UC Berkeley

Title

Lower bias, lower noise CMB lensing with foreground-hardened estimators

Permalink

https://escholarship.org/uc/item/5hw551rf

Journal

ISSN

Authors

Sailer, Noah
Schaan, Emmanuel
Ferraro, Simone

Publication Date

DOI

Peer reviewed

Lower bias, lower noise CMB lensing with foreground-hardened estimators

Noah Sailer,[1, *] Emmanuel Schaan,[2, 1, †] and Simone Ferraro[2, 1, ‡]

[1]*Berkeley Center for Cosmological Physics, Department of Physics,*
University of California, Berkeley, CA 94720, USA
[2]*Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, USA*

Extragalactic foregrounds in temperature maps of the Cosmic Microwave Background (CMB) severely limit the ability of standard estimators to reconstruct the weak lensing potential. These foregrounds are not fully removable by multi-frequency cleaning or masking and can lead to large biases if not properly accounted for. For foregrounds made of a number of unclustered point sources, an estimator for the source amplitude can be derived and deprojected, removing any bias to the lensing reconstruction. We show with simulations that all of the extragalactic foregrounds in temperature can be approximated by a collection of sources with identical profiles, and that a simple bias hardening technique is effective at reducing any bias to lensing, at a minimal noise cost. We compare the performance and bias to other methods such as "shear-only" reconstruction, and discuss how to jointly deproject any arbitrary number of foregrounds, each with an arbitrary profile. In particular, for a Simons Observatory-like experiment foreground-hardened estimators allow us to extend the maximum multipole used in the reconstruction, increasing the overall statistical power by $\sim 50\%$ over the standard quadratic estimator, both in auto and cross-correlation. We conclude that source hardening outperforms the standard lensing quadratic estimator both in auto and cross-correlation, and in terms of lensing signal-to-noise and foreground bias.

I. INTRODUCTION

CMB lensing [1, 2] is rapidly becoming one of the most powerful tools in a cosmologist's toolkit. With the ability to measure matter fluctuations to high redshift, no photometric redshift uncertainties and its sensitivity to crucial cosmological parameters, sub-percent precision measurements have the potential to greatly increase our understanding of the Universe.

Current and upcoming wide-field CMB experiments such as AdvACT [3], SPT-3G [4] and Simons Observatory [5] will heavily rely on reconstruction from CMB temperature fluctuations (as opposed to polarization). For future polarization-dominated experiments like CMB-S4 [6] and CMB-HD [7, 8], temperature lensing will still contribute a non-negligible part of the total signal-to-noise.

Temperature maps contain significant contamination from extragalactic foregrounds such as the thermal and kinematic Sunyaev-Zel'dovich effects (tSZ and kSZ), the Cosmic Infrared Background (CIB), and both radio and infrared point sources. Lensing reconstruction on contaminated maps creates significant biases in both the auto- and cross-correlation of the reconstructed field, which in turn can bias our inference of cosmology [9–14]. These biases can be up to several tens of percent for upcoming surveys, thus severely limiting our ability to use CMB lensing for precision cosmology.

Several techniques have been proposed and implemented to mitigate the impact of extragalactic foregrounds. Some are based on symmetries, such as the recently proposed shear-only reconstruction [13] (and related generalizations such as hybrid and general multipole estimators), while others make use of multi-frequency maps to reduce the impact of particular foregrounds [11, 15].

Here we revisit and extend a bias hardening technique first proposed in [10, 16] and applied to real data in [17], and compare it to shear-only reconstruction and the hybrid estimator of [13]. We show that if we approximate extragalactic foregrounds as a collection of sources with identical profiles, an estimator for their amplitude can be obtained and hardened against in an optimal way to mitigate their impact on the reconstructed lensing convergence. With the use of realistic and correlated simulations of all of the foregrounds, we show that this approximation is excellent and leads to a very large suppression in bias, allowing for the use of considerably smaller scales in the reconstruction.

The remainder of the paper is organized as follows: In Section II we review the general formalism for bias hardening against a single foreground, then generalize the procedure to an arbitrary number of foregrounds, each with arbitrary profiles. In Section III we specialize to the case where foregrounds are point sources, and in Section IV we numerically explore the biases and the performance of mitigation through bias hardening or shear-only estimators. We conclude with Section V.

II. FOREGROUND-HARDENED QUADRATIC ESTIMATORS

In this section, we first review the derivation of the standard lensing quadratic estimator (QE), and rephrase it in terms of the linear response of the map covariance to the lensing convergence. We build on this formalism to construct foreground estimators, and finally null the linear response of the lensing QE to these foregrounds, as in [10, 16].

---

* nsailer@berkeley.edu
† eschaan@lbl.gov
‡ sferraro@lbl.gov

A. Lensing quadratic estimator, bispectrum & linear response

The observed CMB temperature T is the sum of the *lensed* primary CMB T^{CMB} and the foregrounds s, which will be approximated by a collection of sources in this work: $T = T^{\mathrm{CMB}} + s$. The lensed CMB temperature T^{CMB} field is statistically invariant under translations. As a result, the off-diagonal covariance $\langle T_{\boldsymbol{\ell}}^{\mathrm{CMB}} T_{\boldsymbol{L}-\boldsymbol{\ell}}^{\mathrm{CMB}} \rangle$ with $\boldsymbol{L} \neq 0$ is zero. However, fixing the lensing convergence Fourier mode $\kappa_{\boldsymbol{L}}$ breaks this statistical homogeneity, and produces non-zero off-diagonal covariances:

$$\langle T_{\boldsymbol{\ell}}^{\mathrm{CMB}} T_{\boldsymbol{L}-\boldsymbol{\ell}}^{\mathrm{CMB}} \rangle_{\text{at fixed } \kappa_{\boldsymbol{L}}} = f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} \kappa_{\boldsymbol{L}} + \mathcal{O}\left(\kappa_{\boldsymbol{L}}^2\right)$$
$$\text{with} \quad f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} \equiv \frac{2\boldsymbol{L}}{L^2} \cdot \left[\boldsymbol{\ell} C_{\ell}^0 + (\boldsymbol{L}-\boldsymbol{\ell}) C_{|\boldsymbol{L}-\boldsymbol{\ell}|}^0 \right], \tag{1}$$

where C_{ℓ}^0 is the *unlensed* power spectrum. From this, we obtain an unbiased lensing estimator (to first order in κ) from the ratio $T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}} / f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa}$. The standard lensing QE [18] is simply the inverse-variance weighted average of all these "building block" estimators, summing over $\boldsymbol{\ell}$ at fixed $\boldsymbol{L}$. This leads to the usual result:

$$\hat{\kappa}_{\boldsymbol{L}} = N_{\boldsymbol{L}}^{\kappa} \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}}$$
$$(N_{\boldsymbol{L}}^{\kappa})^{-1} = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} \tag{2}$$
$$F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} = \frac{f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa}}{2 C_{\ell}^{\mathrm{tot}} C_{|\boldsymbol{L}-\boldsymbol{\ell}|}^{\mathrm{tot}}}.$$

In order to generalize this construction to foreground quadratic estimators, we consider Eq. (1) and note that $f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa}$ can be seen as the linear response of the two point correlator $\langle T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}} \rangle_{\text{at fixed } \kappa_{\boldsymbol{L}}}$ to the lensing convergence mode $\kappa_{\boldsymbol{L}}$:

$$f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} = \frac{\delta \langle T_{\boldsymbol{\ell}}^{\mathrm{CMB}} T_{\boldsymbol{L}-\boldsymbol{\ell}}^{\mathrm{CMB}} \rangle_{\text{at fixed } \kappa_{\boldsymbol{L}}}}{\delta \kappa_{\boldsymbol{L}}} (\kappa_{\boldsymbol{L}} = 0). \tag{3}$$

Furthermore, by multiplying both sides of Eq. (1) with $\kappa_{-\boldsymbol{L}}$, we see that this linear response can be expressed as the ratio of the following bispectrum and power spectrum:

$$f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} = \frac{\langle T_{\boldsymbol{\ell}}^{\mathrm{CMB}} T_{\boldsymbol{L}-\boldsymbol{\ell}}^{\mathrm{CMB}} \kappa_{-\boldsymbol{L}} \rangle}{\langle \kappa_{\boldsymbol{L}} \kappa_{-\boldsymbol{L}} \rangle}. \tag{4}$$

This final expression is our starting point to derive foreground quadratic estimators.

B. Foreground quadratic estimators

Following the argument above, anytime a field s (e.g. a foreground, the mask, the beam) has a non-zero bispectrum $\langle s_{\boldsymbol{\ell}} s_{\boldsymbol{L}-\boldsymbol{\ell}} s_{-\boldsymbol{L}} \rangle$, we can define the following linear response[1]:

$$f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s} = \frac{\langle s_{\boldsymbol{L}} s_{\boldsymbol{L}-\boldsymbol{\ell}} s_{-\boldsymbol{L}} \rangle}{\langle s_{\boldsymbol{L}} s_{-\boldsymbol{L}} \rangle} \tag{5}$$

such that

$$\langle s_{\boldsymbol{\ell}} s_{\boldsymbol{L}-\boldsymbol{\ell}} \rangle_{\text{at fixed } s_{\boldsymbol{L}}} = f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s} s_{\boldsymbol{L}} + \mathcal{O}\left((s_{\boldsymbol{L}})^2\right). \tag{6}$$

We can then define the "building block" foreground estimator, unbiased to first order in the foreground s, via $T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}} / f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s}$. Finally, we inverse-variance weight these estimators for all $\boldsymbol{\ell}$, at fixed $\boldsymbol{L}$, to obtain the standard minimum variance quadratic estimator for s. In practice, evaluating this estimator requires knowledge of the foreground non-Gaussianity, through its bispectrum. We note however that no assumption about the foreground trispectrum is used.

C. Bias-hardened lensing quadratic estimators

In this section we review the derivation of the bias-hardened (BH) lensing estimator of [10, 16] for a single foreground, relate the noise of the BH estimator to the noise of the standard minimum variance QE, and generalize the procedure for an arbitrary number of foregrounds.

