q(jp)(iq)}) \\ &\quad + \Xi_{EE}(l, l', \tilde{W}^{2Q(jq)(ip)})] + \Xi_{EE}(l, l', \tilde{W}^{QQ(ip)(jq)}) + \Xi_{EE}(l, l', \tilde{W}^{QQ(iq)(jp)}), \end{aligned} \quad (\text{E9})$$

$$\begin{aligned} \langle \Delta \tilde{C}_l^{EE,ij} \Delta \tilde{C}_l^{BB,pq} \rangle &\approx [(C_l^{EE} C_l^{EE})^{1/2} + (C_l^{BB} C_l^{BB})^{1/2}]^2 \frac{1}{2} [\Xi_{EB}(l, l', \tilde{W}^{(ip)(jq)}) + \Xi_{EB}(l, l', \tilde{W}^{(iq)(jp)})] \\ &\quad + (C_l^{EE} C_l^{EE})^{1/2} [\Xi_{EB}(l, l', \tilde{W}^{2Q(ip)(jq)}) + \Xi_{EB}(l, l', \tilde{W}^{2Q(iq)(jp)}) + \Xi_{EB}(l, l', \tilde{W}^{2Q(jp)(iq)}) \\ &\quad + \Xi_{EB}(l, l', \tilde{W}^{2Q(jq)(ip)})] + \Xi_{EB}(l, l', \tilde{W}^{QQ(ip)(jq)}) + \Xi_{EB}(l, l', \tilde{W}^{QQ(iq)(jp)}), \end{aligned} \quad (\text{E10})$$

$$\begin{aligned} \langle \Delta \tilde{C}_l^{TT,ij} \Delta \tilde{C}_l^{TE,pq} \rangle &\approx \frac{1}{2} (C_l^{TT} C_l^{TT})^{1/2} (C_l^{TE} C_l^{TE}) [\Xi_{TT}(l, l', \tilde{W}^{(ip)(jq)}) + \Xi_{TT}(l, l', \tilde{W}^{(iq)(jp)})] \\ &\quad + \frac{1}{2} (C_l^{TE} + C_l^{TE}) [\Xi_{TT}(l, l', \tilde{W}^{2T(ip)(jq)}) + \Xi_{TT}(l, l', \tilde{W}^{2T(jp)(iq)})], \end{aligned} \quad (\text{E11})$$

$$\begin{aligned} \langle \Delta \tilde{C}_l^{EE,ij} \Delta \tilde{C}_l^{TE,pq} \rangle &\approx \frac{1}{2} (C_l^{EE} C_l^{EE})^{1/2} (C_l^{TE} C_l^{TE}) [\Xi_{EE}(l, l', \tilde{W}^{(ip)(jq)}) + \Xi_{EE}(l, l', \tilde{W}^{(iq)(jp)})] \\ &\quad + \frac{1}{2} (C_l^{TE} + C_l^{TE}) [\Xi_{EE}(l, l', \tilde{W}^{2Q(ip)(jq)}) + \Xi_{EE}(l, l', \tilde{W}^{2Q(jp)(iq)})], \end{aligned} \quad (\text{E12})$$

$$\langle \Delta \tilde{C}_l^{TT,ij} \Delta \tilde{C}_l^{EE,pq} \rangle \approx C_l^{TE} C_l^{TE} [\Xi_{TT}(l, l', \tilde{W}^{(ip)(jq)}) + \Xi_{TT}(l, l', \tilde{W}^{(iq)(jp)}), \quad (\text{E13})$$

where the various window functions appearing are determined by the power spectra

$$\tilde{W}_l^{(ij)(pq)} = \frac{1}{2l+1} \sum_m w_{lm}^{ij} w_{lm}^{pq*}, \quad (\text{E14})$$

$$\tilde{W}_l^{TT(ij)(pq)} = \frac{1}{2l+1} \sum_m (w_{lm}^T w_{lm}^{T,pq*}), \quad \tilde{W}_l^{TQ(ij)(pq)} \equiv \tilde{W}_l^{TU(ij)(pq)} = \frac{1}{2l+1} \sum_m (w_{lm}^T w_{lm}^{Q,pq*}), \quad (\text{E15})$$

