FERMILAB-PUB-15-542-A (accepted)
DOI: 10.1103/PhysRevD.93.083525

Extreme data compression for the CMB

Alan Zablocki[1,2] and Scott Dodelson[3,1,2]

[1] *Department of Astronomy and Astrophysics, University of Chicago, Chicago, Illinois 60637, USA*
[2] *Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA*
[3] *Fermilab Center for Particle Astrophysics, Fermi National Accelerator Laboratory, Batavia, Illinois 60510-0500, USA*

(Dated: August 3, 2016)

We apply the Karhunen-Loéve methods to cosmic microwave background (CMB) data sets, and show that we can recover the input cosmology and obtain the marginalized likelihoods in Λ cold dark matter cosmologies in under a minute, much faster than Markov chain Monte Carlo methods. This is achieved by forming a linear combination of the power spectra at each multipole l, and solving a system of simultaneous equations such that the Fisher matrix is locally unchanged. Instead of carrying out a full likelihood evaluation over the whole parameter space, we need evaluate the likelihood only for the parameter of interest, with the data compression effectively marginalizing over all other parameters. The weighting vectors contain insight about the physical effects of the parameters on the CMB anisotropy power spectrum C_l. The shape and amplitude of these vectors give an intuitive feel for the physics of the CMB, the sensitivity of the observed spectrum to cosmological parameters, and the relative sensitivity of different experiments to cosmological parameters. We test this method on exact theory C_l as well as on a Wilkinson Microwave Anisotropy Probe (WMAP)-like CMB data set generated from a random realization of a fiducial cosmology, comparing the compression results to those from a full likelihood analysis using CosmoMC. After showing that the method works, we apply it to the temperature power spectrum from the WMAP seven-year data release, and discuss the successes and limitations of our method as applied to a real data set.

Keywords: Cosmology, WMAP, CMB, MCMC, Parameter Estimation, Data Compression, Data Analysis

arXiv:1512.00072v3 [astro-ph.CO] 1 Aug 2016

I. INTRODUCTION

Modern astrophysical data sets are getting ever larger. This is driven in part by the increased size of the telescopes allowing large astronomical surveys, as well as the increase in the detector number, their sensitivity, and the resolution. Future galaxy surveys like the Large Synoptic Survey Telescope (LSST) and Euclid will observe on order $\sim 10^9$ galaxies, while current cosmic microwave background (CMB) experiments such as Planck, the South Pole Telescope (SPT) and the Atacama Cosmology Telescope (ACT) already map the microwave sky over more than $\sim 10^7$ pixels. Data compression and sophisticated statistical methods applied to these extremely large data sets have ushered us into the era of "precision cosmology", where the data is very well described by the simple six parameter Λ cold dark matter (CDM) model.

The large size of today's data sets often makes it impractical to carry out brute force likelihood calculations. This has therefore motivated a number of data compression methods to be developed for use in statistical analyses of galaxy redshift surveys [1] and CMB maps [2, 3]. A common approach is to compress the data quadratically into a number of power spectrum estimates; for galaxy redshift surveys, the compressed data set is a set of power spectrum estimates $P(k)$ and for CMB experiments, it is the anisotropy power spectrum of fluctuations C_l. To obtain estimates of model parameters, one then performs a Bayesian likelihood analysis using Markov chain Monte Carlo (MCMC) methods.

The Karhunen-Loève (KL) eigenvalue method was previously applied to both CMB maps [4] and redshift surveys [5]. The KL compression method can be generalized to two important examples for data sets with certain noise properties: (i) the case where the mean is known and independent of model parameters and (ii) the case where the covariance is independent of model parameters [6]. Here we consider the second case, when the data vector is the power spectrum, C_l, itself.

This case was applied to galaxy spectra, where the speedup in the likelihood computation was achieved using a set of orthonormal compression vectors [7, 8] (akin to the Gram-Schmidt procedure, for which the order of vectors matters). The same procedure was also applied to mock CMB data for only three parameters, but it excluded experimental noise [9]. This covariance-independent case has been shown to occasionally produce multimodal likelihood peaks, in applications to planetary transit light curves [10] and gravitational wave data analysis [11], though there are ways to mitigate these problems, albeit at an increase in computation time by as much as a factor of 20.

More recently, minimizing the computational cost of an exact CMB likelihood and power spectrum estimation using linear compression was investigated in [12] using Wilkinson Microwave Anisotropy Probe (WMAP) data as an example, while in [13] the authors looked at efficiently summarizing CMB data using two shift parameters and the physical baryon density $\Omega_b h^2$ to obtain dark energy constraints. In [14], the authors showed that a nonlinear transformation of cosmological parameters can also serve as a form of data compression, which yields a set of normal parameters with a Gaussian likelihood distribution, although in that case there is no reduction in the number of parameters.

In this work we create the weighting vectors according

to the prescription found in [6]. Instead of creating a set of orthonormal vectors we create a linear combination of all the data, such that the resulting *mode* holds the most information on the parameter of interest, with the data compression automatically marginalizing over all other parameters. We carry out this procedure for six ΛCDM parameters, although we have tested our methods on extensions to ΛCDM, e.g., by including the tensor-to-scalar ratio r parameter.

In contrast to work carried out in [7, 8, 10, 11] our method uses only one mode, offering a significant speedup in obtaining the marginalized likelihoods, and it does not depend on the order of the parameters. We note that the choice of parametrization will matter when investigating models with known or unknown degeneracies.

The paper is organized as follows: in Sec. II we introduce the extreme compression (EC) method and describe its implementation on CMB spectra. In Sec. III we implement the compression for a single parameter and describe the marginalization procedure for the whole parameter space. In Sec. IV, we derive the compression vectors and discuss their physical characteristics as applied to the CMB. We then test our method on two mock data sets, including experimental noise and compare against results obtained using MCMC. As a further test, we analyze the WMAP seven-year CMB specturm in Sec. V and conclude in Sec. VI.

II. DEVELOPING THE FORMALISM

In this section we briefly review some special cases of data compression presented in [6]. We then develop the case where the covariance of the data is assumed to be known and independent of the model parameters, and apply this method to the CMB power spectrum.

A. Compressing the Fisher information matrix

The log-likelihood $\mathcal{L}$ for a Gaussian probability distribution can be written as

$$-2\mathcal{L} = n\ln 2\pi + \ln\det \mathbf{Cov} + (\mathbf{x}-\boldsymbol{\mu})^t\mathbf{Cov}^{-1}(\mathbf{x}-\boldsymbol{\mu}), \quad (1)$$

where the covariance matrix is $\mathbf{Cov} = \langle(\mathbf{x}-\boldsymbol{\mu})^t(\mathbf{x}-\boldsymbol{\mu})\rangle$ and $\boldsymbol{\mu}$ is the mean $\langle\mathbf{x}\rangle$. The Fisher information matrix is defined as

$$\mathbf{F}_{ij} = -\left\langle \frac{\partial^2\mathcal{L}}{\partial\theta_i\partial\theta_j}\right\rangle = -\langle\mathcal{L}_{,ij}\rangle, \quad (2)$$

and is a measure of the curvature of the likelihood around the maximum likelihood point $\theta_{\rm ML}$. Working through some matrix algebra it can be shown that the Fisher matrix can be written as

$$\mathbf{F}_{ij} = \frac{1}{2}\,\mathrm{Tr}[\mathbf{A}_i\mathbf{A}_j + \mathbf{Cov}^{-1}\mathbf{M}_{ij}], \quad (3)$$

where $\mathbf{A}_i = \mathbf{Cov}^{-1}\mathbf{Cov}_{,i} = (\ln\mathbf{Cov})_{,i}$ and $\mathbf{M}_{ij} = \langle\mathbf{D}_{,ij}\rangle = \boldsymbol{\mu}^t_{,i}\boldsymbol{\mu}_{,j} + \boldsymbol{\mu}^t_{,j}\boldsymbol{\mu}_{,i}$[5].

We can perform a linear compression on our data set $\mathbf{x}$ with

$$\boldsymbol{y} = \mathbf{B}\mathbf{x}, \quad (4)$$

where $\mathbf{B}$ is the compression matrix of size $n'\times n$ and $\boldsymbol{y}$ is the resulting data set of dimension n'. It can be shown [6] that for $n = n'$ and $\mathbf{B}$ invertible, the new Fisher matrix after the linear compression, $\tilde{\mathbf{F}}_{ij}$, is given by

$$\tilde{\mathbf{F}}_{ij} = \frac{1}{2}\,\mathrm{Tr}[\mathbf{B}^{-t}(\mathbf{A}_i\mathbf{A}_j + \mathbf{Cov}^{-1}\mathbf{M}_{ij})\mathbf{B}^t] = \mathbf{F}_{ij}. \quad (5)$$

The Fisher matrix is thus unchanged. For $n' < n$, the matrix $\mathbf{B}$ is not invertible and each row of $\mathbf{B}$ specifies one number in the new data set. For the simplest case where only one linear combination of the data is selected so that $\mathbf{B}$ has just one row, $\mathbf{B} = \mathbf{b}^t$ the diagonal entries of the Fisher matrix are

$$\tilde{\mathrm{F}}_{ii} = \frac{1}{2}\left(\frac{\mathbf{b}^t\mathbf{Cov}_{,i}\mathbf{b}}{\mathbf{b}^t\mathbf{Cov}\ \mathbf{b}}\right)^2 + \frac{\left(\mathbf{b}^t\boldsymbol{\mu}_{,i}\right)^2}{\left(\mathbf{b}^t\mathbf{Cov}\ \mathbf{b}\right)}. \quad (6)$$