1. Hardening against a single foreground

In the presence of a single foreground s, the off-diagonal covariances of the observed temperature map take the form

$$\langle T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}} \rangle = f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} \kappa_{\boldsymbol{L}} + f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s} s_{\boldsymbol{L}} \tag{7}$$

to lowest order in κ and s. In this equation the average is taken at fixed $\kappa_{\boldsymbol{L}}$ and $s_{\boldsymbol{L}}$, which will be implied from here on out for notational economy. Since both $\kappa_{\boldsymbol{L}}$ and $s_{\boldsymbol{L}}$ produce off-diagonal covariances, the standard quadratic estimators for κ and s acquire a bias. These biases can be written in the form [10]:

$$\begin{pmatrix} \langle \hat{\kappa}_{\boldsymbol{L}} \rangle \\ \langle \hat{s}_{\boldsymbol{L}} \rangle \end{pmatrix} = \begin{pmatrix} 1 & N_{\boldsymbol{L}}^{\kappa} \mathcal{R}_{\boldsymbol{L}} \\ N_{\boldsymbol{L}}^{s} \mathcal{R}_{\boldsymbol{L}} & 1 \end{pmatrix} \begin{pmatrix} \kappa_{\boldsymbol{L}} \\ s_{\boldsymbol{L}} \end{pmatrix} \tag{8}$$

where the response $\mathcal{R}_{\boldsymbol{L}}$ is defined to be

$$\mathcal{R}_{\boldsymbol{L}} = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} \frac{f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa} f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s}}{2 C_{\ell}^{\mathrm{tot}} C_{|\boldsymbol{L}-\boldsymbol{\ell}|}^{\mathrm{tot}}}. \tag{9}$$

We see that the bias to $\hat{\kappa}$ is proportional to s, and vice-versa. This allows us to create new bias-hardened estimators, which are the unique linear combinations of $\hat{\kappa}$ and

---

[1] We note that this expression only accounts for the non-Gaussianity of the sources. In principle, the sources are themselves lensed which results in an additional correlation and κ dependence of the mask. For simplicity, we ignore this effect in this work.

$\hat{s}$ that null these biases

$$\begin{pmatrix} \hat{\kappa}_{\boldsymbol{L}}^{\mathrm{BH}} \\ \hat{s}_{\boldsymbol{L}}^{\mathrm{BH}} \end{pmatrix} = \begin{pmatrix} 1 & N_{\boldsymbol{L}}^{\kappa}\mathcal{R}_{\boldsymbol{L}} \\ N_{\boldsymbol{L}}^{s}\mathcal{R}_{\boldsymbol{L}} & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\kappa}_{\boldsymbol{L}} \\ \hat{s}_{\boldsymbol{L}} \end{pmatrix}. \qquad (10)$$

By taking an average of Eq. (10), and inserting Eq. (8) on the right hand side, we see that bias-hardened estimators indeed satisfy $\langle \hat{\kappa}_{\boldsymbol{L}}^{\mathrm{BH}} \rangle = \kappa_{\boldsymbol{L}}$ and $\langle \hat{s}_{\boldsymbol{L}}^{\mathrm{BH}} \rangle = s_{\boldsymbol{L}}$.

2. Noise of the hardened lensing QE

By construction, the weights of the standard QE are chosen to minimize its variance. Since the BH estimator's weights differ from the standard QE's, bias-hardening comes at a price in noise, which we quantify in this section. We start with Eq. (10), the definition of the BH estimator, from which we compute the variance

$$\langle \hat{\kappa}_{\boldsymbol{L}}^{\mathrm{BH}} \hat{\kappa}_{-\boldsymbol{L}}^{\mathrm{BH}} \rangle = \Bigg[\langle \hat{\kappa}_{\boldsymbol{L}} \hat{\kappa}_{-\boldsymbol{L}} \rangle + (N_{\boldsymbol{L}}^{\kappa}\mathcal{R}_{\boldsymbol{L}})^2 \langle \hat{s}_{\boldsymbol{L}} \hat{s}_{-\boldsymbol{L}} \rangle$$
$$- 2N_{\boldsymbol{L}}^{\kappa}\mathcal{R}_{\boldsymbol{L}} \mathrm{Re} \langle \hat{\kappa}_{\boldsymbol{L}} \hat{s}_{-\boldsymbol{L}} \rangle \Bigg] \Bigg/ \left(1 - N_{\boldsymbol{L}}^{\kappa} N_{\boldsymbol{L}}^{s}\mathcal{R}_{\boldsymbol{L}}^2\right)^{-2}. \qquad (11)$$

In this equation we've used the identity $N_{\boldsymbol{L}}^{\kappa} = N_{-\boldsymbol{L}}^{\kappa}$, and we've assumed that $f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s}$ is even in both of its arguments, which forces both the noise of the source estimator and the response to be even functions of $\boldsymbol{L}$. This assumption is valid when modeling the field s as a collection of individual sources with profiles $u_{\boldsymbol{\ell}}$ (since $u_{\boldsymbol{\ell}} = u_{-\boldsymbol{\ell}}$ for any physical emission profile), as in section III.

The cross-correlation of $\hat{\kappa}$ and $\hat{s}$ is proportional to the response function, and is explicitly given by

$$\langle \hat{\kappa}_{\boldsymbol{L}} \hat{s}_{\boldsymbol{L}'} \rangle = (2\pi)^2 \delta_{\boldsymbol{L}+\boldsymbol{L}'}^{D} N_{\boldsymbol{L}}^{\kappa} N_{\boldsymbol{L}}^{s}\mathcal{R}_{\boldsymbol{L}}. \qquad (12)$$

Combining Eq. (11) and (12) results in the simple expression

$$N_{\boldsymbol{L}}^{\kappa^{\mathrm{BH}}} = N_{\boldsymbol{L}}^{\kappa} \left(1 - N_{\boldsymbol{L}}^{\kappa} N_{\boldsymbol{L}}^{s}\mathcal{R}_{\boldsymbol{L}}^2\right)^{-1}. \qquad (13)$$

We see that the noise cost for bias-hardening is completely determined by the determinant of the matrix in Eq. (10). The value of this determinant is plotted in Fig. 1 for the case of point source foregrounds ($f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s} = 1$). Nearly identical calculations yield the noise of the bias-hardened source estimator and the noise cross-correlation of the two bias-hardened estimators. The results are

$$N_{\boldsymbol{L}}^{s^{\mathrm{BH}}} = N_{\boldsymbol{L}}^{s} \left(1 - N_{\boldsymbol{L}}^{\kappa} N_{\boldsymbol{L}}^{s}\mathcal{R}_{\boldsymbol{L}}^2\right)^{-1}$$
$$\langle \hat{\kappa}_{\boldsymbol{L}}^{\mathrm{BH}} \hat{s}_{\boldsymbol{L}}^{\mathrm{BH}} \rangle = -\langle \hat{\kappa}_{\boldsymbol{L}} \hat{s}_{\boldsymbol{L}} \rangle \left(1 - N_{\boldsymbol{L}}^{\kappa} N_{\boldsymbol{L}}^{s}\mathcal{R}_{\boldsymbol{L}}^2\right)^{-1}. \qquad (14)$$

3. Simultaneously hardening against n foregrounds

It is straightforward to generalize the bias-hardening procedure for n foreground fields $s_{1,\boldsymbol{L}}, \cdots, s_{n,\boldsymbol{L}}$. In this case the observed map has off-diagonal covariances

$$\langle T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}} \rangle = f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{\kappa}\kappa_{\boldsymbol{L}} + f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s_1}s_{1,\boldsymbol{L}} + \cdots + f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s_n}s_{n,\boldsymbol{L}} \quad (15)$$
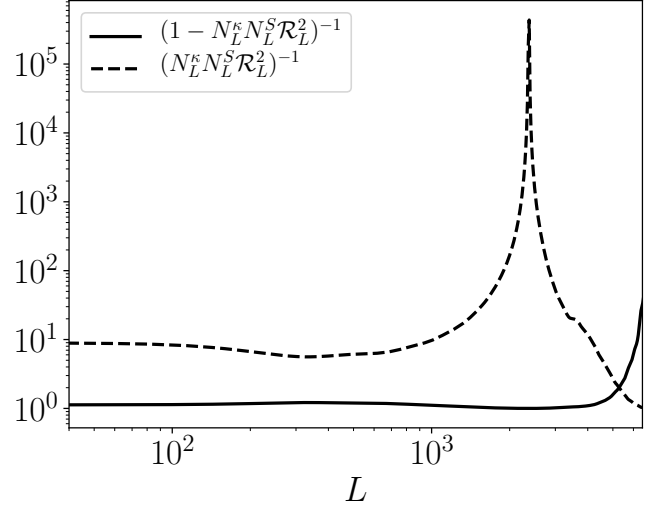

Figure 1. Bias-hardening against point source foregrounds. The solid curve is the determinant of the matrix in Eq. (10), which is the ratio of the bias-hardened estimator's noise to the standard QE's noise. It is close to unity on most scales, implying that the noise penalty for bias hardening is minimal. The dashed curve is the ratio of the SNR of the point source estimator to the relative bias to C_L^{κ} from the source trispectrum. This ratio is roughly 10 at all scales. Thus in the regime where the bias to the standard QE is significant, the source estimator reconstructs the sources with a high SNR, and the source-hardened estimator can remove the bias due to sources with little noise cost. The spike is due to a zero crossing in the response $\mathcal{R}_L$ at $L \sim 2500$.

