The Astrophysical Journal, 922:259 (21pp), 2021 December 1 https://doi.org/10.3847/1538-4357/ac02bb

Optimal Cosmic Microwave Background Lensing Reconstruction and Parameter Estimation with SPTpol Data

M. Millea[1], C. M. Daley[2], T-L. Chou[3,4], E. Anderes[5], P. A. R. Ade[6], A. J. Anderson[7], J. E. Austermann[8,9], J. S. Avva[1], J. A. Beall[8], A. N. Bender[3,10], B. A. Benson[3,7,11], F. Bianchini[12], L. E. Bleem[3,10], J. E. Carlstrom[3,4,10,11,13], C. L. Chang[3,10,11], P. Chaubal[12], H. C. Chiang[14,15], R. Citron[16], C. Corbett Moran[17], T. M. Crawford[3,11], A. T. Crites[3,11,18,19], T. de Haan[1,20], M. A. Dobbs[14,21], W. Everett[22], J. Gallicchio[3,23], E. M. George[1,24], N. Goeckner-Wald[25], S. Guns[1], N. Gupta[12], N. W. Halverson[9,22], J. W. Henning[3,10], G. C. Hilton[8], G. P. Holder[2,21,26], W. L. Holzapfel[1], J. D. Hrubes[16], N. Huang[1], J. Hubmayr[8], K. D. Irwin[25,27], L. Knox[28], A. T. Lee[1,20], D. Li[8,27], A. Lowitz[11], J. J. McMahon[3,4,11], S. S. Meyer[3,4,11,13], L. M. Mocanu[3,11,29], J. Montgomery[14], T. Natoli[3,11], J. P. Nibarger[8], G. Noble[14], V. Novosad[30], Y. Omori[25], S. Padin[3,11,31], S. Patil[12], C. Pryke[32], C. L. Reichardt[12], J. E. Ruhl[33], B. R. Saliwanchik[33,34], K. K. Schaffer[3,13,35], C. Sievers[16], G. Smecher[14,36], A. A. Stark[37], B. Thorne[28], C. Tucker[6], T. Veach[38], J. D. Vieira[2,26], G. Wang[10], N. Whitehorn[39], W. L. K. Wu[3,27], and V. Yefremenko[10]

[1] Department of Physics, University of California, Berkeley, CA 94720, USA
[2] Astronomy Department, University of Illinois at Urbana-Champaign, 1002 W. Green Street, Urbana, IL 61801, USA
[3] Kavli Institute for Cosmological Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[4] Department of Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[5] Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA
[6] Cardiff University, Cardiff CF10 3XQ, UK
[7] Fermi National Accelerator Laboratory, MS209, P.O. Box 500, Batavia, IL 60510, USA
[8] NIST Quantum Devices Group, 325 Broadway Mailcode 817.03, Boulder, CO 80305, USA
[9] Department of Physics, University of Colorado, Boulder, CO 80309, USA
[10] High Energy Physics Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439, USA
[11] Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[12] School of Physics, University of Melbourne, Parkville, VIC 3010, Australia
[13] Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[14] Department of Physics, McGill University, 3600 Rue University, Montreal, Quebec H3A 2T8, Canada
[15] School of Mathematics, Statistics & Computer Science, University of KwaZulu-Natal, Durban, South Africa
[16] University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA
[17] Jet Propulsion Laboratory, Pasadena, CA 91109, USA
[18] Dunlap Institute for Astronomy & Astrophysics, University of Toronto, 50 St George Street, Toronto, ON, M5S 3H4, Canada
[19] Department of Astronomy & Astrophysics, University of Toronto, 50 St George Street, Toronto, ON, M5S 3H4, Canada
[20] Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
[21] Canadian Institute for Advanced Research, CIFAR Program in Gravity and the Extreme Universe, Toronto, ON, M5G 1Z8, Canada
[22] Department of Astrophysical and Planetary Sciences, University of Colorado, Boulder, CO 80309, USA
[23] Harvey Mudd College, 301 Platt Blvd., Claremont, CA 91711, USA
[24] European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching bei München, Germany
[25] Dept. of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA 94305, USA
[26] Department of Physics, University of Illinois Urbana-Champaign, 1110 W. Green Street, Urbana, IL 61801, USA
[27] SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA
[28] Department of Physics and Astronomy, University of California, One Shields Avenue, Davis, CA 95616, USA
[29] Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, NO-0315 Oslo, Norway
[30] Materials Sciences Division, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL 60439, USA
[31] California Institute of Technology, MS 249-17, 1216 E. California Boulevard, Pasadena, CA 91125, USA
[32] School of Physics and Astronomy, University of Minnesota, 116 Church Street S.E. Minneapolis, MN 55455, USA
[33] Physics Department, Center for Education and Research in Cosmology and Astrophysics, Case Western Reserve University, Cleveland, OH 44106, USA
[34] Department of Physics, Yale University, P.O. Box 208120, New Haven, CT 06520-8120, USA
[35] Liberal Arts Department, School of the Art Institute of Chicago, 112 S Michigan Avenue, Chicago, IL 60603, USA
[36] Three-Speed Logic, Inc., Victoria, B.C., V8S 3Z5, Canada
[37] Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
[38] Space Science and Engineering Division, Southwest Research Institute, San Antonio, TX 78238, USA
[39] Department of Physics and Astronomy, Michigan State University, 567 Wilson Road, East Lansing, MI 48824, USA
Received 2020 December 8; revised 2021 May 6; accepted 2021 May 17; published 2021 December 6

