CMB lensing power spectrum without noise bias

Delon Shen[1,2,3,*] Emmanuel Schaan[2,3] and Simone Ferraro[4,5]

[1]*Department of Physics, Stanford University, Stanford, California 94305-4085, USA*
[2]*Kavli Institute for Particle Astrophysics and Cosmology,*
382 Via Pueblo Mall, Stanford, California 94305-4060, USA
[3]*SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA*
[4]*Lawrence Berkeley National Laboratory, Physics Division, Berkeley, California 94720, USA*
[5]*Berkeley Center for Cosmological Physics, Department of Physics, University of California, Berkeley,*
California 94720, USA

Upcoming surveys will measure the cosmic microwave background (CMB) weak lensing power spectrum in exquisite detail, allowing for strong constraints on the sum of neutrino masses among other cosmological parameters. Standard CMB lensing power spectrum estimators aim to extract the connected non-Gaussian trispectrum of CMB temperature maps. However, they are generically dominated by a large disconnected, or Gaussian, noise bias, which thus needs to be subtracted at high accuracy. This is currently done with realistic map simulations of the CMB and noise, whose finite accuracy currently limits our ability to recover CMB lensing on small scales. In this paper, we propose a novel estimator which instead avoids this large Gaussian bias. This estimator relies only on the data and avoids the need for bias subtraction with simulations. Thus, our bias avoidance method is (1) insensitive to misestimates in simulated CMB and noise models and (2) avoids the large computational cost of standard simulation-based methods like "realization-dependent $N^{(0)}$" (RDN$^{(0)}$). We show that our estimator is as robust as standard methods in the presence of realistic inhomogeneous noise (e.g., from scan strategy) and masking. Moreover, our method can be combined with split-based methods, making it completely insensitive to mode coupling from inhomogeneous atmospheric and detector noise. We derive the corresponding expressions for our estimator when estimating lensing from CMB temperature and polarization. Although we specifically consider CMB weak lensing power spectrum estimation in this paper, we illuminate the relation between our new estimator, RDN$^{(0)}$ subtraction, and general optimal trispectrum estimation. Through this discussion, we conclude that our estimator is applicable to analogous problems in other fields that rely on estimating connected trispectra/four-point functions like large-scale structure.

I. INTRODUCTION

Cosmic microwave background (CMB) photons are gravitationally lensed by large-scale structure as they propagate through the Universe. This lensing distorts our images of CMB anisotropies and imprints onto the CMB four-point correlation function[1] a distinct non-Gaussian component [1]. Measurements of this non-Gaussian component, the CMB weak lensing power spectrum $\langle \kappa\kappa \rangle$, provide us with a wealth of information on the growth of structures in our Universe. Thus, a measurement of $\langle \kappa\kappa \rangle$ can provide us with some of the strongest constraints on the properties of neutrinos [2], primordial non-Gaussianity [3], dark matter [4], and dark energy [5]. The first detection of CMB lensing was reported with data from the Wilkinson

Microwave Anisotropy Probe in Ref. [6] and CMB lensing power spectrum in Refs. [5,7]. Since then, many additional detections of CMB lensing spectra have been reported [8–17].

In order to robustly and accurately estimate $\langle \kappa\kappa \rangle$, we must navigate the problem of extracting a non-Gaussian signal, the connected trispectrum, from the four-point function, which is generically dominated by a Gaussian bias $N^{(0)}$ that is typically several orders of magnitude larger than the non-Gaussian signal. In typical analyses, this $N^{(0)}$ is estimated from realistic map simulations of the CMB and noise and then subtracted off. However, since $N^{(0)}$ is orders of magnitude larger than our signal CMB lensing power spectrum, even a 0.1% misestimate of $N^{(0)}$ can lead to a 10% bias in our estimate of the lensing power spectrum on small scales. The level of accuracy needed for robust estimates of $\langle \kappa\kappa \rangle$ on small scales is difficult to achieve with traditional methods since any mismodeling of the

---

[*]Contact author: delon@stanford.edu
[1]In what follows, we use *n*-point correlation function and *n*-point function interchangeably.

maps that are input to traditional methods propagates to errors in estimates of $N^{(0)}$. Two particular concerns are (1) that the complex and inhomogeneous noise structure arising from ground-based CMB experiments is difficult to simulate accurately and (2) uncertainties on the cosmological parameters, both of which can bias our estimates of $\langle \kappa\kappa \rangle$.

In this paper, we propose an estimator of the CMB weak lensing power spectrum that avoids the large $N^{(0)}$ Gaussian bias. This is achieved by examining exactly which terms contribute to the $N^{(0)}$ bias and constructing a scheme that estimates $\langle \kappa\kappa \rangle$ while avoiding these terms. It turns out that this leads to a negligible reduction in signal-to-noise ratio. In addition, this method relies only on data, completely avoiding the need for simulations to avoid $N^{(0)}$ making it (1) insensitive to inaccuracies in the modeling of the CMB and noise and (2) significantly more computationally efficient compared to standard realization-dependent $N^{(0)}$ (RDN$^{(0)}$) methods. We also demonstrate that this estimator is as robust as standard methods in the presence of realistic inhomogeneous noise, such as the noise pattern that may arise from scan strategy, and masking. In addition, we describe how our estimator may be combined with split-based methods proposed in Ref. [18] which make it insensitive to spurious mode couplings from inhomogeneous atmospheric and detector noise. This new estimator was hinted at in Ref. [19] but not implemented or studied there. Instead they focused on an alternative estimator. Crucially, this estimator can be implemented efficiently (at least in the flat-sky limit) with the use of fast Fourier transforms (FFTs), often required in a real analysis.

The rest of this paper is organized as follows. In Sec. II, we briefly review how standard CMB lensing and lensing spectrum estimation is done. In Sec. III, we outline our proposed estimator for the lensing power spectrum which avoids $N^{(0)}$ bias. Following this, we compare our method with the standard RDN$^{(0)}$ subtraction method in Sec. IV. In Sec. V, we comment on the effect of $N^{(0)}$ on the covariance of CMB lensing spectra, as well as how our method and the standard RDN$^{(0)}$ prescription for handling $N^{(0)}$ removes covariances introduced by $N^{(0)}$. We explore an illuminating toy model for optimal trispectrum estimating first presented in Ref. [20] in Sec. VI, which allows us to see how our estimator, the standard RDN$^{(0)}$ estimator, and optimal trispectrum estimation are related and builds intuition for the various features of each method. In Sec. VII, we study the robustness of our proposed estimator in the presence of realistic anisotropic noise which leads to additional complications in CMB weak lensing power spectrum estimation. This motivates a discussion of how one might combine our method with a split-based method in Sec. VIII. In Sec. IX, we study the robustness of our proposed estimator in the presence of masking, which also

leads to additional complications in CMB weak lensing power spectrum estimation. Throughout this paper, we focus mostly on estimating CMB lensing using only temperature anisotropies, but in Sec. X we spell out the general form of our proposed estimator when estimating CMB lensing using both temperature and polarization anisotropies. Finally, we conclude in Sec. XI.

II. STANDARD CMB LENSING POWER SPECTRUM ESTIMATION

The effect of lensing on the CMB temperature maps at the position x can be expressed as

$$T(x) = T^0(x + d(x)) = T^0(x) + d \cdot \nabla T^0(x) + O(d^2). \quad (1)$$

One can then do a Helmholtz decomposition of the displacement field $d = \nabla \psi$, where we have neglected the field rotation component since it is negligible for current sensitivities [21]. From this we can also define the convergence $\kappa \equiv -\nabla^2 \psi/2$. Throughout this paper, we will refer to lensing potential ψ and convergence κ interchangeably but will primarily work with the lensing convergence κ. With this we can rewrite Eq. (1) in Fourier space,

$$T_{\ell} = T^0_{\ell} - \int \frac{d^2\ell'}{(2\pi)^2} \ell' \cdot (\ell - \ell') \frac{2\kappa_{\ell - \ell'}}{(\ell - \ell')^2} T^0_{\ell'} + O(\kappa^2)$$

$$\equiv T^{(0)}_{\ell} + T^{(1)}_{\ell} + O(\kappa^2). \quad (2)$$

From statistical homogeneity and isotropy one can show that for the unlensed CMB

$$\langle T^0_{\ell_1} T^0_{\ell_2} \rangle = (2\pi)^2 \delta^{(D)}(\ell_1 + \ell_2) C^{TT}_{\ell_1}. \quad (3)$$

At fixed realization of the lensing convergence κ_L, lensing introduces nontrivial off-diagonal correlations

$$\langle T_{\ell_1} T_{\ell_2} \rangle = (2\pi)^2 \delta(\ell_1 + \ell_2) C^{TT}_{\ell_1}$$
$$+ \kappa_{\ell_1 + \ell_2} f^{\kappa}_{\ell_1, \ell_2} + O(\kappa^2), \quad (4)$$
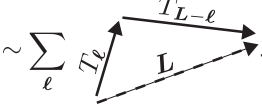
where we defined $f^{\kappa}_{\ell_1, \ell_2}$ in the second line with the relation

$$f^{\kappa}_{\ell_1, \ell_2} \kappa_{\ell_1 + \ell_2} \equiv \langle T^{(0)}_{\ell_1} T^{(1)}_{\ell_2} \rangle + \langle T^{(1)}_{\ell_1} T^{(0)}_{\ell_2} \rangle$$
$$\Rightarrow f^{\kappa}_{\ell_1, \ell_2} = \frac{2(\ell_1 + \ell_2)}{(\ell_1 + \ell_2)^2} \cdot [\ell_1 C^{TT}_{\ell_1} + \ell_2 C^{TT}_{\ell_2}]. \quad (5)$$

This means that CMB lensing breaks statistical homogeneity and isotropy. It is particularly useful to consider

$$\langle T_{\ell} T_{L-\ell} \rangle = f^{\kappa}_{\ell, L-\ell} \kappa_L + \mathcal{O}(\kappa^2). \quad (6)$$

From this we can intuit a quadratic estimator for the lensing potential

$$\hat{\kappa}_{\boldsymbol{L}} = N_L^\kappa \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^\kappa T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}}$$

$$\sim \sum_{\boldsymbol{\ell}} \quad . \qquad (7)$$

The normalization for this estimator is derived from asserting the estimator is unbiased, $\langle\hat{\kappa}_{\boldsymbol{L}}\rangle = \kappa_{\boldsymbol{L}}$, which leads to

$$(N_L^\kappa)^{-1} = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^\kappa f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^\kappa. \qquad (8)$$

Let $\tilde{C}_\ell^{TT}$ be the "total observed" temperature power spectrum, i.e., the power spectra of the "lensed" temperature field which includes instrument noise and foregrounds. Expanding Eq. (7) to linear order in κ allows one to derive the minimum-variance weights at this order,

$$F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^\kappa = \frac{f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^\kappa}{2\tilde{C}_\ell^{TT}\tilde{C}_{|\boldsymbol{L}-\boldsymbol{\ell}|}^{TT}}. \qquad (9)$$

Intuitively, it makes sense that we need two powers of the temperature field to build an estimator. This is because we assume the CMB is isotropic and homogeneous when emitted; therefore, independent modes of T_ℓ^0 are statistically independent. However, as we saw in Eq. (4), weak lensing of CMB introduces correlations between previously independent modes. Thus, it should be possible to extract the lensing potential that causes this lensing by considering pairs of modes $T_\ell, T_{\boldsymbol{L}-\boldsymbol{\ell}}$ and seeing what kind of correlations we have measured.

Using this estimator for the lensing potential $\kappa_{\boldsymbol{L}}$ we can now study the power spectrum of the lensing potential by first considering the two-point function of our estimator for the lensing potential

$$\langle\hat{\kappa}_{\boldsymbol{L}}\hat{\kappa}_{\boldsymbol{L}}^*\rangle = (N_L^\kappa)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F_{\boldsymbol{\ell}_1,\boldsymbol{L}-\boldsymbol{\ell}_1}^\kappa F_{-\boldsymbol{\ell}_2,-\boldsymbol{L}+\boldsymbol{\ell}_2}^\kappa$$
$$\times \langle T_{\boldsymbol{\ell}_1} T_{\boldsymbol{L}-\boldsymbol{\ell}_1} T_{-\boldsymbol{\ell}_2} T_{-\boldsymbol{L}+\boldsymbol{\ell}_2}\rangle. \qquad (10)$$

This estimate for the lensing power spectrum has several significant noise biases which can helpfully be denoted by $N^{(i)}$ biases.

A. $N^{(i)}$ noise biases

There are several noise biases hindering our estimate of the true κ power spectrum:

(i) $N^{(0)}$: large disconnected Gaussian noise bias. This bias is expected even for a Gaussian unlensed CMB map or Gaussian noise and foregrounds.

(ii) $N^{(1)}$: bias arising from an integral over one power of the lensing power spectrum $\iint\langle\kappa^2\rangle$.

In general, we have

(iii) $N^{(i)}$: bias arising from an integral over the ith power of the lensing power spectrum.

We spell out biases of these type in more detail within Appendix B. Moreover, taking into account the non-Gaussianity of the lensing field, there exists a nonzero bispectrum that allows contraction that leads to other noise biases:

(iv) $N^{(i/2)}$: bias arising from intrinsic non-Gaussianity in the lensing potential, involving the ith power of the lensing field [22–25].

The standard way to estimate the Gaussian bias $N^{(0)}$ is to compute what is called a realization-dependent $N^{(0)}$, RDN$^{(0)}$, Monte Carlo (MC) correction as described in Refs. [26,27]. This estimator is constructed to be robust to misestimates of the underlying total power spectrum in comparison with a naive method where one subtracts $N_{\text{theory}} \equiv N^\kappa$ from Eq. (8). We expand on RDN$^{(0)}$ in Sec. IV and the N_{theory} subtraction in Appendix D.

We can also estimate the first relevant higher order $N^{(1)}$ noise bias in two ways: analytic computation of the $N^{(1)}$ bias as described in Refs. [28–31], which we review in Appendix B, or numerical computation from simulations as described in Ref. [16], which we will discuss in Appendix H. In our numerical studies, we utilize both of these methods to estimate $N^{(1)}$, as we will discuss in Appendix H as well.

Higher order biases such as the $N^{(2)}$ bias can still lead to significant misestimates of the lensing power spectrum at low $\boldsymbol{L}$ [29]. However, in Ref. [32], it is shown that one can do a nonperturbative treatment of the lensing potential power spectrum and derive that replacing C_ℓ^{TT} with $C_\ell^{T\nabla T}$ in Eq. (5) significantly reduces the $N^{(2)}$ bias,

$$f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^\kappa = \frac{2\boldsymbol{L}}{L^2}\cdot\left[\boldsymbol{\ell} C_\ell^{T\nabla T} + (\boldsymbol{L}-\boldsymbol{\ell})C_{|\boldsymbol{L}-\boldsymbol{\ell}|}^{T\nabla T}\right]. \qquad (11)$$

We make use of these $C_L^{T\nabla T}$ weights in this paper for our numerical studies.

III. EXACT NOISE BIAS AVOIDANCE

Our proposed method to avoid the Gaussian $N^{(0)}$ noise bias, outlined in Fig. 1, boils down to isolating and subsequently discarding the set of $\{\boldsymbol{\ell}_1,\boldsymbol{\ell}_2\}$ in Eq. (10) which contains the entire $N^{(0)}$ Gaussian noise bias. Formally, this is a measure zero set and, in practice, means ignoring N_{modes} out of the N_{modes}^2 terms $\langle T_{\boldsymbol{\ell}_1} T_{\boldsymbol{L}-\boldsymbol{\ell}_1} T_{\boldsymbol{\ell}_2} T_{-\boldsymbol{L}-\boldsymbol{\ell}_2}\rangle$ used to estimate the lensing power spectrum.
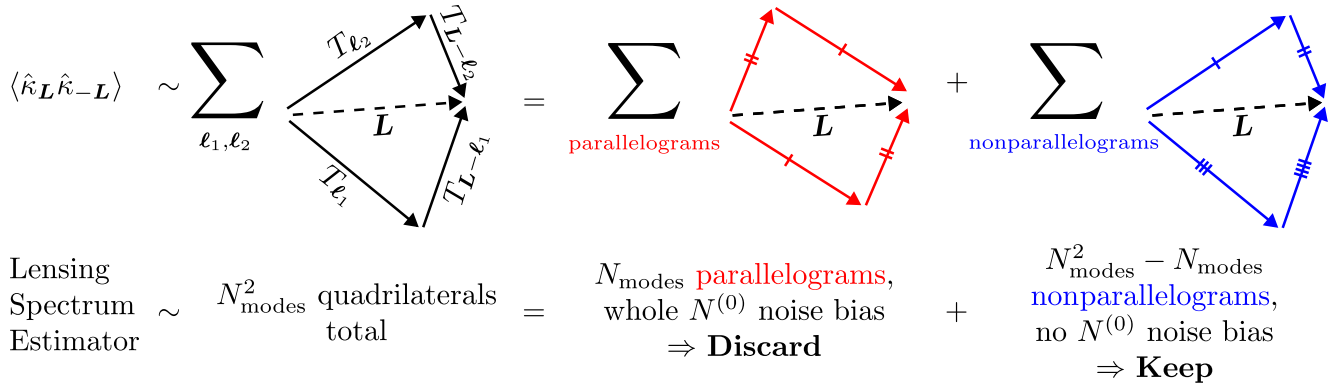
Noise Bias Avoidance Principle

FIG. 1. A visualization of our proposed $\hat{N}$ subtraction method. One can schematically think of the estimator for the CMB lensing power spectrum (10) as a sum over quadrilaterals where each side of the quadrilateral corresponds to a mode of the temperature map. All quadrilaterals contain information about the lensing power spectrum, but only the parallelograms contain information about the disconnected Gaussian bias. We show this explicitly in Eq. (B28). So if we neglect parallelograms in computing $\langle \hat{\kappa}\hat{\kappa} \rangle$ we can remove the large disconnected Gaussian bias. We expand on this more in Sec. III.

This $N^{(0)}$ noise bias avoidance improves over traditional methods (1) by circumventing the need for extensive MC simulations to subtract $N^{(0)}$ and (2) leads to a negligible fractional reduction in the SNR (of order $\sim 1/N_{\text{modes}} \sim 10^{-6}$). Also by nature of (1) our estimator is insensitive to inexact assumptions about the power spectrum C_ℓ^{TT}, which is a source of misestimation for traditional MC-based corrections.

To implement this avoidance of the configurations that source the $N^{(0)}$ bias, we simply keep the usual CMB lensing power spectrum estimator (which contains all the configurations) and subtract from it the problematic configurations from the data. As illustrated schematically in Eq. (7), the CMB lensing quadratic estimator can be thought of as the sum of $\sim N_{\text{modes}}$ elementary quadratic estimators. The noise bias only arises from the auto spectra of these elementary estimators. By keeping only the $N_{\text{modes}}(N_{\text{modes}} - 1)/2$ cross spectra, we avoid the noise bias, while discarding only a fraction $\sim 1/N_{\text{modes}} \sim 10^{-6}$ of the signal-to-noise ratio. Our method thus amounts to estimating the CMB lensing auto spectrum only from cross spectra. We denote these problematic configurations $\hat{N}_L$, defined as[2]

$$\hat{N}_L = 2(N_L^\kappa)^2 \int \frac{d^2\ell}{(2\pi)^2} F^\kappa_{\ell, L-\ell} F^\kappa_{-\ell, -L+\ell} |T_\ell|^2 |T_{L-\ell}|^2. \quad (12)$$

Note that due to the structure of the integral, this can be efficiently implemented with the use of FFTs as shown in

our code LensQuEst. We can thus form a new estimator for the CMB lensing auto spectrum,

$$\hat{C}_L^{\kappa\kappa, (\text{no bias})} \equiv \hat{\kappa}_L \hat{\kappa}_L^* - \hat{N}_L. \quad (13)$$

Crucially, just like the standard lensing quadratic estimator, this quantity can be computed efficiently with FFT on the flat sky (respectively, spherical harmonics transforms on the curved sky). As a result, it does not increase the complexity or the computation time of the analysis. This new estimator subtracts the Gaussian noise bias exactly on average. Indeed, in the absence of lensing signal when $\langle \hat{\kappa}_L \hat{\kappa}_L^* \rangle_{\text{GRF}} = N_L^{(0)}$ only,[3]

$$\langle \hat{\kappa}_L \hat{\kappa}_L^* \rangle_{\text{GRF}} - \langle \hat{N}_L \rangle_{\text{GRF}} = 0. \quad (14)$$

In Fig. 2, we show how our exact noise bias avoidance works on a lensed CMB temperature map with no masking and isotropic noise in comparison with the standard method. In Secs. VII and IX we comment on the robustness of our estimator in the presence of realistic anisotropic noise and masking. In this plot we use a hybrid method to estimate the $N^{(1)}$ term, which we describe in Appendix H.[4] We see from this plot that in this case our method is able to

---

[2]In Eq. (12) we are suppressing a factor of $[(2\pi)^2 \delta^{(D)}(0)]^{-1}$ that turns into a finite area correction when computing this in the discrete case. Namely, $\delta^{(D)}(0)$ becomes the area of the map in the discrete case. See discussion around Eq. (A4).