to lowest order in κ and the foreground fields. Just as before, this average is taken at fixed $\kappa_{\boldsymbol{L}}, s_{1,\boldsymbol{L}}, \cdots, s_{n,\boldsymbol{L}}$. First, we write down the standard minimum variance quadratic estimators for each of the fields:

$$\hat{X}_{\boldsymbol{L}} = N_{\boldsymbol{L}}^{X} \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{X} T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}}$$
$$\left(N_{\boldsymbol{L}}^{X}\right)^{-1} = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{X} f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{X} \qquad (16)$$
$$F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{X} = \frac{f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{X}}{2C_{\ell}^{\mathrm{tot}} C_{|\boldsymbol{L}-\boldsymbol{\ell}|}^{\mathrm{tot}}},$$

where $X = \kappa, s_1, \cdots, s_n$. We then write down the $(n+1) \times (n+1)$ matrix that quantifies the biases to all the estimators, analogous to Eq. (8). Finally, we invert this matrix to obtain the bias-hardened estimators

$$\begin{pmatrix} \hat{\kappa}_{\boldsymbol{L}}^{\mathrm{BH}} \\ \hat{s}_{1,\boldsymbol{L}}^{\mathrm{BH}} \\ \vdots \\ \hat{s}_{n,\boldsymbol{L}}^{\mathrm{BH}} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & N_{\boldsymbol{L}}^{\kappa}\mathcal{R}_{\boldsymbol{L}}^{\kappa,s_1} & \cdots & N_{\boldsymbol{L}}^{\kappa}\mathcal{R}_{\boldsymbol{L}}^{\kappa,s_n} \\ N_{\boldsymbol{L}}^{s_1}\mathcal{R}_{\boldsymbol{L}}^{s_1,\kappa} & 1 & \cdots & N_{\boldsymbol{L}}^{s_1}\mathcal{R}_{\boldsymbol{L}}^{s_1,s_n} \\ \vdots & \vdots & \vdots & \vdots \\ N_{\boldsymbol{L}}^{s_n}\mathcal{R}_{\boldsymbol{L}}^{s_n,\kappa} & N_{\boldsymbol{L}}^{s_n}\mathcal{R}_{\boldsymbol{L}}^{s_n,s_1} & \cdots & 1 \end{pmatrix}^{-1}}_{\equiv M_{\boldsymbol{L}}^{-1}} \begin{pmatrix} \hat{\kappa}_{\boldsymbol{L}} \\ \hat{s}_{1,\boldsymbol{L}} \\ \vdots \\ \hat{s}_{n,\boldsymbol{L}} \end{pmatrix}$$
$$(17)$$

where the generalized response is defined to be

$$\mathcal{R}_{\boldsymbol{L}}^{X,Y} = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} \frac{f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{X} f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{Y}}{2C_{\ell}^{\mathrm{tot}} C_{|\boldsymbol{L}-\boldsymbol{\ell}|}^{\mathrm{tot}}}. \qquad (18)$$

Just as in the previous section, the response $\mathcal{R}_L^{X,Y}$ is proportional to the cross-correlation of the two fields X and Y. When $X = Y$, the response is simply the inverse of the noise, which is manifested in Eq. (17) by the 1's on the diagonal. Writing the matrix $\boldsymbol{M_L}$ as

$$\boldsymbol{M_L} = \begin{pmatrix} 1 & \boldsymbol{w_L^T} \\ \boldsymbol{v_L} & \boldsymbol{A_L} \end{pmatrix}, \qquad (19)$$

where $\boldsymbol{A_L}$ is a $n \times n$ matrix and $\boldsymbol{w_L}, \boldsymbol{v_L}$ are n-dimensional column vectors, we find that the noise of the BH lensing estimator is simply

$$N_L^{\kappa^{\mathrm{BH}}} = \frac{\det(\boldsymbol{A_L})}{\det(\boldsymbol{M_L})} N_L^\kappa. \qquad (20)$$

A derivation of this equation is given in Appendix C.

III. SOURCE QE AND SOURCE-HARDENED LENSING QE

A. Halo model for sources

In the previous section we reviewed how to bias-harden against any number of foregrounds. The only input to this procedure is the linear response $f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^s$ of each foreground, which can be calculated if the foreground's bispectrum and power spectrum are known. To estimate these quantities we adopt a halo model, writing each foreground $s(\boldsymbol{x})$ as a sum of individual sources with profiles $u(\boldsymbol{x})$:

$$s(\boldsymbol{x}) = \sum_i s_i u(\boldsymbol{x} - \boldsymbol{x}_i), \qquad (21)$$

where s_i and $\boldsymbol{x}_i$ are the flux and position of the i'th source, and the profile $u(\boldsymbol{x})$ is assumed to be source independent. These sources are assumed to be a Poisson sampling of the matter density field. Because they trace matter on large scales, the source field s is correlated with the true lensing convergence field. As we show below, this correlation causes a bias in the CMB lensing cross- and auto-correlations. With this model, the linear response takes the form

$$f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^s = \frac{\langle s_\ell s_{\boldsymbol{L}-\boldsymbol{\ell}} s_{-\boldsymbol{L}} \rangle}{\langle s_{\boldsymbol{L}} s_{-\boldsymbol{L}} \rangle} = \frac{\langle s_i^3 \rangle}{\langle s_i^2 \rangle} \frac{u_\ell u_{\boldsymbol{L}-\boldsymbol{\ell}}}{u_{\boldsymbol{L}}}, \qquad (22)$$

where in the second equality we assumed that the sources are sparse, such that the bispectrum and power spectrum are both shot noise dominated, and the source clustering is negligible. In this case the linear response simplifies to a separable function of $\boldsymbol{\ell}$ and $\boldsymbol{L}-\boldsymbol{\ell}$, and is proportional to the ratio of the third to the second moment of the individual source fluxes.

We can choose the normalization of the source estimator $\hat{s}$ to remove the constant factor $\langle s_i^3 \rangle/\langle s_i^2 \rangle$. Doing so makes $\hat{s}$ an estimator of $s' \equiv s\langle s_i^2 \rangle/\langle s_i^3 \rangle$, which has linear response $f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^{s'} = u_\ell u_{\boldsymbol{L}-\boldsymbol{\ell}}/u_{\boldsymbol{L}}$. This simplifies the expression of the source estimator to

$$\hat{s}_{\boldsymbol{L}} = N_{\boldsymbol{L}}^s \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} \frac{T_{\boldsymbol{\ell}} T_{\boldsymbol{L}-\boldsymbol{\ell}}}{2 C_\ell^{\mathrm{tot}} C_{|\boldsymbol{L}-\boldsymbol{\ell}|}^{\mathrm{tot}}} \frac{u_\ell u_{\boldsymbol{L}-\boldsymbol{\ell}}}{u_{\boldsymbol{L}}}$$

$$(N_{\boldsymbol{L}}^s)^{-1} = \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} \frac{1}{2 C_\ell^{\mathrm{tot}} C_{|\boldsymbol{L}-\boldsymbol{\ell}|}^{\mathrm{tot}}} \left[\frac{u_\ell u_{\boldsymbol{L}-\boldsymbol{\ell}}}{u_{\boldsymbol{L}}} \right]^2, \qquad (23)$$

which can be quickly evaluated using FFT. We emphasize that the weights and normalization of the BH estimator are left unchanged by this renormalization of $\hat{s}$. Therefore the bias-hardening procedure requires no knowledge of the source amplitude, and the normalization of the profile u is completely arbitrary. For this reason we will drop the primes from here-on-out, defining $f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^s = u_\ell u_{\boldsymbol{L}-\boldsymbol{\ell}}/u_{\boldsymbol{L}}$.

B. Noise cost and importance of the non-Gaussianity

Consider a single source component s. If s were Gaussian, it would be indistinguishable from map noise, and the source estimator $\hat{s}$ would not be able to pick it out. Here, we show how the non-Gaussianity of the source map controls both the signal-to-noise (SNR) of the source estimator, and the bias from sources to CMB lensing. For simplicity let s be a Poisson distributed point source field ($f_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^s = u_\ell = 1$). In this case the mean number of sources $\bar{n}$ determines the non-Gaussianity; as $\bar{n}$ goes to infinity, the statistics converges to that of a Gaussian.

The point source estimator is an unbiased (to lowest order) estimator of s, so the SNR per Fourier mode is simply :

$$\mathrm{SNR}_L^s = \langle \hat{s}_{\boldsymbol{L}} \hat{s}_{-\boldsymbol{L}} \rangle / N_L^s \qquad (24)$$

$$= \mathcal{T}^s / N_L^s. \qquad (25)$$

Thus the SNR for the source estimator scales with the scale-independent source trispectrum $\mathcal{T}^s$: it is larger for low source densities which corresponds to a higher non-Gaussianity. As the number of sources increases, the field becomes increasingly Gaussian, and can no longer be distinguished from Gaussian noise in the map. The expression above is valid for mildly non-Gaussian source field, as is the case in reality. For the very highly non-Gaussian case, which is not relevant for the CMB, the source non-Gaussianity contributes also to the noise, leading to a saturation of the SNR to a finite value as $\bar{n}$ goes to zero.

Even if the source field and lensing convergence were independent, point sources would cause a bias to the standard QE due to its trispectrum:

$$\frac{\text{bias to } C_L^\kappa \text{ from } \mathcal{T}^s}{N_L^\kappa} = \frac{\mathcal{T}^s}{N_L^\kappa} \left(N_L^\kappa \int \frac{d^2\boldsymbol{\ell}}{(2\pi)^2} F_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}^\kappa \right)^2$$

$$= \mathcal{T}^s N_L^\kappa \mathcal{R}_L^2. \qquad (26)$$

Here again, the source non-Gaussianity controls the bias to lensing via the $1/\bar{n}$ scaling. From equations (25) and (26), we see that the ratio of the SNR of the source estimator to the relative bias to lensing from the source

trispectrum is $(N_L^\kappa N_L^S \mathcal{R}_L^2)^{-1}$. This ratio cancels out the non-Gaussianity of the sources and is around 10 for most scales of interest, as shown in Fig 1. This means that subtracting the lensing bias due to the source trispectrum will always cause only a small loss in lensing SNR. Indeed, as shown in Fig 1 (solid line), the lensing noise for the standard QE and the bias hardened estimators are very similar, meaning that subtracting the source bias causes a negligible increase in the lensing noise. This is consistent with the findings of [19].

Finally, the non-Gaussianity of the point sources also contributes an extra noise term in the lensing reconstruction, which we haven't included. However, this noise is only important when the bias to CMB lensing is also important. Thus, by making sure that the non-Gaussian point sources (and other foregrounds) do not significantly bias the lensing signal, we are making sure that their noise contribution is also small.

IV. BIASES TO CMB LENSING & FOREGROUND HARDENING

In this section we quantify the biases for three different bias-hardening schemes, and compare these results to the standard QE and Shear estimators. We consider two source components: point-sources $f_{\ell,L-\ell}^{\rm PS} = 1$ and sources with tSZ-like profiles $f_{\ell,L-\ell}^{\rm tSZ} = u_\ell u_{L-\ell}/u_L$. The profile u_ℓ is calculated by taking the square root of the tSZ power spectrum, which is measured from the simulations. Our three different bias-hardening schemes harden against one or both of these source components. The Point Source Hardened (PSH) estimator hardenens against point sources; the Profile Hardened (PH) estimator hardens against a tSZ-like cluster profile, as described in App. D; and the Point source and Profile Hardened (PPH) estimator simultaneously hardens against both. We show the noise power spectra for these various estimators in Fig. 2, and performed various checks of our code in App. F.