Abstract

We perform the first simultaneous Bayesian parameter inference and optimal reconstruction of the gravitational lensing of the cosmic microwave background (CMB), using 100 deg^2 of polarization observations from the SPTpol receiver on the South Pole Telescope. These data reach noise levels as low as 5.8 μK arcmin in polarization, which are low enough that the typically used quadratic estimator (QE) technique for analyzing CMB lensing is significantly suboptimal. Conversely, the Bayesian procedure extracts all lensing information from the data and is optimal at any noise level. We infer the amplitude of the gravitational lensing potential to be $A_\phi = 0.949 \pm 0.122$ using the Bayesian pipeline, consistent with our QE pipeline result, but with 17% smaller error bars. The Bayesian analysis also provides a simple way to account for systematic uncertainties, performing a similar job as frequentist "bias hardening" or linear bias correction, and reducing the systematic uncertainty on A_ϕ due to polarization calibration from almost half of the statistical error to effectively zero. Finally, we jointly constrain A_ϕ along with A_L, the amplitude of lensing-like effects on the CMB power spectra, demonstrating that the Bayesian method can be used to easily infer parameters both from an optimal lensing reconstruction and from the delensed CMB, while exactly accounting for the correlation between the two. These results demonstrate the feasibility of the Bayesian approach on real data, and pave the way for future analysis of deep CMB polarization measurements with SPT-3G, Simons Observatory, and CMB-S4, where improvements relative to the QE can reach 1.5 times tighter constraints on A_f and seven times lower effective lensing reconstruction noise.

Unified Astronomy Thesaurus concepts: Cosmology (343); Cosmic microwave background radiation (322); Gravitational lensing (670); Weak gravitational lensing (1797); Bayesian statistics (1900)

1. Introduction

Gravitational lensing of the cosmic microwave background (CMB) occurs as CMB photons traveling to us from the last scattering surface are deflected by the gravitational potentials of intervening matter. This effect has been detected with high significance, allowing inference of the line-of-sight projected gravitational field of the intervening matter and of the late-time expansion history and geometry of the universe (Lewis & Challinor 2006; Planck Collaboration et al. 2020a). Better measurements of the lensing effect are one of the main goals of nearly all future CMB probes, and can help constrain dark matter, neutrinos, modified gravity, and a wealth of other cosmological physics (Benson et al. 2014; Abazajian et al. 2016; The Simons Observatory Collaboration et al. 2019).

Traditionally, analysis of lensed CMB data has relied on the so-called quadratic estimate (QE) of the gravitational lensing potential, f (Zaldarriaga & Seljak 1999; Hu & Okamoto 2002). The QE is a frequentist point estimate of f formed from quadratic combinations of the data. It is conceptually simple and near minimum-variance at noise levels up to and including many present-day experiments. However, it was realized by Hirata & Seljak (2003a, 2003b) and Seljak & Hirata (2004) that when instrumental noise levels drop below ~ 5 μK arcmin, where lensing-induced B-modes begin to be resolved with signal-to-noise greater than one, the QE ceases to be minimum-variance, and better analysis can extract more information from the same data. Hirata & Seljak (2003b) were the first to construct a better estimator, using a method based on the Bayesian posterior for CMB lensing. This included a maximum a posteriori (MAP) estimate of f, which has lower variance than the QE,[40] and a maximum-likelihood estimate (MLE) of the power spectrum of gravitational lensing potential, C_ℓ^{ff}. These results used a number of simplifying approximations, including perfectly white noise and periodic flat-sky boundaries with no masking in the pixel domain. Extending this original work, Carron & Lewis (2017) upgraded this MAP f procedure to work without these approximations, rendering it applicable to realistic instrumental conditions.