How can we use this result to estimate the value of some parameter θ_i and the error $\Delta\theta_i$ associated with it? We wish to define $\mathbf{b}^t$ such that the compressed data set carries as much information about parameter θ_i as possible. That is, we aim to minimize the error on θ_i. To do so, we maximize the element of the Fisher matrix $\tilde{\mathrm{F}}_{ii}$. The solution in general is nonlinear in $\mathbf{b}$. Inspection of Eq. (6) shows that the Fisher matrix now consists of two terms, one of which depends on the derivative of the covariance $\mathbf{Cov}_{,i}$ and another that depends on the derivative of the mean $\boldsymbol{\mu}_{,i}$. Assuming that the CMB covariance matrix is weakly dependent on the parameters, even though this assumption is not quite correct at low multipoles, yields an interesting result. In that case, the Fisher matrix is just

$$\tilde{\mathrm{F}}_{ii} = \frac{\left(\mathbf{b}^t\boldsymbol{\mu}_{,i}\right)^2}{\left(\mathbf{b}^t\mathbf{Cov}\ \mathbf{b}\right)}. \quad (7)$$

Maximizing this leads to the solution $\mathbf{b} = \mathbf{Cov}^{-1}\boldsymbol{\mu}_{,i}$. Our compressed data set, $y = \mathbf{b}^t\mathbf{x}$, now consists of just one number y_i,

$$y_i = \boldsymbol{\mu}^t_{,i}\mathbf{Cov}^{-1}\mathbf{x}. \quad (8)$$

In this case the compressed Fisher matrix is given by

$$\mathrm{F}_{ii} = \boldsymbol{\mu}^t_{,i}\mathbf{Cov}^{-1}\boldsymbol{\mu}_{,i}. \quad (9)$$

B. Applying data compression to the CMB power spectrum

The CMB temperature anisotropies form a scalar 2D field on the sky and are often expanded in spherical harmonics

$$\frac{\Delta T}{T}(\theta,\phi) = \sum_l\sum_m a_{lm}Y_{lm}(\theta,\phi), \quad (10)$$

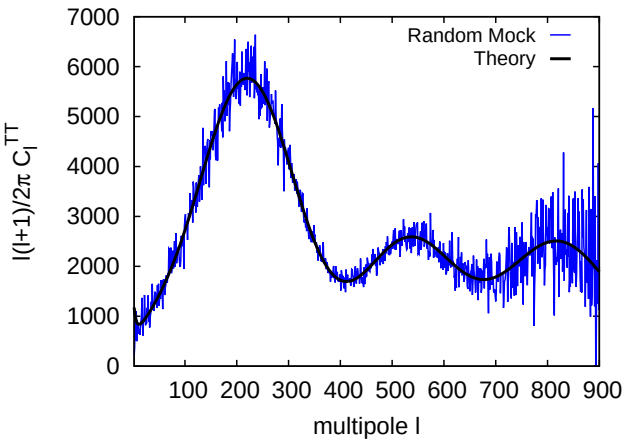

FIG. 1. Temperature power spectrum obtained from a random realization (solid blue) and the theory power spectrum computed with the Boltzmann code CAMB [15] in solid black. We ignore the effect of lensing on the CMB.

where ΔT is the temperature variation from the mean, l is the multipole, $Y_{lm}(\theta,\phi)$ is the spherical harmonic function of degree l and order m, and a_{lm} are the expansion coefficients or multipole moments. The variance $\delta_{ll'}\delta_{mm'}C_l = \langle a^*_{lm}a_{l'm'}\rangle$, where $\delta_{ll'}$ is the Kronecker delta function, contains all the statistical information. Here we use the temperature power spectrum so that the data vector is

$$\mathbf{x} = \frac{1}{2l+1}\sum_{m=-l}^{m=l}|a^T_{lm}|^2, \tag{11}$$

such that $\langle\mathbf{x}\rangle = \boldsymbol{\mu} = C_l$. We are therefore carrying out a quadratic precompression [6]. In Fig. 1 we compare the theory temperature power spectrum with that of a random realization for a WMAP-like experiment. The compressed data set for a given parameter θ_i is a single linear combination of the C_l's:

$$y_i = \sum_l \frac{\partial C_l}{\partial\theta_i}\mathrm{Cov}^{-1}(C_l,C_l)\ \frac{1}{2l+1}\sum_{m=-l}^{m=l}|a_{lm}|^2. \tag{12}$$

The measurement of the angular power spectrum C_l has characteristic uncertainty due to finite beam size and a limit on the number of modes we observe on the sky known as cosmic variance, with the variance at each multipole given by

$$\mathrm{Cov}(C_l,C_l) = \frac{2}{(2l+1)f_{\rm sky}}(C_l+N_l)^2, \tag{13}$$

where $f_{\rm sky}$ is the fraction of the sky covered by the experiment. For maps made with Gaussian beams the noise term N_l has the form [16]

$$N_l = (\sigma\theta)^2 e^{l(l+1)\,\theta^2/8\ln 2}, \tag{14}$$

where σ and θ are the sensitivity $(\Delta T/T)$ and angular resolution in radians respectively.

The expected value $\langle y_i\rangle$ is then

$$\langle y_i\rangle = \sum_l \frac{\partial C_l}{\partial\theta_i}\mathrm{Cov}^{-1}(C_l,C_l)\ C_l, \tag{15}$$

and $\langle y_i\rangle$ carries all the information contained in the data on θ_i. We can define the coefficients α^i_l to be

$$\alpha^i_l = \frac{\partial C_l}{\partial\theta_i}\mathrm{Cov}^{-1}(C_l,C_l), \tag{16}$$

so that

$$\langle y_i\rangle = \sum_l \alpha^i_l\ C_l. \tag{17}$$

For a given parameter θ_i, the coefficients α^i_l describe the combination of multipoles that carry the information about θ_i.

The variance of $\langle y_i\rangle$ is

$$\sigma^2_{\langle y_i\rangle} = \langle y_i^2\rangle - \sum_{l,l'}\alpha^i_l\alpha^i_{l'}C_lC_{l'}. \tag{18}$$

Since the a_{lm} are Gaussian fields, the resulting four-point functions are easily evaluated and

$$\sigma^2_{\langle y_i\rangle} = \sum_{l=1}\alpha^i_l\ \mathrm{Cov}(C_l,C_l)\ \alpha^i_l. \tag{19}$$

Using the expected value and variance of $\langle y_i\rangle$ we can rewrite the compressed Fisher matrix given by Eq. (9) as

$$\mathrm{F}^y_{ii} = \left(\frac{dy}{d\theta_i}\right)^2\frac{1}{\sigma^2_{\langle y_i\rangle}}. \tag{20}$$

We can compare the error bars obtained from the extremely compressed Fisher matrix above to the error bar obtained with Eq. (9), which is identical to the Fisher information matrix for the CMB as

$$\mathrm{F}^{\rm CMB}_{ij} = \sum_l \frac{\partial C_l}{\partial\theta_i}\mathrm{Cov}^{-1}(C_l,C_l)\frac{\partial C_l}{\partial\theta_j}. \tag{21}$$

III. IMPLEMENTATION

A. One parameter example

Using the prescription in the previous section we are now able to compress the CMB temperature power spectrum into just a handful of numbers. To illustrate the procedure we first choose a simple one parameter example focusing on the scalar power spectrum normalization parameter A_s. Using Eq. (16) and choosing a fiducial point at which to compute the derivative of C_l with respect to $\ln(10^{10}A_s)$, we obtain the weighting vector on A_s, which is plotted in Fig. 2. In general, we expect the weights to start with a small amplitude at low l, where

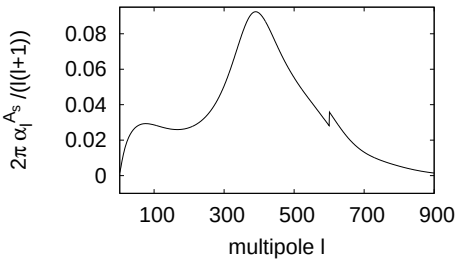

FIG. 2. The compression vector on the scalar power spectrum amplitude A_s. The discontinuity at $l \sim 600$ is due to a drop in the WMAP experimental noise.

cosmic variance is high, then to increase until the experimental noise starts to dominate. For WMAP, this starts at $l \sim 900$, with the weights decreasing to zero between an l of 900–1200. A simple test of this compression is to use the theory C_l's as the *data vector*, and with WMAP-like noise, compute the likelihood for A_s. This is depicted in Fig. 3. The one curve there is actually three curves, (i) the likelihood computed using a single mode y_{A_s}:

$$-2\ln\mathrm{L} = \frac{(y_{A_s} - \bar{y}_{A_s})^2}{2\sigma^2_{\langle y_{A_s}\rangle}}, \tag{22}$$

(ii) the likelihood using the full set of C_l's, and (iii) the Fisher (Gaussian) approximation with the variance obtained from Eq. (20). All three approaches give the same answer, showing that in this simple case, the compression works well.

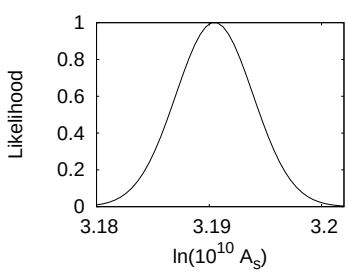

FIG. 3. Unmarginalized likelihood for the log power of the primordial curvature perturbations. The data used here is the exact theory C_l.

B. Two parameter model example

In the previous section we showed how to compress a data set and obtain the likelihood for a single parameter. As can be seen in Fig. 3, the likelihood is quite narrow, and the error on $\ln(10^{10}A_s)$ is very small. In this section, we will show how to compress accounting for a second parameter, obtaining marginalized distributions very quickly.

Each compressed data set y_i, by design, carries all the information on the parameter of interest θ_i. However, it will also have some sensitivity to the other parameter, a sensitivity that we would like to remove. We now show with this simple two-dimensional example how to remove the unwanted sensitivity, essentially marginalizing over the remaining parameter.