$$\tilde{W}_l^{2T(ij)(pq)} = \frac{1}{2l+1} \sum_m (w_{lm}^{ij} w_{lm}^{T,pq*}), \quad \tilde{W}_l^{2Q(ij)(pq)} \equiv \tilde{W}_l^{2U(ij)(pq)} = \frac{1}{2l+1} \sum_m (w_{lm}^{(ij)} w_{lm}^{Q,pq*}), \quad (\text{E16})$$

$$\tilde{W}_l^{QQ(ij)(pq)} \equiv \tilde{W}_l^{QU(ij)(pq)} \equiv \tilde{W}_l^{UU(ij)(pq)} = \frac{1}{2l+1} \sum_m w_{lm}^{Q,ij} w_{lm}^{Q,pq*}, \quad (\text{E17})$$

and the harmonic coefficients are given as sums over pixels with area Ω_s as

$$w_{lm}^{ij} = \sum_s w^i(s) w^j(s) \Omega_s Y_{lm}(s)^*, \quad (\text{E18})$$

$$w_{lm}^{T,ij} = \sum_s (\sigma_s^T)^2 w^i(s) w^j(s) \Omega_s^2 Y_{lm}(s),$$

$$w_{lm}^{Q,ij} \equiv w_{lm}^{U,ij} = \sum_s (\sigma_s^Q)^2 w^i(s) w^j(s) \Omega_s^2 Y_{lm}(s). \quad (\text{E19})$$

At the level of approximation considered here $\Xi_{EE} \sim \Xi_{TT} \sim \Xi_{TE}$, so there is some ambiguity in which particular form to use in the approximations. Note that the contribution of E to the \tilde{C}_l^B covariance is neglected, which is a poor approximation when the noise is not dominant; more accurate approximations are given in Ref. [31]. If the B -polarization contribution to the variance becomes important relative to the noise, the non-Gaussianity of the lensed B -polarization field also becomes an issue (see e.g. Ref. [6]). For Planck noise levels the B -lensing signal is well below the noise and E/B mixing effects are also well below the noise at $l \geq 100$.

The covariance matrix for the \hat{C}_l estimators is determined straightforwardly by applying the inverse coupling matrix to the above results. The covariance of the hybrid estimator is then just a contraction of the full multiestimator covariance with the hybrid mixing matrix.

3. Test simulations

The diagonal of the covariance matrix approximations given above agree very well with simulations at $l \geq 30$ if the weight map does not have too much small-scale power. The covariance approximations are more sensitive to small-scale power in the noise and weights than the coupling matrices; for this reason we use a smoother mask and noise map than is needed to obtain an accurate coupling matrix. This avoids numerical issues in our tests so that we can focus on any errors due to the likelihood approximations. We use a HEALPIX [38,39] pixelization at $N_{\text{side}} = 2048$, upgrading the simulated Planck noise map [32] and convolving it with 7 arcmin Gaussian kernel so that it is smooth on this scale. For the mask we take the WMAP kp2 map, upgrade to $N_{\text{side}} = 2048$ (12×2048^2 pixels), smooth with 7 arcmin kernel, set negative pixels to zero, and smooth again with a 7 arcmin kernel. This gives point source cuts that still go to essentially zero, while having edges smoothly tapering to one. To calculate the pseudo- C_l estimators we take w^1 as uniform weighting (multiplied by the cut), and a regularized inverse-noise weighting given by $w^2(s) \propto 1/(\sigma_s^2 + \min(\sigma_s^2))$, smoothed with a 7 arcmin kernel and then multiplied by the cut. We use the same

weight functions for temperature and polarization, and take $(\sigma_s^Q)^2 = (\sigma_s^U)^2 = 4(\sigma_s^T)^2$ for simplicity. Gaussian simulations are done to $l_{\text{max}} = 2200$ with zero monopole and dipole. The simulation code is available on the web [35].