Although estimates of the f maps are useful, here we are interested in reconstructing not only f but its theory power spectrum as well. A common misconception is that once one has a better estimate of f (e.g., a MAP f estimate), one can take its power spectrum, subtract a noise bias, and obtain the desired estimate of C_ℓ^{ff}. While this does work for the QE, it is only because the QE can be analytically normalized and its power spectrum analytically noise debiased (up to some usually minor Monte Carlo corrections), yielding an unbiased estimate of the theory lensing spectrum. However, this is not generically the case for MAP estimates, for which analytic calculations of normalization and noise biases do not exist. In theory, one could try computing these entirely via Monte Carlo, but this can only be done at a single fiducial cosmological model, and it is unknown to what extent these could be cosmology-dependent or how one might deal with this. If a frequentist estimate is nevertheless desired, a more promising approach may be something akin to the C_ℓ^{ff} MLE proposed by Hirata & Seljak (2003b). However, this has not yet been demonstrated on realistic data.

An alternate approach is based on direct Bayesian inference of cosmological quantities of interest, without the need for explicit normalization and debiasing of any intermediate power spectra. Recent progress was presented in Anderes et al. (2015), who developed a Monte Carlo sampler of the Bayesian posterior of unlensed CMB temperature maps and f maps given fixed cosmological parameters. Millea et al. (2019) began the process of incorporating polarization into this procedure, resulting in a joint MAP estimate of both the f map and the CMB polarization fields. Finally, Millea et al. (2020, hereafter MAW20) extended this to a full Monte Carlo sampler and included cosmological parameters in the sampling, giving the key ingredients needed for the work here. By virtue of directly mapping out the Bayesian posterior for these quantities, this method achieves the goal of fully extracting cosmological information from lensed CMB data and is optimal at all noise levels.

Instrumental noise levels that are low enough at the relevant scales to necessitate anything beyond the QE have only recently been attained. The POLARBEAR collaboration performed the first (and to-date only) beyond-QE analysis of real data (Adachi et al. 2020). This used the Carron & Lewis (2017) MAP f estimate to internally "delense" the data, removing the lensing-induced B-mode polarization. Unlike generic C_ℓ^{ff} estimation, B-mode delensing does not require renormalizing the f estimate, and noise biases can be mitigated via the "overlapping B-mode deprojection" technique.

In this work, we go a step further and perform an optimal lensing reconstruction and full parameter extraction from the lensing potential and from internally delensed bandpowers. Although similar in spirit, our methodology is quite different, however, and it is based on the MAW20 Bayesian sampling procedure rather than on any point estimates. We use the deepest 100 deg^2 of South Pole Telescope polarization data obtained with the SPTpol receiver, restricting ourselves to just this deepest patch since we are mainly interested in the low-noise regime where the Bayesian procedure will outperform the QE. We infer cosmological parameters A_f and A_L, along with a host of systematics parameters. The A_f parameter is a standard parameter scaling the theory lensing spectrum as $C_\ell^{ff} \rightarrow A_f C_\ell^{ff}$. A_f can be considered a proxy for any physical

[40] The MAP f estimate from Hirata & Seljak (2003b) has sometimes been called the "iterative quadratic estimate," but because several methods exist that involve iterating something akin to a quadratic estimate, we do not use this term and instead more precisely refer to individual methods.