A. Bias to CMB lensing from non-Gaussian foregrounds

The presence of a non-Gaussian foreground component s in the temperature map produces three bias terms to CMB lensing, referred to as primary, secondary and trispectrum terms [9, 10, 12, 13]. Indeed, if $T = T^{\rm CMB} + s$, then the quadratic estimator $\mathcal{Q}$ applied to T can be expanded bilinearly as:

$$\mathcal{Q}[T,T] = \mathcal{Q}[T^{\rm CMB}, T^{\rm CMB}] + 2\mathcal{Q}[T^{\rm CMB}, s] + \mathcal{Q}[s,s], \quad (27)$$

assuming the quadratic estimator is symmetric (or has been symmetrized in its arguments).

In cross-correlation with a mass tracer g, this simply leads to the bispectrum bias:

$$\langle g\mathcal{Q}\rangle = \underbrace{\langle g\kappa_{\rm CMB}\rangle}_{\text{lensing signal}} + \underbrace{\langle g\mathcal{Q}[s,s]\rangle}_{\text{bispectrum bias}}. \quad (28)$$
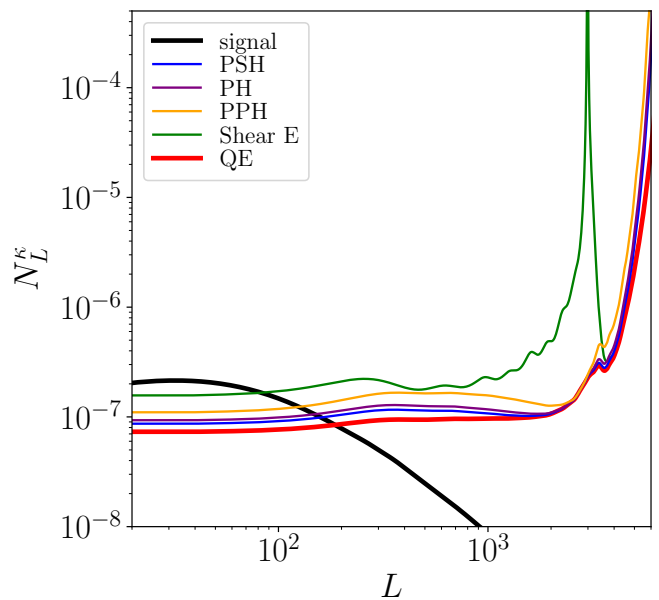

Figure 2. Noises of the Point Source Hardened (PSH), Profile Hardened (PH), Point source and Profile Hardened (PPH), Shear, and standard quadratic (QE) estimators for $\ell_{\rm max,T} = 3500$. In black is the lensing signal C_L^κ. Hardening against a single source (PSH,PH) marginally increases the noise over QE, whereas hardening against multiple sources (PPH) results in a larger noise penalty. We also note that the noise cost of bias-hardening increases with the size of the source profile u_ℓ, which is discussed further in Appendix E.

For the auto-correlation, more terms arise:

$$\begin{aligned}
\langle \mathcal{Q}\mathcal{Q}\rangle = &\underbrace{\langle \kappa_{\rm CMB}\kappa_{\rm CMB}\rangle}_{\text{lensing signal}} \\
&+ \underbrace{2\langle \mathcal{Q}[T^{\rm CMB}, T^{\rm CMB}]\mathcal{Q}[s,s]\rangle}_{\text{Primary bispectrum bias}} \\
&+ \underbrace{4\langle \mathcal{Q}[T^{\rm CMB}, s]\mathcal{Q}[T^{\rm CMB}, s]\rangle}_{\text{Secondary bispectrum bias}} \\
&+ \underbrace{\langle \mathcal{Q}[s,s]\mathcal{Q}[s,s]\rangle}_{\text{Trispectrum bias}}.
\end{aligned} \quad (29)$$

The primary and secondary bias terms are due to the non-Gaussianity of the foreground and its correlation with the true CMB lensing signal. They are integrals of the $\kappa_{\rm CMB}ss$ bispectrum. In particular, predicting them requires knowing the redshift distribution of the foreground. On the other hand, the trispectrum bias is present whether or not the sources are correlated with CMB lensing, and only depend on the statistics (the trispectrum) of the projected 2d foreground map.

In what follows, we compute these biases for the various quadratic lensing estimators of interest. We follow the approach of [13], using the simulations from [20].

B. Effectiveness of foreground hardening: expectations

Here we explain which of these bias terms are nulled with bias-hardening. The only knowledge of foregrounds used in the bias hardening process is the shape of the foreground bispectrum and power spectrum, through f^s. In particular, the procedure does not assume anything about the $ss\kappa_{\rm CMB}$ bispectrum, which causes the primary and secondary biases, nor about the $ssss$ trispectrum, which causes the trispectrum bias. Why would bias hardening help reducing foreground biases at all then? The answer is somewhat subtle.

The bias hardening procedure nulls the linear response of the lensing estimator to the foreground s, i.e. the bispectrum $\langle \mathcal{Q}^{\rm BH}[s,s]\,s\rangle$. This is very similar but not identical to the primary bispectrum bias $\langle \mathcal{Q}^{\rm BH}[s,s]\,\kappa_{\rm CMB}\rangle$. The two bispectra are all the more similar as the foreground map and the true lensing map are highly correlated. Thus in principle, the primary bispectrum bias is only partially reduced by the bias-hardening. In practice though, as we show below with simulated foregrounds, this reduction is very large.

As for the secondary bispectrum bias, there is no reason *a priori* why the bias hardening procedure would reduce it. Indeed, we show below that the various lensing estimators have a similar level of secondary bias.

The case of the trispectrum bias is interesting. In general, if $\mathcal{T}_{\ell_1,\ell_2,\ell_3,\ell_4}$ is the source trispectrum, the trispectrum bias to lensing is given by

$$\langle \mathcal{Q}[s,s]\mathcal{Q}[s,s]\rangle_c =$$
$$N_{\boldsymbol L}^2 \int \frac{d^2\boldsymbol\ell}{(2\pi)^2} F_{\boldsymbol\ell,\boldsymbol L-\boldsymbol\ell} \int \frac{d^2\boldsymbol\ell'}{(2\pi)^2} F_{\boldsymbol\ell',-\boldsymbol L-\boldsymbol\ell'} \mathcal{T}_{\boldsymbol\ell,\boldsymbol L-\boldsymbol\ell,\boldsymbol\ell',-\boldsymbol L-\boldsymbol\ell'}. \tag{30}$$

If the lensing weights of the bias-hardened estimator are denoted as $F_{\boldsymbol\ell,\boldsymbol L-\boldsymbol\ell}^{\rm BH}$, then the zero linear response to the foreground can be translated as:

$$\int \frac{d^2\boldsymbol\ell}{(2\pi)^2} F_{\boldsymbol\ell,\boldsymbol L-\boldsymbol\ell}^{\kappa^{\rm BH}} f_{\boldsymbol\ell,\boldsymbol L-\boldsymbol\ell}^s = 0. \tag{31}$$

This does not null the trispectrum bias in general. However, it does in the case of Poisson sources with identical profiles. Indeed, in this case, the trispectrum $\mathcal{T}_{\ell_1,\ell_2,\ell_3,\ell_4} = \bar n \langle s^4 \rangle u_{\boldsymbol\ell_1} u_{\boldsymbol\ell_2} u_{\boldsymbol\ell_3} u_{\boldsymbol\ell_4}$ is separable, such that the trispectrum bias becomes:

$$\bar n \langle s^4 \rangle N_{\boldsymbol L}^2 \left[\int \frac{d^2\boldsymbol\ell}{(2\pi)^2} F_{\boldsymbol\ell,\boldsymbol L-\boldsymbol\ell}^{\rm BH} u_{\boldsymbol\ell} u_{\boldsymbol L-\boldsymbol\ell} \right]^2. \tag{32}$$

Since the response $f_{\boldsymbol\ell,\boldsymbol L-\boldsymbol\ell}^s = u_{\boldsymbol\ell} u_{\boldsymbol L-\boldsymbol\ell}/u_{\boldsymbol L} \propto u_{\boldsymbol\ell} u_{\boldsymbol L-\boldsymbol\ell}$, we see that Eq. (31) implies that the trispectrum bias is exactly nulled. However, it is no longer generally true if the sources are clustered, or if there is a distribution of profile sizes or shapes. Because our profile hardening assumes a single fiducial source profile, the trispectrum bias is also not exactly nulled when the foreground sources have a range of profiles, as is the case for the tSZ.

In summary, the PH estimator should significantly reduce the tSZ trispectrum bias, while the PSH estimator should exactly null the point source trispectrum. Both should largely reduce the primary bias, with no *a priori* expectation for the secondary bias.

For bias hardening against more complex foregrounds, such as clustered sources, the effectiveness of this simple bias hardening is not guaranteed. However, if the statistical properties of the source distribution is known, it is nonetheless possible to extend our formalism to harden against non-linear clustering. This has been explored in the context of line intensity mapping in [21], and we believe that a similar treatment could be helpful for mitigating biases to CMB lensing.

C. Effectiveness of foreground hardening: simulations

To quantity the biases of each estimator for each of the extragalactic foregrounds, we use simulated maps [20] of the CIB, tSZ, kSZ, radio point sources, and $\kappa_{\rm CMB}$ at 148 GHz. We rescale each of the foregrounds (0.38 for CIB, 0.7 for tSZ, 0.82 for kSZ, 1.1 for radio PS) to match [22], then subtract the mean from each foreground, and finally mask each foreground using a matched filter for point sources with a flux cut of 5 mJy and mask patch radius 3'. The resulting power spectra are plotted in Appendix A.

For CMB measurements, we consider an experiment similar to Simons Observatory [5], with a white noise level of 6 μK-arcmin and a Gaussian beam with full-width at half maximum of $1.4'$.

As in [13], we re-weight the halos from [20] to match the redshift distribution of the LSST gold sample $(dn/dz \propto (z/0.24)^2 e^{-z/0.24}/0.48)$ [23] and obtain a projected galaxy number density count δ_g. Altogether, a set of these maps (foregrounds, $\kappa_{\rm CMB}$, and δ_g) allow us to calculate the biases to the cross-correlation of CMB lensing with galaxy number density counts $C_L^{\kappa\delta_g}$, as well as the primary, secondary, and trispectrum biases to the lensing auto-correlation $C_L^{\kappa\kappa}$. We run the estimators on 81 flat square cutouts from these maps, obtaining the biases plotted in Fig. 3 and 4.

