


Model independent variance cancellation in CMB lensing cross-correlations

Antón Baleato Lizancos^{*} and Simone Ferraro[†]

*Berkeley Center for Cosmological Physics, Department of Physics, University of California,
Berkeley, California 94720, USA
and Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, California 94720, USA*

 (Received 25 January 2023; revised 4 April 2023; accepted 13 June 2023; published 23 June 2023)

Cross-correlations of CMB lensing reconstructions with other tracers of matter constrain primordial non-Gaussianity, neutrino masses and structure growth as a function of cosmic time. We formalize a method to improve the precision of these measurements by using a third tracer to remove structure from the lensing reconstructions. Crucially, our method enjoys the variance reduction benefits of a joint-modeling approach without the need to model the cosmological dependence of the ancillary tracer. We present a first demonstration of variance cancellation using data from Planck and the DESI Legacy Surveys, showing a 10%–20% reduction in both lensing power and cross-correlation variance using the cosmic infrared background (CIB) or DESI Legacy Survey luminous red galaxies (LRGs) as matter tracers.

DOI: [10.1103/PhysRevD.107.123532](https://doi.org/10.1103/PhysRevD.107.123532)

I. INTRODUCTION

The cosmic microwave background (CMB) is the oldest light we can observe; it is made up of photons which (for the most part) last scattered at redshift $z \approx 1100$. The CMB we see has been gravitationally lensed by the distribution of matter—both luminous and dark—that the photons encountered along their trajectory, an effect that can be harnessed to reconstruct maps of that very matter distribution in projection (see [1] for a review).

These reconstructions can in turn be cross-correlated with other tracers of matter to extract insights that cannot be gleaned with either tracer alone. This is one of the most promising ways to measure the growth of cosmic structures, primordial non-Gaussianity, or the sum of the neutrino masses [2]. Moreover, cross-correlations make it possible to isolate contributions from different redshifts, a prized property in times of tantalizing discrepancies between probes of early and late cosmic times (see, e.g., [3–7] and references therein). Heuristically, if $\hat{\kappa}$ is a reconstruction of the CMB lensing convergence¹ and g is some other tracer of the matter distribution—such as a galaxy survey—with redshift support z_g , the “clumpiness” of matter at the time corresponding to z_g can be determined from a ratio of angular spectra, $\sigma_8(z_g) \sim C^{\hat{\kappa}g} / \sqrt{C^{gg}}$, where σ_8 refers to the amplitude of the linear matter power spectrum on a scale of $8 \text{ h}^{-1} \text{ Mpc}$ [8,9]. Similarly, galaxy bias, including any scale dependence induced by

primordial non-Gaussianity, can be extracted from $b(z_g) \sim C^{gg} / C^{\hat{\kappa}g}$. For a typical tracer g , its autocorrelation is measured much more accurately than its cross-correlation with lensing, so the uncertainty on $\sigma_8(z_g)$ and $b(z_g)$ is dominated by the error on $C^{\hat{\kappa}g}$.

Reducing the cross-correlation error requires limiting chance correlations between features in the lensing and galaxy maps. One way to account for these is to introduce a third tracer, I , which correlates with structures in $\hat{\kappa}$ that are not correlated with g , and modeling everything jointly (e.g., [2]). Often times, however, we might be interested in obtaining constraints that are independent of I , be it because the tracer cannot be modeled easily or accurately, or to avoid introducing a dependence on physics from the cosmic era sourcing I .

In this work, we explore an alternative approach to variance reduction that relies solely on measurable quantities and limits the dependence on redshifts different from those we intend to isolate. Our method entails subtracting a filtered version of I from $\hat{\kappa}$ in a procedure analogous to “delensing.” In Sec. II, we derive the optimal form of these filters and forecast potential gains in signal-to-noise. Then, in Sec. III, we demonstrate that the variance reduction seen on real data from Planck and the DESI Legacy Surveys matches theoretical expectations. In an Appendix, we generalize our method to the case where g and I are correlated, explaining how to account for this in real analyses.

II. THEORY

Consider $\hat{C}_l^{\hat{\kappa}g}$, the measured angular cross-correlation of $\hat{\kappa}$ and g . In general, $\hat{C}_l^{\hat{\kappa}g}$ follows a χ^2 distribution with $2l + 1$

^{*}a.baleatolizancos@berkeley.edu

[†]sferraro@lbl.gov

¹We write the reconstruction as $\hat{\kappa}$ to differentiate it from the true κ . Note, however, that g and I are noisy observations.

degrees of freedom, but away from the lowest few multipoles the distribution can be approximated as Gaussian by the central limit theorem. In this regime, the variance of the i th multipole bin of $\hat{C}_l^{\hat{\kappa}g}$ is given by

$$\sigma^2(\hat{C}_l^{\hat{\kappa}g}) = \frac{1}{(2l_i + 1)f_{\text{sky}}\Delta l} [C_i^{\hat{\kappa}\hat{\kappa}}C_i^{gg} + (C_i^{\hat{\kappa}g})^2], \quad (1)$$

where f_{sky} is the fraction of sky covered by the observations, Δl is the width of the bins (which we assume to be uniform for simplicity), and l_i is the central multipole of the i th bin. This expression suggests that, by removing structure from $\hat{\kappa}_{lm}$, we can suppress $C_i^{\hat{\kappa}\hat{\kappa}}$ and thus lower the variance of the measurement.²

With this goal in mind, let us introduce a third tracer, I , which is partially correlated with $\hat{\kappa}$ (and possibly also with g). This can be any tracer of low redshift matter such as a the cosmic infrared background (CIB) or a galaxy density or weak lensing (shear) field. We can use it to obtain a “redshift-cleaned” convergence map as³

$$\hat{\kappa}_{lm}^{\text{cIn}} = \hat{\kappa}_{lm} - f_{lm}I_{lm}, \quad (2)$$

where f is a filter to be optimized shortly; throughout this work, we will assume that I is statistically homogeneous such that the optimal f is isotropic (i.e., independent of m), but the method applies more generally.

Recently, Refs. [10–12] determined the weights that approximately null contributions from a given redshift range. Let us instead determine the choice of f that maximizes the signal-to-noise on a measurement of the cross-correlation. The signal-to-noise ratio of the i th bin is defined as

$$(S/N)_i \equiv \frac{C_i^{\hat{\kappa}g}}{\sigma(\hat{C}_i^{\hat{\kappa}g})} = \sqrt{\frac{(2l_i + 1)f_{\text{sky}}\Delta l}{1 + (\rho_i^{\hat{\kappa}g})^{-2}}}, \quad (3)$$

where $\rho_i^{\hat{\kappa}g}$ is the correlation coefficient between the lensing reconstruction, $\hat{\kappa}$, and tracer g , and is defined as⁴

$$\rho_i^{\hat{\kappa}g} = \frac{C_i^{\hat{\kappa}g}}{\sqrt{C_i^{\hat{\kappa}\hat{\kappa}}C_i^{gg}}}, \quad (4)$$

where we take $C_i^{\hat{\kappa}\hat{\kappa}}$ and C_i^{gg} to include reconstruction noise and shot noise, respectively. It follows that the choice of f_i

that maximizes the signal-to-noise is also that which maximizes $\rho_l^{\hat{\kappa}g}$.

There are two effects we must consider when maximizing $\rho_l^{\hat{\kappa}g}$ with respect to f : on the one hand, any nonzero field we add to $\hat{\kappa}$ will affect the variance (i.e., “noise”) of $\hat{\kappa}^{\text{cIn}}$; on the other, if tracers g and I are correlated, the cross-correlation signal will itself be impacted. Taking both into account, we determine the optimal filter to be

$$f_l = \frac{C_l^{\kappa I}}{C_l^{II}} \left(\frac{\rho_l^{\hat{\kappa}g} - \rho_l^{gI}/\rho_l^{\hat{\kappa}I}}{\rho_l^{\hat{\kappa}g} - \rho_l^{gI}\rho_l^{\hat{\kappa}I}} \right) \equiv \frac{C_l^{\kappa I}}{C_l^{II}} \gamma_l, \quad (5)$$

granted $\rho_l^{gI} \neq 1$ (when $\rho_l^{gI} = 1$, g and I are one and the same tracer, and one must logically set $f_l = 0$). Note that C_l^{II} includes all sources of noise. If tracers g and I are completely uncorrelated, $\rho_l^{gI} = 0$, so $\gamma_l = 1$ and $f \rightarrow C_l^{\kappa I}/C_l^{II}$ —a form familiar from several applications of variance reduction in cosmology [11,13].

The form of f_l given in Eq. (5) guarantees an improvement in the S/N of the cross-correlation. However, the goal of this work is to achieve this while *removing* structure in $\hat{\kappa}$ that correlates with I ; this sets the additional requirement that $f_l > 0$ [cf. Eq. (2)].⁵ It can be shown that $f_l > 0$ if and only if

$$\rho_l^{\hat{\kappa}I} > \frac{\rho_l^{gI}}{\rho_l^{\hat{\kappa}g}} \quad \text{or} \quad \rho_l^{\hat{\kappa}I} > \frac{\rho_l^{\hat{\kappa}g}}{\rho_l^{gI}}. \quad (6)$$

When $\rho_l^{gI} = 0$, the first condition is met automatically. On the other hand, when $\rho_l^{gI} \neq 0$ the situation is more nuanced, as the $\rho_l^{\hat{\kappa}I} - \rho_l^{\hat{\kappa}g}$ plane splits into regions where either one or none of the above conditions are satisfied: for reference, a prerequisite for the first condition to be met is that $\rho_l^{\hat{\kappa}I} > \rho_l^{gI}$; for the second, the proviso is $\rho_l^{\hat{\kappa}g} < \rho_l^{gI}$. We study this general case in detail in the Appendix A.