[3]In Eqs. (13) and (14), there is actually an additional $\ell = L/2$ contraction not included. However, we show in Appendix D that this additional contraction is smaller, again by a factor $\sim 1/N_{\text{modes}} \sim 10^{-6}$.

[4]We also show in Appendix B 1 that our proposed $\hat{N}$ subtraction does not affect $N^{(1)}$ biases, thus allowing us to use standard methods of computing $N^{(1)}$.

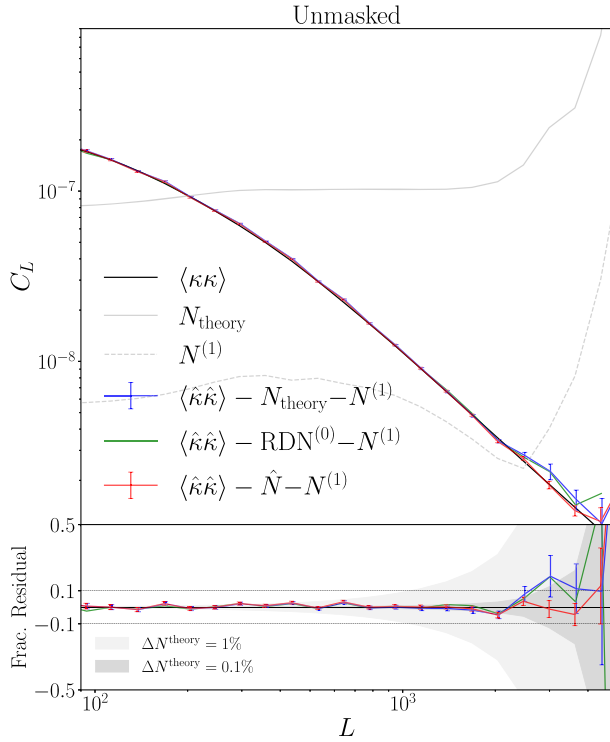FIG. 2. When estimating the lensing power spectrum $\langle \kappa\kappa \rangle$ (black) as a function of angular multipole for a Simons Observatory- (SO) like survey, the $N^{(0)}$ Gaussian bias (gray solid) dominates on small scales. As a result, as the bottom residual plot shows, even a 0.1% misestimate of $N^{(0)}$ at high L leads to a 10% bias in $\langle \kappa\kappa \rangle$. When the input CMB and noise power spectra are perfectly known, standard methods like theory subtraction ($\langle \hat{\kappa}\hat{\kappa} \rangle - N_{\text{theory}} - N^{(1)}$) and RDN$^{(0)}$ subtraction ($\langle \hat{\kappa}\hat{\kappa} \rangle - \text{RDN}^{(0)} - N^{(1)}$) perform equally well as our new bias avoidance method ($\langle \hat{\kappa}\hat{\kappa} \rangle - \hat{N} - N^{(1)}$). This new method continues to work when the input models are inaccurate, as we will see in Secs. VII and IX, and is computationally faster, by not relying on a set of simulations. In fact, since we computed the RDN$^{(0)}$ correction using common pool of simulations, thus introducing nontrivial correlations, it is difficult, though not impossible through semianalytic methods [33,34], to estimate a meaningful error bar on it. So for the RDN$^{(0)}$ subtracted spectra, we do not include error bars. We expand on how we simulate our maps in Appendix C.

estimate the $N^{(0)}$ bias to better than 0.1%, doing just as well as standard methods.

IV. COMPARISON WITH THE STANDARD REALIZATION-DEPENDENT $N^{(0)}$ SUBTRACTION

Currently, the standard way to estimate the Gaussian bias is called realization-dependent $N^{(0)}$ [26,27]. It makes use of the actual measured data d, along with two sets of simulations $\{s\}, \{s'\}$ which are Gaussian random fields with power spectra that are equal to the total lensed CMB power spectrum $\tilde{C}_\ell^{TT}$,
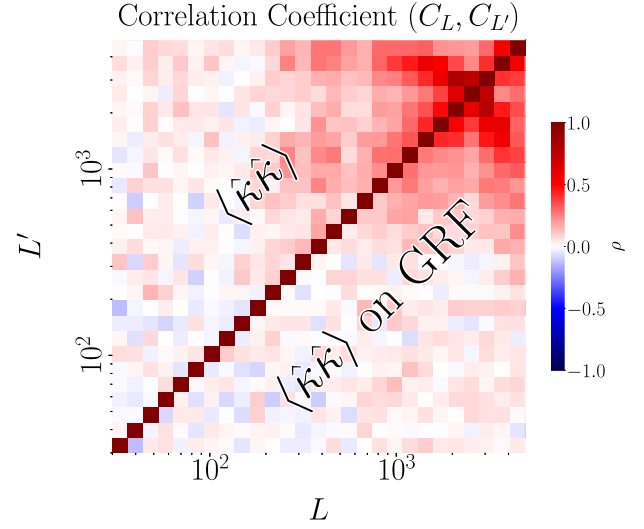

FIG. 3. There is nontrivial covariance structure between different angular multipoles of $\langle \hat{\kappa}\hat{\kappa} \rangle \sim$ CMB lensing spectrum $+ N^{(0)} + \cdots$ as shown in the upper left where we plot the covariance structure of $\langle \hat{\kappa}\hat{\kappa} \rangle$. The dominant contribution to this covariance is from the $N^{(0)}$ bias. This is shown in the lower right where we plot the covariance structure of $\langle \hat{\kappa}\hat{\kappa} \rangle$ run on Gaussian random fields (GRFs) with the same power spectrum as an actually lensed map. Since in the bottom right plot we are only using GRFs, there is only the Gaussian $N^{(0)}$ contribution and no non-Gaussian lensing signal or higher order $N^{(i)}$ within $\langle \hat{\kappa}\hat{\kappa} \rangle_{\text{GRF}}$. Because the dominant covariance structure in the upper left also shows up in the bottom right, we see that the dominant contribution to the covariance in the CMB lensing spectrum is from $N^{(0)}$. We spell out the origin of this covariance structure in Appendix E. Since the dominant contribution to covariances in the CMB lensing spectrum are due to $N^{(0)}$, it should be expected that any prescription to handle the $N^{(0)}$ bias should also handle the covariance introduced by the $N^{(0)}$ bias. We shall see in Fig. 4 that both our proposed method and the standard RDN$^{(0)}$ do this.

$$
\begin{aligned}
\text{RDN}_L^{(0)} \equiv\ &< C_L(\hat{\kappa}^{ds}, \hat{\kappa}^{ds}) + C_L(\hat{\kappa}^{ds}, \hat{\kappa}^{sd}) \\
&+ C_L(\hat{\kappa}^{sd}, \hat{\kappa}^{ds}) + C_L(\hat{\kappa}^{sd}, \hat{\kappa}^{sd}) \\
&- (C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{ss'}) + C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{s's})) >_{s,s'} .
\end{aligned} \tag{15}
$$

This estimator for the disconnected noise bias arises naturally from an Edgeworth expansion of the CMB likelihood, which we show explicitly in Appendix G. It has two advantages over simply evaluating N_{theory} [Eq. (8)] numerically. First, while both methods require modeling the power spectrum of the observed data d, RDN$^{(0)}$ is parametrically less sensitive to inaccuracies in the power spectrum model. Second, as we show below in Fig. 4, it optimally suppresses the covariance of the lensing band powers.

Our method differs from the usual RDN$^{(0)}$ subtraction in CMB lensing in several important ways. First, it does not require running the lensing estimator on Gaussian
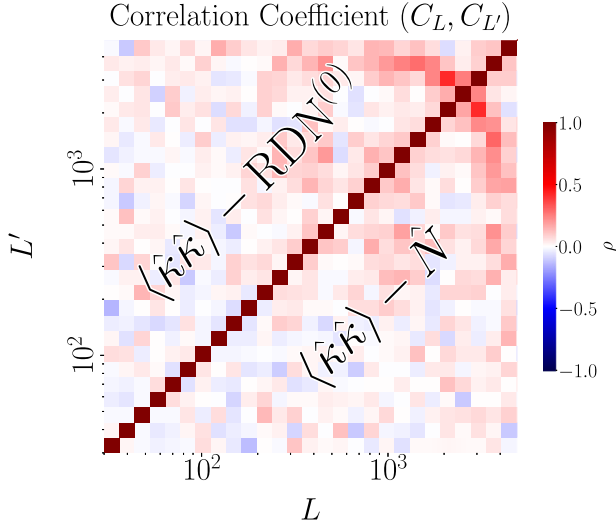
FIG. 4. The nontrivial covariance between different angular multipoles of $\langle \hat\kappa \hat\kappa \rangle \sim$ CMB lensing spectrum $+ N^{(0)} + \cdots$ shown in Fig. 3, which we argue is mostly due to $N^{(0)}$, is removed by both the standard $\mathrm{RDN}^{(0)}$ subtraction (upper left) and our proposed method of $N^{(0)}$ bias avoidance (lower right). Note that doing the naive N^{theory} subtraction would not remove the covariance since N^{theory} is a constant that cannot affect the covariance.

simulations with the same power spectrum as the data. Thus, our estimator is insensitive to errors in the modeling of the total map power spectrum, which is a required input for those Gaussian simulations. Second, by not requiring to run our estimator on many simulations we are also able to significantly reduce the computational run time and memory cost of estimating the $N^{(0)}$ bias.

However, our method shares several limitations with the usual $\mathrm{RDN}^{(0)}$ subtraction. Non-Gaussian foregrounds contribute a non-Gaussian bias to the lensing power spectrum estimator. This contribution is not subtracted automatically in our method, similar to the other $N^{(0)}$ subtraction methods mentioned. Similarly, another bias to the lensing power spectrum can occur from inhomogeneous atmospheric noise, detector noise from a nonuniform scan strategy, or due to mode coupling from the survey mask. We study these in more detail in Secs. VII and IX.

V. EFFECT ON THE CMB WEAK LENSING POWER SPECTRUM COVARIANCE

In this section we argue that the bulk of the covariance in the CMB weak lensing power spectrum is due to the given $N^{(0)}$ realization. Correspondingly, we should expect the bulk of the covariance to be removed if we remove the $N^{(0)}$ noise bias. This is desirable for our measured CMB weak lensing power spectra: making the covariance matrix more diagonal simplifies the problem of its estimation.

Indeed, Fig. 3 shows that the correlation structure of $C_L^{\hat\kappa \hat\kappa}$ is almost identical in two cases: (1) $\hat\kappa$ run on lensed temperature maps (upper left) and (2) $\hat\kappa$ run on an unlensed temperature map (Gaussian random fields) with the same power spectrum. In (1), the resulting $C_L^{\hat\kappa \hat\kappa}$ includes $N^{(0)}$, the true lensing power spectrum, and higher order $N^{(i)}$ biases. In (2), $C_L^{\hat\kappa \hat\kappa}$ includes only $N^{(0)}$. Since the dominant correlation structure in (1) also appears in (2) where there is only $N^{(0)}$, we see that the bulk of the covariance in the CMB weak lensing power spectrum is due to $N^{(0)}$. We analytically explain the origin of these off-diagonal covariances in $N^{(0)}$ in Appendix E.

Subtracting the expected $N^{(0)}$, computed from the power spectrum of the input maps, would not change the correlation structure in Fig. 3. Indeed, this N_{theory} subtraction only removes the mean $N^{(0)}$, not its exact realization. On the other hand, both our $\hat N$ estimator and the $\mathrm{RDN}^{(0)}$ subtraction successfully suppress the off-diagonal covariances (Fig. 4). While both methods are as effective in this respect, the $\hat N$ subtraction is dramatically cheaper computationally, requiring no simulations.

VI. TOY MODEL OF OPTIMAL TRISPECTRUM ESTIMATION

Reference [20] presents an illuminating toy model for optimal trispectrum estimation, which we connect to our method of $N^{(0)}$ subtraction, the standard realization-dependent $N^{(0)}$ subtraction, and the naive N_{theory} subtraction methods discussed in this paper. Consider a weakly non-Gaussian random variable X with zero mean, known variance σ^2, and a small kurtosis $\mathcal{K}$ we wish to estimate,

$$\langle X^2 \rangle = \sigma^2, \tag{16}$$

$$\langle X^4 \rangle = 3\sigma^4 + \mathcal{K}. \tag{17}$$

Let $x_1, \ldots, x_N$ be independent realization of this random variable X,

$$\langle x_i x_j \rangle = \delta_{ij} \sigma^2, \tag{18}$$

$$\langle x_i x_j x_k x_l \rangle = \sigma^4 (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + \mathcal{K}\delta_{ij}\delta_{jk}\delta_{kl}. \tag{19}$$

In this toy model, x_i is analogous to a specific Fourier mode of the CMB temperature anisotropy field T_ℓ and knowing σ^2 is analogous to knowing the total power spectrum of the temperature field.

From Eqs. (17) and (19) one might write down the naive estimator for $\mathcal{K}$,

$$\hat{\mathcal{K}}_{\mathrm{naive}} = \frac{1}{N}\left(\sum_{i=1}^{N} x_i^4\right) - 3\sigma^4. \tag{20}$$

However, if we are searching for a minimum-variance estimator, this is suboptimal. Indeed, one can compute that

$$\text{Var}(\hat{\mathcal{K}}_{\text{naive}}) = \frac{96\sigma^8}{N}. \qquad (21)$$

Furthermore, it suffers from a strong parametric dependence on our estimate of σ^2. Suppose we estimate σ^2 with some error,

$$\sigma_{\text{est}}^2 = \sigma_{\text{true}}^2 - \Delta\sigma^2. \qquad (22)$$

Then the estimate of $\hat{\mathcal{K}}_{\text{naive}}$ is biased as

$$\langle \hat{\mathcal{K}}_{\text{naive}} \rangle = \mathcal{K} - 6\sigma^2(\Delta\sigma^2) + O(\Delta\sigma^2)^2. \qquad (23)$$

This makes the naive estimator fragile to misestimates of σ^2 in comparison to other methods we will discuss shortly.

This naive method is analogous to the naive $N^{(0)}$ subtraction method we described earlier in this paper of subtracting off N_{theory}. So we can glean that the naive N_{theory} subtraction suffers the same problems as this naive kurtosis estimator. It is not the optimal minimum-variance estimator of $C_L^{\kappa\kappa}$ and is vulnerable to misestimates of the total CMB temperature power spectrum ("σ^2").

$\hat{\mathcal{K}}_{\text{opt}}$, the minimum-variance unbiased estimator of this kurtosis given that σ^2 is known, can be shown to be (Ref. [20] and Appendix F)

$$\hat{\mathcal{K}}_{\text{opt}} = \underbrace{\frac{1}{N}\sum_{i=1}^{N} x_i^4 - \frac{6\sigma^2}{N}\sum_{i=1}^{N} x_i^2}_{(A)} + \underbrace{3\sigma^4}_{(B)}. \qquad (24)$$

Another computation will lead to the variance of this estimator,

$$\text{Var}(\hat{\mathcal{K}}_{\text{opt}}) = \frac{24\sigma^8}{N}. \qquad (25)$$

Comparing this to Eq. (21) we see that the naive estimator is significantly suboptimal. We can also perform a similar robustness analysis as in Eq. (22) to find that if we misestimate σ^2 then

$$\langle \hat{\mathcal{K}}_{\text{opt}} \rangle = \mathcal{K} + 3(\Delta\sigma^2)^2 + \dots. \qquad (26)$$

In comparison to Eq. (23) we see that the optimal estimator is also much more robust to misestimates of σ^2 by one power in the error. This is extremely attractive for problems where σ^2 is difficult to estimate.

The optimal kurtosis estimator $\hat{\mathcal{K}}_{\text{opt}}$ is exactly analogous to the standard realization-dependent $N^{(0)}$ subtraction method. We show this rigorously in Appendixes F and G by demonstrating that the derivation of RDN$^{(0)}$ is the multivariate generalization of $\hat{\mathcal{K}}_{\text{opt}}$. But this can also be seen schematically. Indeed, terms (A) and (B) in Eq. (24) can be related to the corresponding terms in RDN$_L^{(0)}$, Eq. (15),

$$\text{RDN}_L^{(0)} = \Big\langle \underbrace{\begin{array}{c} C_L(\hat{\kappa}^{ds}, \hat{\kappa}^{ds}) + C_L(\hat{\kappa}^{ds}, \hat{\kappa}^{sd}) \\ + C_L(\hat{\kappa}^{sd}, \hat{\kappa}^{ds}) + C_L(\hat{\kappa}^{sd}, \hat{\kappa}^{sd}) \end{array}}_{(A)} \\ \underbrace{-(C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{ss'}) + C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{s's}))}_{(B)} \Big\rangle_{s,s'}. \qquad (27)$$

Namely:
  (i) The (A) terms of the form $C_L(\hat{\kappa}^{ds}\hat{\kappa}^{ds})$ are the same as being a product of the sample variance $\sum_i x_i^2/N$ from the dd contractions and the assumed "true" input variance σ^2 from the ss contractions.
  (ii) The (B) terms of the form $C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{ss'})$ only contain information about the assumed "true" variance σ^2 from the ss and $s's'$ contractions.

We put "true" in quotation marks since, in reality, the underlying power spectra is not known, making the optimal kurtosis (and thus trispectrum) estimator sensitive to how well we understand the underlying power spectrum as discussed in Eq. (26). Thus, from this toy model and correspondence between $\hat{\mathcal{K}}_{\text{opt}}$ and RDN$^{(0)}$ we can gleam the fact that RDN$^{(0)}$ is more robust to misestimates of "σ^2" compared to the naive method of N^{theory} as it was designed to be and can be thought of as a generalization of the optimal minimum-variance kurtosis estimator given that σ^2 is known perfectly.

In addition to the optimal estimator, Ref. [20] presents an alternative near-optimal kurtosis estimator $\hat{\mathcal{K}}_{\text{alt}}$ of the form

$$\hat{\mathcal{K}}_{\text{alt}} = \frac{1}{N}\sum_i x_i^4 - \underbrace{\frac{3}{N(N-1)}\sum_{i\neq j} x_i^2 x_j^2}_{(C)}. \qquad (28)$$

As before, you can compute the variance of this estimator,

$$\text{Var}(\hat{\mathcal{K}}_{\text{alt}}) = \frac{24\sigma^8}{N}\left(1 + \frac{3}{N-1}\right) \qquad (29)$$

$$= \text{Var}(\hat{\mathcal{K}}_{\text{opt}}) + O\left(\frac{1}{N}\right)^2. \qquad (30)$$

From the last line where we compare to the variance of the optimal minimum-variance estimator (25), we see that this estimator is *nearly* optimal. But a key difference is that this alternative estimator has no dependence on an assumed σ^2. This means that, unlike the optimal estimator $\hat{\mathcal{K}}_{\text{opt}}$, this alternative estimator $\hat{\mathcal{K}}_{\text{alt}}$ is completely insensitive to misestimates in σ^2.