For the cross-correlation of CMB lensing with galaxy counts (Fig. 3), we find that the standard minimum variance quadratic estimator acquires a $\sim 10\%$ bias from the CIB and tSZ for reasonable values of $\ell_{\rm max,T}$. Hardening against a single source component with a tSZ-like profile (PH) effectively nulls the bias from both tSZ and kSZ. Hardening against just point sources (PSH) is complementary, effectively nulling the radio point source and CIB biases. When simultaneously hardening against both tSZ-like profiles and point sources (PPH), we're able to null the biases from all extragalactic foregrounds, creating a highly robust estimator of the cross-correlation.

For the CMB lensing auto-correlation we again find that the standard QE acquires a $\sim 10\%$ bias, primarily due to the CIB and tSZ. This is consistent with the results found in [10]. In Fig. 4 we see that the PH estimator nulls the primary bias from tSZ, and significantly reduces the remaining primary and trispectrum biases. Likewise the PSH estimator nulls the primary bias from point sources,
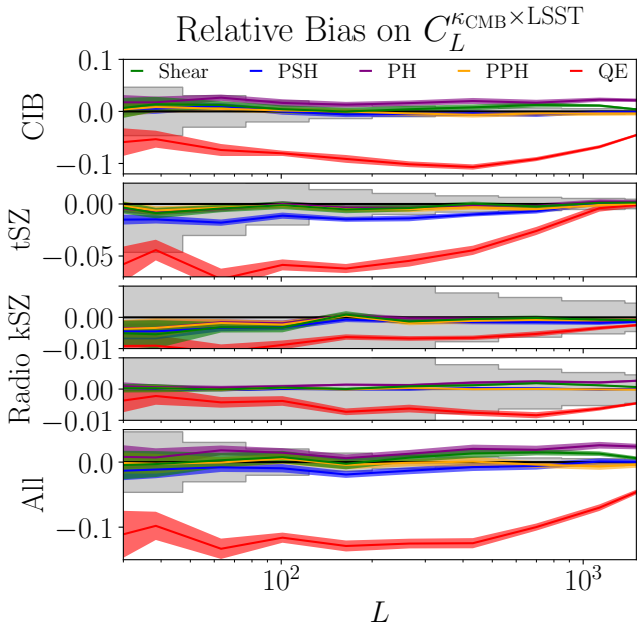
Figure 3. Relative bias on the cross with LSST at $\ell_{\mathrm{max,T}} = 3500$. The gray boxes are the statistical uncertainties when reconstructing $C_L^{\kappa\delta_g}$ with the standard QE. We find that hardening against point sources (PSH) nulls or dramatically reduces the bias from radio point sources and CIB, whereas hardening against a tSZ-like profile (PH) nulls or significantly reduces the bias from tSZ and kSZ. Hardening against both (PPH) effectively nulls the bias from all foregrounds, resulting in a bias that is roughly an order of magnitude smaller than the bias to the standard QE.

while also significantly reducing remaining primary and trispectrum biases. The PPH estimator essentially nulls all primary biases, and reduces the trispectrum bias to a level comparable to Shear. As discussed in the previous section, there is no reason to expect a significant reduction to the secondary biases from bias-hardening. In practice we find that the secondary bias to bias-hardened estimators is slightly worse than (but comparable to) the bias to the standard QE. We note that the secondary bias is negative, whereas the bias from the trispectrum is positive. This results in a cancellation in the overall bias to the lensing power, allowing higher values of $\ell_{\mathrm{max,T}}$ than one might expect from looking at the secondary and trispectrum biases alone.

As a figure of merit for the performance of each estimator, we use the signal-to-noise ratio (SNR) of the lensing amplitude A_{lens}, defined for each L as $C_L^{\kappa,m}/C_L^{\kappa}$, where $C_L^{\kappa,m}$ and C_L^{κ} are the measured and true lensing power spectra respectively. By inverse variance weighting the measurement of A_{lens} for each L, we obtain an estimator for the lensing amplitude:

$$\hat{A}_{\mathrm{lens}} = \int \frac{d^2\boldsymbol{L}}{(2\pi)^2} \frac{C_L^{\kappa,m}}{C_L^{\kappa}} \frac{(C_L^{\kappa})^2}{\sigma_L^2} \bigg/ \int \frac{d^2\boldsymbol{L}}{(2\pi)^2} \frac{(C_L^{\kappa})^2}{\sigma_L^2} \quad (33)$$

$$\text{with } \sigma_L^2 = 2(C_L^{\kappa} + N_L)^2.$$

Since A_{lens} has a fiducial value of 1, the SNR and bias of
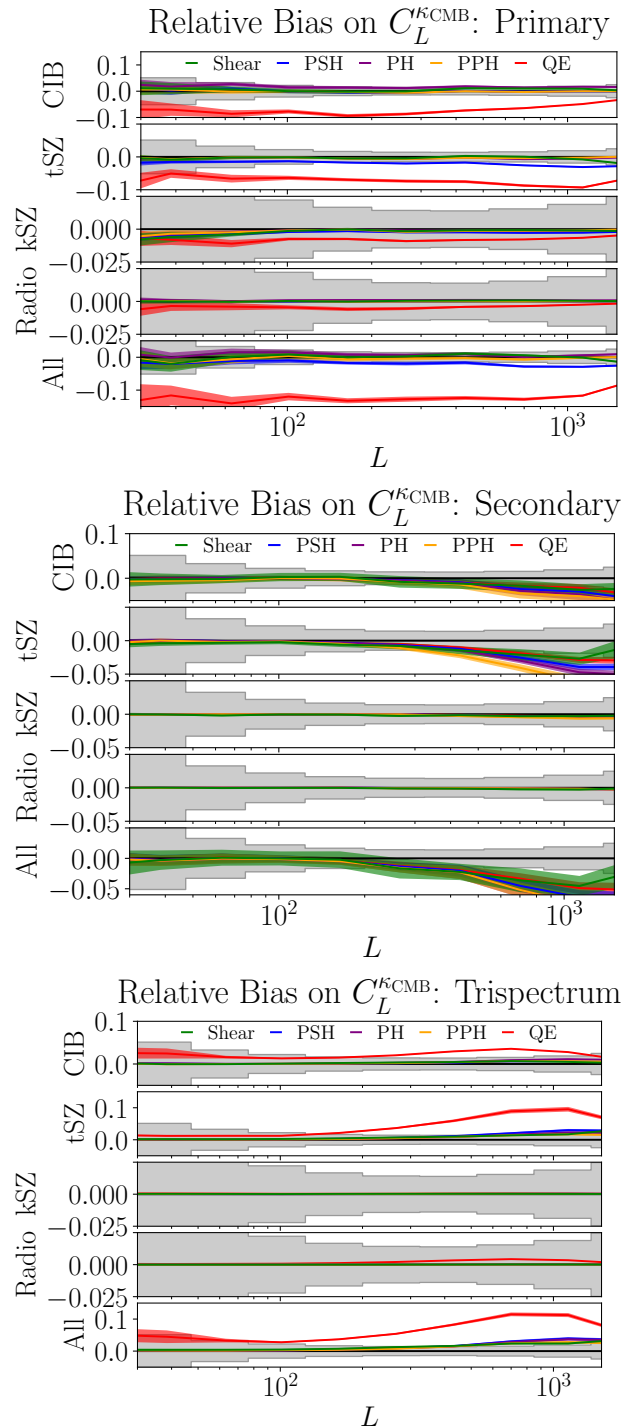
Figure 4. Bias to the lensing auto-correlation at $\ell_{\mathrm{max,T}} = 3500$. The gray boxes are the statistical uncertainties when reconstructing C_L^{κ} with the standard QE. Hardening against point sources (PSH) significantly reduces the primary and trispectrum bias from radio point sources and CIB, whereas hardening against a tSZ-like profile significantly reduces these biases from tSZ. Simultaneously hardening against both (PPH) effectively nulls the primary bias. Bias-hardening has no significant effect on the secondary bias.
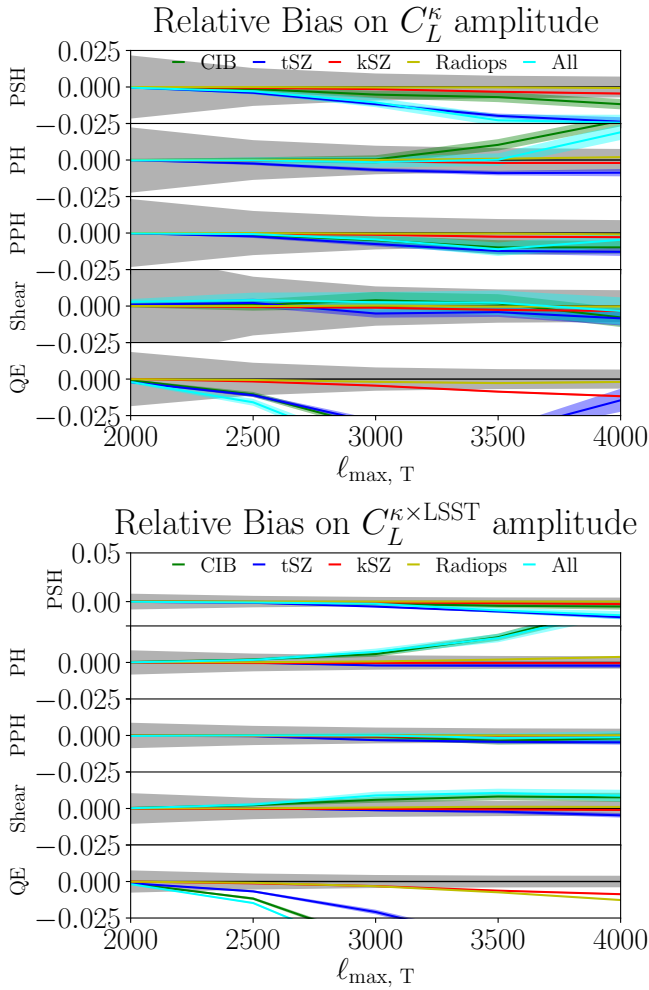
Figure 5. Relative bias on the lensing amplitude and the cross-correlation with LSST as a function of $\ell_{\max,\,T}$. In gray is the 1σ error for each estimator $1/\mathrm{SNR}(A_{\mathrm{lens}})$. CIB and tSZ are the main sources of bias to the standard QE. The Point Source Hardened (PSH) estimator drastically reduces the bias from CIB, whereas the Profile Hardened (PH) estimator reduces the bias from tSZ. The Point source and Profile Hardened (PPH) estimator reduces the bias from both tSZ and CIB to a level comparable to Shear while at a lower noise cost.