Though both of these qualitatively-different scenarios in principle allow for variance cancellation that results in improved S/N, they differ in their practical benefits. When $\rho_l^{gI} \neq 0$, our efforts to cancel variance will to some extent entail a reduction of the signal, making modeling difficult unless the removed signal can be accounted for accurately enough; moreover we will have introduced an undesired dependence on tracer I . In the Appendix A, we show that when $\rho_l^{\hat{\kappa}I} > \rho_l^{gI}/\rho_l^{\hat{\kappa}g}$, measurement uncertainties are typically small enough that the removed signal can either be ignored or restored based on direct measurements of \hat{C}_l^{gI} , with the benefits from variance cancellation outweighing the additional error introduced when characterizing this

²It follows from Eq. (1) that removing the correlated part also leads to lower variance, though the gain is typically subdominant to that stemming from a reduction in $C_i^{\hat{\kappa}\hat{\kappa}}$.

³The same effect can be achieved by including $f_i C_l^{\hat{\kappa}I}$ in the likelihood.

⁴This is not to be confused with the correlation between g and the true lensing convergence, $\rho_i^{\kappa g} = \rho_i^{\hat{\kappa}g}(1 + N_i^{\kappa\kappa}/C_i^{\kappa\kappa})^{1/2}$, where $N_i^{\kappa\kappa}$ is the power spectrum of the reconstruction noise.

⁵Assuming I is positively correlated with κ . The converse holds when they are anticorrelated, as would be the case for a map of the tSZ effect below 217 GHz, for example.

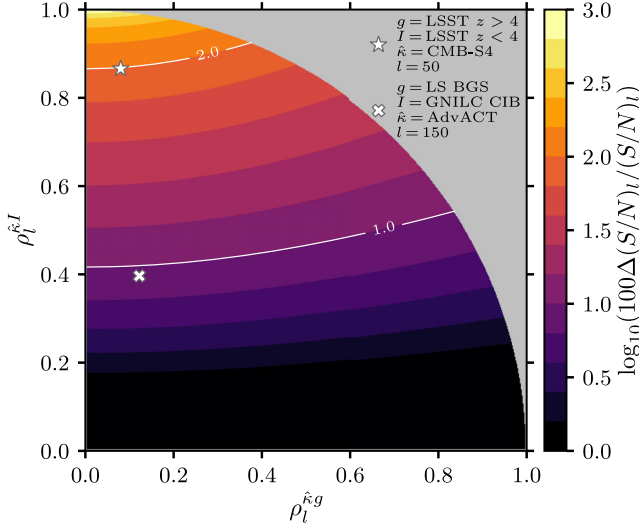


FIG. 1. Fractional change in the signal-to-noise ratio per mode of $\hat{C}_l^{\hat{\kappa}g}$ after cleaning $\hat{\kappa}$ with an optimally filtered tracer I , in the limit that $\rho_l^{gl} = 0$. The gray region corresponds to values of $\rho_l^{\hat{\kappa}g}$ and $\rho_l^{\hat{\kappa}I}$ incompatible with $\rho_l^{gl} = 0$ (see the Appendix B). The star and cross represent two particular combinations of data outlined in the legend and described in the main text.

correction term empirically. However, when $\rho_l^{\hat{\kappa}I} > \rho_l^{\hat{\kappa}g}/\rho_l^{gl}$ the penalty from restoring the signal is higher, and values of $\{\rho_l^{\hat{\kappa}I}, \rho_l^{\hat{\kappa}g}, \rho_l^{gl}\}$ that both satisfy $\rho_l^{\hat{\kappa}I} > \rho_l^{\hat{\kappa}g}/\rho_l^{gl}$ and lead to overall improved precision are in most cases only possible if \hat{C}_l^{gl} can be measured on a patch of sky larger than the one where we carry out the redshift cleaning. In fact, it is true more generally that the balance of factors—and thus the change in S/N after cleaning—depends on the relative sizes of these two patches.

Leaving the caveat of signal restoration to the Appendix A, the fractional change in signal-to-noise per mode of $C_l^{\hat{\kappa}g}$ that can be attained using our optimal weights is

$$\frac{\Delta(S/N)_l}{(S/N)_l} = \left\{ 1 + [1 + (\rho_l^{\hat{\kappa}g})^2]^{-1} \left[\frac{1 + \gamma_l(\gamma_l - 2)(\rho_l^{\hat{\kappa}I})^2}{(1 - \gamma_l \rho_l^{gl} \rho_l^{\hat{\kappa}I} / \rho_l^{\hat{\kappa}g})^2} - 1 \right] \right\}^{-\frac{1}{2}} - 1. \quad (7)$$

Figure 1 shows the improvement in signal-to-noise as a function of $\rho_l^{\hat{\kappa}g}$ and $\rho_l^{\hat{\kappa}I}$ in the limit that g and I are uncorrelated. Improvements of several tens of percent are achievable with realistic tracers, especially whenever $\rho_l^{\hat{\kappa}I}$ is significantly higher than $\rho_l^{\hat{\kappa}g}$. For a given value of $\rho_l^{\hat{\kappa}I}$, the fractional improvement in S/N is larger the smaller $\rho_l^{\hat{\kappa}g}$ is. This is because with lower $\rho_l^{\hat{\kappa}g}$, the term proportional to $C_l^{\hat{\kappa}g}$ dominates the error in Eq. (1) more clearly, amplifying the impact of cleaning. The method is thus especially suited to improve low-significance measurements or to make first detections.

To illustrate the promise of redshift cleaning, we consider as a potential application the use of low-redshift ($z < 4$) measurements from the future Vera Rubin Observatory (VRO) LSST [14] to clean CMB-S4 [15] lensing reconstructions, and subsequently correlating these with the highest-redshift bin of LSST ($z > 4$).⁶ We find that the S/N of the cross-correlation is improved by $\sim 100\%$ at $l \sim 50$, as shown by the star in Fig. 1.⁷ This gain could be important, for instance, when searching for f_{NL} , the signal of which peaks at high redshifts and large angular scales [2]. Since CMB-S4 is overwhelmingly signal dominated at these low l s, the only way to improve measurement significance is by getting around the cosmic variance of κ . Our method does precisely this while bypassing the need to model the low redshift tracers.⁸

Another promising application is to improve measurements of σ_8 . With existing data, the gains are modest: the cross in Fig. 1 shows a 10% gain in precision of $C_{l=150}^{\hat{\kappa}g}$ when using GNILC CIB to clean AdvACT DR6 and cross-correlate with Legacy Survey BGS (assuming $\rho_l^{gl} = 0$; more details about these tracers to come). However, the figure also shows that improvements of a factor of two or greater on the large-scale amplitude of fluctuations are possible with upcoming experiments.

III. DEMONSTRATION

We now present a first demonstration of variance cancellation by cross-correlating CMB lensing data from Planck with galaxy survey data from the DESI Legacy Survey (LS) [16]. Since the benefits of our method are predicated on being able to remove true lensing modes that do not correlate with g , and the Planck lensing noise levels are relatively high—only a handful of modes are signal-dominated [17]—it is at present difficult to find a combination of datasets for which the S/N improves. However, this will soon change dramatically as data from AdvACT [18], SO [19], SPT-3G [20], and CMB-S4 [15] become available. It is therefore useful to demonstrate that variance cancellation can be achieved and understood irrespective of whether the S/N improves for the particular scenario under consideration. Hence, and in order to maximize the variance reduction, we set $\gamma_l = 1$ in the weights for the remainder of this section.

⁶We assume the LSST bins to be disjoint. Further details about our parametrization of the tracers are given in the Appendix C.

⁷Improvements are even larger at lower l , but we do not focus on those scales here as they require a more refined treatment due to the Limber approximation breaking down and systematic effects becoming more important.

⁸Further work is needed to assess the impact of decorrelation between κ and I due to scale-dependent bias and nonlinearities.

A. Data

We work with the minimum variance combination of CMB lensing reconstructions obtained from temperature and polarization data from the fourth data release of Planck [17]. The resulting $\hat{\kappa}$ map covers $\sim 67\%$ of the sky and is signal-dominated on scales $10 \lesssim L \lesssim 70$.

As per large-scale structure tracers, we consider various samples of galaxies photometrically selected from Legacy Survey data for spectroscopic follow-up with DESI. These were spectroscopically calibrated during DESI survey validation. We define a ‘‘BGS’’ sample as the ‘‘bright’’ ($r < 19.5$) subset of targets to be observed by the DESI Bright Galaxy Survey (BGS) [21,22]. This sample is relatively-low redshift: Fig. 18 of Ref. [21] shows its redshift distribution, which is mostly confined to $z < 0.5$. We work also with an ‘‘LRG’’ (Luminous Red Galaxy) sample, using directly the maps provided by Ref. [4], which are split into four redshift bins with redshift distribution given in their Fig. 2; this sample is explained in detail in [4,23]. Since the first bin has significant overlap with the BGS sample, we will exclude it whenever we use the LRGs as tracer I and BGS as tracer g in order to limit the correlation between the two samples; the remaining LRG sample spans $0.5 < z < 1$. The galaxy maps thus produced cover $\sim 50\%$ and $\sim 44\%$ of the sky for BGS and the full LRG sample, respectively, and are largely contained within the Planck $\hat{\kappa}$ footprint. Note that while for now we are restricted to using photometric samples with small but nonzero redshift overlap, spectroscopic data from DESI [24] and Euclid [25] will soon allow us to use samples that are nonoverlapping, better ensuring $\rho^{gl} \approx 0$.⁹

Finally, we work also with the cosmic infrared background (CIB) which is a tracer of star formation and as such originates from a wide range of cosmic times peaking at $z \sim 2$. Since this closely matches the CMB lensing kernel, the CIB is highly correlated with CMB lensing (e.g., [13,26]), making it a paradigmatic candidate to be used as tracer I (as suggested already in [11]). Specifically, we use the map extracted from Planck 353 GHz data using the GNILC algorithm [27], masking modes with $l < 100$ to avoid contamination from spurious artifacts and residual galactic dust on large angular scales. This overlaps with the $\hat{\kappa}$ patch across $\sim 51\%$ of the sky.