We can show that this alternative kurtosis estimator $\hat{\mathcal{K}}_{\text{alt}}$ is analogous to our proposed method of subtracting the $N^{(0)}$ bias. Indeed, the (C) term in Eq. (28) can be identified with the terms in our $\hat{N}$ subtraction scheme in Eq. (12),

$$\hat{N}_L = 2(N_L^\kappa)^2 \int_\ell F_{\ell,L-\ell}^\kappa F_{-\ell,-L+\ell}^\kappa \underbrace{|T_\ell|^2 |T_{L-\ell}|^2}_{(C)}. \quad (31)$$

We can understand the term $\sum_{i \neq j} x_i^2 x_j^2$ as all the contractions where we get information only about the disconnected Gaussian bias from the $(ii)(jj)$ contraction, but never about the connected non-Gaussian kurtosis since x_i and x_j are independent for $i \neq j$. This is analogous to $\int_\ell |T_\ell|^2 |T_{L-\ell}|^2$ since these terms are the terms where we get all our contribution to the disconnected Gaussian bias.[5] Thus, from this toy model and correspondence between $\hat{\mathcal{K}}_{\text{alt}}$ and $\hat{N}$ we can intuit the fact that $\hat{N}$ is completely insensitive to misestimates of "σ^2," as we established earlier, but is still a nearly optimal estimator for the connected trispectrum ("kurtosis").

VII. INHOMOGENEOUS NOISE/DEPTH

The complex scan strategy of CMB experiments produces maps with inhomogeneous noise levels. These are difficult to model, resulting in inaccurate simulations, which can bias the standard $N^{(0)}$ subtraction techniques. While the RDN$^{(0)}$ is somewhat robust to this inaccurate modeling, a very small bias of only $\sim 0.1\%$ on $N^{(0)}$ can result in a large bias of $\sim 10\%$ on the small-scale lensing power spectrum (Fig. 2).

The $\hat{N}$ estimator is insensitive to any mismodeling of the observed power spectrum, since it does not rely on simulations. However, our estimator implicitly assumes statistical isotropy which only approximately holds once noise inhomogeneities arise. This can be put in contrast to the standard RDN$^{(0)}$ method which is sensitive to mismodeling but does not assume statistical isotropy and instead utilizes the full two-point function when estimating $N^{(0)}$. So it is important to understand the trade-off between (1) approximating maps with noise inhomogeneities as statistically isotropic and (2) the sensitivity of RDN$^{(0)}$ to mismodeling of noise inhomogeneities. This trade-off will determine the relative performance between our method and the standard RDN$^{(0)}$ in the presence of inhomogeneous instrument noise. In this section we study

---

[5]You might notice a slight subtlety to this correspondence between our $\hat{N}$ and this toy model. The $|T_\ell|^2 |T_{L-\ell}|^2$ terms in our $\hat{N}$ also contain information about the connected non-Gaussian contribution since $\langle T_\ell T_\ell^* T_{L-\ell} T_{L-\ell}^* \rangle_c \neq 0$ due to mode couplings from lensing, whereas $\langle x_i x_i x_j x_j \rangle_c = 0$ by assumption of the x_i being independent identically distributed (IID) in this toy model. However, the analogy still holds. This is because computing $\langle \hat{\kappa}\hat{\kappa} \rangle - \hat{N}$ removes only N_{modes} of the total N_{modes}^2 contributions to the connected non-Gaussian contribution when we subtract $\hat{N}$. So, to roughly first order in $O(1/N_{\text{modes}})$, we can draw the connection between our $\hat{N}$ method and this $\hat{\mathcal{K}}_{\text{alt}}$ despite the IID assumption in the toy model that is not present in the general case.

ACT DR5 IVar Map

Extreme
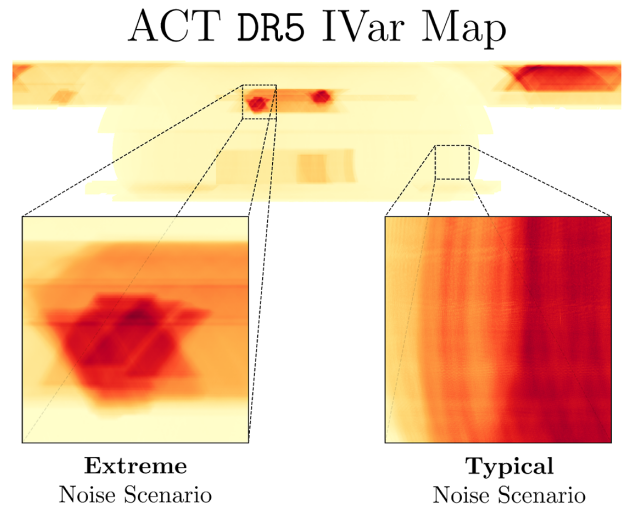Noise Scenario

Typical
Noise Scenario

FIG. 5. To test our proposed $\hat{N}$ subtraction method in the presence realistic anisotropic detector noise, we take two cutouts of the ACT DR5 IVar map: (1) a typical anisotropic noise scenario expected for SO and (2) an extreme anisotropic noise scenario. Both of these anisotropic noise maps are applied multiplicatively to a GRF with the expected noise power spectrum for SO such that $N^{\text{anisotropic}}(\mathbf{x}) \geq N^{\text{isotropic}}(\mathbf{x})$. See Appendix C for more details. Note that in the *typical* noise scenario plot, we have enhanced the colors in order to emphasize that inhomogenous instrument noise can arise due to scan pattern.

the robustness of our method to (1) typical or (2) extreme anisotropic detector noise. For this, we use the actual depth maps from ACT data release 5 (DR5; see Fig. 5). We perform a simulated analysis of the performance of our method and RDN$^{(0)}$. Specifically, we generate anisotropic Gaussian noise realizations by multiplying white noise maps with the typical or extreme inhomogeneous depth maps shown in Fig. 5. We then add an independent lensed CMB realization to each of these and a Gaussian random field representing the level of foregrounds at 150 GHz (neglecting their non-Gaussianity) resulting in 500 simulated maps. Figure 6 shows that our noise bias avoidance method ($\langle \hat{\kappa}\hat{\kappa} \rangle - \hat{N} - N^{(1)}$) is at least as robust to typical noise inhomogeneities as the RDN$^{(0)}$ ($\langle \hat{\kappa}\hat{\kappa} \rangle - \text{RDN}^{(0)} - N^{(1)}$) and N_{theory} ($\langle \hat{\kappa}\hat{\kappa} \rangle - N_{\text{theory}} - N^{(1)}$) subtractions.[6] While the RDN$^{(0)}$ appears to fail at a lower multipole L, this might be improved with more simulations. However, this comes at an increased computational cost which was prohibitive for us. This highlights the benefit of the much lower computing cost of $\hat{N}$.

---

[6]Note that in a real analysis one may use optimal anisotropic Wiener filtering [35] as opposed to the naive quadratic estimator with isotropic weights as we do here. However, we do not believe the use of optimal anisotropic Wiener filtering would change the qualitative picture shown in this section.
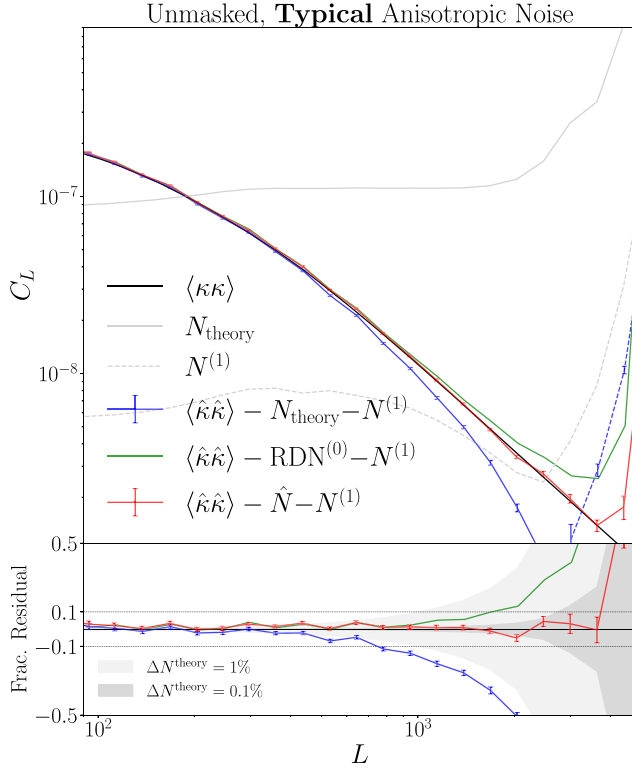
FIG. 6. When CMB temperature anisotropy maps exhibit a typical level of anisotropic noise such as what is displayed in Fig. 5, there are additional mode couplings that one must address when estimating the lensing power spectrum $\langle \kappa\kappa \rangle$ (black) as a function of angular multipole L. The standard method RDN0 subtraction ($\langle \hat{\kappa}\hat{\kappa} \rangle - \mathrm{RDN}^{(0)} - N^{(1)}$) and our new bias avoidance method ($\langle \hat{\kappa}\hat{\kappa} \rangle - \hat{N} - N^{(1)}$) both perform equally well in the presence of typical anisotropic noise levels, whereas the naive N^{theory} subtraction ($\langle \hat{\kappa}\hat{\kappa} \rangle - N_{\mathrm{theory}} - N^{(1)}$) fails at small scales. Dashed lines correspond to the same colored curve when that curve becomes negative. For similar reasons described in Fig. 2 we do not include error bars for the $\mathrm{RDN}^{(0)}$ subtraction. The biased estimate for $\langle \kappa\kappa \rangle$ from $\mathrm{RDN}^{(0)}$ subtraction at small scales may be chalked up to both convergence of the MC correction as well as small errors in the assumed total temperature power spectrum including anisotropic noise needed for MC simulations. Even minor mismodeling can be magnified to percent level biases due to the fact that $N^{(0)}$ is orders of magnitude larger than the lensing signal $\langle \kappa\kappa \rangle$ at small scales.

In this typical noise inhomogeneity scenario, $\mathrm{RDN}^{(0)}$ and $\hat{N}$ both continue to suppress the off-diagonal covariances on the lensing power spectrum (Fig. 7). Indeed, the residual correlation structure is very similar to Fig. 4.

In the extreme noise inhomogeneity scenario, Fig. 8 shows that $\mathrm{RDN}^{(0)}$ works best if accurate noise simulations can be obtained. However, we shall see in the next section that our $\hat{N}$ estimator can be combined with a split-based method [18] that is insensitive to the noise properties and does not rely on detailed understanding of the noise model. While this level of anisotropy was present in some parts of the ACT data, due to the coaddition of maps from multiple seasons with different
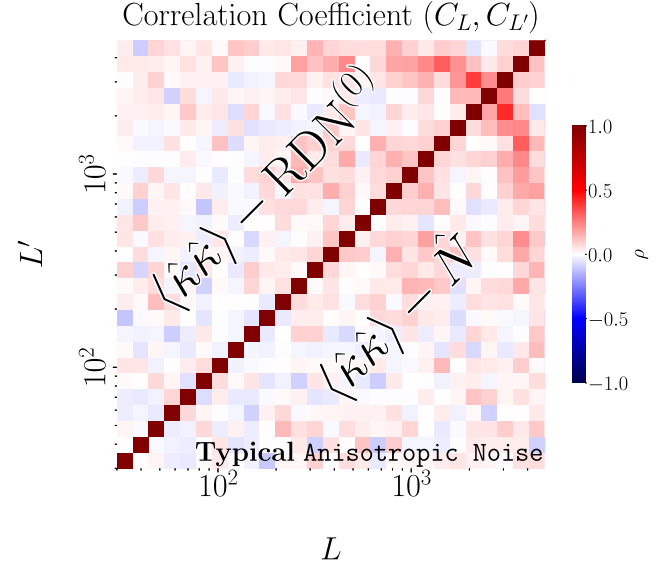

FIG. 7. In the presence of typical anisotropic detector noise, the nontrivial covariance structure between different angular multipoles of $\langle \hat{\kappa}\hat{\kappa} \rangle \sim$ CMB lensing spectrum $+ N^{(0)} + \cdots$ shown in Fig. 3 may be modified. However, this does not spoil the removal of the dominant contributor to these covariances by the standard $\mathrm{RDN}^{(0)}$ (upper left) subtraction and our proposed method of $N^{(0)}$ bias avoidance (lower right).

footprints, we do not expect this level of inhomogeneity in current and upcoming wide-field CMB surveys.

VIII. COMPARISON WITH THE MAP SPLIT-BASED $N^{(0)}$ SUBTRACTION

As we saw in Fig. 8, once detector and atmospheric noise inhomogeneities become large, our method fails to remove the $N^{(0)}$ bias. However, one can combine our proposed method with split-based methods, proposed in Ref. [18], to handle these extreme noise anisotropies. Reference [18] proposed an estimator for the CMB lensing power spectrum with no detector noise bias $\hat{C}_L^{\kappa\kappa,\times}$. This estimator makes use of m splits of the CMB map $T_\ell^{(i)}$ with independent instrument noise. One can then cross-correlate these splits to make an estimator for the lensing potential that is insensitive to modeling of detector noise,

$$\hat{\kappa}_L^{(ij)} \equiv \frac{1}{2} N_L^\kappa \int \frac{d^2\ell}{(2\pi)^2} F_{\ell,L-\ell}^\kappa \left[T_\ell^{(i)} T_{L-\ell}^{(j)} + T_\ell^{(j)} T_{L-\ell}^{(i)} \right], \quad (32)$$

$$\hat{C}_L^{\kappa\kappa,\times} \equiv \frac{1}{m(m-1)(m-2)(m-3)} \sum_{ijkl} \gamma_{ijkl} C_L(\hat{\kappa}_L^{(ij)}, (\hat{\kappa}_L^{(kl)})). \tag{33}$$

For the case of temperature, since $F_{\ell,L-\ell}^\kappa$ is symmetric, this symmetrization in Eq. (32) is trivial, but we use this notation since in general, e.g., when estimating κ from
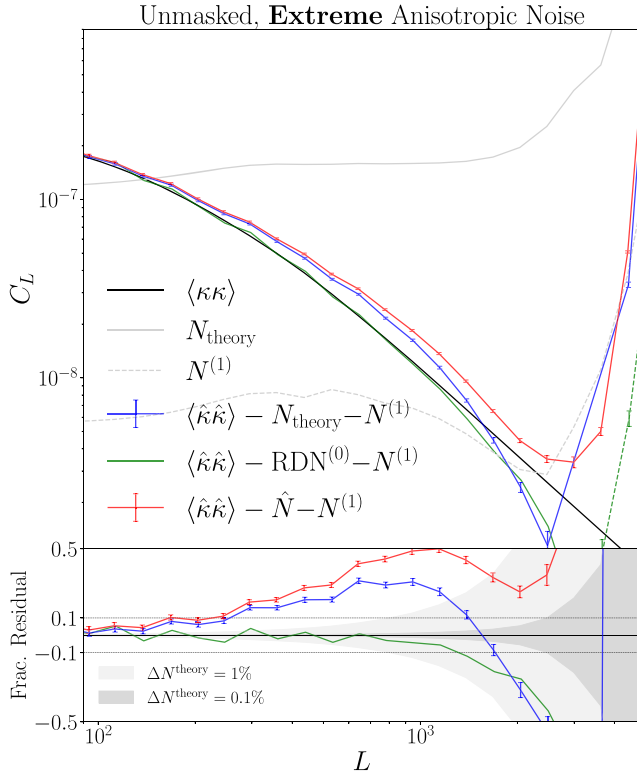
FIG. 8. When CMB temperature anisotropy maps exhibit an *extreme* level of anisotropic noise such as what is displayed in Fig. 5, there are additional strong mode couplings which make it difficult to estimate the lensing power spectrum $\langle \kappa\kappa \rangle$ (black) as a function of angular multipole L. At this level of extreme noise anisotropy, the naive N_{theory} subtraction ($\langle \hat{\kappa}\hat{\kappa} \rangle - N_{\text{theory}} - N^{(1)}$) and our new bias avoidance method ($\langle \hat{\kappa}\hat{\kappa} \rangle - \hat{N} - N^{(1)}$) fail, while the standard RDN$^{(0)}$ subtraction ($\langle \hat{\kappa}\hat{\kappa} \rangle - \text{RDN}^{(0)} - N^{(1)}$), continues to work. Dashed lines correspond to the same colored curve when that curve becomes negative. For similar reasons described in Fig. 2 we do not include error bars for the RDN$^{(0)}$ subtraction. It should be noted that this level of noise anisotropy is exactly what split-based methods described in Sec. VIII and Ref. [18] are designed to handle. We also showed in that section how one might combine split-based methods with our proposed noise bias avoidance method. Thus, if one encounters such extreme aniso-tropic noise in analysis, we prescribe using split-based methods in combination with our proposed noise bias avoidance for split-based lensing power spectrum. In the presence of such extreme anisotropic noise there will also be a mean field that we estimate and subtract off. The origin of this mean field is discussed more in Sec. IX A for the comparable case of masking.

temperature-polarization cross-spectra, F is not symmetric. Following Ref. [18], we have defined γ_{ijkl} as

$$\gamma_{ijkl} = \begin{cases} 1 & \text{if } (i,j,k,l) \text{ all distinct} \\ 0 & \text{otherwise} \end{cases}. \quad (34)$$

By ignoring contributions to the four-point function that repeat a split, this method avoids the noise bias from

atmospheric and detector noise, even if they are spatially inhomogeneous (e.g., due to the scan strategy). This is otherwise a limiting factor for current analyses, even when temperature maps are dominated by the lensed CMB and foregrounds on the relevant scales. This makes the split-based method very attractive.

It is also possible to define a version of our $\hat{N}$ subtraction method for this split-based lensing power spectrum. We can define

$$\hat{N}_L^{ij,kl} = 2(N_L^\kappa)^2 \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F^\kappa_{\boldsymbol{\ell}, L-\boldsymbol{\ell}} F^\kappa_{-\boldsymbol{\ell}, -L+\boldsymbol{\ell}}$$
$$\times \left| T^{(i)}_{\boldsymbol{\ell}} T^{(j)}_{-\boldsymbol{\ell}} \right| \left| T^{(k)}_{L-\boldsymbol{\ell}} T^{(l)}_{-L+\boldsymbol{\ell}} \right|. \quad (35)$$

Our estimate for the $N^{(0)}$ bias can be written as

$$\hat{N}_L^\times = \frac{1}{m(m-1)(m-2)(m-3)} \times \sum_{ijkl} \gamma_{ijkl} \hat{N}_L^{ij,kl}. \quad (36)$$

Thus, we can remove our estimated $N^{(0)}$ bias

$$\hat{C}_L^{\kappa\kappa,(\text{no bias}),\times} = \hat{C}_L^{\kappa\kappa,\times} - \hat{N}_L^\times. \quad (37)$$

Naively this is an $O(m^4)$ computation. However, analo-gous to the fast algorithm proposed by Ref. [18], we can construct a fast $O(m^2)$ algorithm to compute this $\hat{N}_L^\times$ which is described in Appendix I.

IX. IMPACT OF THE MASK

The presence of a mask leads to additional complications in the analysis of the CMB lensing power spectrum by inducing additional difficult-to-model mode couplings and a mean field that must be removed.