4. Lensed simulation

The largest nonlinear effect on intermediate scales is expected to be that of CMB lensing [40]. Detailed modeling of the non-Gaussian distribution induced by this effect is beyond the scope of this paper, however for Planck noise levels the non-Gaussianity can be neglected to good approximation when performing parameter analyses from the lensed CMB power spectra [41]. The effect of lensing on the power spectrum is many percent, and must be included to obtain correct parameters with Planck. We update the LENSPIX code [42] to quickly simulate high-resolution lensed maps accurately. Our simulation method is as follows: (1) we simulate a HEALPIX map of a realization of the lensing deflection angle from a Gaussian realization of the lensing potential; (2) Divide the sphere into a number of slices separated by lines at constant polar angle θ , and assign each slice to a different processor (with some overlap given by the largest θ -deflection); (3) each processor simulates a Gaussian unlensed CMB map over its assigned slice on an equicylindrical grid; (4) interpolate from the equicylindrical grid to the deflected positions corresponding to the center of HEALPIX pixels offset by the deflection angles. Equations used for simulating gradient maps, deflecting points along geodesics, and appropriately rotating Stokes parameters are given in Ref. [42]. Our updated code is publicly available [43].

For our simulation we use $N_{\text{side}} = 2048$, and generate equicylindrical unlensed grids with points at 6144 different θ values ($\text{interp}_{\text{factor}} = 1.5$, effectively the same resolution as HEALPIX at $N_{\text{side}} = 2048$). The number of ϕ -pixels is chosen for each slice to be of the form $2^n 3^m$ (for integer n, m) so that fast Fourier transforms can be performed quickly, with lowest spacing roughly the same as the spacing in θ . To interpolate we use an extended cubic interpolation algorithm TOMS760 [44]; this is significantly slower than a basic bicubic interpolation scheme, but more accurate and stable—it ensures our results converge as the number of equicylindrical pixels is increased. Averaged over simulations our simulated lensed CMB power spectra then agree at the 0.1%-level with theoretical expectations for the same l_{max} [40,45]. Other simulation methods are discussed in Refs. [46–49], though nonlinear evolution effects are minor at Planck noise levels. Since the unlensed CMB is not bandlimited but contains residual power at $l \geq 2000$ our method does not rely on bandlimited interpolations and works directly with maps that con-