We begin by forming a linear combination of y_1 and y_2 for the first parameter as

$$y_1^{'} = c_1 y_1 + c_2 y_2, \tag{23}$$

with $y_1 = \sum_l \alpha_l^1 C_l$ and $y_2 = \sum_l \alpha_l^2 C_l$, where c_1 and c_2 are chosen by the requirement that $y_1^{'}$ does not depend on θ_2. For this to be independent of θ_2 we require that the derivative of $y_1^{'}$ with respect to θ_2 vanishes. We then obtain

$$\frac{\partial y_1^{'}}{\partial \theta_2} = c_1 \left[\sum_l \alpha_l^1 \, \frac{\partial C_l}{\partial \theta_2} \right] + c_2 \left[\sum_l \alpha_l^2 \, \frac{\partial C_l}{\partial \theta_2} \right] = 0. \tag{24}$$

The quantities in square brackets are just the Fisher matrix elements so that the equation for $y_1^{'}$ is

$$\frac{\partial y_1^{'}}{\partial \theta_2} = 0 = c_1 \mathrm{F}_{12} + c_2 \mathrm{F}_{22}. \tag{25}$$

This fixes the ratio of the two coefficients, and c_1 can be set to unity, so that the new, marginalized vector $y_1^{'}$ is

$$y_1^{'} = \sum_l \alpha_l^{'1} \, C_l \tag{26}$$

with

$$\alpha_l^{'1} = \alpha_l^1 - \frac{\mathrm{F}_{12}}{\mathrm{F}_{22}} \alpha_l^2. \tag{27}$$

Repeating the procedure for the second parameter yields the weighting vector

$$\alpha_l^{'2} = \alpha_l^2 - \frac{\mathrm{F}_{12}}{\mathrm{F}_{11}} \alpha_l^1. \tag{28}$$

We note that in two dimensions, this particular example is equivalent to the common approach of creating an orthonormal basis using the Gram-Schmidt process in quantum mechanics. More specifically, the dot product (defined by $\mathbf{b}^t\mathbf{Cov}\ \mathbf{b}$) is only zero for the combinations of $\boldsymbol{\alpha}^{2'}\mathbf{Cov}\ \boldsymbol{\alpha}^1$ and $\boldsymbol{\alpha}^{1'}\mathbf{Cov}\ \boldsymbol{\alpha}^2$ with $\boldsymbol{\alpha}^{2'}\mathbf{Cov}\ \boldsymbol{\alpha}^{1'} \neq 0$.

As an example, consider the compressed data set for n_s and A_s. All the information about each parameteris contained in a single χ^2; e.g.,

$$\chi^2_{n_s} = \frac{(y^{'}_{n_s} - \bar{y}^{'}_{n_s})^2}{2\sigma^2_{\langle y^{'}_{n_s}\rangle}} \tag{29}$$

is a function of n_s only. With information on the other parameter removed, we need explore only one dimension to get the marginalized posterior. This is why the method is much faster than spanning the full two-dimensional likelihood space. If we sample each dimension 20 times,

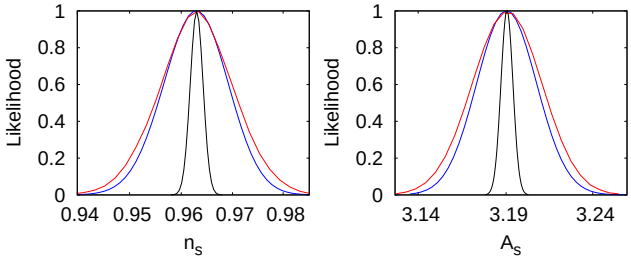

FIG. 4. Marginalized likelihoods on the spectral index n_s and the scalar power spectrum amplitude A_s using extreme compression (solid blue) and the exact result from MCMC (solid red) for the two parameter toy model. Marginalization is achieved using the solutions for $\alpha_l^{'1}$ and $\alpha_l^{'2}$ and Eqs. (27) and (28). The unmarginalized case is shown in solid black for reference. The data used is the exact theory C_l.

the full likelihood is obtained with only $2 \times 20 = 40$ samples instead of $20^2 = 400$. And of course, as the parameter space gets larger, the difference becomes much more pronounced. In Fig. 4, we show the n_s and A_s marginalized likelihoods for exact theory C_l from the full likelihood and the compression given by Eqs. (27) and (28).

C. Generalizing to higher dimensions

Based on the results of the previous section we now present the general problem for n parameters along with the solutions. The most general linear combination of all the data in a model with n parameters can be written as

$$y_1^{'} = c_1 y_1 + c_2 y_2 + \cdots + c_n y_n, \tag{30}$$

such that the compressed mode $y_1^{'}$ carries all the information on the first parameter θ_1, with information on all other parameters removed. To obtain the extreme compressed θ_1 mode, $y_1^{'}$, we must solve the matrix problem

$$\begin{pmatrix} \mathrm{F}_{22} & \mathrm{F}_{23} & \cdots & \mathrm{F}_{2n} \\ \mathrm{F}_{32} & \mathrm{F}_{33} & \cdots & \mathrm{F}_{3n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{F}_{n2} & \mathrm{F}_{n3} & \cdots & \mathrm{F}_{nn} \end{pmatrix} \begin{pmatrix} c_2 \\ c_3 \\ \vdots \\ c_n \end{pmatrix} = \begin{pmatrix} -\mathrm{F}_{12} \\ -\mathrm{F}_{13} \\ \vdots \\ -\mathrm{F}_{1n} \end{pmatrix}. \tag{31}$$

This yields $n-1$ unique constants on the $n-1$ coefficients $c_i (i > 1)$ and c_1 can be set to unity. The same procedure holds for all other modes: for mode $i = \alpha$, the coefficients are determined by the general equation

$$\mathrm{F}'_{\alpha,ij} c_j = -\mathrm{F}_{\alpha i}, \tag{32}$$

where F'_α is the Fisher matrix with row and column α removed.

In the next section we calculate the weighting vectors for a WMAP-like experiment, and apply the compression method to mock WMAP data sets.

IV. TESTS ON A WMAP-LIKE EXPERIMENT

We now apply this formalism to obtain marginalized likelihoods from synthetic data from a WMAP-like experiment (mock data sets with WMAP noise) to see how well we can recover the parameters using extreme compression. We use the same parametrization as CosmoMC, with $100 * \theta_{\rm MC}$, an approximation for $r_s(z_\star)/D_A(z_\star)$, the angular scale of the sound horizon at last scattering, replacing Ω_Λ or H_0 due to a known geometric degeneracy in the CMB (see Appendix A). The fiducial cosmology assumed is: $\omega_c = \Omega_c h^2 = 0.1109$, $\omega_b = \Omega_b h^2 = 0.02258$, $100 * \theta_s = 1.039485$, $n_s = 0.963$, $\ln(10^{10} A_s) = 3.1904$ and $\tau = 0.088$. We first obtain the posterior distributions assuming that the data vector is the exact theory C_l, and then test on a more realistic mock data set using a random realization of the fiducial cosmology.

A. WMAP weighting vectors

In Sec. II B, we showed that to achieve locally lossless compression of our CMB data set we need to compute the covariance of the data (where data is the spectrum C_l) and the derivative of the data with respect to the cosmological parameters in the ΛCDM model. To calculate the weighting vectors for the CMB power spectrum, we obtain the six derivatives of the power spectrum with respect to the parameter vector $\Theta = \{\omega_c, \omega_b, 100\theta_s, n_s, A_s, \tau\}$. We use a double sided derivative formula with a step size of 3% (we use 0.5% for the derivative with respect to θ_s).

In Fig. 5 we show the compression vectors for all the parameters. Due to cosmic variance the data at lower multipoles is given a low weight, while for $l > 900$ the amplitude of the vectors tends to zero due to the experimental noise. For a WMAP-like experiment, therefore, the vectors all peak in the range $l \approx 330 - 440$, with WMAP being cosmic variance limited up to around $l \sim 550$. The jump at $l \sim 600$ is due to a discontinuity in the WMAP noise.

We have already seen that the mode that captures the amplitude A_s is as shown in the middle bottom panel: uniformly positive, but weighing the higher signal to noise modes most heavily. The mode that captures the baryon density differences the heights of the first and second peaks, as expected. The sound horizon angle is captured by its alternating effect on peaks and troughs. The dark matter density leaves its imprint on the first peak, normalized to the most constrained value at $l \sim 400$. The optical depth is essentially the inverse vector of the amplitude because it enters the temperature spectrum via $A_s e^{-2\tau}$, while the mode that captures the spectral index n_s is sensitive to the decrease in C_l amplitude as the spectral index increases, up to the first peak.

In Fig. 6 we show the marginalized vectors with other parameters removed and compare them to vectors from Fig. 5. We find that many of the qualitative features

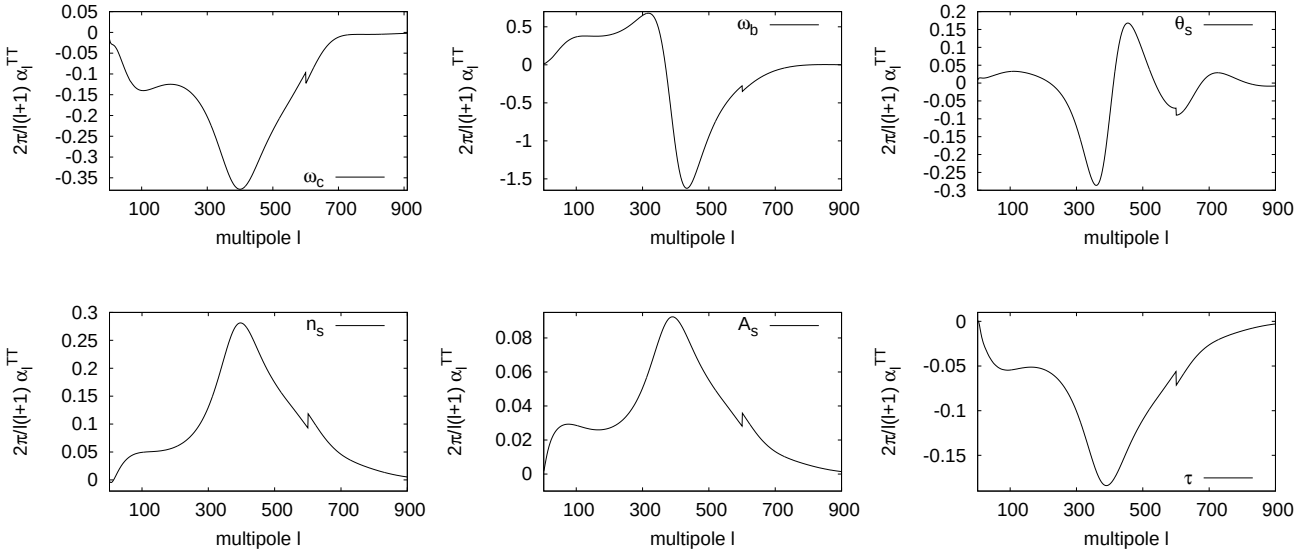

FIG. 5. The six ΛCDM weighting vectors α_l^i for a CMB experiment with WMAP noise and sky coverage. Each vector is used to compress the temperature power spectrum C_l^{TT}, into a single number y_i that carries all the information on each parameter θ_i. A general feature of these vectors, is that their amplitudes are small at low-l, where cosmic variance is large [Eq. (13)], and at high-l, where experimental noise dominates. The weights go down to zero between $l = 900$ and $l = 1200$. All six vectors reach their maximum amplitude between l of 330-440. The jump at $l \sim 600$ is due to a discontinuity in WMAP noise.