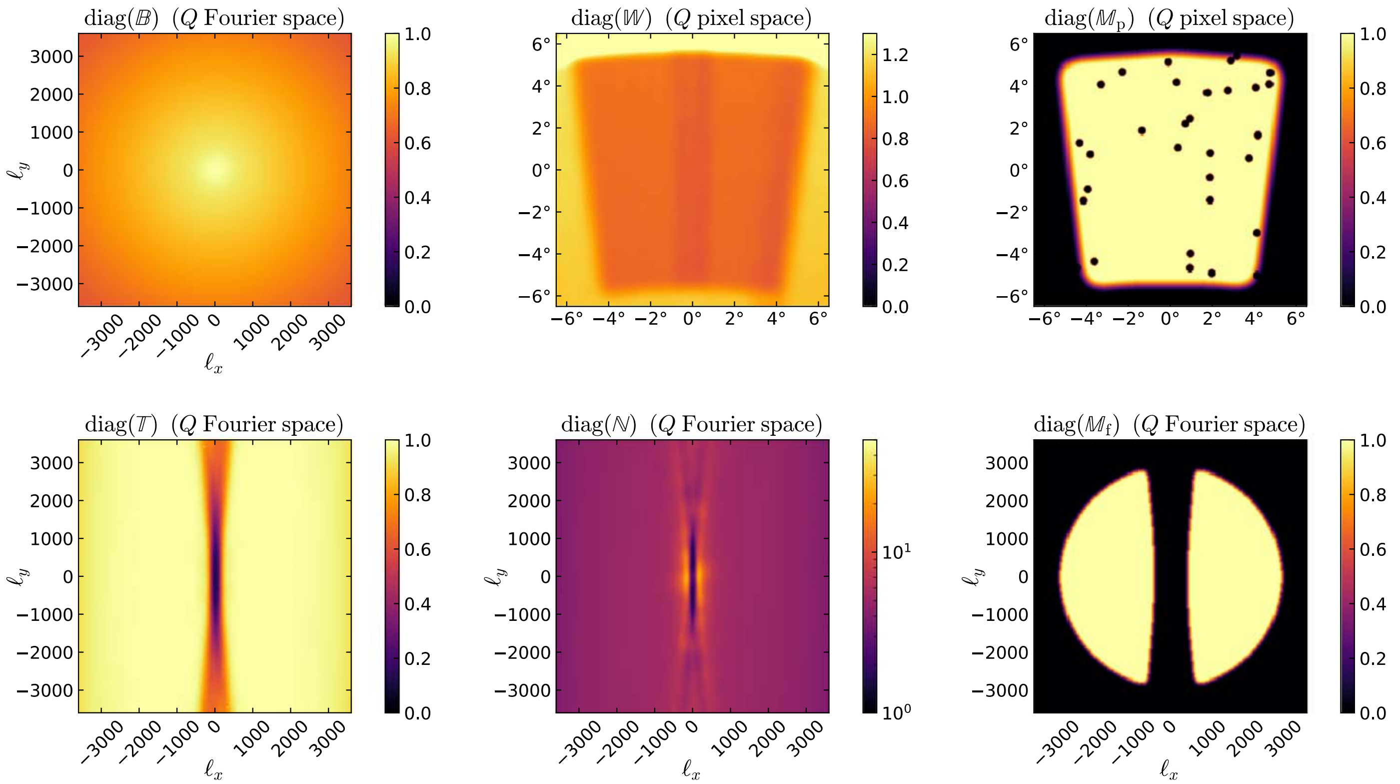

Figure 1. To help orient the reader, a visualization of the various linear operators that enter the CMB lensing posterior in Equation (8) is presented. The operators $\mathbb{B}$ and $\mathbb{T}$ are the beams and transfer functions, respectively; $\mathbb{W}$ and $\mathbb{N}$ together form the noise covariance as $\mathbb{C}_n = \mathbb{W}\mathbb{N}\mathbb{W}^\dagger$, and $\mathbb{M}_\mathrm{p}$ and $\mathbb{M}_\mathrm{f}$ are the pixel-space and Fourier-space masks, respectively (see Section 3 for a full description). These operators correspond to $N_\mathrm{pix} \times N_\mathrm{pix}$ matrices, which act on the N_pix-dimensional vector space of spin-2 (i.e., polarization) 2D maps or 2D Fourier transforms (here $N_\mathrm{pix} = 2 \cdot 260^2$). The quantities plotted above are the Q component of the diagonal of these matrices when represented in the basis labeled in each plot. For $\mathbb{T}$, $\mathbb{M}_\mathrm{p}$, and $\mathbb{M}_\mathrm{f}$, the Q and U components are taken to be identical, while for $\mathbb{B}$, $\mathbb{W}$, and $\mathbb{N}$, they are allowed to be different (but qualitatively end up very similar, and hence only Q is shown).

parameter that is constrained by the lensing potential, such as the matter density or the sum of neutrino masses. We choose to estimate A_f here for simplicity, but in the future, the method could easily be extended to estimate more physical parameters instead. The A_L parameter scales the lensing-like contribution to the model CMB power spectrum, and is defined such that $A_\mathrm{L} = 1$ if the underlying cosmological model is correct. Unlike frequentist estimates, the Bayesian procedure requires a self-consistent data model that includes both A_f and A_L, and we develop one here. Finally, we include several systematics parameters, noting that it is particularly easy to incorporate systematic errors into the Bayesian approach. The final output of this procedure is a Markov Chain Monte Carlo (MCMC) composed of samples of these parameters along with samples of the f maps and unlensed CMB polarization maps, for a total of 202,808 dimensions sampled. Ultimately, we demonstrate a 17% improvement of the Bayesian constraint on A_f as compared to the QE.