A_{lens} are

$$\mathrm{SNR}^2(A_{\mathrm{lens}}) = \int \frac{d^2\boldsymbol{L}}{(2\pi)^2} \frac{(C_L^\kappa)^2}{\sigma_L^2}$$

$$\mathrm{bias}(A_{\mathrm{lens}}) = \int \frac{d^2\boldsymbol{L}}{(2\pi)^2} \frac{\mathrm{bias}(C_L^\kappa)}{C_L^\kappa} \frac{(C_L^\kappa)^2}{\sigma_L^2} \bigg/ \int \frac{d^2\boldsymbol{L}}{(2\pi)^2} \frac{(C_L^\kappa)^2}{\sigma_L^2}.$$

(34)

Similar definitions for the amplitude of the CMB-LSST cross-correlation can be made by replacing the lensing power C_L^κ with $C_L^{\kappa\delta_g}$ and by setting $\sigma_L^2 = (C_L^\kappa + N_L^\kappa)(C_L^{\delta_g} + N_L^{\delta_g}) + (C_L^{\kappa\delta_g})^2$ in the equations above. We show the best SNR achievable in auto and cross-correlation for the various estimators, while maintaining the foreground bias below 1σ, in Tab. I.

For the auto-correlation, we find that the Profile Hardened (PH) estimator achieves the highest SNR be-
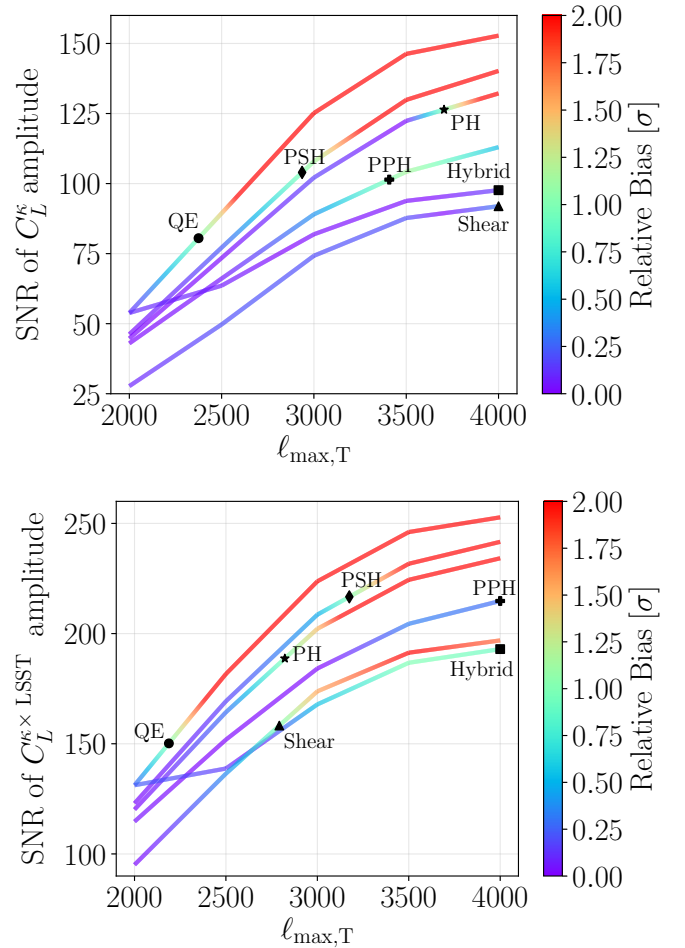
Figure 6. Total signal-to-noise of the lensing amplitude (top) and the cross correlation with LSST (bottom). The black markers denote the highest $\ell_{\max,\,T}$ where the bias is less than 1σ. We find that the Profile Hardened (PH) estimator reconstructs the lensing amplitude with the highest signal to noise: a $\sim 50\%$ improvement over the standard QE. We find that the Point Source Hardened (PSH) and Point source and Profile Hardened (PPH) reconstruct the cross with similar SNRs: a $\sim 50\%$ improvement over the standard QE. The Hybrid estimator is a combination of Shear and QE, and is described in greater detail in Appendix B. We note that the color-scale saturates at 2σ, and that the bias to the standard QE can be $> 10\sigma$ at $\ell_{\max,T} \gtrsim 3000$.

fore $\mathrm{bias}(A_{\mathrm{lens}})$ becomes larger than the noise $\sigma = 1/\mathrm{SNR}(A_{\mathrm{lens}})$, as shown in Fig. 6. This is a $\sim 50\%$ improvement over the standard QE. The PH estimator's 1σ bias threshold occurs at $\ell_{\max,T} \sim 3700$. From Fig. 5, we see that this high value of $\ell_{\max,T}$ is due to a cancellation in the tSZ and CIB biases. A more conservative stopping point would be when any one of the foregrounds crosses the 1σ threshold, which would be around $\ell_{\max,T} \sim 3200$ for the PH estimator. We note that even at this lower $\ell_{\max,T}$, the PH estimator still has the highest SNR and is still a $\sim 30\%$ improvement over QE, which crosses the 1σ bias threshold at $\ell_{\max,T} \sim 2400$.

For the cross-correlation, the PSH and PPH achieve similar SNRs at the 1σ bias threshold, improving upon

QE by $\sim 40\%$. We note that the PPH has a stable bias across the entire range of realistic $\ell_{\max,\mathrm{T}}$'s, making it a conservative yet competitive option for reconstructing the cross-correlation.

We conclude by noting that for an experiment with $\ell_{\max,\mathrm{T}} \sim 3000$, the PH and PPH estimators reconstruct the auto- and cross-correlations respectively with the highest SNR while achieving a sub-percent level bias, as shown in Tab. II.

	SNR in auto ($\ell_{\max,\mathrm{T}}$)	SNR in cross ($\ell_{\max,\mathrm{T}}$)
PSH	104 (2936)	217 (3175)
PH	126 (3705)	189 (2822)
PPH	101 (3408)	215 (4000)
Shear	92 (4000)	158 (2792)
QE	81 (2375)	150 (2188)

Table I. Comparison of the total signal-to-noise ratio on the amplitude of the CMB lensing auto-spectrum (first column) and cross-spectrum with LSST-like galaxies (second column). In all cases, the maximum temperature multipole $\ell_{\max,\,\mathrm{T}}$ is selected to guarantee a foreground bias smaller than 1σ.

	Rel. bias to auto (cross) [%]	Rel. bias to auto (cross) [σ]	SNR on C_L^κ ($C_L^{\kappa\times\mathrm{LSST}}$)
PSH	-1.1 (-0.2)	-1.14 (-0.46)	108 (209)
PH	-0.1 (0.7)	-0.15 (1.47)	102 (202)
PPH	-0.4 (0.1)	-0.39 (0.01)	89 (184)
Shear	0.3 (0.9)	0.19 (1.55)	74 (173)
QE	-4.9 (-5.0)	-6.15 (-11.16)	125 (223)

Table II. Relative bias (in both % and σ units) and SNR for each estimator at $\ell_{\max,\mathrm{T}} = 3000$. The PH and PPH estimators reconstruct the auto- and cross-correlations respectively with the highest SNR while achieving a sub-percent level bias.

V. DISCUSSION AND CONCLUSIONS

In this study we revisited foreground-hardened lensing quadratic estimators, described how to harden against arbitrary profiles, and extended the formalism to simultaneously harden against an arbitrary number of extragalactic foregrounds. We showed that the point source hardened estimator (PSH) reduces not only the bias from radio point sources, but also from the cosmic infrared background, the thermal and the kinematic Sunyaev-Zel'dovich effects.

Motivated by the extended nature of tSZ clusters, we modify the point source hardening technique to deproject the lensing bias from uncorrelated clusters (PH estimator), or simultaneously point sources and extended clusters (PPH estimator). We propose a simple and sufficient method to approximate the input cluster profile from the data itself, by considering the observed tSZ power spectrum from the same data.

The signal-to-noise cost from deprojecting the point source bias is small ($\sim 15\%$ for Point Source Harden-

ing), and is more than compensated by enabling the use of higher multipoles: overall, the SNR on the lensing power spectrum is increased by $\sim 56\%$ when using Profile Hardening (PH), while for cross-correlations, both PSH and PPH provide a $\sim 45\%$ increase in SNR compared to QE. Part of the improvement of PH and PPH estimators comes from the increase in the maximum multipole that can be included in the lensing reconstruction.

The PSH, PH and PPH estimators improve the lensing SNR over the shear and hybrid shear estimators of [13], although at the cost of a slightly larger bias in cross-correlation. They also provide consistency checks for the standard quadratic estimator, having a smaller foreground bias for a given $\ell_{\max}$ in temperature. Overall, these estimators outperform the standard lensing quadratic estimator both in auto and cross-correlation, both in terms of the lensing SNR and in terms of the foreground bias. We recommend their use in the CMB lensing analyses of upcoming temperature-dominated CMB data such as the Simons Observatory. Future surveys that are dominated by polarization reconstruction such as CMB-S4 [6], will still benefit from improved robustness, since temperature reconstruction will contribute a non-trivial fraction to the SNR. We note that a realistic analysis will require modeling of the higher order biases $N^{(i)}$ [19] that appear in lensing reconstruction and will likely be non-negligible for the next generation of surveys.

Finally, foreground hardening can and should be implemented in conjunction with other methods of foreground mitigation, such as masking [9] and multi-frequency cleaning in one or both legs of the estimator. We shall report on the optimal combination of these methods in an upcoming paper.