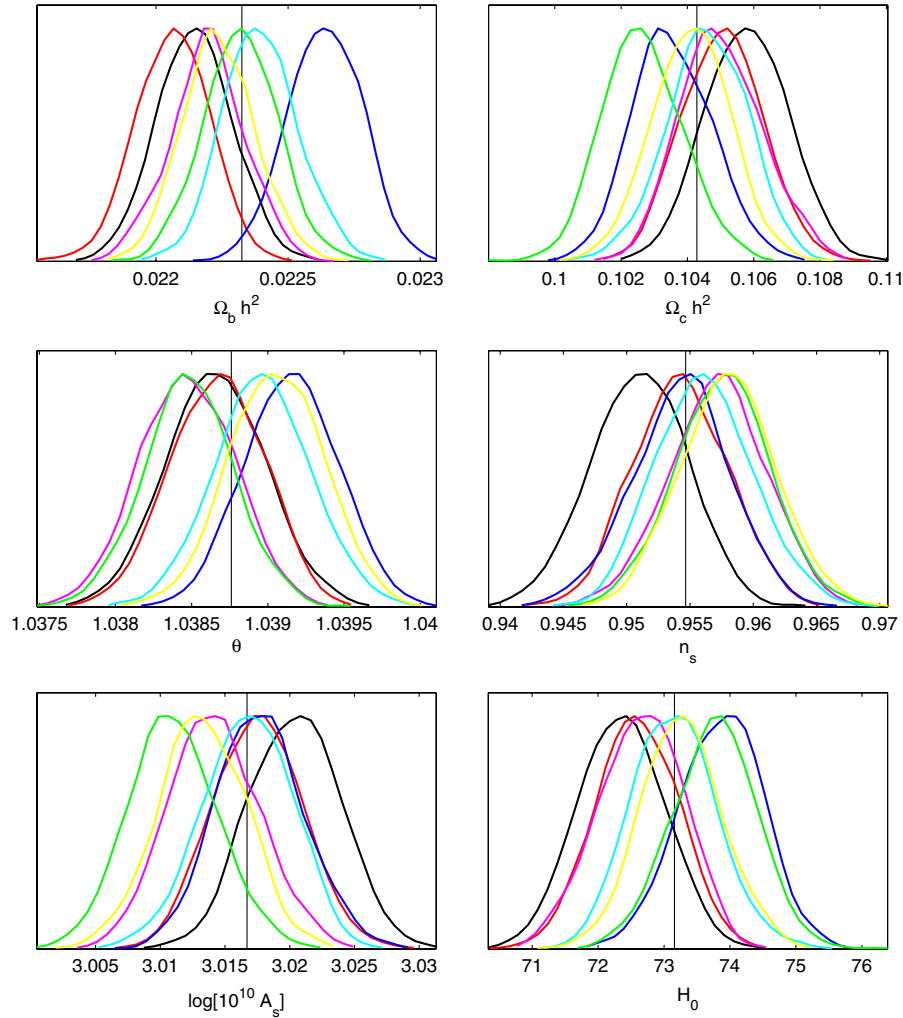


FIG. 11 (color online). Simulated parameter constraints from eight lensed CMB realizations using the new likelihood approximation with E and T (and cross) hybrid pseudo- C_l estimators at $30 \leq l \leq 2000$. The covariance was calculated using only the lensed power spectra. Input parameter values are marked with vertical lines, and the reionization optical depth was fixed. Including B estimators has virtually no effect at Planck noise levels.

tain power up to the highest simulated l_{\max} . On a modern few-node cluster lensed maps with polarization can be simulated in a few minutes.

Figure 11 shows parameter estimation constraints generated using a set of simulated lensed maps with Planck-like noise, and modeling the covariance as in the unlensed case simply by using the lensed power spectra instead of

the unlensed ones. A more optimal analysis would use the non-Gaussian information in the lensed field to indirectly constrain the lensing potential and hence cosmological parameters (see e.g. references in [40]), though it is unclear how much can be gained in the presence of real-world complications.

-
- [1] B.D. Wandelt, D.L. Larson, and A. Lakshminarayanan, Phys. Rev. D **70**, 083511 (2004).
 [2] D.L. Larson *et al.*, Astrophys. J. **656**, 653 (2007).
 [3] M. Chu *et al.*, Phys. Rev. D **71**, 103002 (2005).
 [4] J.R. Bond, A.H. Jaffe, and L.E. Knox, Astrophys. J. **533**, 19 (2000).

- [5] L. Verde *et al.*, Astrophys. J. Suppl. Ser. **148**, 195 (2003).
 [6] S. Smith, A. Challinor, and G. Rocha, Phys. Rev. D **73**, 023517 (2006).
 [7] W.J. Percival and M.L. Brown, Mon. Not. R. Astron. Soc. **372**, 1104 (2006).
 [8] K.M. Górski, Astrophys. J. Lett. **430**, L85 (1994).