A. Biases from the mask mode coupling

To illustrate these additional mode couplings, consider a temperature field with a masking function applied to it,

$$T^m(\boldsymbol{x}) = M(\boldsymbol{x}) T(\boldsymbol{x}). \quad (38)$$

Thus, in Fourier space we have a convolution

$$T^m_{\boldsymbol{\ell}} = \int \frac{d^2\boldsymbol{\ell}'}{(2\pi)^2} T_{\boldsymbol{\ell}'} M_{\boldsymbol{\ell}-\boldsymbol{\ell}'}, \quad (39)$$

where we have introduced the Fourier transform of the mask. Now consider the mask

$$M(\boldsymbol{x}) = 1 - m(\boldsymbol{x}) \quad (40)$$

$$\Rightarrow M_{\boldsymbol{\ell}} = \delta^{(D)}(\boldsymbol{\ell}) - m_{\boldsymbol{\ell}}. \quad (41)$$

This tells us that

$$T_\ell^m = T_\ell - \int \frac{d^2\ell'}{(2\pi)^2} T_{\ell'} m_{\ell-\ell'}. \tag{42}$$

To leading order, this modifies the mode couplings from Eq. (6),

$$\langle T_\ell^m T_{L-\ell}^m \rangle \approx \langle T_\ell T_{L-\ell} \rangle - m_L(C_\ell^{TT} + C_{|L-\ell|}^{TT})$$

$$\equiv \langle T_\ell T_{L-\ell} \rangle + m_L f_{\ell,L}^m, \tag{43}$$

where in the last line we implicitly defined

$$f_{\ell,L}^m = -C_\ell^{TT} - C_{|L-\ell|}^{TT}. \tag{44}$$

Namely, we get additional mode couplings due to masking. There are several methods to account for these mode couplings such as the pseudo-C_ℓ algorithm described in Refs. [36,37] with modifications such as those done in Ref. [16] or, alternatively, the mask can be propagated through the quadratic estimator as shown in Ref. [38]. We refer the reader to Ref. [16] for details on how the latest measurement of the CMB lensing power spectrum by ACT accounts for mode couplings due to masking. Masking also leads to a mean-field bias in our quadratic estimator for κ, Eq. (7),

$$\langle \hat{\kappa}_L^m \rangle \approx \langle \hat{\kappa}_L \rangle + \underbrace{N_L^\kappa \left(\int \frac{d^2\ell}{(2\pi)^2} F_{\ell,L-\ell}^\kappa f_{\ell,L}^m \right) m_L}_{\text{mean-field bias}}. \tag{45}$$

Typically this mean field from a fixed mask is estimated from many simulations with independent CMB and lensing realizations and subtracted off,

$$\text{Mean-field subtraction:} \hat{\kappa} \to \hat{\kappa} - \langle \hat{\kappa} \rangle. \tag{46}$$

In this section we make use of this mean-field subtraction. Since our $\hat{N}$ [Eq. (12)] is also derived from data, we must apply a mean-field correction $\Delta\hat{N}_{\text{mf}}$ to it as well

$$\text{Mean-field subtraction:} \hat{N} \to \hat{N}/f_{\text{sky}}^2 - \Delta\hat{N}_{\text{mf}}. \tag{47}$$

This correction can be computed from the same simulations used to compute the mean field bias $\langle \hat{\kappa} \rangle$ by averaging the difference between $\hat{N}$ run on masked and unmasked maps,

$$\Delta\hat{N}_{\text{mf}} = \langle \hat{N}(\text{Masked}\,T\text{'s})/f_{\text{sky}}^2 - \hat{N}(\text{Unmasked}\,T\text{'s}) \rangle, \tag{48}$$

where we discuss the origin of the f_{sky}^2 factor in Sec. IX B. In this section we also make use of this mean-field correction to $\hat{N}$.

For a lensed CMB temperature map with the mask shown in Fig. 9, we show how our exact noise bias avoidance performs in comparison to the standard method in Fig. 10. We see from this plot that our proposed method
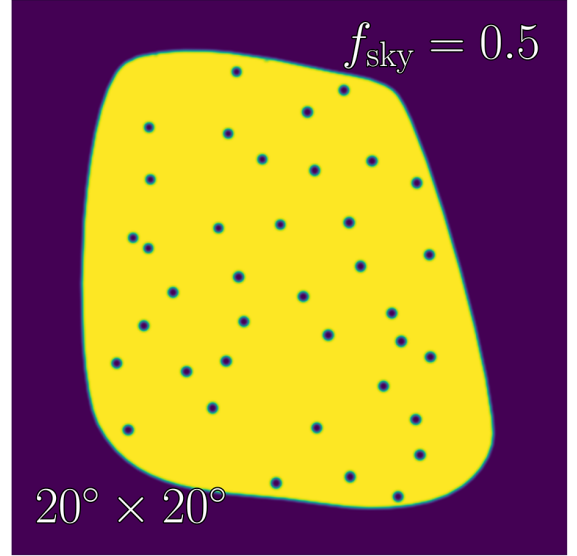

FIG. 9. In our numerical studies in Sec. IX we use the above apodized $20° \times 20°$ mask which includes point sources of radius ~10 arc min to roughly match SPT [39]. Apodization is done using a Gaussian filter where the standard deviation for the Gaussian kernel is 3 pixels. The unapodized mask and point sources were lovingly hand drawn.

is able to match the performance of the standard RDN$^{(0)}$ method.

In practice, the mask induces additional coupling between the observed Fourier modes of the temperature maps. The full treatment of these mode coupling effects is beyond the scope of this paper (see Ref. [16] for a thorough treatment of the effect of masking on CMB lensing reconstruction).

In Fig. 11 we plot the correlation structure if we apply our $\hat{N}$ subtraction in comparison to the standard RDN$^{(0)}$ subtraction in the presence of a mask. This is analogous to Figs. 4 and 7 where we plotted the same thing for isotropic and typical anisotropic noise, respectively. We can see that, in this case, we are still able to remove covariances as before.

B. f_{sky} corrections for masking

In the presence of a mask, the lensing power spectrum from $\langle \hat{\kappa}\hat{\kappa} \rangle$ will be underestimated by approximately f_{sky} and our $\hat{N}$ will underestimate the $N^{(0)}$ bias by approximately f_{sky}^2. In this paper, we adopt the simple approximate correction in the presence of masking by modifying Eq. (13),

$$\hat{C}_L^{\kappa\kappa,(\text{no bias})} = \frac{\hat{\kappa}_L \hat{\kappa}_L^*}{f_{\text{sky}}} - \frac{\hat{N}_L}{f_{\text{sky}}^2}. \tag{49}$$

We note that a Monte Carlo normalization or the use of pseudo-C_ℓ will automatically include this correction which will not need to be applied separately. We now illustrate why this is with the same toy model we considered in Sec. VI.
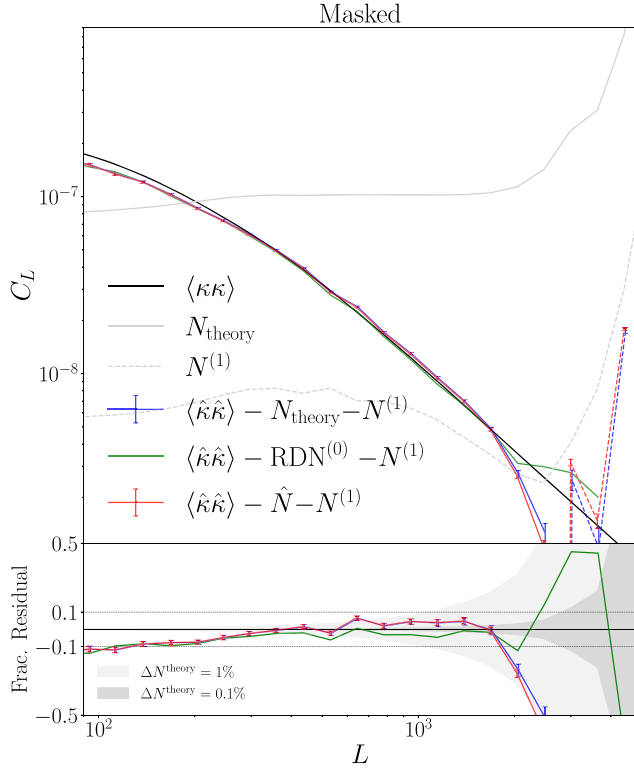
FIG. 10. When masks such as the one shown in Fig. 9 are applied to CMB temperature anisotropy maps, there are additional mode couplings that one must address when estimating the lensing power spectrum $\langle \kappa\kappa \rangle$ (black) as a function of angular multipole L. We spell out roughly how these appear in Sec. IX. The standard method RDN0 subtraction ($\langle \hat{\kappa}\hat{\kappa} \rangle - \text{RDN}^{(0)} - N^{(1)}$), the naive N^{theory} subtraction ($\langle \hat{\kappa}\hat{\kappa} \rangle - N_{\text{theory}} - N^{(1)}$), and our new bias avoidance method ($\langle \hat{\kappa}\hat{\kappa} \rangle - \hat{N} - N^{(1)}$) all perform comparably in the presence of masking. Dashed lines correspond to the same colored curve when that curve becomes negative. For similar reasons described in Fig. 2 we do not include error bars for the RDN$^{(0)}$ subtraction. Since handling additional mask induced mode couplings are beyond the scope of this paper, we do not apply any corrections to account for these mask induced mode couplings. Instead the key result here is that our proposed method has comparable performance to the standard RDN$^{(0)}$ subtraction in the presence of masking.

As in Sec. VI, consider N independent measurements of a random variable X: $\{x_1, \ldots, x_N\}$. This X is again the 1D analog of the T_ℓ maps. The 1D analog of the lensing power spectrum and our $\hat{N}$ estimator may be written as we did before in Eq. (28),

$$\left\langle \frac{1}{N} \sum_i x_i^4 \right\rangle = \langle X^4 \rangle, \tag{50}$$

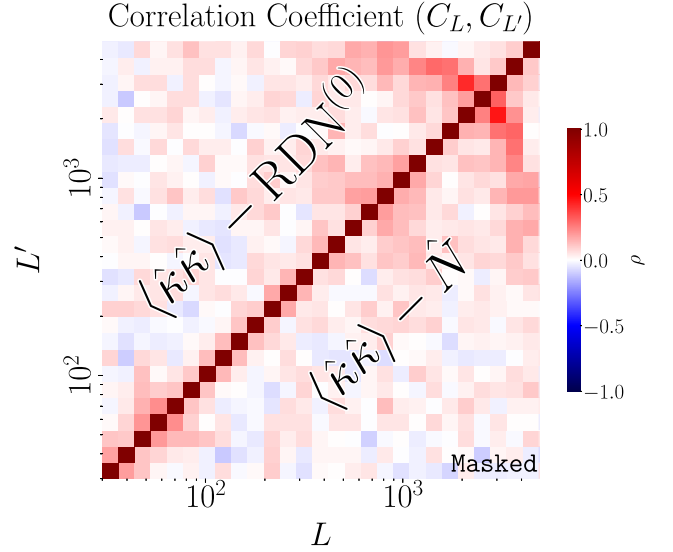$$\left\langle \frac{3}{N(N-1)} \sum_{i \neq j} x_i^2 x_j^2 \right\rangle = 3 \langle X^2 \rangle^2. \tag{51}$$

FIG. 11. In the presence of masking, the nontrivial covariance structure between different angular multipoles of $\langle \hat{\kappa}\hat{\kappa} \rangle \sim$ CMB lensing spectrum $+ N^{(0)} + \cdots$ shown in Fig. 3 may be modified. However, this does not completely spoil the removal of the dominant contributor to these covariances by the standard RDN$^{(0)}$ (upper left) subtraction and our proposed method of $N^{(0)}$ bias avoidance (lower right).

Now suppose that we apply a mask that selects a fraction f of our measurements x_i. Namely, our masked dataset will have $x_i = 0$ for $i > Nf$. This f is analogous to f_{sky}.[7] Then Eq. (50) would be modified as

$$\left\langle \frac{1}{N} \sum_i x_i^4 \right\rangle = f \langle X^4 \rangle. \tag{52}$$

Similarly, if we approximate $N(N-1) \approx N^2$ then Eq. (51) would be modified as

$$\left\langle \frac{3}{N(N-1)} \sum_{i \neq j} x_i^2 x_j^2 \right\rangle \approx f^2 \times 3 \langle X^2 \rangle^2. \tag{53}$$

Thus, to properly normalize our estimator we must include appropriate factors of f^2. For the general case of $\langle \hat{\kappa}\hat{\kappa} \rangle$ this of course is only an approximation, but we believe errors due to this approximation are subdominant to errors from unaccounted for mask induced mode couplings in our numerical studies and so made use of Eq. (49) in this section.

---

[7]A subtlety in this claim is that the number of measurements in the toy model N is analogous to N_{modes}. Since $N_{\text{modes}} \propto$ Area, the analogy between f and f_{sky} holds.

X. POLARIZATION-BASED LENSING

One can estimate lensing statistics from other CMB random fields, such as polarization. Indeed, for $\{X, Y\} \in \{T, E, B\}$ one can write a generalization of Eq. (6),

$$\langle X_\ell Y_{L-\ell} \rangle = f^{XY}_{\ell, L-\ell} \kappa_L + O(\kappa_L^2). \tag{54}$$

We refer readers to Refs. [40,41] for the explicit form of these f^{XY} and subsequent F^{XY} which are generalizations of Eq. (9). Because of lensing's imprint on these correlations one may define a quadratic estimator using any pair of $\{T, E, B\}$. Thus, one may write an estimator for the lensing power spectrum using $X, Y, W, Z \in \{T, E, B\}$ which we will call $C^{XY,WZ}_L$.

The lensing power spectrum estimated from $X, Y, W, Z \in \{T, E, B\}$ will also have its own $N^{(0)}$ bias that our proposed $\hat{N}$ subtraction can be generalized to remove. Namely, we can write as a generalization of Eq. (12),

$$\hat{N}^{XY,WZ}_L = (N^\kappa_L)^2 \int \frac{d^2\ell}{(2\pi)^2} F^{XY}_{\ell, L-\ell} F^{WZ}_{-\ell, -L+\ell}$$
$$\times \left(|X_\ell W_{-\ell}||Y_{L-\ell} Z_{-L+\ell}| \right.$$
$$\left. + |X_\ell Z_{-\ell}||Y_{L-\ell} W_{-L+\ell}| \right). \tag{55}$$

Correspondingly, Eq. (13) becomes

$$\hat{C}^{XY,WZ(\text{no bias})}_L \equiv \hat{C}^{XY,WZ}_L - \hat{N}^{XY,WZ}_L. \tag{56}$$

We leave the numerical exploration of the performance of $\hat{N}^{XY,WZ}$ subtraction to future work.

XI. CONCLUSIONS

In this paper, we have presented an estimator of the CMB lensing power spectrum that avoids the Gaussian $N^{(0)}$ noise bias. This is done by isolating which terms contribute to the $N^{(0)}$ bias when estimating the lensing power spectrum $\langle \kappa\kappa \rangle$ and avoiding these terms. We showed that avoiding these terms is computationally efficient and negligibly reduces the signal-to-noise ratio. This estimator is run only on data, thus avoiding the need for bias subtraction using simulations. Since our estimator avoids simulations, we avoid (1) sensitivity to misestimates in simulated CMB and noise models and (2) the large computational cost that current simulation-based methods like RDN$^{(0)}$ require.

Our estimator is as robust as the standard RDN$^{(0)}$ to (1) masking of the CMB map and (2) the level of noise inhomogeneity expected for current and upcoming wide-field CMB surveys. To handle extreme levels of noise inhomogeneity, we also show how our estimator may be combined with split-based methods (Ref. [18]) which make it insensitive to any assumptions made in modeling or simulating the instrument noise. In addition, even though we focused mostly on estimating lensing with CMB temperature, we show how our estimator can be generalized to include polarization information as well.

We also discuss the connection between our estimator RDN$^{(0)}$ and optimal trispectrum/four-point function estimation in the context of an illuminating toy model presented in Ref. [20]. This toy model allowed us to argue that our proposed estimator has the same variance as the optimal minimum-variance unbiased estimator of the connected trispectrum/four-point function to first order in $1/N_{\text{modes}}$, but with no parametric dependence on an assumed power spectrum/two-point function. Thus, we believe our estimator is applicable to analogous problems in other fields where estimation of connected trispectra/four-point functions are of interest, such as in large-scale structure.

There are several key directions for further explorations of our proposed method. First, for the use of our method in a full-sky analysis of CMB lensing power spectra, a generalization of this method to curved sky must be worked out. Additionally, though we presented the theoretical framework for how our method may be naturally combined with split-based methods such as those presented in Ref. [18], numerical studies of combining our proposed estimator and split-based methods are needed to verify their compatibility. The use of bias hardening techniques [26,42], optimal anisotropic Wiener filtering [35], and global-minimum-variance lensing quadratic estimators [43] are also often employed to enhance CMB weak lensing analyses. We do not explore how our proposed method may be combined with such techniques and leave this to future work. Finally, there is the interesting work of considering how our proposed estimator, which, in general, may be thought of as an estimator for the connected trispectrum/four-point function, could feed into analysis of large-scale structure data where higher order statistics have been an area of recent interest.

The code implementing our proposed estimator and numerical experiments can be found in Ref. [44].

APPENDIX A: CONVENTION

Throughout this paper we use the convention

$$f(\boldsymbol{x}) = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot\boldsymbol{x}} f_{\boldsymbol{\ell}}, \tag{A1}$$

$$f_{\boldsymbol{\ell}} = \int d^2\boldsymbol{x} e^{-i\boldsymbol{\ell}\cdot\boldsymbol{x}} f(\boldsymbol{x}), \tag{A2}$$

$$\Rightarrow \delta^D(\boldsymbol{\ell}) = \int \frac{d^2\boldsymbol{x}}{(2\pi)^2} e^{i\boldsymbol{\ell}\cdot\boldsymbol{x}}. \tag{A3}$$

Our square CMB maps have finite area A and resolution. When Fourier transforming these maps, there is a discrete set of modes we can resolve. These modes are integer multiples of the smallest fundamental mode $k_F = 2\pi/\sqrt{A}$ that we can measure. We may rewrite the Dirac δ for these discrete set of modes. For two modes $\boldsymbol{k}_i = k_F\mathbf{i}$ and $\boldsymbol{k}_j = k_F\mathbf{j}$, where $\mathbf{i}$ and $\mathbf{j}$ are a vector of integers,

$$\delta^{(D)}(\boldsymbol{k}_i - \boldsymbol{k}_j) = \delta^{(D)}(k_F(\mathbf{i} - \mathbf{j})) = \frac{A}{(2\pi)^2}\delta^{(K)}_{\mathbf{i},\mathbf{j}}, \tag{A4}$$

where $\delta^{(K)}$ is the Kronecker δ.