B. Methods

In order to determine the weights f_l , and also to later model the variance reduction, we need fiducial spectra for all the auto- and cross- spectra between $\hat{\kappa}$, I and g . We follow the approach in [28] to fitting the spectra involving galaxies, CIB and lensing on the patch where they overlap, but use the PYCCL code [29] and the galaxy redshift

⁹There will be a small contribution from magnification bias; if significant, it could be handled by following the Appendix A.

distributions given in the LS target selection papers cited above. Note that this does not introduce any unwanted dependence on cosmology or physics at a different redshift, since we do not utilize the best-fit parameters for anything else. The fitting form we use to get f_l need not be the actual physically correct model: all that is required to avoid bias is a smooth fit, and any deviation of the fiducial spectrum from the truth will only result in suboptimal variance reduction [13,28]. Moreover, since the form of f_l we are using was obtained from an optimization procedure, the cleaning performance has no linear-order response to inaccuracies in determining f_l .

When coadding LRG maps from different redshift bins to then use them as tracer I , we first re-weight the galaxy bins using Eq. 20 of [13] to better match the CMB lensing kernel and thus maximize the cross-correlation with lensing. Admittedly, this only improves performance marginally for the tracers we consider.

We measure the angular pseudo- cross-spectrum on scales $10 < l < 1000$ on the same sky patch before and after redshift cleaning, and estimate the variance on estimates of the bandpowers in each of the two cases from the scatter of the measured \hat{C}_l 's within bins of width $\Delta l = 90$. When doing this, it is important that the underlying spectrum be flat over the size of the bin. We ensure this by first dividing the \hat{C}_l 's by a smooth fiducial spectrum—the appropriate model for each measurement given in the Appendix E—and multiply the result by the binned version of this fiducial.

We verify that the pipeline is unbiased by applying it to 1000 Gaussian simulations of all the tracers sharing a degree of correlation that matches what is seen in the data.¹⁰ The scatter of these outputs gives us our error bars.

C. Results

Figure 2 shows that we are able to effectively remove structure from the lensing maps, and that we do so by an amount that agrees with theoretical expectations. To state this more quantitatively, we fit for an amplitude parameter A , which linearly rescales the fiducial model of equation (D2) (the solid lines in the figure). Proceeding by least-squares minimization,¹¹ we obtain best-fit parameter values and associated uncertainties that can be interpreted as highly-significant detections of lensing removal: 41σ when using GNILC CIB maps to clean Planck PR4 $\hat{\kappa}$, and 25σ when using LRG bins 1–4. We test goodness-of-fit by calculating the reduced- χ^2 of a binned model with the best-fit value of A and translating this to a probability-to-exceed (PTE); the latter values, quoted in the figure legend,

¹⁰See Appendix F of [30] for details on how to generate such simulations.

¹¹Since the bins are wide ($\Delta l = 90$), we ignore correlations between them.

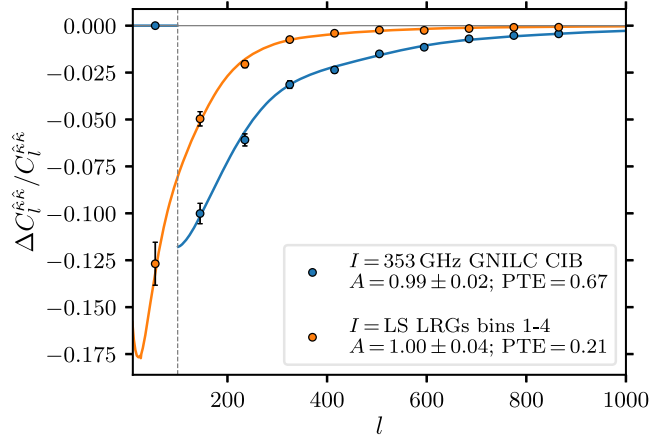


FIG. 2. Fractional change in lensing power after cleaning a Planck PR4 $\hat{\kappa}$ map with different filtered tracers (setting $\gamma_l = 1$). To reduce sample variance, we bin the numerator only after taking differences between spectra. The data are in excellent agreement with the model (shown as solid lines for a fiducial $A = 1$).

suggest the fits are excellent. Moreover, constraints on A are consistent with the fiducial value of unity.

The reduction is greatest at low l because it is on those large scales that the Planck lensing reconstructions have highest S/N per mode. In addition, whenever we clean with a galaxy map, such as our LRG sample, this is compounded with the fact that it is at low- l that contributions from low redshift structures make up a higher fraction of the total lensing power (see, e.g., [1]). It is reassuring that below $l < 100$, where we masked modes in the CIB map to avoid the impact of systematics, the variance is unchanged.

This removal of structure from the CMB lensing map translates to lower variance when cross-correlating it with other tracers. In Fig. 3, we demonstrate this for two tracer combinations: in the left column, we clean the Planck $\hat{\kappa}$ map with the GNILC CIB map, and subsequently correlate this with the combination of all four LRG bins; on the right, we clean it with LRG bins 2–4 (recall that this sample spans $0.5 < z < 1$), weighted to match the lensing kernel, and cross-correlate with the BGS sample ($z < 0.5$).

We consider two estimators of the variance change. The first (top row) shows the fractional difference in variance before and after redshift cleaning. Applying the same significance test as above, we report reductions in the variance of $C_l^{\hat{\kappa}g}$ with confidence slightly above 3σ . This effect is well captured by our fiducial models, which are consistent with the constrained A within 1σ , and yield very plausible PTEs at the best fit values of A .

This first estimator makes intuitive sense, but it is rather noisy. In the Appendix E, we derive an alternative one that agrees with the first in the mean when $\gamma_l = 1$ and $\rho_l^{gI} = 0$ but is less noisy in general; heuristically, it quantifies the variance associated with structures that get removed during the cleaning process. These conditions are not grossly

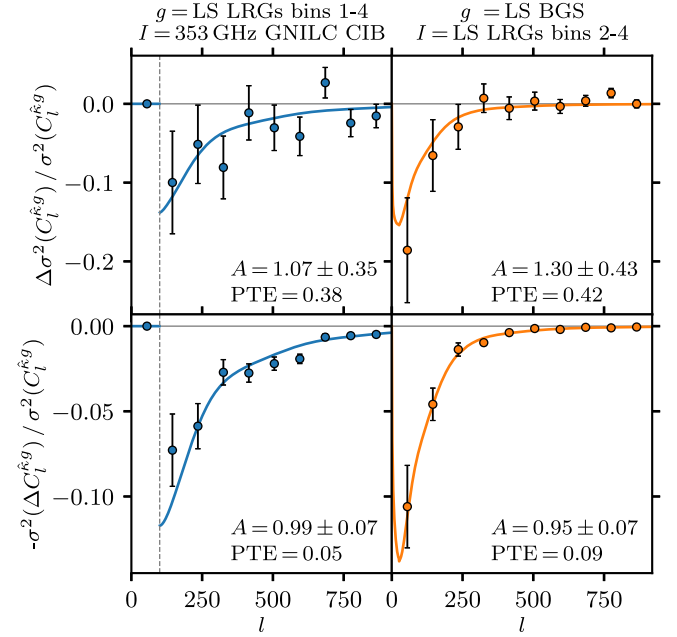


FIG. 3. Variance cancellation in the cross-correlation of tracer g with a Planck PR4 $\hat{\kappa}$ map after cleaning the latter with a filtered tracer I (setting $\gamma_l = 1$). Top row: fractional change in the variance per mode. Bottom row: (minus) the variance associated with the structures that cleaning removes, as a fraction of the original variance. According to the Appendix E, top and bottom rows are equal in the mean in the limit that $\rho_l^{gI} = 0$, but the former is more noisy. Solid lines show fiducial models with $A = 1$, while the best-fit values are annotated in each panel.

violated in the cases we consider here, particularly in the right column of Fig. 3: $\gamma_l = 1$ by design, and though ρ_l^{gI} is strictly speaking not zero, it is small. In this situation, the two estimators differ slightly—in the Appendix E, we model exactly by how much—but the second is still a reasonable estimator of the variance change given the statistical errors. Results from this estimator are reported in the bottom two panels, showing a detection significance of approximately 14σ in the two cases we consider.

For completeness, we show in the Appendix E, Fig. 6, that we can also successfully reduce the variance of the CMB lensing power spectrum by the expected amount, using either the CIB or the LRG samples. We detect this effect with up to 4σ confidence.