- [9] A. Slosar, U. Seljak, and A. Makarov, *Phys. Rev. D* **69**, 123003 (2004).
- [10] L. Page *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **170**, 335 (2007).
- [11] M. Kamionkowski, A. Kosowsky, and A. Stebbins, *Phys. Rev. D* **55**, 7368 (1997).
- [12] A. Gupta and D. Nagar, *Matrix Variate Distributions* (Chapman and Hall, London, 1999), ISBN 1584880465.
- [13] <http://cosmologist.info/cosmomc/>.
- [14] Planck Collaboration, arXiv:astro-ph/0604069.
- [15] A. Lewis and S. Bridle, *Phys. Rev. D* **66**, 103511 (2002).
- [16] J. Martin and C. Ringeval, *Phys. Rev. D* **69**, 083515 (2004).
- [17] D. J. Mortlock, A. D. Challinor, and M. P. Hobson, *Mon. Not. R. Astron. Soc.* **330**, 405 (2002).
- [18] A. Lewis, A. Challinor, and N. Turok, *Phys. Rev. D* **65**, 023505 (2001).
- [19] G. Efstathiou, *Mon. Not. R. Astron. Soc.* **349**, 603 (2004).
- [20] G. Hinshaw *et al.* (WMAP Collaboration), *Astrophys. J. Suppl. Ser.* **170**, 288 (2007).
- [21] M. Tegmark, *Phys. Rev. D* **55**, 5895 (1997).
- [22] B. D. Wandelt, E. Hivon, and K. M. Gorski, *Phys. Rev. D* **64**, 083003 (2001).
- [23] E. Hivon *et al.*, *Astrophys. J.* **567**, 2 (2002).
- [24] F. K. Hansen, K. M. Gorski, and E. Hivon, *Mon. Not. R. Astron. Soc.* **336**, 1304 (2002).
- [25] M. L. Brown, P. G. Castro, and A. N. Taylor, *Mon. Not. R. Astron. Soc.* **360**, 1262 (2005).
- [26] G. Efstathiou, *Mon. Not. R. Astron. Soc.* **370**, 343 (2006).
- [27] K. M. Smith and M. Zaldarriaga, *Phys. Rev. D* **76**, 043001 (2007).
- [28] I. Szapudi, S. Prunet, D. Pogosyan, A. S. Szalay, and J. R. Bond, *Astrophys. J. Lett.* **548**, L115 (2001).
- [29] G. Chon, A. Challinor, S. Prunet, E. Hivon, and I. Szapudi, *Mon. Not. R. Astron. Soc.* **350**, 914 (2004).
- [30] G. Hinshaw *et al.*, *Astrophys. J. Suppl. Ser.* **148**, 135 (2003).
- [31] A. Challinor and G. Chon, *Mon. Not. R. Astron. Soc.* **360**, 509 (2005).
- [32] M. A. J. Ashdown *et al.*, arXiv:astro-ph/0702483.
- [33] C. L. Bennett *et al.* (WMAP), *Astrophys. J. Suppl. Ser.* **148**, 1 (2003).
- [34] <http://lambda.gsfc.nasa.gov/>.
- [35] See new CMB likelihood module at <http://cosmologist.info/cosmomc/CMBLike.html>.
- [36] A. Kogut *et al.*, *Astrophys. J. Suppl. Ser.* **148**, 161 (2003).
- [37] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (World Scientific, Singapore, 1988), ISBN 9971509962.
- [38] <http://www.eso.org/science/healpix/>.
- [39] K. M. Gorski *et al.*, *Astrophys. J.* **622**, 759 (2005).
- [40] A. Lewis and A. Challinor, *Phys. Rep.* **429**, 1 (2006).
- [41] W. Hu, *Phys. Rev. D* **64**, 083005 (2001).
- [42] A. Lewis, *Phys. Rev. D* **71**, 083008 (2005).
- [43] <http://cosmologist.info/lenspix>.
- [44] H. Akima, *ACM Trans. Math. Softw.* **22**, 357 (1996).
- [45] A. Challinor and A. Lewis, *Phys. Rev. D* **71**, 103010 (2005).
- [46] S. Das and P. Bode, arXiv:0711.3793.
- [47] C. Carbone, V. Springel, C. Baccigalupi, M. Bartelmann, and S. Matarrese, arXiv:0711.2655.
- [48] K. M. Smith, O. Zahn, and O. Dore, *Phys. Rev. D* **76**, 043510 (2007).
- [49] C. M. Hirata, N. Padmanabhan, U. Seljak, D. Schlegel, and J. Brinkmann, *Phys. Rev. D* **70**, 103501 (2004).