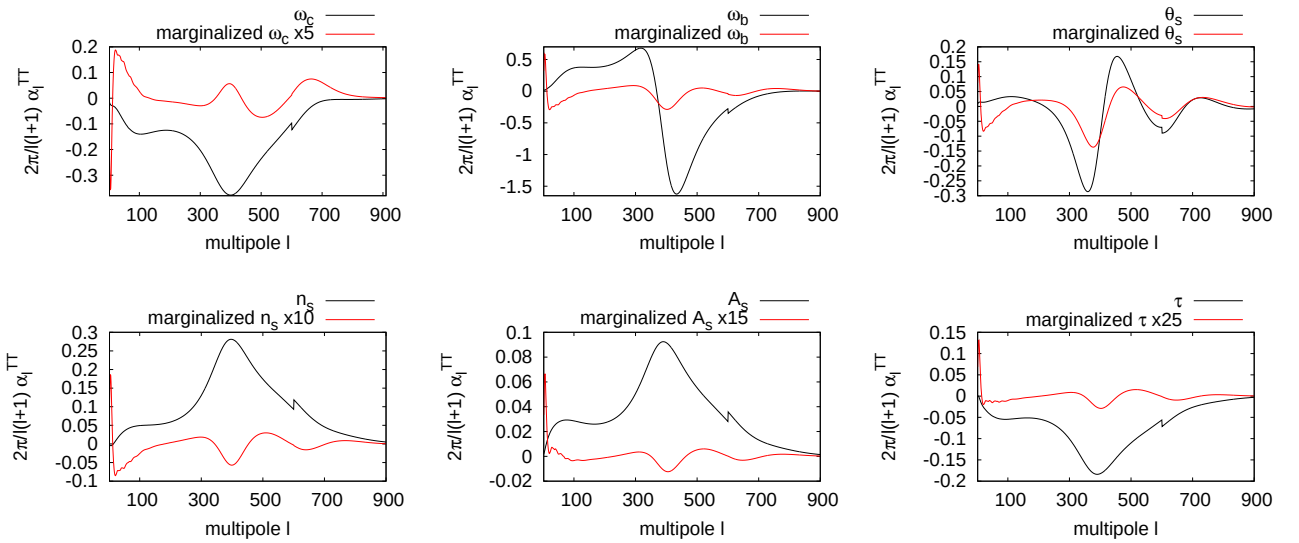

FIG. 6. Comparison of the compression vectors α_l^i for WMAP before (solid black lines) and after marginalization (solid red lines) using the prescription in Sec. III C. Some of the marginalized vectors have been multiplied by a scale factor to ease the comparison. Note the apparent decrease in the amplitude in each vector, once we take out the information on all the other parameters.

remain, but the vectors are reduced in amplitude. This is the cost of removing the information about other parameters: information degenerate with those parameters about the parameter of interest is also removed.

These vectors can be a useful tool to determine the relative importance of obtaining cosmic variance limited measurement of the power spectrum versus a higher sensitivity measurement C_l at smaller scales. A recent example is the apparent need for a precise measurement of the reionization bump in order to break parameter degeneracies and obtain the best constraints on the sum of neutrino masses from a stage 4 CMB experiment [17].

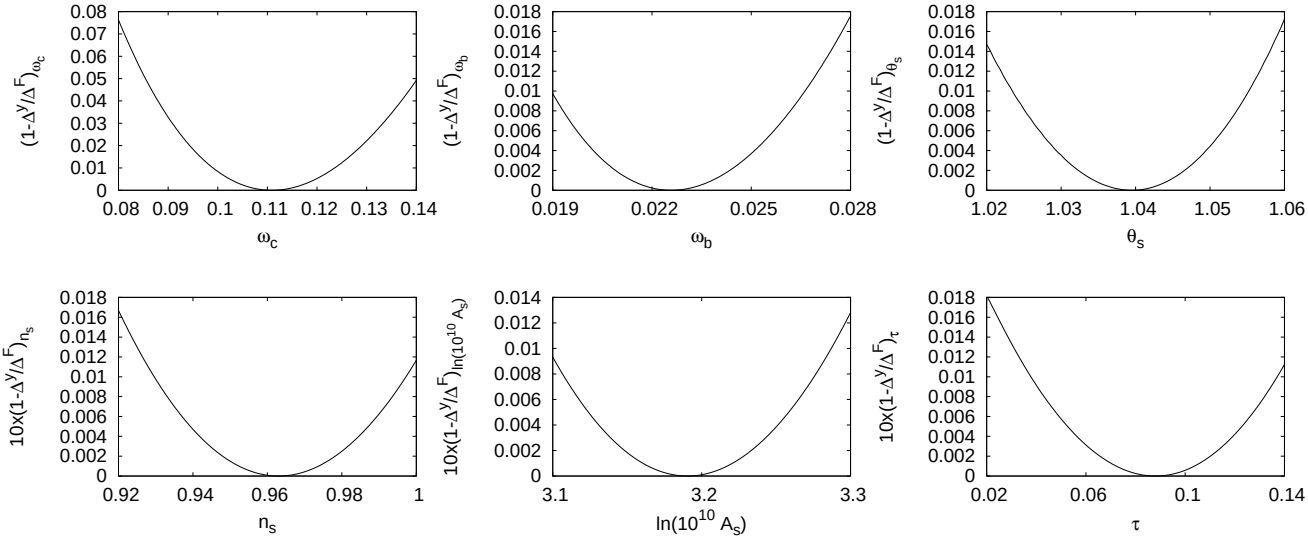

FIG. 7. How well can we recover the Fisher matrix using the new compressed data set? We show that the compressed Fisher matrix (Sec. II A) is the same as the full matrix in Eq. (21). Shown is the ratio between the error from the compressed Fisher matrix and the Fisher matrix using all the data as a function of the value of the parameter assumed when computing the coefficients α_l^i. By construction the Fisher matrix is unchanged at the fiducial value of a parameter (and in this case the maximum likelihood point $\theta_{\rm ML}$). Since the Fisher matrix remains the same, the compression is locally lossless. Some plots have been scaled since the ratios are very small. When computing the C_l derivative with respect to θ_s, we keep ω_c and ω_b constant (Ω_Λ and H_0 are changed to keep a flat universe). In general, when computing the derivatives with respect to ω_c and ω_b, we hold θ_s constant. However in the plots on ω_c and ω_b above, we do not keep θ_s constant.

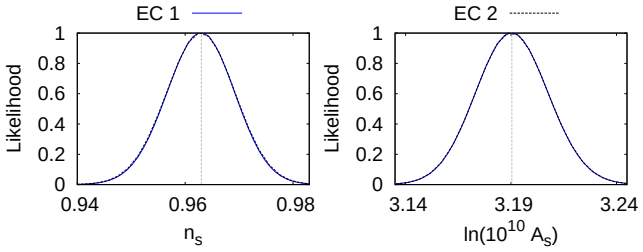

FIG. 8. We plot the recovered likelihoods for the case where the weighting vectors are computed with a different fiducial cosmology, and note the excellent agreement between the two compressions shown in dashed black and solid blue.

B. Sensitivity to fiducial choice

The vectors shown in the previous section are solutions to an eigenvalue problem that minimizes the error on each parameter, and leaves the Fisher matrix locally unchanged. At the fiducial point, at which the derivatives and the covariance are computed, we expect the errors from the compressed Fisher matrix using the extreme compression to equal those from the full Fisher matrix. But how well can we recover the parameters if the coefficients α_l^i are chosen away from the fiducial point, and how much does the error bar increase?

In Fig. 7, we show the ratio of the error from the Fisher matrix obtained with the extreme compression F_{ii}^y to the error obtained using the full data set from Eq. (21). Even over a wide range of parameter space (roughly the same as the expected width of the marginalized posteriors from WMAP) we find that $\Delta\theta_i$ changes by less than 0.2% for parameters n_s, A_s and τ, while the errors increase by at most 2% for ω_b and θ_s. For the physical cold dark matter density ω_c, the error change is less than 8%. At the fiducial point, the compression is locally lossless.

Another important question that we address is whether the fiducial cosmology used in the compression affects the results. To test whether the choice of the fiducial point matters, we created a new set of compression vectors α_l^i computed at a different cosmology, denoted as EC 2, with the following values for the cosmological parameters: $\omega_c = 0.12$, $\omega_b = 0.0235$, $100 * \theta_s = 1.0485995$, $n_s = 0.98$, $\ln(10^{10}A_s) = 3.258$, and $\tau = 0.085$. We then marginalized over all other parameters and used the new marginalized vectors to compress an exact theory C_l data set with WMAP-like noise.