APPENDIX B: $N^{(i)}$ BIASES

In this appendix, we will spell out explicitly the origin of the $N^{(i)}$ biases. These noise biases fall out when using Eq. (2) to expand the four-point function within the integral of Eq. (10) and applying Wick's theorem to the Gaussian fields that fall out. First let us expand Eq. (1) out to another order,

$$T(\boldsymbol{x}) = T^0(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x})) = T^0(\boldsymbol{x}) + (\partial_i\psi(\boldsymbol{x}))(\partial^i T^0(\boldsymbol{x}))$$
$$+ \frac{1}{2}(\partial^i\psi(\boldsymbol{x}))(\partial^j\psi(\boldsymbol{x}))(\partial_i\partial_j T^0(\boldsymbol{x})) + O(\boldsymbol{d}^3). \tag{B1}$$

From here we can rewrite Eq. (2) as

$$T_{\boldsymbol{\ell}} = T^{(0)}_{\boldsymbol{\ell}} + T^{(1)}_{\boldsymbol{\ell}} + T^{(2)}_{\boldsymbol{\ell}} + O(\kappa^3), \tag{B2}$$

where we have defined

$$T^{(0)}_{\boldsymbol{\ell}} \equiv T^0_{\boldsymbol{\ell}}, \tag{B3}$$

$$T^{(1)}_{\boldsymbol{\ell}} \equiv -\int \frac{d^2\boldsymbol{\ell}'}{(2\pi)^2} \boldsymbol{\ell}' \cdot (\boldsymbol{\ell} - \boldsymbol{\ell}') \frac{2\kappa_{\boldsymbol{\ell}-\boldsymbol{\ell}'}}{(\boldsymbol{\ell} - \boldsymbol{\ell}')^2} T^0_{\boldsymbol{\ell}'}, \tag{B4}$$

which is the same as in Eq. (2) and

$$T^{(2)}_{\boldsymbol{\ell}} \equiv -\frac{1}{2}\int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2}\frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} T^0_{\boldsymbol{\ell}_1} \frac{2\kappa_{\boldsymbol{\ell}_2}}{\boldsymbol{\ell}_2^2} \frac{2\kappa^*_{\boldsymbol{\ell}_1+\boldsymbol{\ell}_2-\boldsymbol{\ell}}}{(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 - \boldsymbol{\ell})^2}$$
$$\times [\boldsymbol{\ell}_1 \cdot \boldsymbol{\ell}_2][\boldsymbol{\ell}_1 \cdot (\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 - \boldsymbol{\ell})]. \tag{B5}$$

Now, instead of keeping κ_L fixed as we did in Sec. II, let us promote κ_L to a Gaussian random field with power spectrum $C^{\kappa\kappa}$,

$$\kappa_L \sim \mathcal{N}(0, C^{\kappa\kappa}_{\boldsymbol{\ell}}). \tag{B6}$$

We will be computing $\langle TTTT \rangle$, which at each order has terms like

$$\langle T_{\boldsymbol{\ell}_1} T_{\boldsymbol{\ell}_2} T_{\boldsymbol{\ell}_3} T_{\boldsymbol{\ell}_4} \rangle \approx \underbrace{\langle T^{(0)}_{\boldsymbol{\ell}_1} T^{(0)}_{\boldsymbol{\ell}_2} T^{(0)}_{\boldsymbol{\ell}_3} T^{(0)}_{\boldsymbol{\ell}_4} \rangle}_{O(\kappa^0)} + \underbrace{\langle T^{(1)}_{\boldsymbol{\ell}_1} T^{(0)}_{\boldsymbol{\ell}_2} T^{(0)}_{\boldsymbol{\ell}_3} T^{(0)}_{\boldsymbol{\ell}_4} \rangle + 3\,\text{perms}}_{O(\kappa^1)} + \langle T^{(2)}_{\boldsymbol{\ell}_1} T^{(0)}_{\boldsymbol{\ell}_2} T^{(0)}_{\boldsymbol{\ell}_3} T^{(0)}_{\boldsymbol{\ell}_4} \rangle + 3\,\text{perms}$$
$$+ \langle T^{(1)}_{\boldsymbol{\ell}_1} T^{(1)}_{\boldsymbol{\ell}_2} T^{(0)}_{\boldsymbol{\ell}_3} T^{(0)}_{\boldsymbol{\ell}_4} \rangle + \langle T^{(0)}_{\boldsymbol{\ell}_1} T^{(0)}_{\boldsymbol{\ell}_2} T^{(1)}_{\boldsymbol{\ell}_3} T^{(1)}_{\boldsymbol{\ell}_4} \rangle + \langle T^{(1)}_{\boldsymbol{\ell}_1} T^{(0)}_{\boldsymbol{\ell}_2} T^{(1)}_{\boldsymbol{\ell}_3} T^{(0)}_{\boldsymbol{\ell}_4} \rangle + \langle T^{(0)}_{\boldsymbol{\ell}_1} T^{(1)}_{\boldsymbol{\ell}_2} T^{(0)}_{\boldsymbol{\ell}_3} T^{(1)}_{\boldsymbol{\ell}_4} \rangle$$
$$+ \underbrace{\langle T^{(0)}_{\boldsymbol{\ell}_1} T^{(1)}_{\boldsymbol{\ell}_2} T^{(1)}_{\boldsymbol{\ell}_3} T^{(0)}_{\boldsymbol{\ell}_4} \rangle + \langle T^{(1)}_{\boldsymbol{\ell}_1} T^{(0)}_{\boldsymbol{\ell}_2} T^{(0)}_{\boldsymbol{\ell}_3} T^{(1)}_{\boldsymbol{\ell}_4} \rangle}_{O(\kappa^2)}. \tag{B7}$$

We could use Wick contractions to compute each of these terms. For example, for the $O(\kappa^0)$ term,

$$\langle T^{(0)}_{\boldsymbol{\ell}_1} T^{(0)}_{\boldsymbol{\ell}_2} T^{(0)}_{\boldsymbol{\ell}_3} T^{(0)}_{\boldsymbol{\ell}_4} \rangle$$
$$= (2\pi)^4[\delta^{(D)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2)\delta^{(D)}(\boldsymbol{\ell}_3 + \boldsymbol{\ell}_4)C^{TT}_{\boldsymbol{\ell}_1}C^{TT}_{\boldsymbol{\ell}_3}$$
$$+ \delta^{(D)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_3)\delta^{(D)}(\boldsymbol{\ell}_2 + \boldsymbol{\ell}_4)C^{TT}_{\boldsymbol{\ell}_1}C^{TT}_{\boldsymbol{\ell}_2}$$
$$+ \delta^{(D)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_4)\delta^{(D)}(\boldsymbol{\ell}_2 + \boldsymbol{\ell}_3)C^{TT}_{\boldsymbol{\ell}_1}C^{TT}_{\boldsymbol{\ell}_2}]. \tag{B8}$$

These terms are what will lead to the $N^{(0)}$ bias once plugged into Eq. (10) as we shall see shortly. Since $\langle\kappa\rangle = 0$

and $\langle T^{(1)}T^{(0)}T^{(0)}T^{(0)}\rangle$ will generically be of the form $\int\langle\kappa\rangle\langle T^{(0)}T^{(0)}T^{(0)}T^{(0)}\rangle$, we have

$$\langle T^{(1)}_{\boldsymbol{\ell}_1} T^{(0)}_{\boldsymbol{\ell}_2} T^{(0)}_{\boldsymbol{\ell}_3} T^{(0)}_{\boldsymbol{\ell}_4} \rangle = 0. \tag{B9}$$

At $O(\kappa^2)$ the Wick contraction machinery alone becomes unwieldy. There is a much more conceptually transparent method to understand higher order terms and consequently $N^{(i)}$ biases through the use of Feynman diagrams. We will connect our Wick contraction machinery stated so far to this diagrammatic language. This diagrammatic approach

was derived in Refs. [30,31] and used to understand the CMB lensing bispectrum in Ref. [45]. To start, we will restate the Feynman rules derived in Refs. [30,31] here:

  (i) Each lensed CMB temperature field $T_{\boldsymbol{\ell}}$ corresponds to a vertex with a momentum $\boldsymbol{\ell}$ flowing into that vertex.

 (ii) Two vertices connected by a solid or dashed line with momentum $\boldsymbol{\ell}$ come with a factor of C_{ℓ}^{TT} or $C_{\ell}^{\kappa\kappa}$, respectively. These are our propagators,

$$\xrightarrow{\hspace{1cm}\boldsymbol{\ell}\hspace{1cm}} = (2\pi)^2 C_{\ell}^{TT} \qquad \text{(B10)}$$

$$\dashrightarrow{\hspace{1cm}\boldsymbol{\ell}\hspace{1cm}} = (2\pi)^2 C_{\ell}^{\kappa\kappa} \qquad \text{(B11)}$$

 (iii) Uncontracted external legs come with a factor of their corresponding field,

$$\xrightarrow{\hspace{0.6cm}\boldsymbol{\ell}\hspace{0.6cm}}\bullet = T_{\ell}^0 \qquad \text{(B12)}$$

$$\dashrightarrow{\hspace{0.6cm}\boldsymbol{\ell}\hspace{0.6cm}}\bullet = \kappa_{\boldsymbol{\ell}} \qquad \text{(B13)}$$

 (iv) Coming out of each vertex $T_{\boldsymbol{\ell}}$ we can draw one solid line corresponding to the unlensed CMB temperature field $T_{\boldsymbol{\ell}'}$ and an arbitrary number n of dashed lines with momentum $\boldsymbol{k}_i$ where $i = 1, \ldots, n$. These dashed lines correspond to the order κ^n to which we are expanding. For such a vertex we must assert momentum conservation $\sum_k \boldsymbol{k}_i + \boldsymbol{\ell}' = \boldsymbol{\ell}$. Each vertex also comes with a factor $\prod_{i=1}^{n} -2\boldsymbol{k}_i \cdot \boldsymbol{\ell}'/\boldsymbol{k}_i^2$. Such a factor can be seen for the $n = 1$ case in Eq. (B4) and $n = 2$ in Eq. (B5),

$$\boldsymbol{\ell} \rightarrow \quad \begin{matrix} \boldsymbol{k}_1 \\ \vdots \\ \boldsymbol{k}_n \\ \boldsymbol{\ell}' \end{matrix} = \delta^{(D)}\left(\sum_i \boldsymbol{k}_i + \boldsymbol{\ell}' - \boldsymbol{\ell}\right)\prod_{i=1}^{n} -\frac{2\boldsymbol{k}_i \cdot \boldsymbol{\ell}'}{\boldsymbol{k}_i^2} \qquad \text{(B14)}$$

  (v) Unconstrained momenta $\boldsymbol{k}$ are integrated over with $\int d^2\boldsymbol{k}/(2\pi)^2$.

We can now understand the terms in Eq. (B7) in a more transparent way:

  To start, we can consider the contraction of purely Gaussian random field, Eq. (B8),

$$\langle T_{\boldsymbol{\ell}_1}^{(0)} T_{\boldsymbol{\ell}_2}^{(0)} T_{\boldsymbol{\ell}_3}^{(0)} T_{\boldsymbol{\ell}_4}^{(0)}\rangle = \qquad \qquad + \qquad \qquad + \qquad \qquad . \qquad \text{(B15)}$$

which recovers exactly what is expected. Similarly, the $\langle T^{(1)} T^{(0)} T^{(0)} T^{(0)}\rangle$ term, which we know evaluates to 0 after taking average $\langle\kappa\rangle$,

$$\langle T_{\boldsymbol{\ell}_1}^{(1)} T_{\boldsymbol{\ell}_2}^{(0)} T_{\boldsymbol{\ell}_3}^{(0)} T_{\boldsymbol{\ell}_4}^{(0)}\rangle = \qquad \qquad + \qquad \qquad + \qquad \qquad . \qquad \text{(B16)}$$

Here we get to the real meat of the usefulness of this diagrammatic approach. First the $\langle T^{(2)} T^{(0)} T^{(0)} T^{(0)}\rangle$ terms, which include a single loop,

$$\langle T^{(2)}_{\ell_1} T^{(0)}_{\ell_2} T^{(0)}_{\ell_3} T^{(0)}_{\ell_4} \rangle = \quad \text{(B17)}$$

Through working out the Wick contraction or from looking at these diagrams, one can see that these terms go like $\langle T^{(2)} T^{(0)} T^{(0)} T^{(0)} \rangle \sim \langle T^{(0)} T^{(0)} T^{(0)} T^{(0)} \rangle \times \int C^{\kappa\kappa}$. As stated in Sec. II A, the $N^{(1)}$ bias arises from integrals of the lensing power spectrum. So, naively, one might think this term and its permutations would contribute to $N^{(1)}$. However, it turns out that, if one replaces the unlensed CMB temperature power spectrum C^{TT}_ℓ with the lensed spectrum $\tilde{C}^{TT}_\ell$ in Eqs. (4) and (B15), terms like $\langle T^{(2)} T^{(0)} T^{(0)} T^{(0)} \rangle$ are automatically taken care of and do not contribute to $N^{(1)}$. Roughly, this is because the lensed spectra $\tilde{C}^{TT}_\ell$, which we will denote with a double line, expanded to $O(\kappa^2)$ looks like

$$\tilde{C}^{TT}_\ell = \quad \text{(B18)}$$

Terms like Eq. (B17) are accounted for by contributions coming from the second line of Eq. (B18) if we reorganize our expansion with $\tilde{C}^{TT}_\ell$. We again refer the reader to Refs. [30,31] for more details on this point. Given this point, we will neglect these $\langle T^{(2)} T^{(0)} T^{(0)} T^{(0)} \rangle$ terms going forward.

The diagrams contributing to the final terms at $O(\kappa^2)$ in Eq. (B7) are[8]

$$\langle T^{(1)}_{\ell_1} T^{(1)}_{\ell_2} T^{(0)}_{\ell_3} T^{(0)}_{\ell_4} \rangle + \langle T^{(0)}_{\ell_1} T^{(0)}_{\ell_2} T^{(1)}_{\ell_3} T^{(1)}_{\ell_4} \rangle = \quad \text{(B19)}$$

$$\langle T^{(1)}_{\ell_1} T^{(0)}_{\ell_2} T^{(1)}_{\ell_3} T^{(0)}_{\ell_4} \rangle + \langle T^{(0)}_{\ell_1} T^{(1)}_{\ell_2} T^{(0)}_{\ell_3} T^{(1)}_{\ell_4} \rangle = \quad \text{(B20)}$$

---

[8]Here, we neglect some diagrams that contain a loop like the ⟋⟍ diagram contributing to $\langle T^{(1)}_{\ell_1} T^{(0)}_{\ell_2} T^{(1)}_{\ell_3} T^{(0)}_{\ell_4} \rangle$. We neglect these diagrams for the same reason we neglect Eq. (B17). The neglected diagrams are handled by the second term in the first line of Eq. (B18) when replacing the unlensed with the lensed temperature power spectrum.

$$\langle T^{(0)}_{\ell_1} T^{(1)}_{\ell_2} T^{(1)}_{\ell_3} T^{(0)}_{\ell_4}\rangle + \langle T^{(1)}_{\ell_1} T^{(0)}_{\ell_2} T^{(0)}_{\ell_3} T^{(1)}_{\ell_4}\rangle = \quad \text{(B21)}$$

Adding all these diagrams together allows us to make a very nifty simplification. We follow Refs. [30,31] in defining a composite vertex which restates Eq. (5) diagrammatically,

$$f^{\kappa}_{\ell_1,\ell_2} = \qquad = \qquad + \qquad . \qquad \text{(B22)}$$

The slashes on the line denote the fact that these lines do not produce a temperature power spectrum when translated into equations. Now note that all of the diagrams in Eqs. (B19)–(B21) can better organized in three diagrams that utilize this composite vertex,

$$= \qquad + \qquad + \qquad + \qquad \text{(B23)}$$

$$\underbrace{\qquad}_{\subset \langle T^{(1)}T^{(0)}T^{(1)}T^{(0)}\rangle} \quad \underbrace{\qquad}_{\subset \langle T^{(0)}T^{(1)}T^{(0)}T^{(1)}\rangle} \quad \underbrace{\qquad}_{\subset \langle T^{(0)}T^{(1)}T^{(1)}T^{(0)}\rangle} \quad \underbrace{\qquad}_{\subset \langle T^{(1)}T^{(0)}T^{(0)}T^{(1)}\rangle}$$

$$= \qquad + \qquad + \qquad + \qquad \text{(B24)}$$

$$\underbrace{\qquad}_{\subset \langle T^{(1)}T^{(1)}T^{(0)}T^{(0)}\rangle} \quad \underbrace{\qquad}_{\subset \langle T^{(0)}T^{(0)}T^{(1)}T^{(1)}\rangle} \quad \underbrace{\qquad}_{\subset \langle T^{(0)}T^{(1)}T^{(1)}T^{(0)}\rangle} \quad \underbrace{\qquad}_{\subset \langle T^{(1)}T^{(0)}T^{(0)}T^{(1)}\rangle}$$

$$(B25)$$

The sum of Eqs. (B19)–(B21) can be written as

$$\begin{aligned}
&\langle T^{(1)}_{\ell_1} T^{(1)}_{\ell_2} T^{(0)}_{\ell_3} T^{(0)}_{\ell_4}\rangle + \langle T^{(0)}_{\ell_1} T^{(0)}_{\ell_2} T^{(1)}_{\ell_3} T^{(1)}_{\ell_4}\rangle \\
&+ \langle T^{(1)}_{\ell_1} T^{(0)}_{\ell_2} T^{(1)}_{\ell_3} T^{(0)}_{\ell_4}\rangle + \langle T^{(0)}_{\ell_1} T^{(1)}_{\ell_2} T^{(0)}_{\ell_3} T^{(1)}_{\ell_4}\rangle = \\
&+ \langle T^{(0)}_{\ell_1} T^{(1)}_{\ell_2} T^{(1)}_{\ell_3} T^{(0)}_{\ell_4}\rangle + \langle T^{(1)}_{\ell_1} T^{(0)}_{\ell_2} T^{(0)}_{\ell_3} T^{(1)}_{\ell_4}\rangle
\end{aligned}$$

$$(B26)$$

$$= (2\pi)^2 \delta^{(D)}(\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 + \boldsymbol{\ell}_3 + \boldsymbol{\ell}_4) \begin{pmatrix} f^{\kappa}_{\ell_1,\ell_2} C^{\kappa\kappa}_{\ell_1+\ell_2} f^{\kappa}_{\ell_3,\ell_4} \\ + f^{\kappa}_{\ell_1,\ell_3} C^{\kappa\kappa}_{\ell_1+\ell_3} f^{\kappa}_{\ell_2,\ell_4} \\ + f^{\kappa}_{\ell_1,\ell_4} C^{\kappa\kappa}_{\ell_1+\ell_4} f^{\kappa}_{\ell_2,\ell_3} \end{pmatrix}. \qquad (B27)$$

With all of this in hand, we can now compute the $N^{(0)}$ and $N^{(1)}$ bias. As stated earlier, the $N^{(i)}$ come out when we plug Eq. (B7) into Eq. (10). At $O(\kappa^0)$ when we plug Eq. (B8) [equivalently, Eq. (B15)] we will get the $N^{(0)}$ bias,

$$\begin{aligned}
N^{(0)}_L &= (N^{\kappa}_L)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F^{\kappa}_{\ell_1, L-\ell_1} F^{\kappa}_{-\ell_2, -L+\ell_2} \times \langle T^{(0)}_{\ell_1} T^{(0)}_{L-\ell_1} T^{(0)}_{-\ell_2} T^{(0)}_{-L+\ell_2}\rangle \\
&= (2\pi)^4 (N^{\kappa}_L)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F^{\kappa}_{\ell_1, L-\ell_1} F^{\kappa}_{-\ell_2, -L+\ell_2} \left[\underbrace{\delta^{(D)}(L)^2 C_{\ell_1} C_{\ell_2}}_{=0 \quad \text{since } L \neq 0} + \underbrace{2\delta^{(D)}(\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2)\delta^{(D)}(\boldsymbol{\ell}_2 - \boldsymbol{\ell}_1) C^{TT}_{\ell_1} C^{TT}_{L-\ell_1}}_{\boldsymbol{\ell}_1 = \boldsymbol{\ell}_2 \text{ (parallelograms)}} \right] \\
&= (2\pi)^2 \delta^{(D)}(0) \times 2(N^{\kappa}_L)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} (F^{\kappa}_{\ell_1, L-\ell_1})^2 C^{TT}_{\ell_1} C^{TT}_{L-\ell_1},
\end{aligned} \qquad (B28)$$

where in the second line we used the $\boldsymbol{\ell}_2 \to L - \boldsymbol{\ell}_2$ symmetry of the integral. A key point of this calculation is that, of all the $\{\boldsymbol{\ell}_1, \boldsymbol{\ell}_2\}$ used to compute $\langle \hat{\kappa}\hat{\kappa}\rangle$, only the $\boldsymbol{\ell}_1 = \boldsymbol{\ell}_2$ contribute to the $N^{(0)}$ noise bias. This explicitly shows the motivation for our $N^{(0)}$ noise bias avoidance principle outlined in Fig. 1 and Sec. III.