Appendix A: Foreground power spectra

Our processing of the Sehgal simulations follows [13]. After masking the sources detected in the foreground maps at 5 mJy, which are detected at about 5σ given the noise level assumed, the power spectrum of each foreground at 148 GHz is shown in Fig. 7.
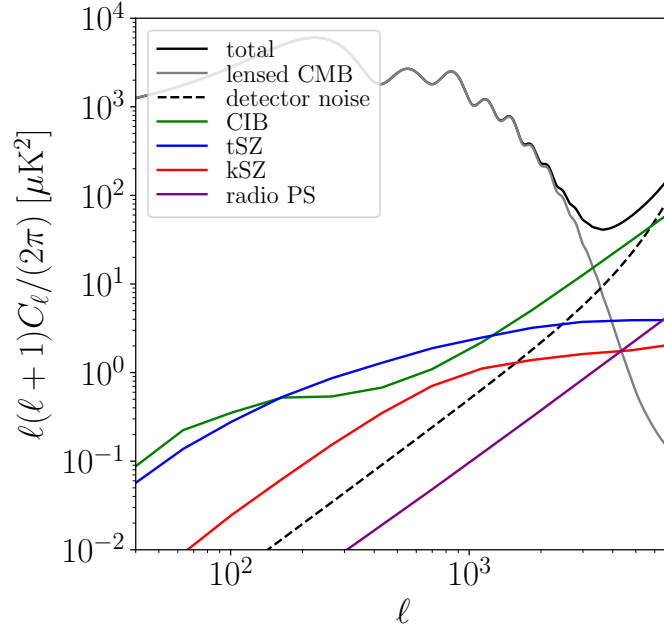

Figure 7. Foreground power spectra at 148 GHz after being masked with a match-filter with a flux cut of 5 mJy and a mask patch radius of 3'. We assume a 6 μK-arcmin noise, and a 1.4' beam.

Appendix B: Hybrid estimator

The hybrid estimator is defined to be

$$\hat{\kappa}_L^{\mathrm{H}} = N_L^{\mathrm{H}} \left(\hat{\kappa}_L^{\mathrm{QE}}/N_L^{\mathrm{QE}} + \hat{\kappa}_L^{\mathrm{S}}/N_L^{\mathrm{S}} \right) \tag{B1}$$

where the standard estimator $\hat{\kappa}_L^{\mathrm{QE}}$ is calculated using $\ell_{\mathrm{min,T}} \le \ell \le 2000$ and the shear estimator $\hat{\kappa}_L^{\mathrm{S}}$ is calculated using $2000 \le \ell \le \ell_{\mathrm{max,T}}$. The noise of the hybrid estimator is simply

$$N_L^{\mathrm{H}} = \frac{1}{1/N_L^{\mathrm{QE}} + 1/N_L^{\mathrm{S}}}.$$

assuming that shear and QE are uncorrelated. The bias B_L to the lensing signal is defined by

$$C_L^{\mathrm{QE}} = C_L^{\kappa} + N_L^{\mathrm{QE}} + B_L^{\mathrm{QE}} \tag{B2}$$
$$C_L^{\mathrm{S}} = C_L^{\kappa} + N_L^{\mathrm{S}} + B_L^{\mathrm{S}}. \tag{B3}$$

Assuming that the noises of $\hat{\kappa}_L^{\mathrm{QE}}$ and $\hat{\kappa}_L^{\mathrm{S}}$ are uncorrelated, the bias to the Hybrid estimator is

$$B_L^{\mathrm{H}} = \left(N_L^{\mathrm{H}} \right)^2 \left(\frac{B_L^{\mathrm{QE}}}{(N^{\mathrm{QE}})^2} + \frac{B_L^{\mathrm{S}}}{(N^{\mathrm{S}})^2} \right). \tag{B4}$$

Here B_L^{QE} is the bias when the standard quadratic estimator is calculated using $\ell_{\mathrm{min,T}} \le \ell \le 2000$. We approximate B_L^{S} as the bias from Shear when calculated with $\ell_{\mathrm{min,T}} \le \ell \le \ell_{\mathrm{max,T}}$. This approximation should be OK since the contribution from $\ell_{\mathrm{min,T}} \le \ell \le 2000$ is negligible. The bias to the hybrid estimator's cross-correlation with LSST is calculated by inverse noise weighting the biases to QE and Shear.

Appendix C: Derivation of Eq. (20)

We start with

$$
M_L = \begin{pmatrix}
1 & N_L^{\kappa}\mathcal{R}_L^{\kappa,s_1} & \cdots & N_L^{\kappa}\mathcal{R}_L^{\kappa,s_n} \\
N_L^{s_1}\mathcal{R}_L^{s_1,\kappa} & 1 & \cdots & N_L^{s_1}\mathcal{R}_L^{s_1,s_n} \\
\vdots & \vdots & \vdots & \vdots \\
N_L^{s_n}\mathcal{R}_L^{s_n,\kappa} & N_L^{s_n}\mathcal{R}_L^{s_n,s_1} & \cdots & 1
\end{pmatrix} \equiv \begin{pmatrix} 1 & w_L^T \\ v_L & A_L \end{pmatrix},
\tag{C1}
$$

where A_L is a $n \times n$ matrix and w_L, v_L are n-dimensional column vectors. By splitting the matrix into blocks, we can express its inverse and determinant as

$$
M_L^{-1} = \begin{pmatrix}
(1 - w_L^T A_L^{-1} v_L)^{-1} & -(1 - w_L^T A_L^{-1} v_L)^{-1} w_L^T A_L^{-1} \\
-A_L^{-1} v_L (1 - w_L^T A_L^{-1} v_L)^{-1} & A_L^{-1} + A_L^{-1} v_L (1 - w_L^T A_L^{-1} v_L)^{-1} w_L^T A_L^{-1}
\end{pmatrix}
\tag{C2}
$$

$$
\det(M_L) = \det(A_L)(1 - w_L^T A_L^{-1} v_L).
\tag{C3}
$$

Using this notation, the bias-hardened lensing estimator takes the form

$$
\hat{\kappa}_L^{\mathrm{BH}} = \frac{\hat{\kappa}_L - w_L^T A_L^{-1} \hat{s}_L}{1 - w_L^T A_L^{-1} v_L},
\tag{C4}
$$

where $\hat{s}_L \equiv (\hat{s}_{1,L}, \hat{s}_{2,L}, \cdots, \hat{s}_{n,L})^T$. The variance of the bias-hardened lensing estimator is simply

$$
\langle \hat{\kappa}_L^{\mathrm{BH}} \hat{\kappa}_{-L}^{\mathrm{BH}} \rangle = (1 - w_L^T A_L^{-1} v_L)^{-2} \Bigg[\langle \hat{\kappa}_L \hat{\kappa}_{-L} \rangle + \sum_{ij} \sum_{ab} (w_L)_i (A_L^{-1})_{ij} (w_L)_a (A_L^{-1})_{ab} \langle \hat{s}_{j,-L} \hat{s}_{b,L} \rangle
$$
$$
- 2 \sum_{ij} (w_L)_i (A_L^{-1})_{ij} \mathrm{Re}\langle \hat{\kappa}_{-L} \hat{s}_{j,L} \rangle \Bigg],
\tag{C5}
$$

from which we get the noise

$$
N_L^{\kappa^{\mathrm{BH}}} = (1 - w_L^T A_L^{-1} v_L)^{-2} \Bigg[N_L^{\kappa} + \sum_{ij} (w_L)_i (A_L^{-1})_{ij} N_L^{s_j} \sum_{ab} (w_L)_a (A_L^{-1})_{ab} N_L^{s_b} \mathcal{R}_L^{s_b,s_j}
$$
$$
- 2 N_L^{\kappa} \sum_{ij} (w_L)_i (A_L^{-1})_{ij} N_L^{s_j} \mathcal{R}_L^{\kappa,s_j} \Bigg].
\tag{C6}
$$

In the past two equations we've assumed that the noises and responses are even functions of L, which is true so long that the sources are modeled as halos with profiles u_L. Let's focus on the second term of the RHS of the equation above. Note that $N_L^{s_b}\mathcal{R}_L^{s_b,s_j}$ is the b'th component of the j'th column vector of A_L. Therefore

$$
\sum_b (A_L^{-1})_{ab} N_L^{s_b} \mathcal{R}_L^{s_b,s_j} = \delta_{aj}^K.
\tag{C7}
$$

From this we find

$$
\sum_{ab} (w_L)_a (A_L^{-1})_{ab} N_L^{s_b} \mathcal{R}_L^{s_b,s_j} = \sum_a (w_L)_a \sum_b (A_L^{-1})_{ab} N_L^{s_b} \mathcal{R}_L^{s_b,s_j} = \sum_a (w_L)_a \delta_{aj}^K = (w_L)_j.
\tag{C8}
$$

Recall that the j'th component of w_L is just $N_L^{\kappa}\mathcal{R}_L^{\kappa,s_j}$. Plugging this result back into our expression for the noise gives

$$
N_L^{\kappa^{\mathrm{BH}}} = (1 - w_L^T A_L^{-1} v_L)^{-2} \Bigg[1 - \sum_{ij} (w_L)_i (A_L^{-1})_{ij} \underbrace{N_L^{s_j} \mathcal{R}_L^{\kappa,s_j}}_{=(v_L)_j} \Bigg] N_L^{\kappa}
$$
$$
= (1 - w_L^T A_L^{-1} v_L)^{-1} N_L^{\kappa}
$$
$$
= \frac{\det(A_L)}{\det(M_L)} N_L^{\kappa}.
\tag{C9}
$$

Appendix D: tSZ-like profile

If the tSZ foreground map is modeled as $s_{\boldsymbol{\ell}} = \sum_i s_i u_{\boldsymbol{\ell}}$, then the tSZ power spectrum is proportional to $|u_{\boldsymbol{\ell}}|^2$. Recall that the bias-hardened estimator is insensitive to the normalization of the profile. Therefore we take the square root of the tSZ power spectrum (measured from the simulations) as our tSZ-like profile, which is plotted in Fig. 8. In practice, determining the exact profile from the measured tSZ power spectrum may be challenging. However, we have checked that the residual biases from tSZ and other foregrounds vary slowly when the assumed profile is changed, such that a precise knowledge of the profile is not needed. This is encouraging, and suggests that this method will be useful in practice.
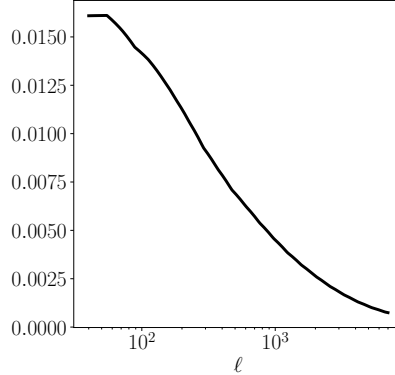

Figure 8.   tSZ-like profile $u_{\boldsymbol{\ell}}$. For $\ell \lesssim 60$ we approximate the profile as a flat curve.