IV. CONCLUSIONS

We have explored a method to reduce cosmic variance in CMB lensing maps by invoking an ancillary tracer which, importantly, does not need to be modeled in detail. This is particularly useful when the tracers are poorly understood, as is the case with the CIB or low-redshift tracers in the non-linear regime. Since our formalism is built around correlation coefficients determined from measurable quantities, it automatically accounts for the gravitational lensing

of g and I , as well as their decorrelation with each other and with κ due to gravitational nonlinearities and nonlinear bias. With some minor modifications, it is likely that the method could be applicable to galaxy weak lensing as well.

We identified simple conditions that need to be met for the technique to result in improvements in the S/N of CMB lensing cross-correlations. These are automatically satisfied when $\rho_l^{gl} = 0$; hence, the method promises to be useful when spectroscopic data are available. On the other hand, when $\rho_l^{gl} \neq 0$, some amount of signal is removed during the process, and any benefits depend on the hierarchy between ρ_l^{gl} , $\rho_l^{\hat{\kappa}l}$ and $\rho_l^{\hat{\kappa}g}$ (though when $\rho_l^{gl} \ll 1$, the bias from signal suppression is typically negligible). In the Appendix A, we explored ways to partially restore the removed signal, should it be needed, including direct measurements of \hat{C}_l^{gl} . When the correlation between g and I is significant, the method is especially suited to enhance deep CMB lensing measurements limited to compact sky areas, especially when \hat{C}_l^{gl} can be measured over larger regions. Users interested in the technique can simply evaluate Eqs. (A9) or (A10) to assess whether benefits are accessible to them.

We then demonstrate that variance cancellation can already be achieved using existing tracers. First of all, we report a reduction in Planck PR4 CMB lensing power with 41σ confidence when using GNILC CIB, or 25σ using LS LRGs. These lead to 4σ and 2σ detections of variance reduction in the CMB lensing autospectrum. We then demonstrate variance cancellation in the cross-correlation of CMB lensing with galaxy surveys at 3σ confidence with standard estimators, or 14σ with a tailor-made estimator. The theoretical framework we develop proves excellent when it comes to faithfully modeling empirical results.

ACKNOWLEDGMENTS

We thank M. White, G. Farren, F. McCarthy, T. Namikawa, F. Qu and E. Schaan for their insightful comments on an early draft. In addition, we are grateful for fruitful conversations with A. Raichoor, R. Zhou, N. Sailer, J. DeRose, N. Weaverdyck and B. Sherwin. A. B. L. is a BCCP fellow at UC Berkeley and Lawrence Berkeley National Laboratory. S. F. is funded by the Physics Division of Lawrence Berkeley National Laboratory. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. We have also made use of NASA's Astrophysics Data System, the arXiv preprint server, the PYTHON programming language and packages NumPy, MATPLOTLIB, SciPy, ASTROPY, and HEALPY [31,32]. This work was carried out on the territory of xučyun (Huichin), the ancestral and unceded land of the Chochoyeno speaking Ohlone people, the successors of the sovereign Verona Band of Alameda County.

APPENDIX A: GENERALIZATION

In Eq. (6), we identified the two conditions which, if met, guarantee that $f_l > 0$, meaning that variance cancellation is possible and will in theory result in improved S/N on $C_l^{\hat{\kappa}g}$. When $\rho_l^{gl} = 0$, the first condition is always satisfied, but the situation is more nuanced when $\rho_l^{gl} \neq 0$.

In general, when $\rho_l^{gl} \neq 0$, the $\rho_l^{\hat{\kappa}l}$ - $\rho_l^{\hat{\kappa}g}$ plane can be divided into regions where f_l is positive or negative. This is illustrated in the left column of Fig. 4: in the green regions, $f_l > 0$, and conversely for the red ones. The curves separating these regions are given by the limit where the inequalities in Eq. (5) become equalities.¹²

In the regions where $f_l < 0$, there can be no gain in S/N from variance cancellation, so Eq. (2) is essentially telling us to fold I in as a tracer of g . In fact, a derivation of f_l using a Lagrange multiplier that constrains f_l to be non-negative returns $f_l = 0$ in this regime.

The regions we are interested in, however, are those where $f_l > 0$, so we now characterize them in detail. When $\rho_l^{\hat{\kappa}g} > \rho_l^{gl}$, either the first of the two conditions is met—namely $\rho_l^{\hat{\kappa}l} > \rho_l^{gl} / \rho_l^{\hat{\kappa}g}$ —or none are. This is telling us that when g_l is more correlated with $\hat{\kappa}_l$ than with I_l , it is necessary (though not sufficient) that I_l itself be more correlated with $\hat{\kappa}_l$ than with g_l . Qualitatively, the type of variance cancellation we can obtain in this branch is a natural extension of that we see in the $\rho_l^{gl} = 0$ limit. At the dividing line between regions, $\rho_l^{\hat{\kappa}l} = \rho_l^{gl} / \rho_l^{\hat{\kappa}g}$, the weights are zero because this equality can only be satisfied when $\hat{\kappa}_l = I_l$, in which case we are better off not doing anything.

On the other hand, when $\rho_l^{\hat{\kappa}g} < \rho_l^{gl}$, only the second condition—that is, $\rho_l^{\hat{\kappa}l} > \rho_l^{\hat{\kappa}g} / \rho_l^{gl}$ —can be satisfied. This time, the necessary (but, again, insufficient) condition is that when g_l is less correlated with $\hat{\kappa}_l$ than with I_l , $\hat{\kappa}_l$ must be significantly more correlated with I_l than with g_l (linearly more so). In this case, the weights are undefined at the boundary between regimes because the one requirement in our derivation of f_l was that $\rho_l^{\hat{\kappa}l} \neq \rho_l^{\hat{\kappa}g} / \rho_l^{gl}$.

These conditions are in fact quite idealized. Both they and Eq. (7) tell us about the S/N in a way we cannot directly work with, because when $\rho_l^{gl} \neq 0$ we need to know what has happened to the cross-correlation signal through the redshift cleaning process. Whenever tracers g and I are correlated and $f_l > 0$, the signal decreases as

$$C_l^{\hat{\kappa}g} = C_l^{\hat{\kappa}g} - f_l C_l^{gl}. \quad (\text{A1})$$

¹²In the discussion that follows, it will be useful to note that the physically allowed region—as determined in the Appendix B—is tangent to the top axis where $\rho_l^{\hat{\kappa}g} = \rho_l^{gl}$, and to the right axis where $\rho_l^{\hat{\kappa}l} = \rho_l^{gl}$; these are also the points where the thresholds in Eq. (6) intersect the axes.