Since we have only vanishing terms at $O(\kappa^1)$ we go straight to $O(\kappa^2)$, which gives us both $C^{\kappa\kappa}_L$ and the $N^{(1)}$ bias. Plugging Eq. (B27) into Eq. (10) yields

$$\begin{aligned}
(2\pi)^2 \delta^{(D)}(0) \times (N^{\kappa}_L)^2 &\left(C^{\kappa\kappa}_L \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} F^{\kappa}_{\ell_1, L-\ell_1} f^{\kappa}_{\ell_1, L-\ell_1} \times \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F^{\kappa}_{-\ell_2, -L+\ell_2} f^{\kappa}_{-\ell_2, -L+\ell_2} \right. \\
&\left. + \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F^{\kappa}_{\ell_1, L-\ell_1} F^{\kappa}_{-\ell_2, -L+\ell_2} \times [f^{\kappa}_{\ell_1, -\ell_2} C^{\kappa\kappa}_{\ell_1-\ell_2} f^{\kappa}_{L-\ell_1, -L+\ell_2} + f^{\kappa}_{\ell_1, \ell_2-L} C^{\kappa\kappa}_{\ell_1+\ell_2-L} f^{\kappa}_{L-\ell_1, -\ell_2}] \right).
\end{aligned} \qquad (B29)$$

From here we see two things. The first line when combined with Eq. (8) gives us $(2\pi)^2\delta^{(D)}(0)C^{\kappa\kappa}_L$, the CMB lensing power spectrum, as promised. The second line gives us the $N^{(1)}$ bias,

$$N_L^{(1)} = (2\pi)^2 \delta^{(D)}(0) \times (N_L^\kappa)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F^\kappa_{\boldsymbol{\ell}_1, L-\boldsymbol{\ell}_1} F^\kappa_{-\boldsymbol{\ell}_2, -L+\boldsymbol{\ell}_2}$$

$$\times \left[f^\kappa_{\boldsymbol{\ell}_1, -\boldsymbol{\ell}_2} C^{\kappa\kappa}_{\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2} f^\kappa_{L-\boldsymbol{\ell}_1, -L+\boldsymbol{\ell}_2} + f^\kappa_{\boldsymbol{\ell}_1, \boldsymbol{\ell}_2 - L} C^{\kappa\kappa}_{\boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 - L} f^\kappa_{L-\boldsymbol{\ell}_1, -\boldsymbol{\ell}_2} \right]$$

$$= (2\pi)^2 \delta^{(D)}(0) \times (N_L^\kappa)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F^\kappa_{\boldsymbol{\ell}_1, L-\boldsymbol{\ell}_1} F^\kappa_{-\boldsymbol{\ell}_2, -L+\boldsymbol{\ell}_2} \times 2 \left[f^\kappa_{\boldsymbol{\ell}_1, -\boldsymbol{\ell}_2} C^{\kappa\kappa}_{\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2} f^\kappa_{L-\boldsymbol{\ell}_1, -L+\boldsymbol{\ell}_2} \right], \quad \text{(B30)}$$

where in the final line we used the $\boldsymbol{\ell}_2 \to L - \boldsymbol{\ell}_2$ symmetry of the integral to combine the two terms in the brackets. This result is known in the literature and first computed by Ref. [28] and also reproduced via the use of Feynman diagrams in Refs. [30,31]. From this equation, we can explicitly see the fact that $N^{(1)} \sim \iint \langle \kappa\kappa \rangle$ as stated in Sec. II A. Higher order biases such as $N^{(2)}$ are also computed in Refs. [30,31] and we refer the interested reader to these sources.

1. Why $\hat{N}$ subtraction has no effect on $N^{(1)}$

Our $\hat{N}$'s effect on $N_L^{(1)}$, $\Delta N_L^{(1)}$ can be computed by inserting a $\delta^{(D)}(\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2)$ into the integrand,

$$\Delta N_L^{(1)} = (2\pi)^2 \delta^{(D)}(0) \times 2(N_L^\kappa)^2$$

$$\times \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} (F^\kappa_{\boldsymbol{\ell}_1, L-\boldsymbol{\ell}_1})^2 \left[f^\kappa_{\boldsymbol{\ell}_1, -\boldsymbol{\ell}_1} C^{\kappa\kappa}_0 f^\kappa_{L-\boldsymbol{\ell}_1, -L+\boldsymbol{\ell}_1} \right]$$

$$= 0, \quad \text{(B31)}$$

where in the second line we used Eq. (5) to see that $f^\kappa_{\boldsymbol{\ell}, -\boldsymbol{\ell}} = 0$. This means that $\hat{N}$ does not affect the $N^{(1)}$ biases. In general, when our proposed noise bias avoidance method removes a small subset of terms in any integration, that results in a $N^{(i \geq 1)}$ bias. So, we believe that our proposed bias avoidance method will also have a negligible effect on $N^{(i>1)}$ biases similar to its effect on the $N^{(1)}$. Thus, standard methods of avoiding or estimating higher order noise bias such as the use of $T\nabla T$ weights and Monte Carlo computations of $N^{(1)}$ can be naturally combined with our proposed method.

APPENDIX C: DETAILS ON MAP SIMULATIONS

In this appendix, we describe how we generate our simulated maps used for our numerical studies in this paper. All of our maps span $20° \times 20°$ on the sky and have 1200×1200 pixels. We start by computing the unlensed, lensed, and lensed gradient CMB temperature power spectrum C_ℓ^{TT}, $C_\ell^{TT,L}$, and $C_\ell^{T\nabla T}$ as well as the lensing potential spectrum $C_\ell^{\kappa\kappa}$ with CAMB (Refs. [32,46–48]). Then using LensQuEst (Ref. [49]) we compute the expected foreground C_ℓ^F and noise spectrum C_ℓ^N for a SO-like survey.[9]

With these spectra in hand, our recipe to generate a lensed temperature map is:

(1) Generate a Gaussian random field with the unlensed temperature field power spectra C_ℓ^{TT},

$$T_\ell^0 \sim \mathcal{N}(0, C_\ell^{TT}). \quad \text{(C1)}$$

(2) Generate a Gaussian random field with the lensing potential power spectra $C_\ell^{\kappa\kappa}$,

$$\kappa_\ell \sim \mathcal{N}(0, C_\ell^{\kappa\kappa}). \quad \text{(C2)}$$

(3) Generate the lensed temperature map T_ℓ by first computing the deflection field $\boldsymbol{d}(\boldsymbol{x})$ due to the lensing potential κ_ℓ with

$$\boldsymbol{d}_\ell = -\frac{2i\boldsymbol{\ell}}{\ell^2} \kappa_\ell. \quad \text{(C3)}$$

Then generate the lensed temperature T_ℓ^L by computing $T^L(\boldsymbol{x})$ as in Eq. (1),

$$T^L(\boldsymbol{x}) = T^0(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x})). \quad \text{(C4)}$$

(4) Generate two Gaussian random fields corresponding to foregrounds and detector noise with the corresponding spectra C_ℓ^F and C_ℓ^N,

$$F_\ell \sim \mathcal{N}(0, C_\ell^F), \quad \text{(C5)}$$

$$N_\ell \sim \mathcal{N}(0, C_\ell^N). \quad \text{(C6)}$$

(5) Generate the total lensed CMB temperature map T_ℓ by adding all these maps together,

$$T_\ell = T_\ell^L + F_\ell + N_\ell. \quad \text{(C7)}$$

This is the recipe we used to generate 500 lensed CMB temperature maps used in the numerical studies for Figs. 2–4. Note that the above recipe implies that the total CMB temperature power spectrum is

$$\tilde{C}_\ell^{TT} = C_\ell^{TT,L} + C_\ell^F + C_\ell^N. \quad \text{(C8)}$$

In Eq. (14) we claimed that $\hat{N}_{\text{GRF}} = \langle \hat{\kappa}\hat{\kappa}^* \rangle_{\text{GRF}} \sim N_{\text{theory}}$ [see Eq. (D4)]. With Eq. (C8) we are able to confirm this numerically by generating 500 Gaussian random fields $T_\ell^{\text{GRF}} \sim \mathcal{N}(0, \tilde{C}_\ell^{TT})$ and computing $\hat{N}$ on these fields. We plot the results in Fig. 12 and find that indeed Eq. (14) holds. Something to note is that, as we will argue in
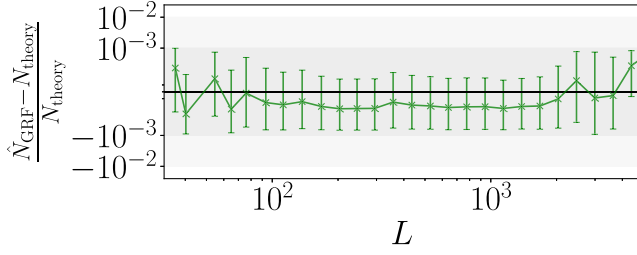
FIG. 12. In Eq. (14) we claimed that our estimator $\hat{N}$ when run on a Gaussian random field $\hat{N}^{\mathrm{GRF}}$ would equal $\langle\hat{\kappa}\hat{\kappa}^*\rangle_{\mathrm{GRF}} \sim N_{\mathrm{theory}}$ [see Eq. (D4)]. In this plot, we numerically check this claim. We also point out that $\hat{N}$ is highly correlated between different L, L'. This is explicitly shown in Fig. 13. This off-diagonal correlation makes χ^2 by eye hard to do here. We explain this correlation structure in Appendix E.

Appendix E, $\hat{N}$ is highly correlated between different L, L', which makes doing χ^2 by eye tests deceptive for this plot.

In Sec. VII we study the effect of anisotropic instrument noise on our proposed CMB lensing spectrum estimator. To generate CMB maps with realistic anisotropic detector noise, we use cutouts from the ACT DR5 IVar map as shown in Fig. 5. Let the anisotropic noise pattern we extract from the ACT DR5 maps be called $A(\boldsymbol{x})$. Using this, we modify our noise map computed in Eq. (C6) as

$$N^{\mathrm{aniso}}(\boldsymbol{x}) = N(\boldsymbol{x}) \times \frac{A(\boldsymbol{x})}{\min[A(\boldsymbol{x})]}. \tag{C9}$$

In this way $N_{\mathrm{aniso}}(\boldsymbol{x}) \geq N(\boldsymbol{x})$. From here Eq. (C7) is then modified as

$$T_{\boldsymbol{\ell}}^{\mathrm{aniso}} = T_{\boldsymbol{\ell}}^L + F_{\boldsymbol{\ell}} + N_{\boldsymbol{\ell}}^{\mathrm{aniso}}. \tag{C10}$$

Note that this means Eq. (C8) is no longer true. Instead, we estimate $\tilde{C}_{\ell}^{TT,\mathrm{aniso}}$ by generating 500 maps using Eq. (C10) and averaging the power spectra of these maps.

In Sec. IX we study the effect of masking on our proposed CMB lensing spectrum estimator. To generate CMB maps with masking we use the mask shown in Fig. 9. Let this mask be called $M(\boldsymbol{x})$. After Eq. (C7) we apply this mask,

$$T^{\mathrm{masked}}(\boldsymbol{x}) = M(\boldsymbol{x}) \times T(\boldsymbol{x}). \tag{C11}$$

We perform our numerical studies in Sec. IX using these maps.

APPENDIX D: NEGLECTED ADDITIONAL CONTRACTIONS IN $\hat{N}$ ARE NEGLIGIBLE

In footnote 3 we commented that Eq. (14) is not exact since we do not include a $\boldsymbol{\ell} = \boldsymbol{L}/2$ term, but that this term is suppressed like $1/N_{\mathrm{modes}}$. In this appendix, we shall show this suppression, thus concluding that the neglected additional contraction is negligible.

Recall from Eq. (10)

$$\langle\hat{\kappa}_L\hat{\kappa}_L^*\rangle = (N_L^\kappa)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F_{\boldsymbol{\ell}_1,\boldsymbol{L}-\boldsymbol{\ell}_1}^\kappa F_{-\boldsymbol{\ell}_2,-\boldsymbol{L}+\boldsymbol{\ell}_2}^\kappa$$
$$\times \langle T_{\boldsymbol{\ell}_1} T_{\boldsymbol{L}-\boldsymbol{\ell}_1} T_{-\boldsymbol{\ell}_2} T_{-\boldsymbol{L}+\boldsymbol{\ell}_2}\rangle. \tag{D1}$$

For a Gaussian random field with no lensing, but with a power spectrum equal to the total CMB power spectrum $T_\ell^G \sim \mathcal{N}(0, \tilde{C}_\ell^{TT})$, we can evaluate the four-point function in the integral above,

$$\langle T_{\boldsymbol{\ell}_1}^G T_{\boldsymbol{L}-\boldsymbol{\ell}_1}^G T_{-\boldsymbol{\ell}_2}^G T_{-\boldsymbol{L}+\boldsymbol{\ell}_2}^G\rangle = \underbrace{\langle T_{\boldsymbol{\ell}_1}^G T_{\boldsymbol{L}-\boldsymbol{\ell}_1}^G\rangle\langle T_{-\boldsymbol{\ell}_2}^G T_{-\boldsymbol{L}+\boldsymbol{\ell}_2}^G\rangle}_{=0\text{ since we assume }L\neq0} + \langle T_{\boldsymbol{\ell}_1}^G T_{-\boldsymbol{\ell}_2}^G\rangle\langle T_{\boldsymbol{L}-\boldsymbol{\ell}_1}^G T_{-\boldsymbol{L}+\boldsymbol{\ell}_2}^G\rangle + \langle T_{\boldsymbol{\ell}_1}^G T_{-\boldsymbol{L}+\boldsymbol{\ell}_2}^G\rangle\langle T_{\boldsymbol{L}-\boldsymbol{\ell}_1}^G T_{-\boldsymbol{\ell}_2}^G\rangle$$
$$+ \underbrace{\langle T_{\boldsymbol{\ell}_1}^G T_{\boldsymbol{L}-\boldsymbol{\ell}_1}^G T_{-\boldsymbol{\ell}_2}^G T_{-\boldsymbol{L}+\boldsymbol{\ell}_2}^G\rangle_c}_{=0\text{ for a GRF}}. \tag{D2}$$

We know from Eq. (3) that this then evaluates to

$$\langle T_{\boldsymbol{\ell}_1}^G T_{\boldsymbol{L}-\boldsymbol{\ell}_1}^G T_{-\boldsymbol{\ell}_2}^G T_{-\boldsymbol{L}+\boldsymbol{\ell}_2}^G\rangle = (2\pi)^4 \tilde{C}_{\ell_1}^{TT}\tilde{C}_{\boldsymbol{L}-\boldsymbol{\ell}_1}^{TT}[(\delta^{(D)}(\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2))^2 + (\delta^{(D)}(\boldsymbol{\ell}_1 - \boldsymbol{L} + \boldsymbol{\ell}_2))^2]. \tag{D3}$$

Plugging this into Eq. (10) to get

$$\langle\hat{\kappa}_L\hat{\kappa}_L^*\rangle_{\mathrm{GRF}} = (N_L^\kappa)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} \int \frac{d^2\boldsymbol{\ell}_2}{(2\pi)^2} F_{\boldsymbol{\ell}_1,\boldsymbol{L}-\boldsymbol{\ell}_1}^\kappa F_{-\boldsymbol{\ell}_2,-\boldsymbol{L}+\boldsymbol{\ell}_2}^\kappa (2\pi)^4 \tilde{C}_{\ell_1}^{TT}\tilde{C}_{\boldsymbol{L}-\boldsymbol{\ell}_1}^{TT} \times 2(\delta^{(D)}(\boldsymbol{\ell}_1 - \boldsymbol{\ell}_2))^2$$

$$= (2\pi)^2\delta^{(D)}(0) \times (N_L^\kappa)^2 \int \frac{d^2\boldsymbol{\ell}_1}{(2\pi)^2} F_{\boldsymbol{\ell}_1,\boldsymbol{L}-\boldsymbol{\ell}_1}^\kappa f_{\boldsymbol{\ell}_1,\boldsymbol{L}-\boldsymbol{\ell}_1}^\kappa$$

$$= (2\pi)^2 \delta^{(D)}(0) \times N_L^\kappa \text{ [using Eq. (8)]}$$

$$\equiv (2\pi)^2 \delta^{(D)}(0) \times N_{\text{theory}}, \tag{D4}$$

where in the second line we have a factor of 2 by exploiting the $\boldsymbol{\ell}_2 \to \boldsymbol{L} - \boldsymbol{\ell}_2$ symmetry of the integral, allowing us to combine both Dirac δ's. Now, recall our definition of $\hat{N}$,

$$\hat{N}_L = \frac{2(N_L^\kappa)^2}{(2\pi)^2 \delta^{(D)}(0)} \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell}, \boldsymbol{L}-\boldsymbol{\ell}}^\kappa F_{-\boldsymbol{\ell}, -\boldsymbol{L}+\boldsymbol{\ell}}^\kappa |T_{\boldsymbol{\ell}}|^2 |T_{\boldsymbol{L}-\boldsymbol{\ell}}|^2, \tag{D5}$$

where in the numerator of the prefactor we include the additional finite area correction we suppressed in Eq. (12) as pointed out in footnote 2. Let us also evaluate $\langle \hat{N}_L \rangle$ on the same Gaussian random field $T_{\boldsymbol{\ell}}^G \sim \mathcal{N}(0, \tilde{C}_{\boldsymbol{\ell}}^{TT})$. To start, we can evaluate the four-point function in the integral,

$$\langle T_{\boldsymbol{\ell}}^G T_{-\boldsymbol{\ell}}^G T_{\boldsymbol{L}-\boldsymbol{\ell}}^G T_{-\boldsymbol{L}+\boldsymbol{\ell}}^G \rangle = (2\pi)^4 \tilde{C}_{\boldsymbol{\ell}}^{TT} \tilde{C}_{\boldsymbol{L}-\boldsymbol{\ell}}^{TT} [(\delta^{(D)}(0))^2 + (\delta^{(D)}(2\boldsymbol{\ell} - \boldsymbol{L}))^2]. \tag{D6}$$

Plugging this into Eq. (12) gives us

$$\langle \hat{N}_L \rangle_{\text{GRF}} = \frac{(N_L^\kappa)^2}{(2\pi)^2 \delta^{(D)}(0)} \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell}, \boldsymbol{L}-\boldsymbol{\ell}}^\kappa f_{-\boldsymbol{\ell}, -\boldsymbol{L}+\boldsymbol{\ell}}^\kappa (2\pi)^4 [(\delta^{(D)}(0))^2 + (\delta^{(D)}(2\boldsymbol{\ell} - \boldsymbol{L}))^2]$$

$$= (2\pi)^2 \times \frac{(N_L^\kappa)^2}{\delta^{(D)}(0)} [(N_L^\kappa)^{-1} (\delta^{(D)}(0))^2 \tag{D7}$$

$$+ \frac{1}{4(2\pi)^2} \delta^{(D)}(0) F_{\boldsymbol{L}/2, \boldsymbol{L}/2}^\kappa f_{-\boldsymbol{L}/2, -\boldsymbol{L}/2}^\kappa] \tag{D8}$$

$$\equiv (2\pi)^2 \delta^{(D)}(0) \times N_{\text{theory}}(\boldsymbol{L})(1 + \mathcal{C}). \tag{D9}$$

So we see that there is an extra $\boldsymbol{L}/2$ contraction that leads to an extra term $\mathcal{C}$ which makes Eq. (14) not exact. However, note that $\mathcal{C} \sim N_{\text{theory}} \sim [\int (C_{\boldsymbol{\ell}}/\tilde{C}_{\boldsymbol{\ell}})^2]^{-1}$. If we approximate $C_{\boldsymbol{\ell}} \approx \tilde{C}_{\boldsymbol{\ell}}$ then the integral evaluates to N_{modes}, meaning $\mathcal{C}$ is suppressed by a factor of $1/N_{\text{modes}} \sim 10^{-6}$,

$$\mathcal{C} \sim N_{\text{theory}} \sim \frac{1}{N_{\text{modes}}} \ll 1. \tag{D10}$$

Thus, we neglect this extra contraction coming from $\delta^{(D)}(2\boldsymbol{\ell} - \boldsymbol{L})$.

APPENDIX E: ORIGIN OF OFF-DIAGONAL COVARIANCES OF THE LENSING POWER SPECTRUM

In Sec. V we point out the fact that the $N^{(0)}$ bias and, correspondingly, our $\hat{N}$ has significant off-diagonal correlation $\text{cov}[\hat{N}_L, \hat{N}_{L'}]$. We explicitly plot this strong off-diagonal correlation structure in Fig. 13. In this appendix, we shall explain the reason for this strong off-diagonal correlation.