In Fig. 8, we plot the recovered likelihoods when compressing the data with our fiducial cosmology denoted as EC 1 (solid blue lines), and the new cosmology as EC 2 (solid black lines). Figure 8 shows that no matter what the fiducial point we choose, we still get back the correct answer.

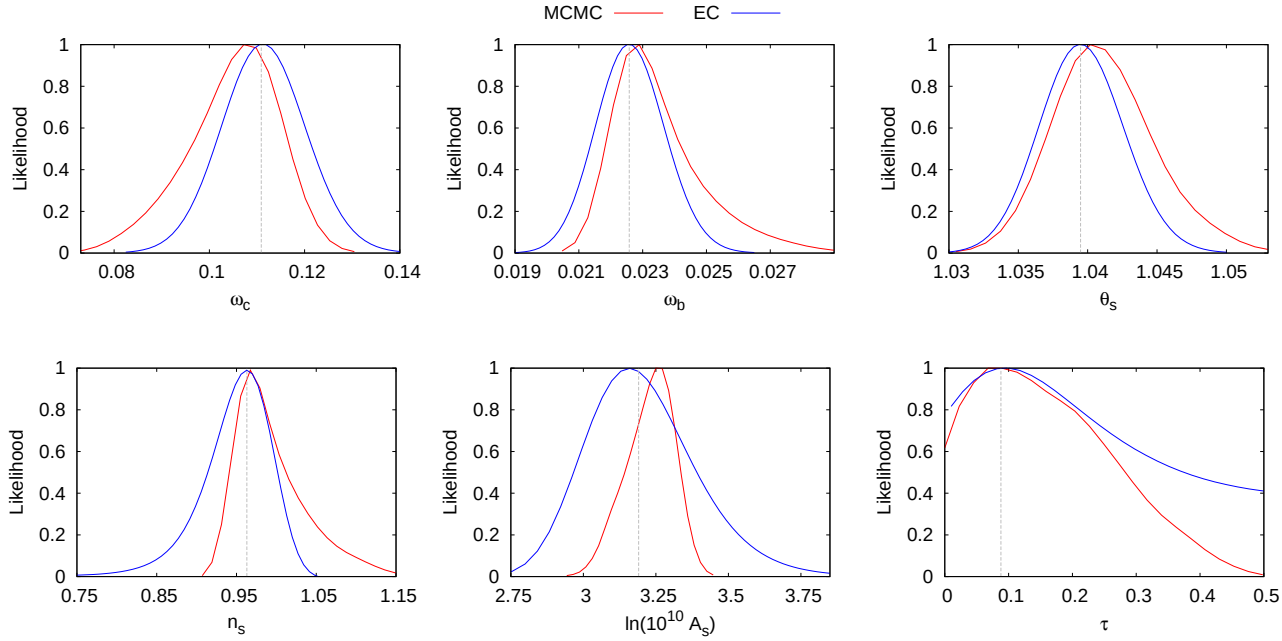

FIG. 9. Comparison between the recovered WMAP posteriors from a single compression vector in blue and the marginalized likelihoods from a MCMC analysis in red. The data used here is the exact theory temperature power spectrum C_l, with the fiducial cosmology set to WMAP seven-year best-fit parameter values (gray vertical lines). The constraints on the optical depth τ from the temperature spectrum C_l alone are weak, which is reflected in the wide likelihood distribution, although even in this case the recovered likelihood peaks at the fiducial value of $\tau = 0.088$. Since the parameter combination $A_s e^{-2\tau}$ determines the overall amplitude of the observed CMB anisotropy, the recovered value of $\ln(10^{10} A_s)$, the log power of the primordial curvature perturbations is slightly biased. Here, we reach the hard limit in the sampler inserted for the redshift of reionization $z_{\rm re} = 40$, which corresponds to $\tau \sim 0.6$. This signals that the temperature data alone does not constrain the full six parameter ΛCDM model very well.

C. How does extreme compression compare to a MCMC calculation?

Once we form the compression vectors, it is easy and very fast to compute the likelihood for each parameter, with a typical time of less than a minute. To test the method and to see how well we can recover the posterior probabilities, we first analyze a mock data set, where the observed data set is the set of theory C_l's. In this case, we expect an unbiased estimate of the input cosmology from both our method and the MCMC.

Since the spherical harmonic coefficients a_{lm} are Gaussian random variates and are statistically isotropic, the likelihood function for the temperature power spectrum is a Wishart distribution with $\mathcal{P}(\hat{\boldsymbol{C}}_l|\boldsymbol{C}_l) \propto \mathcal{L}(\boldsymbol{C}_l|\hat{\boldsymbol{C}}_l)$ and

$$\chi^2_{\rm eff} = -2\ln\mathcal{L}(\boldsymbol{C}_l|\hat{\boldsymbol{C}}_l) = \sum_{l=2}^{lmax}(2l+1)\left(\frac{\hat{\boldsymbol{C}}_l}{\boldsymbol{C}_l^{th}} + \ln\left(\frac{\boldsymbol{C}_l^{th}}{\hat{\boldsymbol{C}}_l}\right) - 1\right), \tag{33}$$

where $\hat{\boldsymbol{C}}_l$ is the observed data set [18–20]. The above likelihood is a general case for an experiment with no noise and a full-sky coverage. In practice, experiments have noise and observe only a fraction of the sky. We modify Eq. (33) by replacing $\boldsymbol{C}_l^{th}$ with $\boldsymbol{C}_l^{th} + \boldsymbol{N}_l$, and by decreasing the number of modes on the sky from $(2l+1)$ to $(2l+1)f_{\rm sky}$. Both $\boldsymbol{C}_l^{th}$ and $\boldsymbol{N}_l$ appear in the likelihood because they are both Gaussian random fields. Note that the likelihood above is normalized such that $\chi^2_{\rm eff} = 0$, when $\hat{\boldsymbol{C}}_l = \boldsymbol{C}_l^{th}$.

In our WMAP mock MCMC likelihood calculations we assume that the fraction of the remaining sky after applying the WMAP mask KQ85y7 is 78.3% [21]. When analyzing the WMAP seven-year data however, we use the sky fraction contained in the WMAP likelihood code, which varies with the multipole l.

We show our results in Fig. 9, where we plot the MCMC posteriors in solid red and the result using our compressed vectors in solid blue. Because the Thomson scattering optical depth due to reionization is not well constrained by the temperature spectrum alone, the MCMC posterior has a wide, non-Gaussian distribution and the 95% C.L. upper limit for τ is 0.36. The extreme compression formalism implicitly assumes Gaussian distributions for the parameters, so the τ distribution offers a nice test of the impact of the breakdown of this assumption on the full analysis. Figure 9 shows that the impact falls mainly on the parameter A_s with which τ is degenerate (recall that the amplitude of the pertur-

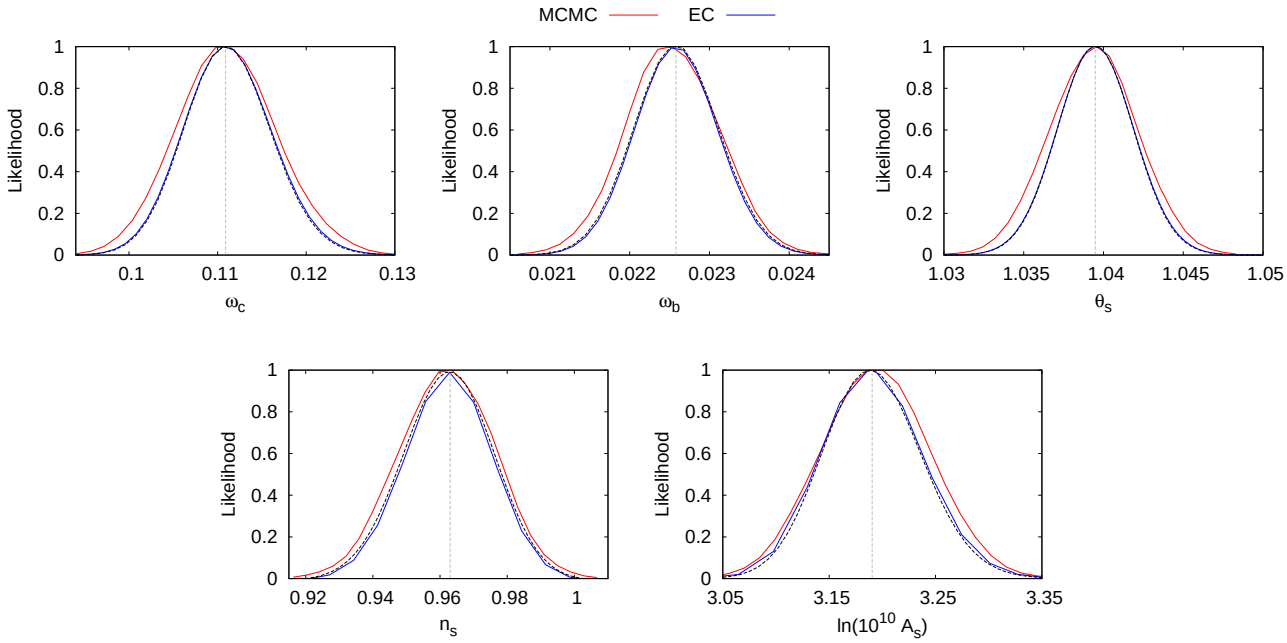

FIG. 10. Same as Fig. 9 but with τ held fixed. The posterior distributions from the MCMC (solid red lines) agree well with the distributions from the EC analysis (solid blue lines). We also plot the likelihood obtained with weighting vectors which are computed with a different fiducial cosmology (dashed black lines), and note the excellent agreement between the two compressions. To compute a second set of weighting vectors we use the following set of parameters: $\omega_c = 0.12$, $\omega_b = 0.0235$, $100 * \theta_s = 1.0485995$, $n_s = 0.98$, $\ln(10^{10} A_s) = 3.258$, and $\tau = 0.085$.