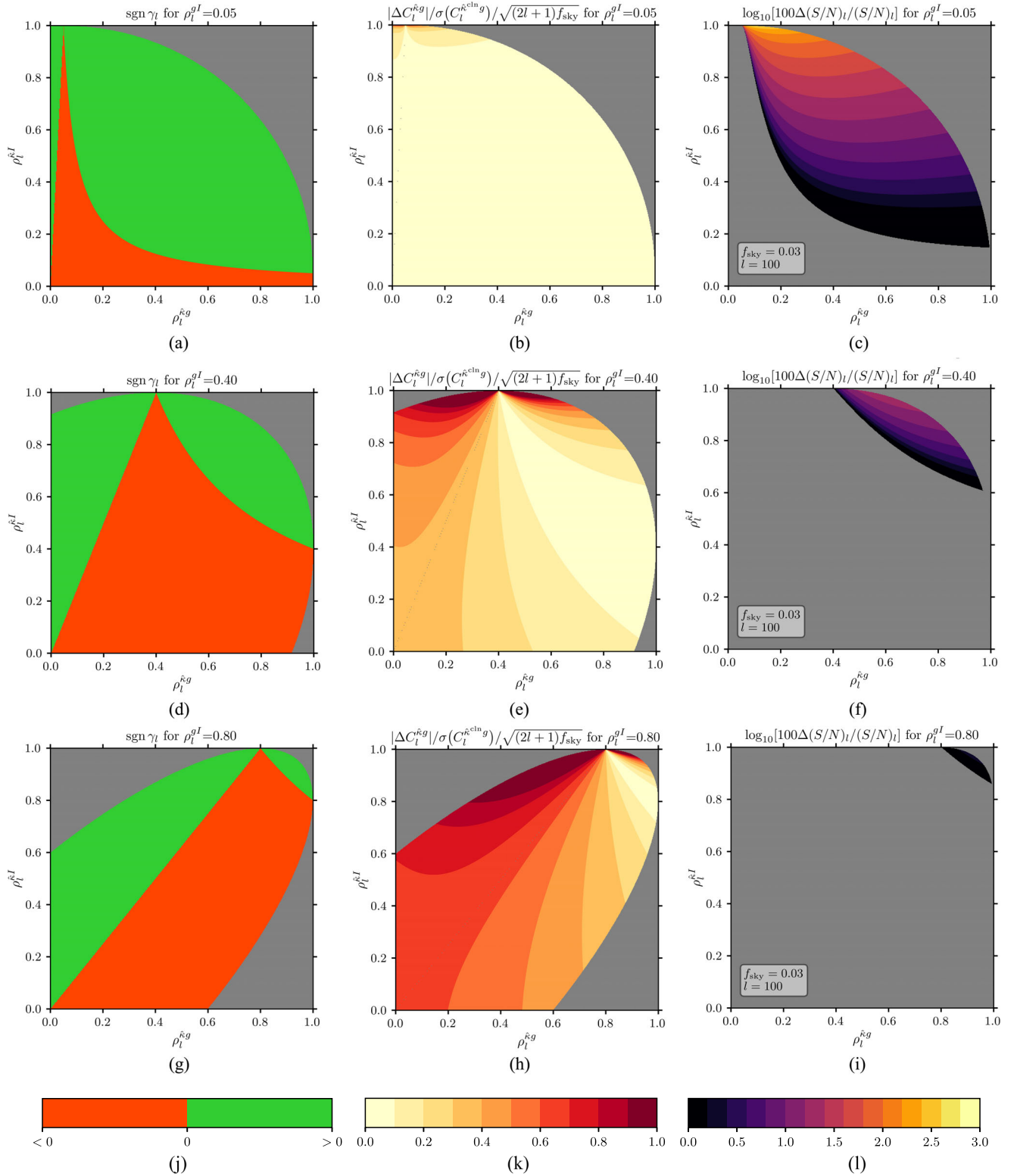


FIG. 4. Generalization of the theory of variance cancellation to the case of $\rho_l^{gl} \neq 0$. Left column: sign of f_l ; where positive (green) variance cancellation can in theory lead to improved S/N. However, the two green branches differ by the extent to which the signal is reduced. Middle column: ratio of bias from signal suppression to statistical uncertainty after redshift cleaning. The bias-variance tradeoff can be optimized by minimizing the MSE after a direct measurement of \hat{C}_l^{gl} is used to partially restore the signal. Right panel: fractional change in S/N per mode of C_l^{kg} at the MSE optimum after cleaning and restoring the signal from a measurement of \hat{C}_l^{gl} on the same patch, showing only regions where there is a gain in S/N. Animations at intermediate values of ρ_l^{gl} can be found online [33].

As long as the optimal weights in Eq. (5) are used, it is in principle still advantageous to pursue variance cancellation. However, attempts to fit the data with a theoretical model for $C_l^{\kappa g}$ will be biased, all the while that the variance is reduced. Moreover, the data will have acquired a dependence on l , and thus on physics at redshifts different from that which we would like to isolate.

Let us formalize our definition of this bias as

$$\begin{aligned} \Delta C_l^{\hat{\kappa}g} &\equiv \langle \hat{C}_l^{\hat{\kappa}^{\text{cln}g}g} \rangle - C_l^{\hat{\kappa}g} \\ &= -f_l C_l^{gI}. \end{aligned} \quad (\text{A2})$$

The ratio of the bias magnitude to the statistical uncertainty after cleaning is then

$$\begin{aligned} \frac{|\Delta C_l^{\hat{\kappa}g}|}{\sigma(\hat{C}_l^{\hat{\kappa}^{\text{cln}g}g})} &= \sqrt{(2l+1)f_{\text{sky}}} \left\{ (\gamma_l \rho_l^{gI} \rho_l^{\hat{\kappa}I})^{-2} [1 + (\rho_l^{\hat{\kappa}I})^2 \gamma_l (\gamma_l - 2)] \right. \\ &\quad \left. + \left(\frac{\rho_l^{\hat{\kappa}g}}{\gamma_l \rho_l^{gI} \rho_l^{\hat{\kappa}I}} \right)^2 (1 - \gamma_l \rho_l^{gI} \rho_l^{\hat{\kappa}I} / \rho_l^{\hat{\kappa}g})^2 \right\}^{-\frac{1}{2}}, \end{aligned} \quad (\text{A3})$$

which is larger the more modes are observed—hence the factor of $\sqrt{(2l+1)f_{\text{sky}}}$, where f_{sky} tracks the size of the patch where cleaning is performed (small patches offer greater tolerance to bias due to their higher sample variance). This expression is evaluated in the central column of Fig. 4 for three values of ρ_l^{gI} , immediately revealing a significant qualitative difference between the two variance cancellation branches identified in Eq. (6): the fractional bias is smaller when $\rho_l^{\hat{\kappa}I} > \rho_l^{gI} / \rho_l^{\hat{\kappa}g}$ than in the alternative branch where $\rho_l^{\hat{\kappa}I} > \rho_l^{\hat{\kappa}g} / \rho_l^{gI}$. In fact, when $\rho_l^{gI} \ll 1$, the bias can be safely ignored across the former branch for most reasonable values of l and f_{sky} (as expected, since this is the natural continuation of the $\rho_l^{gI} = 0$ case studied in the main text). On the other hand, when $\rho_l^{\hat{\kappa}I} > \rho_l^{\hat{\kappa}g} / \rho_l^{gI}$ the bias is generally larger than the statistical uncertainty except at the largest angular scales and smallest sky patches, for which $\sqrt{(2l+1)f_{\text{sky}}} \lesssim 1$.

If the bias amplitude is unacceptably large for the application at hand, there are ways to mitigate it. The approach we consider here is to actively try to restore the signal we have removed, fitting our theory model to a corrected cross-spectrum

$$\tilde{C}_l^{\hat{\kappa}^{\text{cln}g}g} \equiv \hat{C}_l^{\hat{\kappa}^{\text{cln}g}g} + f_l C_l^{\text{corr}}. \quad (\text{A4})$$

For example, if we somehow knew the true C_l^{gI} with no uncertainty, we could set $C_l^{\text{corr}} = C_l^{gI}$, and this would eliminate all bias while also retaining all variance suppression. The gain in S/N would then simply be given by Eq. (7). Though C_l^{gI} is unlikely to be known perfectly, it could conceivably be predicted from theory rather

accurately if the redshift distributions of the samples are known. Alternatively, it could come from a fit to a wide range of scales and maybe even a larger sky patch, ideally in a way that remains cosmology-independent despite having to assume a fitting form.

If none of these approaches are viable, C_l^{gI} can still be determined from the data on a multipole-by-multipole basis, but at the cost of increased variance. If we set C_l^{corr} to the measured g - l correlation on the *same* patch where we do the cleaning, $C_l^{\text{corr}} = \hat{C}_l^{gI}$, we will have achieved exact unbiasedness but also restored all of the variance we had originally set out to remove.

We therefore seek solutions that lie between the two limits we have seen, compromises between bias and variance reduction.¹³ These entail setting $C_l^{\text{corr}} = \mathcal{W}_l \hat{C}_l^{gI}$, where \mathcal{W}_l is a filter that we must optimize such that this correction moves us in the direction of unbiasedness but only insofar as we are willing to see the variance increase. One could, for example, choose the form of \mathcal{W}_l that results in the largest amounts of variance reduction subject to the ratio of bias-to-variance,

$$\frac{|\langle \tilde{C}_l^{\hat{\kappa}^{\text{cln}g}g} \rangle - C_l^{\hat{\kappa}g}|}{\sigma(\tilde{C}_l^{\hat{\kappa}^{\text{cln}g}g})} = \frac{|f_l C_l^{gI} (\mathcal{W}_l - 1)|}{\sigma(\hat{C}_l^{\hat{\kappa}^{\text{cln}g}g} + f_l \mathcal{W}_l \hat{C}_l^{gI})}, \quad (\text{A5})$$

being smaller than some threshold. Note that this expression differs from (A3) in that both the numerator and denominator are calculated *after* applying the correction. In particular, the denominator depends on a number of factors, including the size of the patches where cleaning is performed and where \hat{C}_l^{gI} is measured. These need not be the same; if the latter is larger, the uncertainty introduced when restoring the signal will be comparatively smaller; and if the patches are disjoint, the covariance between them will be greatly reduced. In regions where (A3) is already below the threshold, the constraint should be inactive and $\mathcal{W}_l = 0$: no signal needs to be restored, so variance reduction should be maximal. On the other hand, when the constraint is active, the desired threshold value should be exactly enforced.

Yet another possibility is to choose the form of \mathcal{W}_l that minimizes the mean-squared-error,¹⁴

¹³The metrics we will be introducing in Eqs. (A5) and (A6) are not adequate in the regime where the optimal weights are negative— $f < 0$, where, heuristically, l is included as a tracer of g to boost the cross-correlation signal. Since the focus of this work is the regime where $f > 0$, we show only results appropriate for this case.

¹⁴Though we do not include it here, the impact of systematic effects in the measurement of \hat{C}_l^{gI} , such as inhomogeneities in galaxy samples (e.g., [34,35]), could be incorporated at this point if its contribution to the error budget after redshift cleaning is known.

$$\text{MSE} \equiv \langle (\tilde{C}_l^{\hat{\kappa}^{\text{cin}}g} - C_l^{\hat{\kappa}g})^2 \rangle, \quad (\text{A6})$$

thus reducing bias and/or variance according to their relevance in the scenario at hand. This will limit the gains in S/N promised by Eq. (7), but it will also offer the key benefit of limiting the bias amplitude to being at worst of the order of the statistical error. It can be shown that the weights that accomplish this are

$$\mathcal{W}_l = \frac{(C_l^{gI})^2 - \text{Cov}[\hat{C}_l^{\hat{\kappa}^{\text{cin}}g}|_{\hat{\kappa}g \cap I}, \hat{C}_l^{gI}|_{g \cap I}]/f_l}{(C_l^{gI})^2 + \sigma^2(\hat{C}_l^{gI}|_{g \cap I})}, \quad (\text{A7})$$

where we have differentiated between measurements made on the patch where \hat{C}_l^{gI} is measured (which we denote as $g \cap I$) or in the patch where we carry out the redshift cleaning ($\hat{\kappa} \cap g \cap I$).