Let us consider the covariance between $\hat{N}_L$ and $\hat{N}_{L'}$. Recall the language of summing over quadrilaterals to estimate the lensing potential spectrum that we introduced in Eq. (7) and Fig. 1. In this language, as we stated in Fig. 1, $\hat{N}$ can be thought of summing over only parallelograms,
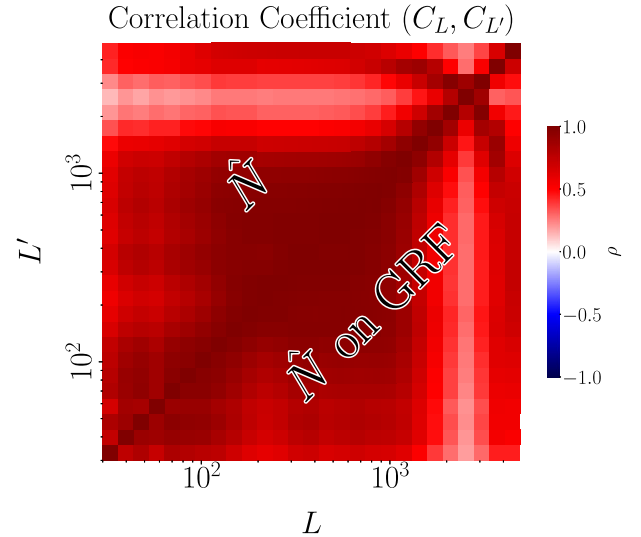

FIG. 13. We saw in Sec. V that there is strong correlation between $\hat{N}_L$ and $\hat{N}_{L'}$. Here we explicitly show this off-diagonal correlation for $\hat{N}$ run on lensed maps (upper left) and $\hat{N}$ run on Gaussian random fields with the same total power spectrum as a lensed map (lower right). We analytically explain the origin of this correlation structure in Appendix E.

$$\hat{N}_{\boldsymbol{L}} \sim \sum_{\boldsymbol{\ell}} \quad \text{[diagram]} \tag{E1}$$

In this language, we can write $\mathrm{cov}[\hat{N}_{\boldsymbol{L}}, \hat{N}_{\boldsymbol{L}'}]$ as

$$\mathrm{cov}[\hat{N}_{\boldsymbol{L}}, \hat{N}_{\boldsymbol{L}'}] \sim \sum_{\boldsymbol{\ell}, \boldsymbol{\ell}'} \quad \text{[diagram]} - \text{[diagram]} \tag{E2}$$

$$= \sum_{\boldsymbol{\ell}, \boldsymbol{\ell}'} \mathrm{cov}\left[T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}} T_{-\boldsymbol{\ell}} T_{\boldsymbol{\ell}-\boldsymbol{L}}, T_{\boldsymbol{\ell}'} T_{\boldsymbol{L}'-\boldsymbol{\ell}'} T_{-\boldsymbol{\ell}'} T_{\boldsymbol{\ell}'-\boldsymbol{L}'} \right],$$

where we implicitly take expectation values over each disjoint diagram. Since we argued in Fig. 3 the dominant contribution to the covariance is $N^{(0)}$, we will approximate $T_{\boldsymbol{\ell}}$ to be a Gaussian random field since this approximation still captures the $N^{(0)}$'s effect on the covariance structure. The only time the term we are summing over in Eq. (E2) is nonzero[10] is when both parallelograms share a leg: $\boldsymbol{\ell} = \boldsymbol{\ell}'$ (or symmetrically $\boldsymbol{\ell} = \boldsymbol{L}' - \boldsymbol{\ell}'$). All other $\{\boldsymbol{\ell}, \boldsymbol{\ell}'\}$ force the parallelograms to be statistically independent and thus do not contribute to the covariance. So we have

$$\mathrm{cov}[\hat{N}_{\boldsymbol{L}}, \hat{N}_{\boldsymbol{L}'}] \sim \sum_{\boldsymbol{\ell}} \quad \text{[diagram]} - \text{[diagram]} \tag{E3}$$

$$= \sum_{\boldsymbol{\ell}} \mathrm{var}\left[T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}} T_{-\boldsymbol{\ell}} T_{\boldsymbol{\ell}-\boldsymbol{L}}, T_{\boldsymbol{\ell}} T_{\boldsymbol{L}'-\boldsymbol{\ell}} T_{-\boldsymbol{\ell}} T_{\boldsymbol{\ell}-\boldsymbol{L}'} \right],$$

Since for a fixed $\boldsymbol{L}$, there is a unique parallelogram associated with having a leg of length $\boldsymbol{\ell}$, we see that each diagram contributing to $\hat{N}_{\boldsymbol{L}}$ is correlated with exactly one diagram contributing to $\hat{N}_{\boldsymbol{L}'}$ and vice versa.[11] This is the origin of the strong off-diagonal covariance due to $N^{(0)}$ and consequently removed by our $\hat{N}$ as well as $\mathrm{RDN}^{(0)}$.

---

[10]Once again neglecting additional contributions that are similar to $\boldsymbol{\ell} = \boldsymbol{L}/2$ in the eight-point function, since we believe they are negligible just like they were in the four-point function as we argued for in Appendix D.

[11]This is true if we had access to all modes. Some modes we cannot see due to finite resolution/size measurements, but for the bulk of the modes this is a negligible effect.

APPENDIX F: DERIVING THE OPTIMAL KURTOSIS ESTIMATOR

In Sec. VI we were considering how to estimate the small kurtosis of a weakly non-Gaussian random variable. This was used as a toy model to understand optimal trispectrum estimation. During this discussion we recalled a result from Ref. [20] that stated the optimal unbiased and minimum-variance estimator for the kurtosis in this toy model is

$$\hat{\mathcal{K}}_{\mathrm{opt}} = \frac{1}{N} \sum_i x_i^4 - \frac{6\sigma^2}{N} \sum_i x_i^2 + 3\sigma^4. \tag{F1}$$

In this appendix, we will spell out how to arrive at this result. Obtaining the form of $\mathrm{RDN}^{(0)}$ described in Eq. (15)

requires a similar argument and we will spell this out in Appendix G. Our derivation in this appendix is directly inspired by a similar discussion for a skewness estimator in Ref. [50].

The optimal kurtosis estimator $\hat{\mathcal{K}}_{\text{opt}}$ stated in Eq. (24) is derived via Edgeworth expansion. Let us add a small kurtosis $\mathcal{K}$ to a normal random variable and call this perturbed random variable X. Then its moment generating function is

$$M_X(J) = \exp\left\{-\frac{1}{2}J^2\sigma^2 + \frac{1}{4!}J^4\mathcal{K}\right\} \tag{F2}$$

$$\approx \exp\left\{-\frac{1}{2}J^2\sigma^2\right\}\left(1 + \frac{1}{4!}J^4\mathcal{K} + \frac{1}{2 \times 4!4!}J^8\mathcal{K}^2\right). \tag{F3}$$

This reconstructs the random variable we were considering in Sec. VI,

$$\langle X^2 \rangle = \sigma^2, \tag{F4}$$

$$\langle X^4 \rangle = 3\sigma^4 + \mathcal{K}. \tag{F5}$$

Before we continue, recall the definition of the Hermite polynomials,

$$h_n(x) = (-1)^n p_{\mathcal{N}}^{-1}(x)\frac{\partial^n}{\partial x^n}p_{\mathcal{N}}(x). \tag{F6}$$

Here, $p_{\mathcal{N}}(x)$ is the probability density function of the Gaussian random variable with zero mean and variance σ^2,

$$p_{\mathcal{N}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{x^2}{\sigma^2}\right\}. \tag{F7}$$

Inverse transforming the moment generating function (F3) gives us the PDF

$$\frac{p_X(x)}{p_{\mathcal{N}}(x)} = 1 + \frac{\mathcal{K}}{4!}h_4(x) + \frac{\mathcal{K}^2}{2 \times 4!4!}h_8(x), \tag{F8}$$

where we have

$$h_4(x) = \frac{x^4 - 6\sigma^2 x^2 + 3\sigma^4}{\sigma^8}, \tag{F9}$$

$$h_8(x) = \frac{x^8 - 28\sigma^2 x^6 + 210\sigma^4 x^4 - 420\sigma^6 x^2 + 105\sigma^8}{\sigma^{16}}. \tag{F10}$$

If we have N independent realization of this random variable $\{x_i\}$, where each $x_i \sim X$, then the log likelihood $\mathcal{L} = \sum \log p_X(x_i)$ becomes

$$\mathcal{L} - \mathcal{L}_{\mathcal{N}} = \sum_i\left[\log\left(1 + \frac{\mathcal{K}}{4!}h_4(x_i) + \frac{\mathcal{K}^2}{2 \times 4!4!}h_8(x_i)\right)\right]$$

$$\approx \sum_i\left[\frac{\mathcal{K}}{4!}h_4(x_i) + \frac{\mathcal{K}^2}{2 \times 4!4!}(h_8(x_i) - h_4^2(x_i))\right]. \tag{F11}$$

Asserting $\partial\mathcal{L}/\partial\mathcal{K} = 0$ we find the maximum likelihood estimate for $\mathcal{K}$ is

$$\hat{\mathcal{K}}_{\text{opt}} = -\frac{4!\sum_i h_4(x_i)}{\sum_i[h_8(x_i) - h_4^2(x_i)]}. \tag{F12}$$

Now let us approximate

$$\sum_{i=1}^{N}x_i^n = N\sigma^n \times (n-1)!! \tag{F13}$$

for even n to derive a relation between h_4 and h_8. Essentially, we are saying we know the variance exactly and so can replace anything where we are estimating the variance with the true variance. Non-Gaussian corrections will be subleading for the purpose of deriving this relation between h_4 and h_8. With this approximation we get

$$\sum_{i=1}^{N}h_8(x_i) = 0, \tag{F14}$$

$$\sum_{i=1}^{N}h_4^2(x_i) = \frac{4! \times N}{\sigma^8}, \tag{F15}$$

$$\Rightarrow \sum_i[h_8(x_i) - h_4^2(x_i)] = -\frac{4! \times N}{\sigma^8}. \tag{F16}$$

Plugging this into Eq. (F12) gives us the optimal estimator asserted by Ref. [20],

$$\hat{\mathcal{K}}_{\text{opt}} = \frac{1}{N}\sum_i x_i^4 - \frac{6\sigma^2}{N}\sum_i x_i^2 + 3\sigma^4. \tag{F17}$$

A slight curiosity is that this estimator looks very similar to $h_4(x)$ written in Eq. (F9).

APPENDIX G: DERIVING RDN$^{(0)}$

The derivation of RDN$^{(0)}$ in Eq. (15) follows from a multivariate Edgeworth expansion of the CMB likelihood (Refs. [26,51,52]). In this appendix, we shall spell out this calculation. This derivation contains the multivariate generalization of the derivation we performed in Appendix F for the optimal kurtosis estimator, Eq. (24). Because of this, we shall parallel the structure of Appendix F in this appendix for clarity.

For discrete data, we can write an arbitrary Fourier mode ℓ_i as a vector of integers times the fundamental ℓ_F,

$$\boldsymbol{\ell}_i = \boldsymbol{i}\ell_F, \qquad (G1)$$

so in this appendix for brevity we shall use the notation $T_{\boldsymbol{\ell}_i} \equiv T_i$. Suppose our lensed temperature map $T_{\boldsymbol{\ell}}$ is approximately described by a Gaussian random field $\mathcal{N}(0, \tilde{C}_{ij})$ but perturbed such that it has small connected four-point function $\mathcal{T}$,

$$\mathcal{T}_{ijkl} = \langle T_{\boldsymbol{\ell}_i} T_{\boldsymbol{\ell}_j} T_{\boldsymbol{\ell}_k} T_{\boldsymbol{\ell}_l} \rangle_c. \qquad (G2)$$

In this case, the moment generating function M of $T_{\boldsymbol{\ell}}$ is

$$M(\boldsymbol{J}) = \exp\left\{-\frac{1}{2} J_i \tilde{C}_{ij} J_j + \frac{1}{4!} J_i J_j J_k J_l \mathcal{T}_{ijkl}\right\}$$

$$\approx \exp\left\{-\frac{1}{2} J_i \tilde{C}_{ij} J_j\right\}\left(1 + \frac{1}{4!} J_i J_j J_k J_l \mathcal{T}_{ijkl}\right.$$

$$\left. + \frac{1}{2 \times 4!4!} J_i J_j J_k J_l \mathcal{T}_{ijkl} \times J_a J_b J_c J_d \mathcal{T}_{abcd}\right), \quad (G3)$$

where repeated indices are summed over. Before we continue, we can define the multivariate generalization of the Hermite polynomial (F8), the Hermite tensor $h_{ij\ldots}$ which has r subscripts,

$$h_{ij\ldots} = (-1)^r p_{\mathcal{N}}^{-1}(T_{\boldsymbol{\ell}})\left[\frac{\partial}{\partial T_i}\frac{\partial}{\partial T_j}\ldots\right] p_{\mathcal{N}}(T_{\boldsymbol{\ell}}). \quad (G4)$$

Here $p_{\mathcal{N}}(T_{\boldsymbol{\ell}})$ is the probability density function of a Gaussian random field with zero mean and covariance matrix $\tilde{C}_{ij}$,

$$p_{\mathcal{N}}(T_{\boldsymbol{\ell}}) = \frac{1}{\sqrt{(2\pi)^N \det[\tilde{C}_{ij}]}} \exp\left\{-\frac{1}{2} T_i \tilde{D}_{ij} T_j\right\}, \quad (G5)$$

where N is the number of $T_{\boldsymbol{\ell}}$ we have and we have defined the inverse covariance matrix $\tilde{D} \equiv (\tilde{C})^{-1}$.

Inverse transforming the moment generating function (G3) gives us the PDF

$$\frac{p(T_{\boldsymbol{\ell}})}{p_{\mathcal{N}}(T_{\boldsymbol{\ell}})} = 1 + \frac{\mathcal{T}_{ijk\ell}}{4!} h_{ijk\ell} + \frac{\mathcal{T}_{ijk\ell}\mathcal{T}_{abcd}}{2 \times 4!4!} h_{ijk\ell abcd}, \quad (G6)$$

where we have

$$h_i = T_a \tilde{D}_{ai}, \qquad (G7)$$

$$h_{ijk\ell} = h_i h_j h_k h_\ell - [h_i h_j \tilde{D}_{k\ell} + (5 \text{ perms})]$$
$$+ [\tilde{D}_{ij}\tilde{D}_{k\ell} + (2 \text{ perms})], \qquad (G8)$$

$$h_{ijk\ell abcd} = h_i h_j h_k h_\ell h_a h_b h_c h_d$$
$$- [h_i h_j h_k h_\ell h_a h_b \tilde{D}_{cd} + (27 \text{ perms})]$$
$$+ [h_i h_j h_k h_\ell \tilde{D}_{ab}\tilde{D}_{cd} + (209 \text{ perms})]$$
$$- [h_i h_j \tilde{D}_{k\ell}\tilde{D}_{ab}\tilde{D}_{cd} + (419 \text{ perms})]$$
$$+ [\tilde{D}_{ij}\tilde{D}_{k\ell}\tilde{D}_{ab}\tilde{D}_{cd} + (104 \text{ perms})]. \qquad (G9)$$

This then yields a log likelihood $\mathcal{L} = \log p(T_{\boldsymbol{\ell}})$ of

$$\mathcal{L} - \mathcal{L}_{\mathcal{N}} = \log\left(1 + \frac{1}{4!}\mathcal{T}_{ijk\ell} h_{ijk\ell} + \frac{\mathcal{T}_{ijk\ell}\mathcal{T}_{abcd}}{2 \times 4!4!} h_{ijk\ell abcd}\right)$$

$$\approx \frac{1}{4!}\mathcal{T}_{ijk\ell} h_{ijk\ell} + \frac{\mathcal{T}_{ijk\ell}\mathcal{T}_{abcd}}{2 \times 4!4!} h_{ijk\ell,abcd}, \qquad (G10)$$

where we defined

$$h_{ijk\ell,abcd} \equiv h_{ijk\ell abcd} - h_{ijk\ell} h_{abcd}. \qquad (G11)$$

Asserting $\partial\mathcal{L}/\partial\mathcal{T}_{ijk\ell} = 0$ to get the ML estimate for the connected trispectrum yields

$$\hat{\mathcal{T}}_{ijk\ell} = -4! h_{abcd} \times (h^{-1})_{abcd,ijk\ell}. \qquad (G12)$$

From here, to find a useful form of $h_{ijk\ell,abcd}$, the generalization of Eq. (F16), we make the generalization of the approximation we made in Eq. (F13) and replace products of T with the appropriate Wick contracted power spectra,

$$T_i T_j \ldots T_k T_\ell = \sum_{\text{Wick}} \tilde{C}_{ab} \times \ldots \times \tilde{C}_{cd}, \qquad (G13)$$

where $\{a, b, \ldots, c, d\}$ is a permutation of $\{i, j, \ldots, k, \ell\}$ and the summation in Eq. (G13) is over all permutations. This approximation allows us to derive[12] a generalization of the relation found in Eq. (F16),

$$h_{abcd,ijk\ell} = -[\tilde{D}_{ai}\tilde{D}_{bj}\tilde{D}_{ck}\tilde{D}_{d\ell} + (23 \text{ perms})], \qquad (G14)$$

where the permutations come from pairings of $\{a, b, c, d\}$ with $\{i, j, k, \ell\}$. Now note in the context of a summation, such as in Eq. (G12), all the terms in Eq. (G14) are the same under relabeling of dummy indices. So Eq. (G12) turns into

$$\hat{\mathcal{T}}_{ijk\ell} = -4! h_{abcd} \times [-24\tilde{D}_{ai}\tilde{D}_{bj}\tilde{D}_{ck}\tilde{D}_{d\ell}]^{-1}$$
$$= h_{abcd} \times \tilde{C}_{ai}\tilde{C}_{bj}\tilde{C}_{ck}\tilde{C}_{d\ell}. \qquad (G15)$$

Plugging in Eq. (G8) then leads to the connected four-point function estimator,

---

[12]With the heavy reliance on *Mathematica* to make this a tractable calculation.

$$\hat{T}_{ijk\ell} = T_i T_j T_k T_\ell - [T_i T_j \tilde{C}_{k\ell} + (5 \text{ perms})] + [\tilde{C}_{ij} \tilde{C}_{jk} + (3 \text{ perms})]. \tag{G16}$$

This is the generalization of Eq. (24).

Now we can specialize this to CMB lensing power spectrum estimation to derive RDN$^{(0)}$. Note from Eq. (10) that our estimate of the CMB lensing power spectrum is not from a general connected four-point function but specifically the combination $\int \langle T_{\ell_1} T_{L-\ell_1} T_{-\ell_2} T_{-L+\ell_2} \rangle_c = \int \mathcal{T}_{\ell_1, L-\ell_1, -\ell_2, -L+\ell_2}$. So our estimator, Eq. (G16), for this specific combination yields

$$\hat{T}_{\ell_1, L-\ell_1, -\ell_2, -L+\ell_2} = T_{\ell_1} T_{L-\ell_1} T_{-\ell_2} T_{-L+\ell_2} - \left[T_{\ell_1} T_{L-\ell_1} \tilde{C}_{-\ell_2, -L+\ell_2} + T_{\ell_1} T_{-\ell_2} \tilde{C}_{L-\ell_1, -L+\ell_2} + T_{\ell_1} T_{-L+\ell_2} \tilde{C}_{L-\ell_1, -\ell_2} \right.$$
$$\left. + T_{L-\ell_1} T_{-\ell_2} \tilde{C}_{\ell_1, -L+\ell_2} + T_{L-\ell_1} T_{-L+\ell_2} \tilde{C}_{\ell_1, -\ell_2} + T_{-\ell_2} T_{-L+\ell_2} \tilde{C}_{\ell_1, L-\ell_1} \right]$$
$$+ \tilde{C}_{\ell_1, L-\ell_1} \tilde{C}_{-\ell_2, -L+\ell_2} + \tilde{C}_{\ell_1, -\ell_2} C_{L-\ell_1, -L+\ell_2} + \tilde{C}_{\ell_1, -L+\ell_2} \tilde{C}_{L-\ell_1, -\ell_2}. \tag{G17}$$

Each temperature map T corresponds to the actual measured data d, whereas the $\tilde{C}$ come from the two sets of simulations $\{s\}, \{s'\}$ needed to compute RDN$^{(0)}$ that we discussed in Sec. IV. Note that the covariance matrix $\tilde{C}_{ij}$ is defined in terms of the two-point function of Gaussian random fields that have the same total power spectrum as the lensed CMB, $\tilde{C}_\ell^{TT}$, not a lensed temperature map. In this case, we can rewrite the covariance matrix in terms of the power spectrum,

$$\tilde{C}_{ij} = \langle T_{\ell_i} T_{\ell_j} \rangle = (2\pi)^2 \delta^{(D)}(\boldsymbol{\ell}_i + \boldsymbol{\ell}_j) \tilde{C}_{\ell_i}^{TT}. \tag{G18}$$

So if we assume $\boldsymbol{L} \neq 0$, then

$$\tilde{C}_{\ell_1, L-\ell_1} = \tilde{C}_{-\ell_2, -L+\ell_2} \propto \delta^{(D)}(\boldsymbol{L}) = 0. \tag{G19}$$

So, the connected four-point function estimator (G16) specialized to CMB lensing power spectrum estimation, Eq. (G17), reduces to

$$\hat{T}_{\ell_1, L-\ell_1, -\ell_2, -L+\ell_2} = \underbrace{T_{\ell_1} T_{L-\ell_1} T_{-\ell_2} T_{-L+\ell_2}}_{\subset C_L(\hat{\kappa}_L, \hat{\kappa}_L), \text{Eq.}(10)} - \left[\underbrace{T_{\ell_1} T_{-\ell_2} \tilde{C}_{L-\ell_1, -L+\ell_2}}_{\subset C_L(\hat{\kappa}^{ds}, \hat{\kappa}^{ds})} + \underbrace{T_{\ell_1} T_{-L+\ell_2} \tilde{C}_{L-\ell_1, -\ell_2}}_{\subset C_L(\hat{\kappa}^{ds}, \hat{\kappa}^{sd})} \right.$$
$$\left. + \underbrace{T_{L-\ell_1} T_{-\ell_2} \tilde{C}_{\ell_1, -L+\ell_2}}_{\subset C_L(\hat{\kappa}^{sd}, \hat{\kappa}^{ds})} + \underbrace{T_{L-\ell_1} T_{-L+\ell_2} \tilde{C}_{\ell_1, -\ell_2}}_{\subset C_L(\hat{\kappa}^{sd}, \hat{\kappa}^{sd})} \right] + \underbrace{\tilde{C}_{\ell_1, -\ell_2} C_{L-\ell_1, -L+\ell_2}}_{\subset C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{ss'})} + \underbrace{\tilde{C}_{\ell_1, -L+\ell_2} \tilde{C}_{L-\ell_1, -\ell_2}}_{\subset C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{s's})}. \tag{G20}$$

From this we can read off the terms in Eq. (15) that define RDN$^{(0)}$. So we have shown how RDN$^{(0)}$ naturally arises from a Edgeworth expansion of the CMB likelihood that we claimed in Sec. IV. We have also established the correspondence between RDN$^{(0)}$ and the optimal Kurtosis estimator (24) for the toy model discussed in Sec. VI and derived in Appendix F.