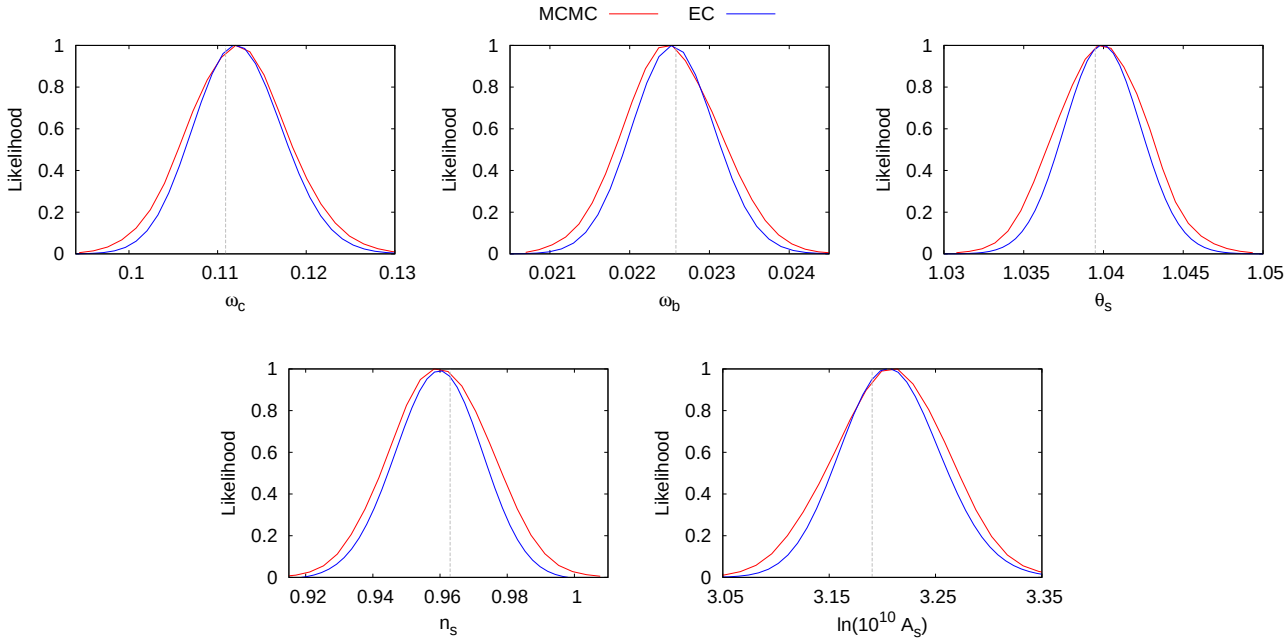

FIG. 11. Same as Fig. 10 but for a WMAP-like experiment where the CMB power spectrum data set is generated from a random realization of a fiducial cosmology. In Appendix B we show how we generate our random data set. In the case above, we do not expect the posterior distributions (solid blue or solid red lines) to peak at the fiducial parameter input values shown with dashed gray lines.

bations is roughly $A_s e^{-2\tau}$). The ensuing bias on A_s is small: relative to the mean μ from MCMC, the value of $\ln(10^{10} A_s)$ is biased low by 0.88σ, where the error on $\ln(10^{10} A_s)$ is $\sigma = 0.0814$. Note that in general the en-

TABLE I. Bias in the recovered values of $\ln(10^{10}A_s)$ using extreme compression relative to both the mean μ and the best fit (the maximum likelihood point $\theta_{\rm ML}$) from MCMC.

Model & Parameters	Bias relative to μ	Bias relative to $\theta_{\rm ML}$	Standard Deviation σ
TH τ free	-0.88σ	-0.43σ	0.0814
TH τ fixed	-0.07σ	0.15σ	0.0492
RN τ fixed	-0.05σ	0.20σ	0.0508

suing biases are smaller when the maximum likelihood is used, as opposed to the mean likelihood. In Table I, we show the bias on $\ln(10^{10}A_s)$ for exact theory C_l and a random catalog.

If we fix the optical depth to its fiducial value of $\tau = 0.088$, we obtain the results shown in Fig. 10, and then the likelihood results from the MCMC and EC are in very good agreement. In this case the MCMC means and the estimates from EC coincide with the input cosmology.

Figure 10 also illustrates that the EC method is insensitive to the choice of fiducial parameters. The dashed black curves show the likelihoods when the coefficients α_l^i are chosen assuming the nonfiducial parameter set: $\omega_c = 0.12$, $\omega_b = 0.0235$, $100 * \theta_s = 1.0485995$, $n_s = 0.98$, $\ln(10^{10}A_s) = 3.258$ and $\tau = 0.085$. The figure shows that shifts of this order leave no imprint on the final likelihood.

Before analyzing real data, we investigate how our method performs on a random mock. We create a realistic mock for a full-sky CMB experiment with WMAP noise. We discuss random mock generation in Appendix B. Figure 11 shows the posteriors in a ΛCDM model with τ fixed at its fiducial value. Again the two distributions agree very well.

In the next section we apply the methods discussed so far to the seven-year WMAP temperature spectrum, and compress the temperature spectrum to estimate the cosmological parameters with WMAP precision.

V. RESULTS

In the previous section we analyzed mock data to see how well we can recover the input cosmology, and we compared the results of the extreme compression to the MCMC means and best-fit (maximum likelihood) MCMC results. In this section we apply the methods to a real data set and as an example choose the seven-year WMAP temperature spectrum. Although this is not the most up-to-date CMB data set, it is a useful test which will inform further development of the EC method. For this analysis, we formulate the vectors that compress the WMAP spectrum using the same WMAP noise and fraction of the sky observed as in the WMAP likelihood. Since the WMAP likelihood is not a simple Gaussian, and consists of a number of components, we review the likelihood briefly in the next section. We discuss how this will affect our results in Sec. V B.

A. WMAP likelihood

The full WMAP likelihood is made up of ten components, four of which form part of the temperature analysis. The analysis is split up into low-l and high-l components. For multipoles $l \leq 32$, there is a choice between a direct evaluation of the likelihood in pixel space and one using Gibbs sampling (see [22] and the references therein). The default is Gibbs sampling, where the spectrum is obtained using a Blackwell-Rao estimator applied to a chain of Gibbs samples. For multipoles $l \geq 33$, the likelihood uses the spectrum derived from the MASTER pseudo-C_l quadratic estimator and a covariance matrix [23, 24]. In addition, there are terms in the likelihood due to uncertainty in determining the WMAP beam and the error in the extragalactic point source removal (for details see the appendix of [24]).

For a large l, Eq. (33) can be approximated as Gaussian $\ln \mathcal{L}_{\rm Gauss}$, but since the likelihood function for the power spectrum is slightly non-Gaussian, this gives a biased estimator. Although [18] suggest using a log-normal distribution $\mathcal{L}_{\rm LN}$, both the Gaussian and the log-normal distributions are found to be biased estimators for WMAP [25]. The approximation for the C_l likelihood used in the WMAP analysis, consists of a Gaussian and a log-normal distribution, where

$$\ln \mathcal{L} = \frac{1}{3} \ln \mathcal{L}_{\rm Gauss} + \frac{2}{3} \ln \mathcal{L}'_{\rm LN} \ . \tag{34}$$

Clearly the likelihood in the real analysis is not trivial and since we do not account for such corrections, we expect that our results will differ from those obtained with MCMC. An interesting question is by how much? How well does a simple method fare against the full, more complex likelihood? We explore these questions in the next section.

B. Analyzing WMAP seven-year data

We analyze the WMAP seven-year temperature power spectrum, using the vectors shown in solid red, in Fig. 6. This analysis differs slightly from those in previous sections, in that here we use the sky fraction contained in the WMAP likelihood, which varies with l, rather than a fixed value of $f_{\rm sky} = 0.783$. The spectrum range included in the analysis is $2 - 1200$, and we neglect the effect of lensing on the CMB. We fix the Sunyaev-Zel'dovich (SZ) amplitude parameter in the MCMC, and we hold the helium fraction constant and equal to $Y_{\rm He} = 0.24$.

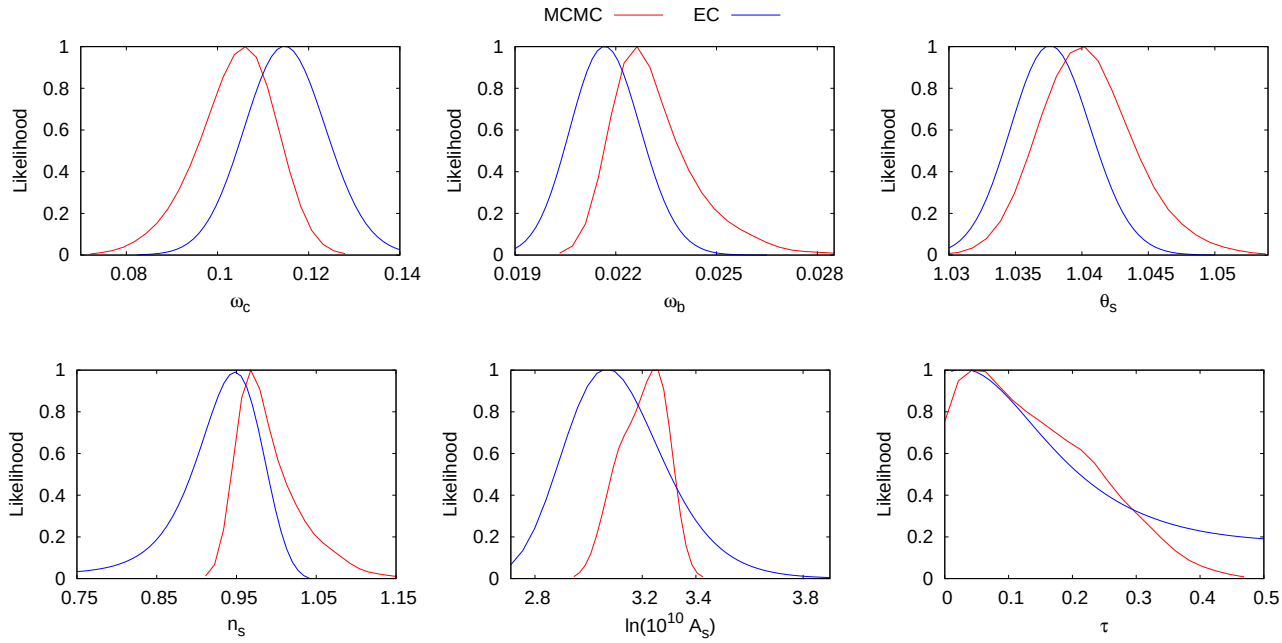

FIG. 12. Results of the EC analysis on the WMAP seven-year temperature data only (solid blue lines) and the full likelihood evaluations with MCMC (solid red lines) in a ΛCDM model. Since the temperature data is not constraining enough to measure τ, the posterior for the optical depth is wide, resulting in a biased result for $\ln(10^{10}A_s)$.