When the two patches are one and the same, the weights reduce to

$$\mathcal{W}_l = 1 - \left[\frac{\rho_l^{\hat{\kappa}g}}{\gamma_l \rho_l^{gI} \rho_l^{\hat{\kappa}I}} + \frac{1}{\gamma_l (\rho_l^{gI})^2} \right] \times \left[1 + \frac{1}{(\rho_l^{gI})^2} + (2l+1)f_{\text{sky}} \right]^{-1}. \quad (\text{A8})$$

As expected, $\mathcal{W}_l \rightarrow 0$ when $\rho_l^{gI} \rightarrow 0$, indicating that the bias is best left untreated, as it is small to begin with. At the other extreme, $\mathcal{W}_l \rightarrow 1$ as $\rho_l^{gI} \rightarrow 1$ —the bias is so large that mitigating it costs us practically all variance reduction.

$$\begin{aligned} \frac{\Delta(S/N)_l}{(S/N)_l} &= [1 - (1 - \mathcal{W}_l)\gamma_l \rho_l^{\hat{\kappa}I} \rho_l^{gI} / \rho_l^{\hat{\kappa}g}] [1 + (\rho_l^{\hat{\kappa}g})^2]^{-\frac{1}{2}} \\ &\times \left\{ 1 + (\rho_l^{\hat{\kappa}I})^2 \gamma_l (\gamma_l - 2) + (\rho_l^{\hat{\kappa}g} - \gamma_l \rho_l^{gI} \rho_l^{\hat{\kappa}I})^2 + (\mathcal{W}_l \gamma_l \rho_l^{\hat{\kappa}I})^2 \left(\frac{f_{\text{sky}}^{\hat{\kappa}g \cap I}}{f_{\text{sky}}^{g \cap I}} \right) [1 + (\rho_l^{gI})^2] \right\}^{-\frac{1}{2}} - 1. \end{aligned} \quad (\text{A10})$$

Let us use this equation to assess the extent to which being able to measure \hat{C}_l^{gI} on a larger, disjoint patch improves prospects for redshift cleaning. Suppose the CMB lensing reconstructions are obtained by a telescope on the South Pole while g and I are both measured on larger footprints by telescopes on the Atacama desert and/or space. For example, most of the 1500 deg² ($f_{\text{sky}}^{\hat{\kappa}g \cap I} \approx 0.03$) covered by SPT-3G are contained within the 5000 deg² observed by the Dark Energy Survey (DES) [36] in its Y3 data release. If g and I are both drawn from DES observations, then $f_{\text{sky}}^{g \cap I} \approx 0.09$ (after

To see this more explicitly, we calculate the change in S/N after minimizing the MSE when the $\hat{\kappa} \cap g \cap I$ and $g \cap I$ patches are actually the same:

$$\frac{\Delta(S/N)_l}{(S/N)_l} = \left\{ 1 + [1 + (\rho_l^{\hat{\kappa}g})^2]^{-1} \left[\frac{1 + \gamma_l'(\gamma_l' - 2)(\rho_l^{\hat{\kappa}I})^2}{(1 - \gamma_l' \rho_l^{gI} \rho_l^{\hat{\kappa}I} / \rho_l^{\hat{\kappa}g})^2} - 1 \right] \right\}^{-\frac{1}{2}} - 1, \quad (\text{A9})$$

where $\gamma_l' \equiv \gamma_l(1 - \mathcal{W}_l)$; notice that this expression reduces to Eq. (7) when $\mathcal{W}_l = 0$, and gives zero when $\mathcal{W}_l = 1$. The right column of Fig. 4 evaluates it for $l = 100$ and $f_{\text{sky}} = 0.03$, a sky fraction similar to that surveyed by SPT-3G. These values help us illustrate some important features without loss of generality. Note, in particular, that gains in S/N are limited to a subregion of the branch where $\rho_l^{\hat{\kappa}I} > \rho_l^{gI} / \rho_l^{\hat{\kappa}g}$.

Consider, on the other hand, the case where patches $\hat{\kappa} \cap g \cap I$ and $g \cap I$ are disjoint. When this is the case, the covariance in the numerator of Eq. (A7) can be ignored.¹⁵ Furthermore, the precision with which \hat{C}_l^{gI} can be measured increases with the size of the $g \cap I$ patch; in the limit where the error in determining C_l^{gI} is made small this way, \mathcal{W}_l approaches unity as the signal can be accurately restored without adding significant amounts of variance. To state this more quantitatively, we once again look at the changes in S/N that can be achieved, this time specifying that the patches be disjoint¹⁶:

excising the region of overlap with $\hat{\kappa} \cap g \cap I$). More futuristically, if we assume that the 18000 deg² coverage of LSST fully contains the SPT-3G patch, $f_{\text{sky}}^{g \cap I} \approx 0.4$. Finally, if g and I come from satellite observations on 80% of the sky (we exclude regions near the Galactic plane), then $f_{\text{sky}}^{g \cap I} \approx 0.8$. Figure 5 shows the gain in S/N associated with these three scenarios for a fixed $\rho_l^{gI} = 0.4$ and $l = 100$, so that they can be readily compared to Fig. 4(f), for which a single patch with $f_{\text{sky}}^{\hat{\kappa}g \cap I} = f_{\text{sky}}^{g \cap I} \approx 0.03$ is used. Notice that the prospects are significantly improved both in terms of the gain in S/N for fixed values of the correlation coefficients, and also in terms of the situations where gains are at all possible, as certain regions

¹⁵Away from the lowest few multipoles.

¹⁶It is straight forward to verify that Eqs. (7), (A9), and (A10) all equal each other when $\mathcal{W}_l = 0$.

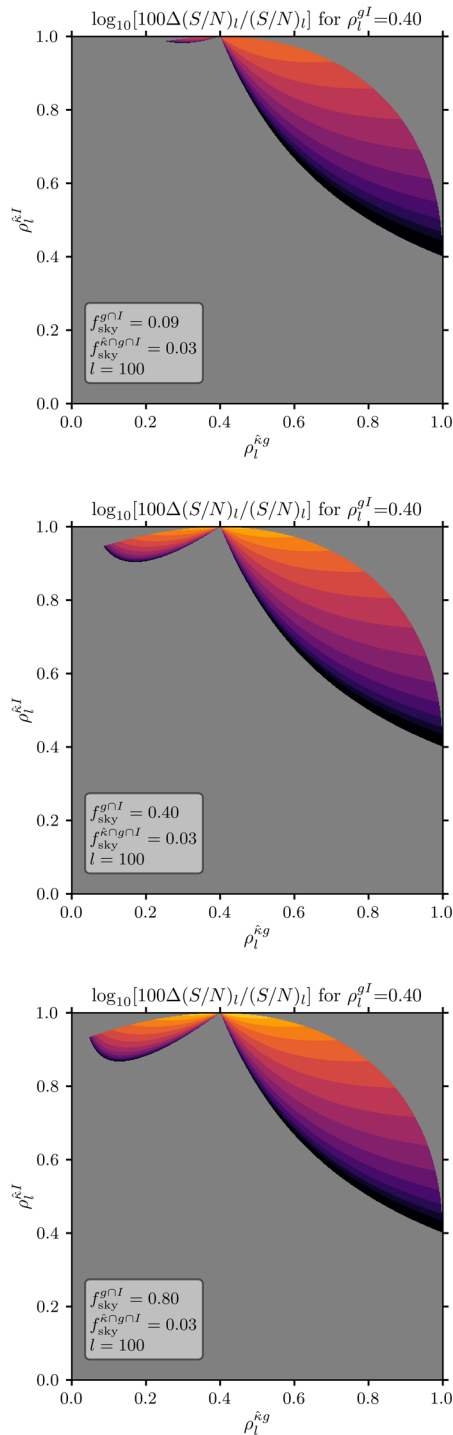


FIG. 5. Impact of the size of the patch where \hat{C}_I^{gI} is measured ($g \cap I$, disjoint from $\hat{c} \cap g \cap I$) on the possible gains in S/N per mode of \hat{C}_I^{kg} when $\rho_I^{gI} = 0.4$, at the MSE optimum. (See text of the Appendix A for what scenarios these f_{sky} values correspond to.) This is to be contrasted with Fig. 4(f), which assumes a single patch with $f_{\text{sky}}^{g \cap I} = f_{\text{sky}}^{\hat{c} \cap g \cap I} = 0.03$ but the same color range given by the color bar in 4(i). Measuring \hat{C}_I^{gI} on a patch of sky larger than that where we redshift-clean increases both the magnitude of gains in S/N and the domain where they are possible.

within the $\rho_I^{\hat{c}I} > \rho_I^{\hat{c}g} / \rho_I^{gI}$ branch now lend themselves to improved precision via redshift cleaning.