Appendix E: Noise price of bias-hardening increases with profile size

We found that the noise price for hardening against a tSZ-like profile is larger than hardening against point sources. To gain an intuition for why this is the case, we plot the noise when hardening against a single Gaussian profile $u_{\boldsymbol{\ell}} = e^{-\sigma^2 \ell^2/2}$ for different values of σ in Fig. 9. We find that the noise increases with σ; that is, the larger the profile the larger the cost in noise. As shown in Fig. 1, in the regime where the response is small the noise cost goes as $\mathcal{R}_L^2$. The response $\mathcal{R}_L \propto \int d^2\ell f^s_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}} f^\kappa_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}/C^{\text{tot}}_\ell C^{\text{tot}}_{|\boldsymbol{L}-\boldsymbol{\ell}|}$ is larger when the linear response to sources $f^s_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}} = e^{\sigma^2 L^2/2} e^{-\sigma^2 \ell^2/2} e^{-\sigma^2 |\boldsymbol{L}-\boldsymbol{\ell}|^2/2}$ looks more like $f^\kappa_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}$. For point sources $f^s_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}$ is flat, however, as the size of the profile gets larger $f^s_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}}$ looks slightly more similar to the linear response to lensing, making it more difficult for the source-hardened estimator to remove the sources, and results in a higher noise price.
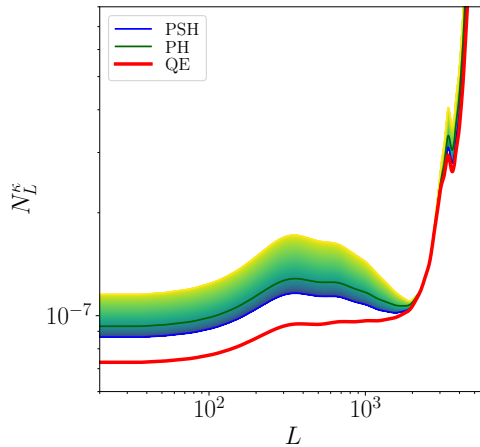

Figure 9.   **Shaded (blue to yellow) region:** Noise when bias hardening against a single Gaussian profile $f^s_{\boldsymbol{\ell},\boldsymbol{L}-\boldsymbol{\ell}} = e^{\sigma^2 L^2/2} e^{-\sigma^2 \ell^2/2} e^{-\sigma^2 |\boldsymbol{L}-\boldsymbol{\ell}|^2/2}$ for $\sigma \in [0', 2.4']$. σ is monotonically increasing from blue to yellow. Note that $\sigma = 0'$ corresponds to hardening against point sources (PSH). **Red:** Noise of the standard QE.

Appendix F: Noise and cross-correlation checks

As a simple pipeline check we verify our calculation of the noise and response when bias-hardening against point sources (PSH). These checks are shown in the plots below.
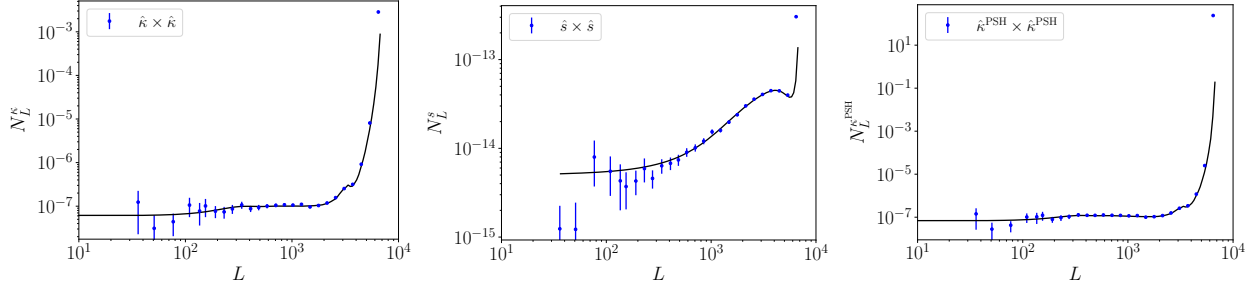

Figure 10. Verification of the noises of each estimator. The auto-correlations (blue) of the estimators $\hat{\kappa}, \hat{s}$, and $\hat{\kappa}^{\mathrm{PSH}}$ when run on a Gaussian random field with power spectrum C_L^{tot}. In black are the theory curves.
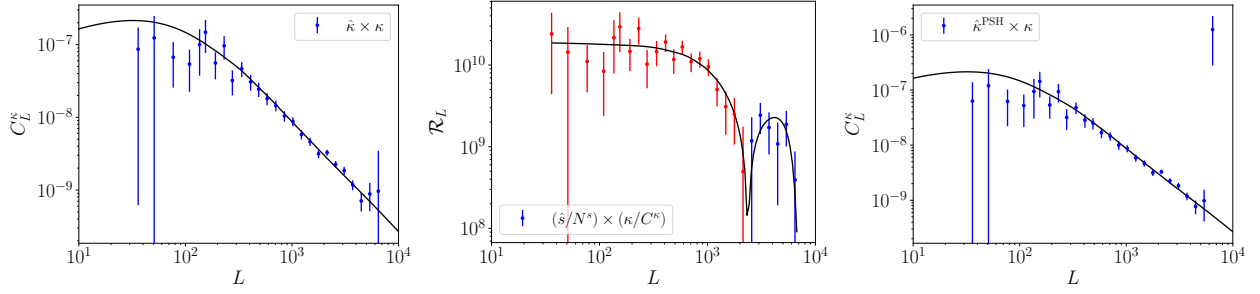

Figure 11. Cross-correlating the true convergence κ with $\hat{\kappa}$, $\hat{s}$, and $\hat{\kappa}^{\mathrm{BH}}$ when run on a lensed CMB map. We see that both the standard QE $\hat{\kappa}$ and the point source hardened estimator $\hat{\kappa}^{\mathrm{PSH}}$ recover the lensing signal C_L^{κ} when correlated with the true convergence. The cross correlation of the point source estimator $\hat{s}$ with the true lensing convergence gives $N_L^s C_L^{\kappa} \mathcal{R}_L$, which is verified in the middle plot. The blue points are positive, whereas the red points are negative. The response has a zero crossing at $L \sim 2500$.
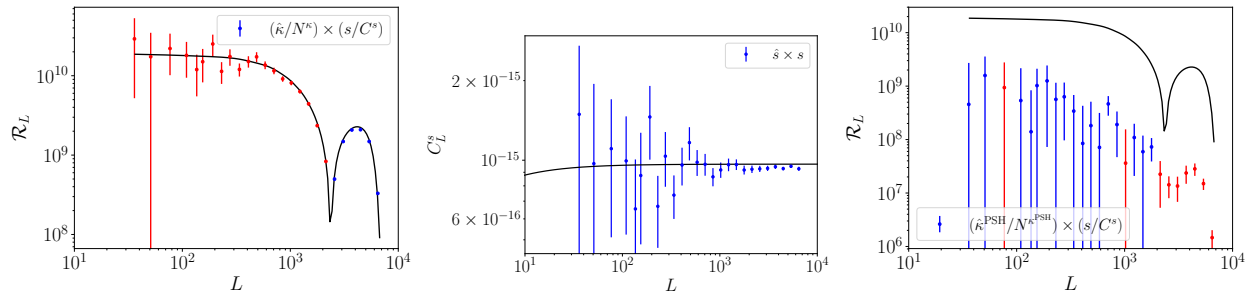

Figure 12. Cross-correlating the true source map s with $\hat{\kappa}$, $\hat{s}$, and $\hat{\kappa}^{\mathrm{PSH}}$ when run on the true point source map s. The cross-correlation of $\hat{\kappa}$ with s gives $N_L^{\kappa} C_L^s \mathcal{R}_L$, which is verified in the plot on the left. $\hat{s}$ recovers the source signal C_L^s when cross correlated with the true source map. Since the point-source hardened estimator is designed to have zero response to point sources, the cross-correlation of $\hat{\kappa}^{\mathrm{PSH}}$ and s should be zero, which is consistent with the result found in the rightmost plot.

[1] A. Lewis and A. Challinor, Phys. Rep. **429**, 1 (2006).

[2] D. Hanson, A. Challinor, and A. Lewis, General Relativity and Gravitation **42**, 2197 (2010).

[3] S. W. Henderson *et al.*, Journal of Low Temperature Physics **184**, 772 (2016).

[4] B. A. Benson *et al.*, **9153**, 91531P (2014).

[5] P. Ade *et al.*, J. Cosm. Astrop. Phys. **2019**, 056 (2019).

[6] K. Abazajian *et al.*, arXiv e-prints arXiv:1907.04473 (2019).

[7] N. Sehgal *et al.*, CMB-HD: An Ultra-Deep, High-Resolution Millimeter-Wave Survey Over Half the Sky, 2019.

[8] N. Sehgal *et al.*, CMB-HD: Astro2020 RFI Response, 2020.

[9] A. van Engelen *et al.*, Astrophys. J.**786**, 13 (2014).

[10] S. J. Osborne, D. Hanson, and O. Doré, J. Cosm. Astrop. Phys. **2014**, 024 (2014).

[11] M. S. Madhavacheril and J. C. Hill, Phys. Rev. D**98**, 023534 (2018).

[12] S. Ferraro and J. C. Hill, Phys. Rev. D**97**, 023512 (2018).

[13] E. Schaan and S. Ferraro, Phys. Rev. Lett.**122**, 181301 (2019).

[14] E. J. Baxter *et al.*, Phys. Rev. D**99**, 023508 (2019).

[15] O. Darwish *et al.*, arXiv e-prints arXiv:2004.01139 (2020).

[16] T. Namikawa, D. Hanson, and R. Takahashi, Mon. Not. R. Astron. Soc. **431**, 609 (2013).

[17] P. A. R. Ade *et al.*, Astronomy & Astrophysics **571**, A17 (2014).

[18] W. Hu and T. Okamoto, Astrophys. J.**574**, 566 (2002).

[19] G. Fabbian, A. Lewis, and D. Beck, Journal of Cosmology and Astroparticle Physics **2019**, 057–057 (2019).

[20] N. Sehgal *et al.*, Astrophys. J.**709**, 920 (2010).

[21] S. Foreman, P. D. Meerburg, A. van Engelen, and J. Meyers, J. Cosm. Astrop. Phys. **2018**, 046 (2018).

[22] J. Dunkley *et al.*, J. Cosm. Astrop. Phys. **2013**, 025 (2013).

[23] LSST Science Collaboration *et al.*, arXiv e-prints arXiv:0912.0201 (2009).