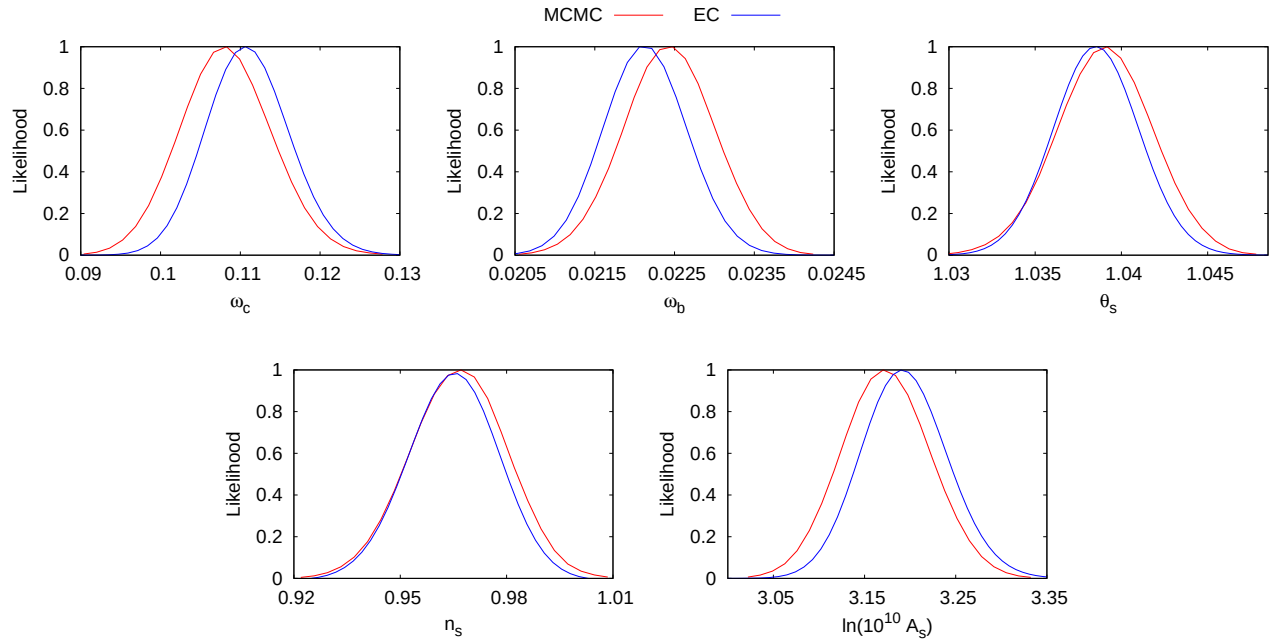

FIG. 13. Same as Fig. 12 but with τ held fixed. Although the distributions are a closer match, there is still some residual bias in the estimated parameters. We quote the bias on each parameter in Table II.

1. *ΛCDM including optical depth τ*

In Fig. 12, we compare the results from extreme compression with MCMC assuming the WMAP likelihood in Eq. (34). As we showed in Fig. 9, we do not expect that the posteriors from both methods will agree exactly, in part because of the degeneracies due to poor constraints on the optical depth τ. We also do not expect to obtain parameter estimates equal to those of the base WMAP+SZ+LENS model, since we do not include po-

larization data. In this sense, we are using compression vectors without assuming a "correct" fiducial model (as was done in Sec. IV, Figs. 9 and 10). Further, we saw when analyzing mock data that the non-Gaussianity of the τ likelihood leads to a bias in A_s in the EC method. Nonetheless, the biases shown in Fig. 12 are still relatively small, with those estimated from the maximum of the likelihood significantly less than the statistical error. We show the bias between the EC method and the MCMC results in Table II, where we calculate the difference between the peak in the EC likelihood and the MCMC mean μ and the best-fit ($\theta_{\rm ML}$) point, relative to the standard deviation σ from MCMC.

2. *ΛCDM and fixed optical depth τ*

In Fig. 13 we show constraints from the compressed data set and MCMC results using the entire WMAP CMB temperature anisotropy power spectrum. The agreement is best for n_s and θ_s, with the other parameters experiencing a bias of less than $\sim 0.5\sigma$. We show the results from the EC method and any bias in determining the posterior mean and the maximum likelihood (ML) point in Table II. As pointed out in Sec. V A, the likelihood used in the full WMAP analysis is not a simple Gaussian. In addition, we do not take into account in our compression method the intricacies involved with beam corrections and point source subtraction. Neither do we account for non-Gaussianity of the data at the lowest multipoles. The fact that WMAP uses Gibbs sampling for the lowest multipoles also means that our results will not be the same. Crucially, if we modify the code to either model the likelihood as a full Gaussian, by discarding log-normal part in Eq. (34) or do not use Gibbs sampling and restrict the analysis to modes with $l > 30$, the resulting shifts in each of the parameter posteriors cause much larger differences than the ones quoted above. So, the biases introduced in the EC method are smaller than those that emanate from much milder assumptions about the likelihood.

VI. CONCLUSION

We have shown that a locally lossless extreme compression of modern CMB data sets gains significant speedup in the computation of marginalized likelihoods in ΛCDM models. By requiring that the Fisher information matrix is unchanged, we derived the weighting vectors for the CMB that can estimate cosmological parameters in less than a minute, much faster than MCMC. The method requires computations of the likelihood for one parameter at a time, instead of having to explore the whole parameter space with MCMC. We therefore achieve extreme data compression by (i) compressing the entire data set into just a few numbers, and (ii) reducing the dimensionality of the parameter space that needs to be explored.

The compression vectors for the CMB are also very useful since their shape and amplitude provide an intuitive feel for the physics of the CMB, the sensitivity of the observed spectrum to cosmological parameters. They can also inform about the relative sensitivity of different experiments to cosmological parameters.

We have tested our method on exact theory C_l as well as on a WMAP-like CMB data set generated from a random realization of a fiducial cosmology. By comparing our results to those from full likelihood analyses using CosmoMC, we have been able to show that the method

TABLE II. Bias in the recovered parameter values using extreme compression relative to the MCMC results. We compare the parameter value at the peak of the likelihood obtained with the EC method, to both the mean μ and the best fit (the maximum likelihood point $\theta_{\rm ML}$) from MCMC. The fourth column, is the standard deviation σ of the MCMC samples.

Model & Parameters	Bias relative to μ	Bias relative to $\theta_{\rm ML}$	Standard Deviation σ
Theory τ free			
ω_c	0.67σ	0.01σ	0.0095
ω_b	-0.66σ	0.09σ	0.0014
θ_s	-0.48σ	0.03σ	0.0036
n_s	-0.71σ	0.04σ	0.0460
$\ln(10^{10}A_s)$	-0.88σ	-0.43σ	0.0814
τ	-0.69σ	0.08σ	0.1033
Theory τ fixed			
ω_c	-0.02σ	0.10σ	0.0054
ω_b	0.06σ	-0.17σ	0.0006
θ_s	0.04σ	-0.10σ	0.0027
n_s	0.08σ	-0.12σ	0.0137
$\ln(10^{10}A_s)$	-0.07σ	0.15σ	0.0492
Random τ fixed			
ω_c	-0.04σ	0.24σ	0.0054
ω_b	0.00σ	-0.18σ	0.0006
θ_s	0.01σ	-0.15σ	0.0027
n_s	-0.03σ	-0.21σ	0.0142
$\ln(10^{10}A_s)$	-0.05σ	0.20σ	0.0508
Data τ free			
ω_c	1.29σ	0.35σ	0.0084
ω_b	-1.19σ	-0.43σ	0.0012
θ_s	-0.82σ	-0.44σ	0.0035
n_s	-1.09σ	-0.25σ	0.0399
$\ln(10^{10}A_s)$	-1.56σ	0.06σ	0.0858
τ	-1.21σ	0.16σ	0.0972
Data τ fixed			
ω_c	0.45σ	0.40σ	0.0055
ω_b	-0.56σ	-0.63σ	0.0006
θ_s	-0.19σ	-0.19σ	0.0027
n_s	-0.08σ	-0.09σ	0.0131
$\ln(10^{10}A_s)$	-0.37σ	0.34σ	0.0470

performs very well, and is able to recover the maximum likelihood estimates for parameters even if the posterior is not Gaussian. If the posterior is Gaussian, then the extreme compression method can recover the posterior means to better than 0.1σ.

We have applied the compression method to the temperature power spectrum from the WMAP seven-year data release, and have found that even though the likelihood for WMAP is nontrivial and non-Gaussian, our method is in good agreement with the posteriors from a full MCMC analysis. The biases in our estimates of cosmological parameters, compared to the mean are: ω_b bias is -0.56σ, ω_c bias is 0.45σ, θ_s bias is -0.19σ, n_s is -0.08σ, A_s is -0.37σ. The biases relative to the best fit (the maximum likelihood) are comparable.

Furthermore, given the nontrivial nature of the likelihood, it is possible that the method may also work well with newer data and a more complicated Bayesian analysis, e.g., the Planck likelihood. We will address this in a future investigation.