APPENDIX B: RELATING $\rho^{\kappa g}$, $\rho^{\kappa I}$, AND ρ^{gI}

In order to derive a relation between the correlation coefficients of tracers κ , g , and I , we will exploit an analogy between the cosine of the angle between vectors X and Y ,

$$\cos \theta_{XY} = \frac{X \cdot Y}{|X||Y|}, \quad (\text{B1})$$

and the correlation coefficient between projected fields X and Y (we drop the multipole dependence for notational convenience),

$$\rho^{XY} = \frac{C^{XY}}{\sqrt{C^{XX}}\sqrt{C^{YY}}}. \quad (\text{B2})$$

Following this analogy, we promote our tracers to vectors in three-dimensional space, and seek an expression for $\cos \theta_{\kappa g}$ in terms of $\cos \theta_{gI}$ and $\cos \theta_{\kappa I}$, and similarly for $\cos \theta_{\kappa I}$. Such a relation exists only if the three vectors are coplanar; fortunately, this limit suffices to place an upper bound on $\cos \theta_{\kappa g}$, once we realize that for fixed θ_{gI} and $\theta_{\kappa I}$, the configuration that maximizes $\cos \theta_{\kappa g}$ is the one where all the vectors are coplanar. Consequently,

$$\begin{aligned} \rho^{\kappa g} &= \cos \theta_{\kappa g} \\ &\leq \cos(\theta_{gI} - \theta_{\kappa I}) \\ &= \cos \theta_{gI} \cos \theta_{\kappa I} + \sin \theta_{gI} \sin \theta_{\kappa I} \\ &= \rho^{gI} \rho^{\kappa I} + \sqrt{[1 - (\rho^{gI})^2][1 - (\rho^{\kappa I})^2]}, \end{aligned} \quad (\text{B3})$$

where, in the second line, equality holds only when the vectors are coplanar. By the same logic,

$$\rho^{\kappa I} \leq \rho^{gI} \rho^{\kappa g} + \sqrt{[1 - (\rho^{gI})^2][1 - (\rho^{\kappa g})^2]}. \quad (\text{B4})$$

APPENDIX C: PARAMETRIZING TRACERS FOR FORECASTS

In this section, we describe the tracer modeling underlying the forecasts of Fig. 1.

Our parametrization of the VRO LSST follows [28]. We assume the galaxies can be split into disjoint tomographic bins with edges at redshifts 0, 0.5, 1, 2, 3, 4 and 7. The first five bins ($0 < z < 4$) are then combined using the weights of [13] to more closely match the galaxy kernel to that of CMB lensing. This constitutes our tracer I —we also retain the last bin ($4 < z < 7$) to be used as our g tracer. As per GNILC CIB and DESI BGS (see Sec. III A for details about

these tracers), our models come from theory-inspired fits to the data, once again following [28].

On the CMB lensing side, we consider experiments with the characteristics of CMB-S4 and ACT [37] DR6. We follow [38] in treating the CMB-S4 Gaussian lensing reconstruction noise as being white on signal-dominated scales. We take the white noise level to be $\Delta_\kappa = 0.2$ arcmin to match Fig. 3 of [2], which shows the reconstruction noise for a minimum-variance combination of temperature and polarization reconstructions—the former up to $l_{\max}^T = 3000$, the latter to $l_{\max}^{E,B} = 3500$ and applied iteratively—for a possible Stage-4 experiment with a symmetric, Gaussian beam with FWHM of 1 arcmin and 1 μK arcmin white noise. On the other hand, the noise curve for ACT DR6 was generated by assuming a 1.4 arcmin beam and a white noise level of 10 μK arcmin.

APPENDIX D: MODELING CHANGES IN SPECTRA

When the weights in Eq. (5) are employed, the angular power spectrum of the redshift cleaned convergence map is

$$C_l^{\hat{\kappa}^{\text{cln}}\hat{\kappa}^{\text{cln}}} = C_l^{\hat{\kappa}\hat{\kappa}} \left[1 - 2f_l \frac{C_l^{\kappa I}}{C_l^{\hat{\kappa}\hat{\kappa}}} + f_l^2 \frac{C_l^{II}}{C_l^{\hat{\kappa}\hat{\kappa}}} \right], \quad (\text{D1})$$

such that the fractional change in lensing power after variance cancellation is

$$\frac{\Delta C_l^{\hat{\kappa}\hat{\kappa}}}{C_l^{\hat{\kappa}\hat{\kappa}}} = -2f_l \frac{C_l^{\kappa I}}{C_l^{\hat{\kappa}\hat{\kappa}}} + f_l^2 \frac{C_l^{II}}{C_l^{\hat{\kappa}\hat{\kappa}}}. \quad (\text{D2})$$

If the fiducials are perfectly matched to the truth, these expressions simplify to

$$\hat{C}_l^{\hat{\kappa}^{\text{cln}}\hat{\kappa}^{\text{cln}}} = \hat{C}_l^{\hat{\kappa}\hat{\kappa}} [1 + (\rho_l^{\hat{\kappa}I})^2 \gamma_l (\gamma_l - 2)], \quad (\text{D3})$$

and

$$\frac{\Delta C_l^{\hat{\kappa}\hat{\kappa}}}{C_l^{\hat{\kappa}\hat{\kappa}}} = (\rho_l^{\hat{\kappa}I})^2 \gamma_l (\gamma_l - 2). \quad (\text{D4})$$

On the other hand, the cross-correlation of $\hat{\kappa}^{\text{cln}}$ with g gives

$$C_l^{\hat{\kappa}^{\text{cln}}g} = C_l^{\hat{\kappa}g} - f_l C_l^{gI}. \quad (\text{D5})$$

If the fiducials match the truth, then

$$C_l^{\hat{\kappa}^{\text{cln}}g} = C_l^{\hat{\kappa}g} (1 - \gamma_l \rho_l^{gI} \rho_l^{\hat{\kappa}I} / \rho_l^{\hat{\kappa}g}). \quad (\text{D6})$$

All in all, in the limit that the fiducials match the truth, the correlation of the cleaned convergence map with tracer g becomes

$$\rho_l^{\hat{\kappa}^{\text{cln}}g} = \rho_l^{\hat{\kappa}g} \left[\frac{1 - \gamma_l \rho_l^{gI} \rho_l^{\hat{\kappa}I} / \rho_l^{\hat{\kappa}g}}{\sqrt{1 + (\rho_l^{\hat{\kappa}I})^2 \gamma_l (\gamma_l - 2)}} \right]. \quad (\text{D7})$$

Furthermore, if g and I are completely uncorrelated, then

$$\rho_l^{\hat{\kappa}^{\text{cln}}g} = \frac{\rho_l^{\hat{\kappa}g}}{\sqrt{1 - (\rho_l^{\hat{\kappa}I})^2}}. \quad (\text{D8})$$

APPENDIX E: MODELING VARIANCE REDUCTION

We now model the change in variance after applying our method. In doing so, let us follow a route that highlights the different contributions to the variance of our estimator. We begin with the general expression for the variance of the difference of two correlated random variables,

$$\sigma^2(C_l^{\hat{\kappa}^{\text{cln}}g}) - \sigma^2(C_l^{\hat{\kappa}g}) = -\sigma^2(C_l^{\hat{\kappa}^{\text{cln}}g} - C_l^{\hat{\kappa}g}) \quad (\text{E1})$$

$$+ 2\sigma^2(C_l^{\hat{\kappa}^{\text{cln}}g}) \quad (\text{E2})$$

$$- 2\text{Cov}(C_l^{\hat{\kappa}^{\text{cln}}g}, C_l^{\hat{\kappa}g}). \quad (\text{E3})$$

Taking the fields to be Gaussian, we find that

$$\frac{\Delta \sigma^2(C_l^{\hat{\kappa}g})}{\sigma^2(C_l^{\hat{\kappa}g})} = -\frac{\sigma^2(\Delta C_l^{\hat{\kappa}g})}{\sigma^2(C_l^{\hat{\kappa}g})} + \frac{2f_l (C_l^{\hat{\kappa}\hat{\kappa}})^{-1}}{1 + (\rho_l^{\hat{\kappa}g})^2} \left[C_l^{\hat{\kappa}I} \left(f_l \frac{C_l^{II}}{C_l^{\hat{\kappa}\hat{\kappa}}} - 1 \right) \right] \quad (\text{E4})$$

$$f_l \frac{C_l^{gI}}{C_l^{gg}} (C_l^{gI} - C_l^{\hat{\kappa}g}) \Big]. \quad (\text{E5})$$

Heuristically, the first term captures the variance reduction due to having removed structure common to $\hat{\kappa}$ and I . It can be modeled as

$$\frac{\sigma^2(\Delta C_l^{\hat{\kappa}g})}{\sigma^2(C_l^{\hat{\kappa}g})} = f_l^2 \frac{C_l^{II}}{C_l^{\hat{\kappa}\hat{\kappa}}} \left[\frac{1 + (\rho_l^{gI})^2}{1 + (\rho_l^{\hat{\kappa}g})^2} \right]. \quad (\text{E6})$$

In the limit that $\gamma_l = 1$ and the fiducial $C_l^{gI,\text{fid}}$ and $C_l^{II,\text{fid}}$ used to build f_l are well matched to the truth, the term in the second line of Eq. (E4) is zero in the mean. The same goes for the term in the third line whenever $\rho_l^{gI} = 0$. In these circumstances—which are very similar to some of the cases considered in the main text, where $\gamma_l = 1$ by construction, and ρ_l^{gI} is small—Eq. (E6) becomes a less noisy estimator of the variance change.