APPENDIX H: COMPUTING $N^{(1)}$ FROM SIMULATIONS

In addition to the analytic expression for $N^{(1)}$ we derived in Eq. (B30), one can also compute $N^{(1)}$ from simulation as we mentioned in Sec. II A. This makes use of four sets of simulations. First, we have two sets of simulations $\{s_\phi\}$ and $\{s'_\phi\}$ that share a common lensing potential ϕ but different realizations of the unlensed Gaussian CMB. The last two sets of simulations $\{s\}, \{s'\}$ have different realizations of both unlensed Gaussian CMB and lensing potential. From these sets of simulations, one can form an estimator based on simulation for $N^{(1)}$ (Refs. [10,16]),

$$N_L^{(1),\text{MC}} = < C_L(\hat{\kappa}^{s_\phi s'_\phi}, \hat{\kappa}^{s_\phi s'_\phi}) + C_L(\hat{\kappa}^{s_\phi s'_\phi}, \hat{\kappa}^{s'_\phi s_\phi}) $$
$$- C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{ss'}) - C_L(\hat{\kappa}^{ss'}, \hat{\kappa}^{s's}) >_{s, s', s_\phi, s'_\phi}. \tag{H1}$$

To show why this works, let us recall the language of Feynman diagrams that we introduced in Appendix B. Because of how we construct $\{s_\phi\}$ and $\{s'_\phi\}$, the only diagrams contributing to the four-point functions in the first line are the diagrams that contribute to $N^{(0)}$ and $N^{(1)}$,

$$\tag{H2}$$

We have no diagrams that contribute to $\langle \kappa\kappa \rangle$ since those diagrams would require the unlensed temperature map in s_ϕ to be correlated with the unlensed temperature map in s'_ϕ. So, to $O(\kappa^2)$ the first line yields an estimate for $N^{(0)} + N^{(1)}$. The terms on the second line yield only diagrams[13] contributing $N^{(0)}$ at $O(\kappa)$ and no diagrams at $O(\kappa^2)$ since the lensing potential κ is no longer shared. So, the second line shares only a $N^{(0)}$ contribution with the first line. Thus, the difference of the two lines yields an simulation-based estimate of $N^{(1)}$.

We make use of a hybrid approach to estimate $N^{(1)}$ in our numerical studies throughout this paper. For a full-sky analysis, Eq. (H1) is sufficient to estimate $N^{(1)}$ for all scales. However, in our numerical studies we work on small patches where the flat-sky approximation holds. In this case, Eq. (H1) cannot easily converge to an estimate of $N^{(1)}$ for very large scales where the number of modes is small. However, one can see from, e.g., Fig. 2 that large scales are also the regime where the $N^{(1)}$ is order of magnitudes smaller than the signal $\langle \kappa\kappa \rangle$. So, our estimate of $N^{(1)}$ does not need to be perfect for these regimes. Because of this, we use the analytical result, Eq. (B30), to compute $N_L^{(1)}$ for $L < 512$. For $L > 512$, we use the simulation-based computation (H1) to compute $N_L^{(1)}$, since on small scales we have sufficient statistics to estimate $N^{(1)}$ from simulations.

APPENDIX I: FAST ALGORITHM TO COMPUTE $\hat{N}^\times$

In Sec. VIII, Eqs. (35)–(37) we described how our proposed estimator could be combined with split-based methods recently proposed in Ref. [18] to build estimators insensitive to modeling of instrument noise. Suppose we have m splits of the CMB map with independent instrument noise. To combine our proposed noise avoidance with the split-based method, we must compute Eq. (36). However, Eq. (36) is naively a $O(m^4)$ computation. In this appendix, we will construct a fast $O(m^2)$ algorithm to compute the $\hat{N}_L^\times$ described in Eq. (36) similar to the fast algorithm discussed in Ref. [18].

To start, we use an identity stated by Ref. [18],

$$
\gamma_{ijkl} = \gamma_i \gamma_j \gamma_k \gamma_l - [\delta_{ij}^{(K)} \gamma_k \gamma_l + (5 \text{ perms})]
$$
$$
+ 2[\delta_{ijk}^{(K)} \gamma_l + (3 \text{ perms})]
$$
$$
+ [\delta_{ij}^{(K)} \delta_{kl}^{(K)} + (2 \text{ perms})] - 6\delta_{ijkl}^{(K)}. \quad \text{(I1)}
$$

Here $\delta^{(K)}$ is a Kronecker δ,

$$
\delta_{i_1 \cdots i_n}^{(K)} = \begin{cases} 1 & i_1 = \ldots = i_n \\ 0 & \text{otherwise} \end{cases}. \quad \text{(I2)}
$$

For future convenience, let us define several things,

$$
Q_\ell^{ij} = |T_\ell^{(i)} T_{-\ell}^{(j)}|, \quad \text{(I3)}
$$

$$
Q^{i+} = \frac{1}{m} \sum_{j=1}^m Q^{ij}, \quad \text{(I4)}
$$

$$
Q^{i\times} = Q^{i+} - \frac{1}{m} Q^{ii}, \quad \text{(I5)}
$$

$$
Q^{++} = \frac{1}{m^2} \sum_{i,j=1}^m Q^{ij}, \quad \text{(I6)}
$$

$$
Q^{\times\times} = Q^{++} - \frac{1}{m^2} \sum_i Q^{ii}. \quad \text{(I7)}
$$

This allows us to define

$$
\hat{N}_L^{\alpha,\beta} = 2(N_L^\kappa)^2 \int_\ell F_{\ell,L-\ell}^\kappa F_{-\ell,-L+\ell}^\kappa \times Q_\ell^\alpha Q_{L-\ell}^\beta. \quad \text{(I8)}
$$

Here, α and β can be any pair of the superscripts $i, +, \times$ appearing in Eqs. (I3)–(I7). Note that we already introduced a special case of this, $\hat{N}^{ij,kl}$, in Eq. (35).

To create a fast algorithm to compute Eq. (36) we can apply the identity in Eq. (I1),

$$
\sum_{ijkl} \gamma_i \gamma_j \gamma_k \gamma_l \hat{N}_L^{ij,kl} = m^4 \hat{N}_L^{++,++}
$$

$$
-\sum_{ijkl} \begin{bmatrix} \delta_{ij}^{(K)} \gamma_k \gamma_l \\ +(5 \text{ perm}) \end{bmatrix} \hat{N}_L^{ij,kl} = -2m^2 \sum_i [2\hat{N}_L^{i+,i+} + \hat{N}_L^{ii,++}]
$$

$$
2\sum_{ijkl} \begin{bmatrix} \delta_{ijk}^{(K)} \gamma_l \\ +(3 \text{ perm}) \end{bmatrix} \hat{N}_L^{ij,kl} = 8m \sum_i \hat{N}_L^{+i,ii}
$$

$$
\sum_{ijkl} \begin{bmatrix} \delta_{ij}^{(K)} \delta_{kl}^{(K)} \\ +(2 \text{ perm}) \end{bmatrix} \hat{N}_L^{ij,kl} = \sum_{ij} [\hat{N}_L^{ii,jj} + 2\hat{N}^{ij,ij}]
$$

$$
-6\sum_{ijkl} \delta_{ijkl}^{(K)} \hat{N}^{ik,jl} = -6 \sum_i \hat{N}^{ii,ii}. \quad \text{(I9)}
$$

We can simplify things tremendously by combining terms,

$$
\hat{N}_L^\times = \frac{1}{m(m-1)(m-2)(m-3)}
$$
$$
\times \left[m^4 \hat{N}_L^{\times\times,\times\times} - 4m^2 \sum_i \hat{N}_L^{i\times,i\times} + 4\sum_{i<j} \hat{N}_L^{ij,ij} \right]. \quad \text{(I10)}
$$

Using this, we are able to compute $\hat{N}_L^\times$ [Eq. (36)] in $O(m^2)$.

---

[13]Neglecting the same diagrams we neglected in footnote 8 for the same reasons.

[1] Antony Lewis and Anthony Challinor, Weak gravitational lensing of the CMB, Phys. Rep. **429**, 1 (2006).

[2] R. Allison, P. Caucal, E. Calabrese, J. Dunkley, and T. Louis, Towards a cosmological neutrino mass detection, Phys. Rev. D **92**, 123535 (2015).

[3] Marcel Schmittfull and Uros Seljak, Parameter constraints from cross-correlation of CMB lensing with galaxy clustering, Phys. Rev. D **97**, 123540 (2018).

[4] Zack Li, Vera Gluscevic, Kimberly K. Boddy, and Mathew S. Madhavacheril, Disentangling dark physics with cosmic microwave background experiments, Phys. Rev. D **98**, 123524 (2018).

[5] Blake D. Sherwin *et al.*, Evidence for dark energy from the cosmic microwave background alone using the Atacama Cosmology Telescope lensing measurements, Phys. Rev. Lett. **107**, 021302 (2011).

[6] Kendrick M. Smith, Oliver Zahn, and Olivier Dore, Detection of gravitational lensing in the cosmic microwave background, Phys. Rev. D **76**, 043510 (2007).

[7] Sudeep Das *et al.*, Detection of the power spectrum of cosmic microwave background lensing by the Atacama Cosmology Telescope, Phys. Rev. Lett. **107**, 021301 (2011).

[8] A. van Engelen *et al.*, A measurement of gravitational lensing of the microwave background using South Pole Telescope data, Astrophys. J. **756**, 142 (2012).

[9] P. A. R. Ade *et al.* (POLARBEAR Collaboration), Measurement of the cosmic microwave background polarization lensing power spectrum with the POLARBEAR experiment, Phys. Rev. Lett. **113**, 021301 (2014).

[10] K. T. Story *et al.* (SPT Collaboration), A measurement of the cosmic microwave background gravitational lensing potential from 100 square degrees of SPTpol data, Astrophys. J. **810**, 50 (2015).

[11] P. A. R. Ade *et al.* (BICEP2, Keck Array Collaborations), BICEP2 / Keck Array VIII: Measurement of gravitational lensing from large-scale B-mode polarization, Astrophys. J. **833**, 228 (2016).

[12] Blake D. Sherwin *et al.*, Two-season Atacama Cosmology Telescope polarimeter lensing power spectrum, Phys. Rev. D **95**, 123529 (2017).

[13] Y. Omori *et al.*, A 2500 deg^2 CMB lensing map from combined South Pole Telescope and Planck data, Astrophys. J. **849**, 124 (2017).

[14] W. L. K. Wu *et al.*, A measurement of the cosmic microwave background lensing potential and power spectrum from 500 deg^2 of SPTpol temperature and polarization data, Astrophys. J. **884**, 70 (2019).

[15] F. Bianchini *et al.* (SPT Collaboration), Constraints on cosmological parameters from the 500 deg^2 SPTpol lensing power spectrum, Astrophys. J. **888**, 119 (2020).

[16] Frank J. Qu *et al.* (ACT Collaboration), The Atacama Cosmology Telescope: A measurement of the DR6 CMB lensing power spectrum and its implications for structure growth, Astrophys. J. **962**, 112 (2024).

[17] Z. Pan *et al.* (SPT Collaboration), Measurement of gravitational lensing of the cosmic microwave background using SPT-3G 2018 data, Phys. Rev. D **108**, 122005 (2023).

[18] Mathew S. Madhavacheril, Kendrick M. Smith, Blake D. Sherwin, and Sigurd Naess, CMB lensing power spectrum estimation without instrument noise bias, J. Cosmol. Astropart. Phys. 05 (2021) 028.

[19] Blake D. Sherwin and Sudeep Das, CMB lensing—power without bias, arXiv:1011.4510.

[20] Kendrick M. Smith, Leonardo Senatore, and Matias Zaldarriaga, Optimal analysis of the CMB trispectrum, arXiv:1502.00635.

[21] Christopher M. Hirata and Uros Seljak, Reconstruction of lensing from the cosmic microwave background polarization, Phys. Rev. D **68**, 083002 (2003).

[22] Vanessa Böhm, Marcel Schmittfull, and Blake D. Sherwin, Bias to CMB lensing measurements from the bispectrum of large-scale structure, Phys. Rev. D **94**, 043519 (2016).

[23] Vanessa Böhm, Blake D. Sherwin, Jia Liu, J. Colin Hill, Marcel Schmittfull, and Toshiya Namikawa, Effect of non-Gaussian lensing deflections on CMB lensing measurements, Phys. Rev. D **98**, 123510 (2018).

[24] Dominic Beck, Giulio Fabbian, and Josquin Errard, Lensing reconstruction in post-Born cosmic microwave background weak lensing, Phys. Rev. D **98**, 043512 (2018).

[25] Giulio Fabbian, Antony Lewis, and Dominic Beck, CMB lensing reconstruction biases in cross-correlation with large-scale structure probes, J. Cosmol. Astropart. Phys. 10 (2019) 057.

[26] Toshiya Namikawa, Duncan Hanson, and Ryuichi Takahashi, Bias-hardened CMB lensing, Mon. Not. R. Astron. Soc. **431**, 609 (2013).

[27] Planck Collaboration, Planck 2018 results. VIII. Gravitational lensing, Astron. Astrophys. **641**, A8 (2020).

[28] Michael H. Kesden, Asantha Cooray, and Marc Kamionkowski, Lensing reconstruction with CMB temperature and polarization, Phys. Rev. D **67**, 123507 (2003).

[29] Duncan Hanson, Anthony Challinor, George Efstathiou, and Pawel Bielewicz, CMB temperature lensing power reconstruction, Phys. Rev. D **83**, 043005 (2011).

[30] Elizabeth E. Jenkins, Aneesh V. Manohar, Wouter J. Waalewijn, and Amit P. S. Yadav, Higher-order gravitational lensing reconstruction using Feynman diagrams, J. Cosmol. Astropart. Phys. 09 (2014) 024.

[31] Elizabeth E. Jenkins, Aneesh V. Manohar, Wouter J. Waalewijn, and Amit P. S. Yadav, Gravitational lensing of the CMB: A Feynman diagram approach, Phys. Lett. B **736**, 6 (2014).

[32] Antony Lewis, Anthony Challinor, and Duncan Hanson, The shape of the CMB lensing bispectrum, J. Cosmol. Astropart. Phys. 03 (2011) 018.

[33] P. A. R. Ade *et al.* (Planck Collaboration), Planck 2015 results. XV. Gravitational lensing, Astron. Astrophys. **594**, A15 (2016).

[34] N. Aghanim *et al.* (Planck Collaboration), Planck 2018 results. VIII. Gravitational lensing, Astron. Astrophys. **641**, A8 (2020).

[35] Mark Mirmelstein, Julien Carron, and Antony Lewis, Optimal filtering for CMB lensing reconstruction, Phys. Rev. D **100**, 123509 (2019).

[36] David Alonso, Javier Sanchez, and Anže Slosar (LSST Dark Energy Science Collaboration), A unified pseudo-C_ℓ framework, Mon. Not. R. Astron. Soc. **484**, 4127 (2019).

[37] E. Hivon, K. M. Gorski, C. B. Netterfield, B. P. Crill, S. Prunet, and F. Hansen, Master of the cosmic microwave background anisotropy power spectrum: A fast method for statistical analysis of large and complex cosmic microwave background data sets, Astrophys. J. **567**, 2 (2002).

[38] Noah Sailer, Gerrit Farren, Simone Ferraro, and Martin White (to be published).

[39] M. Millea *et al.*, Optimal cosmic microwave background lensing reconstruction and parameter estimation with SPTpol data, Astrophys. J. **922**, 259 (2021).

[40] Wayne Hu and Takemi Okamoto, Mass reconstruction with cosmic microwave background polarization, Astrophys. J. **574**, 566 (2002).

[41] Noah Sailer, Simone Ferraro, and Emmanuel Schaan, Foreground-immune CMB lensing reconstruction with polarization, Phys. Rev. D **107**, 023504 (2023).

[42] Noah Sailer, Emmanuel Schaan, and Simone Ferraro, Lower bias, lower noise CMB lensing with foreground-hardened estimators, Phys. Rev. D **102**, 063517 (2020).

[43] Abhishek S. Maniyar, Yacine Ali-Haïmoud, Julien Carron, Antony Lewis, and Mathew S. Madhavacheril, Quadratic estimators for CMB weak lensing, Phys. Rev. D **103**, 083524 (2021).

[44] https://github.com/DelonShen/LensQuEst

[45] Alba Kalaja, Giorgio Orlando, Aleksandr Bowkis, Anthony Challinor, P. Daniel Meerburg, and Toshiya Namikawa, The reconstructed CMB lensing bispectrum, J. Cosmol. Astropart. Phys. 04 (2023) 041.

[46] Antony Lewis, Anthony Challinor, and Anthony Lasenby, Efficient computation of CMB anisotropies in closed FRW models, Astrophys. J. **538**, 473 (2000).

[47] Antony Lewis and Anthony Challinor, CAMB: Code for anisotropies in the microwave background, Astrophysics Source Code Library, record ascl:1102.026 (2011).

[48] Cullan Howlett, Antony Lewis, Alex Hall, and Anthony Challinor, CMB power spectrum parameter degeneracies in the era of precision cosmology, J. Cosmol. Astropart. Phys. 04 (2012) 027.

[49] Emmanuel Schaan and Simone Ferraro, Foreground-immune cosmic microwave background lensing with shear-only reconstruction, Phys. Rev. Lett. **122**, 181301 (2019).

[50] Luca Amendola, Non-Gaussian likelihood function and COBE data, Mon. Not. R. Astron. Soc. **283**, 983 (1996).

[51] D. M. Regan, E. P. S. Shellard, and J. R. Fergusson, General CMB and primordial trispectrum estimation, Phys. Rev. D **82**, 023520 (2010).

[52] Marcel M. Schmittfull, Anthony Challinor, Duncan Hanson, and Antony Lewis, Joint analysis of CMB temperature and lensing-reconstruction power spectra, Phys. Rev. D **88**, 063012 (2013).