Additionally as a bonus, including polarization data and extending the parameter space is not going to increase the computational costs. The vectors can be precomputed and stored, and the calculation of the likelihood is limited only by the speed of one call to CAMB, times the number of samples we wish to obtain. The increase in parameter space, can be accommodated by running each compression separately, one after another, or at the same time using n nodes. In this case, the time for the likelihood computation for the entire parameter space is no longer than a computation for a single parameter, which takes less than a minute

ACKNOWLEDGMENTS

The authors would like to thank the anonymous referee for the valuable comments on the manuscript. A.Z. would like to thank Wayne Hu for the useful discussions. This work was completed in part with resources provided by the University of Chicago Research Computing Center as well as the Joint Fermilab - KICP Supercomputing Cluster, supported by grants from Fermilab, the Kavli Institute for Cosmological Physics, and the University of Chicago. A.Z. acknowledges support from KICP, the Brinson Foundation and U.S. Department of Energy Contract No. DE-FG02-13ER41958. The work of S.D. is supported by the U.S. Department of Energy, including Grant No. DE-FG02-95ER40896.

Appendix A: CHOOSING THE RIGHT PARAMETRIZATION IN A MODEL

If there exist known degeneracies in the data, e.g., the geometric degeneracy in the CMB, then the choice of parametrization will matter. For the CMB, we find that a bad parametrization may have an adverse effect on the compression and therefore the recovered posterior distributions may be non-Gaussian and/or multimodal. In the specific case of the CMB, we found that using Ω_Λ instead of θ_s results in a bimodal distribution for Ω_Λ with all other parameters not affected (that is, their posteriors were all correct). The root of the problem can be seen in Fig. 14, where we plot the geometric degeneracy between H_0 and Ω_Λ. The color coding shows various values of $100\,\theta_s$. The optimal parameter vector is $\Theta = \{\omega_c, \omega_b, 100\theta_s, n_s, A_s, \tau\}$.

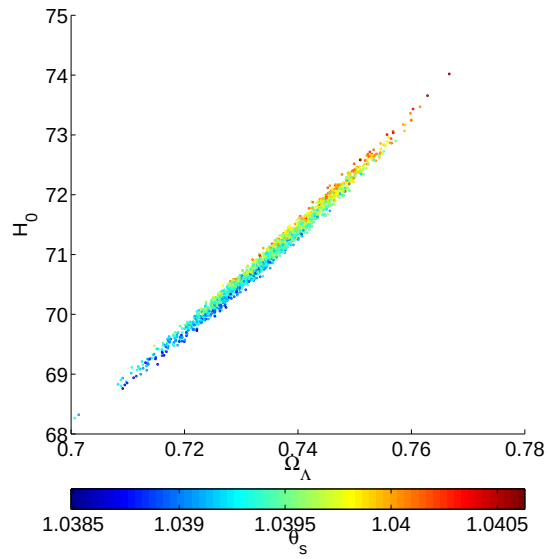

FIG. 14. The extent of the geometric degeneracy in MCMC samples between the cosmological constant density parameter Ω_Λ and the Hubble expansion rate H_0. The color scale shows the values of $100\theta_{\rm MC}$, an approximation for $r_s(z_\star)/D_A(z_\star)$, the angular scale of the sound horizon at last scattering. The data used in this case were exact theory C_l.

Appendix B: CMB DATA GENERATION

We generate two kinds of data sets using the Boltzmann code CAMB, computing the temperature power spectrum C_l up to $l = 1200$. For the first data set (referred to as exact theory C_l), we assume white isotropic noise and Gaussian beams, and add the noise N_l given by Eq. (14) to C_l. In the MCMC analysis, we use the likelihood in Eq. (33) to get parameter constraints. This is because the likelihood is a function of $C_l + N_l$ and not just C_l [see Eq. (33)]. The EC calculation assumes the data is exact theory C_l, with the noise N_l included in the covariance in Eq. (13).

The second data set that we use in our analysis makes use of a random realization of the underlying theory C_l. To create a random mock data set we generate four sets of Gaussian random deviates a, b, c and d, with $\mu = 0$ and $\sigma^2 = 1$. We use these random deviates to create two complex Gaussian fields, $g_{lm} = \frac{1}{\sqrt{2}}(a + \imath b)$ and

$h_{lm} = \frac{1}{\sqrt{2}}(c + \imath d)$ with $\langle h^*_{lm} h_{lm} \rangle = 1$ and $\langle g^*_{lm} g_{lm} \rangle = 1$. For completeness, we include the generation of both a^T_{lm} and a^E_{lm}, such that $C^{TT}_l C^{EE}_l - (C^{TE}_l)^2 > 0$ and $a^{*XX}_{l,m} = (-1)^m a^{XX}_{l,-m}$. The spherical harmonic coefficients for temperature are

$$a^T_{lm} = \sqrt{C^{TT}_l + N^{TT}_l}\ g_{lm} \tag{B1}$$

and the polarization coefficients are given by

$$a^E_{lm} = \frac{C^{TE}_l}{C^{TT}_l + N^{TT}_l} \sqrt{C^{TT}_l + N_l}\ g_{lm} + h_{lm} \sqrt{(C^{EE}_l + N^{EE}_l) - \frac{(C^{TE}_l)^2}{(C^{TT}_l + N^{TT}_l)}}. \tag{B2}$$

The random mock can then be generated using the full-sky power spectra estimators for the temperature, the E-mode polarization, and the cross spectrum between the temperature and the E-mode polarization given by

$$\hat{C}^{TT}_l = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \langle a^{*TT}_{lm} a^{TT}_{lm} \rangle \tag{B3}$$

$$\hat{C}^{EE}_l = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \langle a^{*EE}_{lm} a^{EE}_{lm} \rangle \tag{B4}$$

$$\hat{C}^{TE}_l = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \langle a^{*TT}_{lm} a^{TE}_{lm} \rangle. \tag{B5}$$

We have tested this prescription using MCMC, and find that on average seven out of ten times the estimate of θ_i is within 1σ of the fiducial input value.

[1] M. Davis and P. J. E. Peebles, ApJ **267**, 465 (1983).
[2] U. Seljak and E. Bertschinger, ApJL **417**, L9 (1993), astro-ph/9309003.
[3] K. M. Gorski, ApJL **430**, L85 (1994), astro-ph/9403066.
[4] J. R. Bond, R. Crittenden, R. L. Davis, G. Efstathiou, and P. J. Steinhardt, Physical Review Letters **72**, 13 (1994), astro-ph/9309041.
[5] M. S. Vogeley and A. S. Szalay, ApJ **465**, 34 (1996), astro-ph/9601185.
[6] M. Tegmark, A. N. Taylor, and A. F. Heavens, ApJ **480**, 22 (1997), astro-ph/9603021.
[7] A. F. Heavens, R. Jimenez, and O. Lahav, MNRAS **317**, 965 (2000), astro-ph/9911102.
[8] C. Reichardt, R. Jimenez, and A. F. Heavens, MNRAS **327**, 849 (2001), astro-ph/0101074.
[9] S. Gupta and A. F. Heavens, MNRAS **334**, 167 (2002), astro-ph/0108315.
[10] P. Protopapas, R. Jimenez, and C. Alcock, MNRAS **362**, 460 (2005), astro-ph/0502301.
[11] P. Graff, M. P. Hobson, and A. Lasenby, MNRAS **413**, L66 (2011), arXiv:1010.5907 [astro-ph.IM].
[12] E. Gjerløw, L. P. L. Colombo, H. K. Eriksen, K. M. Górski, A. Gruppuso, J. B. Jewell, S. Plaszczynski, and I. K. Wehus, ApJS **221**, 5 (2015), arXiv:1506.04273 [astro-ph.IM].
[13] Y. Wang and S. Wang, Phys.Rev.D **88**, 043522 (2013), arXiv:1304.4514 [astro-ph.CO].
[14] M. Chu, M. Kaplinghat, and L. Knox, ApJ **596**, 725 (2003), astro-ph/0212466.
[15] A. Lewis, A. Challinor, and A. Lasenby, Astrophys. J. **538**, 473 (2000), astro-ph/9911177.
[16] L. Knox, Phys.Rev.D **52**, 4307 (1995), astro-ph/9504054.
[17] R. Allison, P. Caucal, E. Calabrese, J. Dunkley, and T. Louis, ArXiv e-prints (2015), arXiv:1509.07471.
[18] J. R. Bond, A. H. Jaffe, and L. Knox, ApJ **533**, 19 (2000), astro-ph/9808264.
[19] S. Hamimeche and A. Lewis, Phys.Rev.D **77**, 103013 (2008), arXiv:0801.0554.
[20] S. Hamimeche and A. Lewis, Phys.Rev.D **79**, 083012 (2009), arXiv:0902.0674 [astro-ph.CO].
[21] D. Larson, J. Dunkley, G. Hinshaw, E. Komatsu, M. R. Nolta, C. L. Bennett, B. Gold, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright, ApJS **192**, 16 (2011), arXiv:1001.4635 [astro-ph.CO].
[22] J. Dunkley, E. Komatsu, M. R. Nolta, D. N. Spergel, D. Larson, G. Hinshaw, L. Page, C. L. Bennett, B. Gold, N. Jarosik, J. L. Weiland, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, E. Wollack, and E. L. Wright, ApJS **180**, 306 (2009), arXiv:0803.0586.
[23] E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen, ApJ **567**, 2 (2002), astro-ph/0105302.
[24] G. Hinshaw, M. R. Nolta, C. L. Bennett, R. Bean, O. Doré, M. R. Greason, M. Halpern, R. S. Hill, N. Jarosik, A. Kogut, E. Komatsu, M. Limon, N. Odegard, S. S. Meyer, L. Page, H. V. Peiris, D. N. Spergel, G. S. Tucker, L. Verde, J. L. Weiland, E. Wollack, and E. L. Wright, ApJS **170**, 288 (2007), astro-ph/0603451.
[25] L. Verde, H. V. Peiris, D. N. Spergel, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, E. Wollack, and E. L. Wright, ApJS **148**, 195 (2003), astro-ph/0302218.