Let us also quote, for completeness, the expected fractional change in variance of the convergence bandpowers:

$$\frac{\sigma^2(\Delta C_l^{\hat{\kappa}\hat{\kappa}})}{\sigma^2(C_l^{\hat{\kappa}\hat{\kappa}})} = \left(1 - 2f_l \frac{C_l^{\hat{\kappa}I}}{C_l^{\hat{\kappa}\hat{\kappa}}} + f_l^2 \frac{C_l^{II}}{C_l^{\hat{\kappa}\hat{\kappa}}}\right)^2 - 1. \quad (\text{E7})$$

If the fiducial spectra are accurate, this gives

$$\frac{\sigma^2(\Delta C_l^{\hat{\kappa}\hat{\kappa}})}{\sigma^2(C_l^{\hat{\kappa}\hat{\kappa}})} = (\rho_l^{\hat{\kappa}I})^2 \gamma_l [-4 + 2\gamma_l + (\rho_l^{\hat{\kappa}I})^2 \gamma_l [4(1 - \gamma_l) + \gamma_l]]. \quad (\text{E8})$$

In Fig. 6, we show with actual data that the variance of estimates of the convergence bandpowers is reduced after redshift cleaning. The data we work with are the same that went into producing Fig. 2: we clean Planck PR4 lensing with the GNILC CIB map, or with the combination of all four LRG bins. To gauge the statistical significance of this demonstration, we proceed as in Sec. III C and fit the data with a rescaled version of the model in Eq. (E8). The best-fit values for this amplitude parameter A are quoted in the figure. Thus analyzed, variance cancellation is detected with approximately 4.3σ confidence when cleaning with the CIB, and 2.3σ with LS LRGs. The models provide good fits to the data, as evidenced by the PTE values.

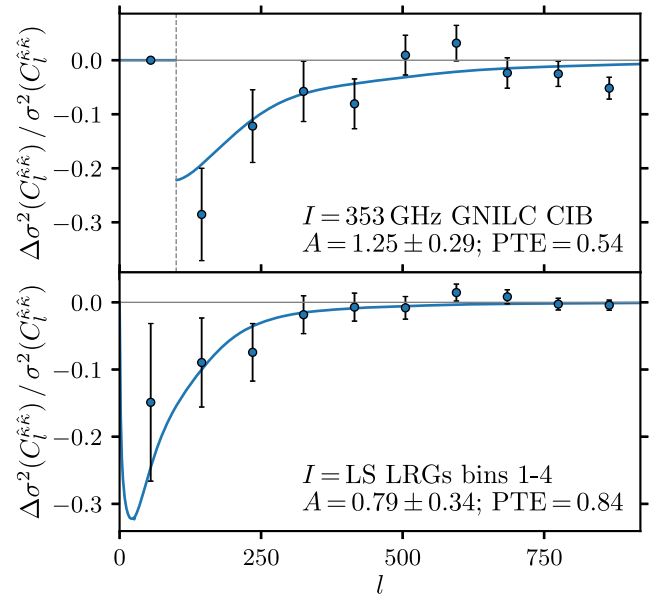


FIG. 6. Reduction in variance of the CMB lensing bandpowers after cleaning a Planck PR4 $\hat{\kappa}$ map with a filtered tracer I (setting $\gamma_l = 1$). Solid lines show the fiducial models with $A = 1$, while the best-fit values are annotated in each panel.

- [1] A. Lewis and A. Challinor, Weak gravitational lensing of the CMB, *Phys. Rep.* **429**, 1 (2006).
- [2] M. Schmittfull and U. Seljak, Parameter constraints from cross-correlation of CMB lensing with galaxy clustering, *Phys. Rev. D* **97**, 123540 (2018).
- [3] E. Abdalla *et al.*, Cosmology intertwined: A review of the particle physics, astrophysics, and cosmology associated with the cosmological tensions and anomalies, *J. High Energy Astrophys.* **34**, 49 (2022).
- [4] M. White *et al.*, Cosmological constraints from the tomographic cross-correlation of DESI luminous red galaxies and Planck CMB lensing, *J. Cosmol. Astropart. Phys.* **02** (2022) 007.
- [5] A. Krolewski, S. Ferraro, and M. White, Cosmological constraints from unWISE and Planck CMB lensing tomography, *J. Cosmol. Astropart. Phys.* **12** (2021) 028.
- [6] T. M. C. Abbott *et al.* (DES Collaboration), Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing, *Phys. Rev. D* **105**, 023520 (2022).
- [7] C. Heymans *et al.*, KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints, *Astron. Astrophys.* **646**, A140 (2021).
- [8] B. Yu, R. Z. Knight, B. D. Sherwin, S. Ferraro, L. Knox, and M. Schmittfull, Towards neutrino mass from cosmology without optical depth information, *arXiv:1809.02120*.
- [9] B. Yu, S. Ferraro, Z. R. Knight, L. Knox, and B. D. Sherwin, The physical origin of dark energy constraints from Rubin observatory and CMB-S4 lensing tomography, *Mon. Not. R. Astron. Soc.* **513**, 1887 (2022).
- [10] F. McCarthy, S. Foreman, and A. van Engelen, Avoiding baryonic feedback effects on neutrino mass measurements from CMB lensing, *Phys. Rev. D* **103**, 103538 (2021).
- [11] F. J. Qu, B. D. Sherwin, O. Darwish, T. Namikawa, and M. S. Madhavacheril, Probing early structure and model-independent neutrino mass with high-redshift CMB lensing mass maps, *arXiv:2208.04253*.
- [12] A. S. Maniyar, E. Schaan, and A. R. Pullen, New probe of the high-redshift Universe: Nulling CMB lensing with interloper-free line intensity mapping pair lensing, *Phys. Rev. D* **105**, 083509 (2022).
- [13] B. D. Sherwin and M. Schmittfull, Delensing the CMB with the cosmic infrared background, *Phys. Rev. D* **92**, 043005 (2015).
- [14] Ž. Ivezić *et al.*, LSST: From science drivers to reference design and anticipated data products, *Astrophys. J.* **873**, 111 (2019).
- [15] CMB-S4 collaboration, CMB-S4 science book, first edition, *arXiv:1610.02743*.
- [16] A. Dey *et al.*, Overview of the DESI legacy imaging surveys, *Astron. J.* **157**, 168 (2019).
- [17] J. Carron, M. Mirmelstein, and A. Lewis, CMB lensing from Planck PR4 maps, *J. Cosmol. Astropart. Phys.* **09** (2022) 039.
- [18] S. W. Henderson *et al.*, Advanced ACTPol cryogenic detector arrays and readout, *J. Low Temp. Phys.* **184**, 772 (2016).

- [19] Simons Observatory Collaboration, The Simons Observatory: Science goals and forecasts, *J. Cosmol. Astropart. Phys.* **02** (2019) 056.
- [20] SPT-3G Collaboration, SPT-3G: A next-generation cosmic microwave background polarization experiment on the South Pole telescope, *Proc. SPIE Int. Soc. Opt. Eng.* **9153**, 91531P (2014).
- [21] C. Hahn *et al.*, DESI bright galaxy survey: Final target selection, design, and validation, *Astron. J.* **165**, 253 (2023).
- [22] A. D. Myers *et al.*, The target selection pipeline for the dark energy spectroscopic instrument, *Astron. J.* **165**, 50 (2023).
- [23] R. Zhou *et al.*, Target selection and validation of DESI luminous red galaxies, *Astron. J.* **165**, 58 (2023).
- [24] DESI Collaboration, The DESI experiment part I: Science, targeting, and survey design, [arXiv:1611.00036](https://arxiv.org/abs/1611.00036).
- [25] L. Amendola *et al.*, Cosmology and fundamental physics with the euclid satellite, *Living Rev. Relativity* **16**, 6 (2013).
- [26] Planck Collaboration, Planck 2013 results. XXX. Cosmic infrared background measurements and implications for star formation, *Astron. Astrophys.* **571**, A30 (2014).
- [27] Planck Collaboration, Planck intermediate results. XLVIII. Disentangling Galactic dust emission and cosmic infrared background anisotropies, *Astron. Astrophys.* **596**, A109 (2016).
- [28] B. Yu, J. C. Hill, and B. D. Sherwin, Multitracer CMB delensing maps from Planck and WISE data, *Phys. Rev. D* **96**, 123511 (2017).
- [29] LSST Dark Energy Science Collaboration, Core cosmology library: Precision cosmological predictions for LSST, *Astrophys. J. Suppl. Ser.* **242**, 2 (2019).
- [30] A. Baleato Lizancos, A. Challinor, B. D. Sherwin, and T. Namikawa, Delensing the CMB with the cosmic infrared background: The impact of foregrounds, *Mon. Not. R. Astron. Soc.* **514**, 5786 (2022).
- [31] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, HEALPix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere, *Astrophys. J.* **622**, 759 (2005).
- [32] A. Zonca, L. Singer, D. Lenz, M. Reinecke, C. Rosset, E. Hivon, and K. Gorski, HEALPY: Equal area pixelization and spherical harmonics transforms for data on the sphere in python, *J. Open Source Software* **4**, 1298 (2019).
- [33] https://abaleato.github.io/kappa_delensing/.
- [34] D. Huterer, C. E. Cunha, and W. Fang, Calibration errors unleashed: Effects on cosmological parameters and requirements for large-scale structure surveys, *Mon. Not. R. Astron. Soc.* **432**, 2945 (2013).
- [35] A. Baleato Lizancos and M. White, The impact of anisotropic redshift distributions on angular clustering, [arXiv:2305.15406](https://arxiv.org/abs/2305.15406).
- [36] DES Collaboration, Dark energy survey year 3 results: Photometric dataset for cosmology, *Astrophys. J. Suppl. Ser.* **254**, 24 (2021).
- [37] O. Darwish and ACT Collaboration, The Atacama Cosmology Telescope: A CMB lensing mass map over 2100 square degrees of sky and its cross-correlation with BOSS-CMASS galaxies, *Mon. Not. R. Astron. Soc.* **500**, 2250 (2021).
- [38] A. Baleato Lizancos, A. Challinor, and J. Carron, Limitations of CMB B -mode template delensing, *Phys. Rev. D* **103**, 023518 (2021).