The results here are new in three regards:

1. The first time a parameter (A_f) is estimated from an optimal lensing reconstruction.
2. The first joint inference of parameters controlling the lensing potential (A_f) and controlling the CMB band-powers (A_L), while fully and exactly accounting for correlation between the reconstruction and the delensed CMB.
3. The first application of a fully Bayesian method to CMB lensing data.

These demonstrate important pieces of the type of fully optimal beyond-QE analysis, which will be a requirement if next-generation experiments such as SPT-3G, Simons Observatory, and CMB-S4 are to reach their full (and expected) potential (Benson et al. 2014; Abazajian et al. 2016; The Simons Observatory Collaboration et al. 2019).

The organization of the paper is as follows. The reader who wishes to skip the details of the MCMC sampling procedure and simply trust that it yields samples from the exact CMB lensing posterior can jump to the main results in Section 6 and discussion in Section 7. The earlier sections give the technical details of the data modeling and sampling. In Section 2, we describe the data and simulations used in this work. These data have been previously vetted in Story et al. (2015) and Wu et al. (2019b), and we refer the reader to these works for various null tests, here choosing instead to concentrate on the lensing analysis. Most of the focus of this work is on the Bayesian pipeline in particular, and Section 3 lays out the forward model necessary to construct the posterior for CMB lensing given the South Pole Telescope (SPT) data. Section 4 describes the Bayesian and QE lensing pipelines, and Section 5 provides validation of the procedures, including on a suite of realistic simulations of the actual data.

2. Data and Simulations

2.1. Data

In this work, we use data from the 150 GHz detectors from the SPTpol receiver on the SPT (Padin et al. 2008; Carlstrom

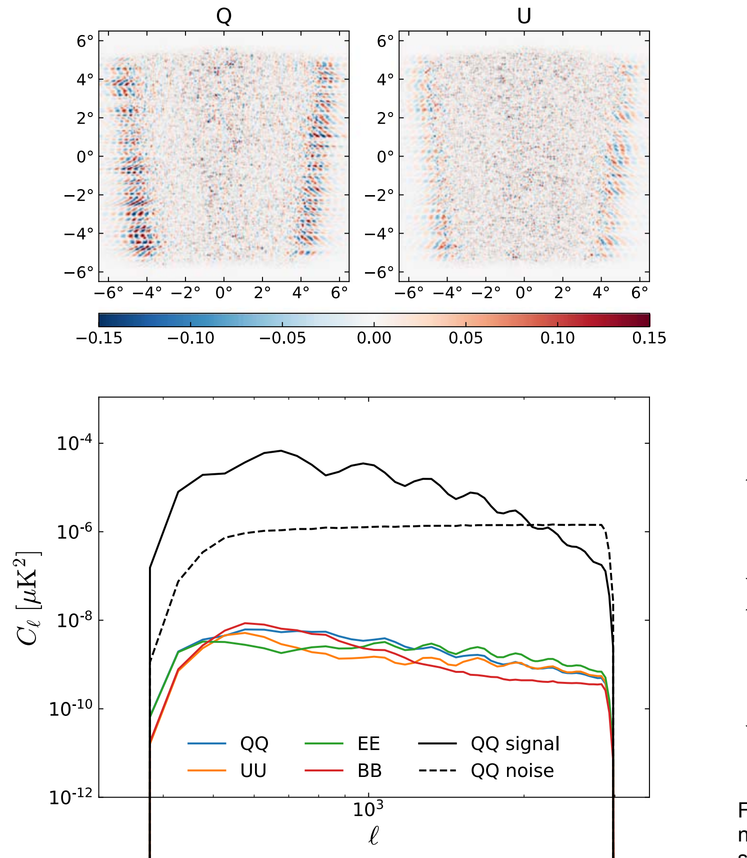

Figure 2. Validation of the approximations underlying our estimate of the transfer function, $\mathcal{T}$ (see Section 3.6). The top plots shows the Q and U components of the difference between (1) a full 100d TOD-level noise-free pipeline simulation and (2) a simple projection of the same realization then multiplication by $\mathcal{T}$. The differences arise from mode coupling induced by the TOD filtering and Monte Carlo error in the transfer function estimation procedure. The bottom plot shows the power spectrum of these difference maps, averaged over several realizations, as well as of the QQ signal and noise for comparison. Differences are one to four orders of magnitude below the noise power spectrum, hence negligible. We note that in both the top and bottom plots, the full Fourier and pixel mask, $\mathcal{M}$, has been applied, so as to pick out the modes that are actually relevant in the analysis.

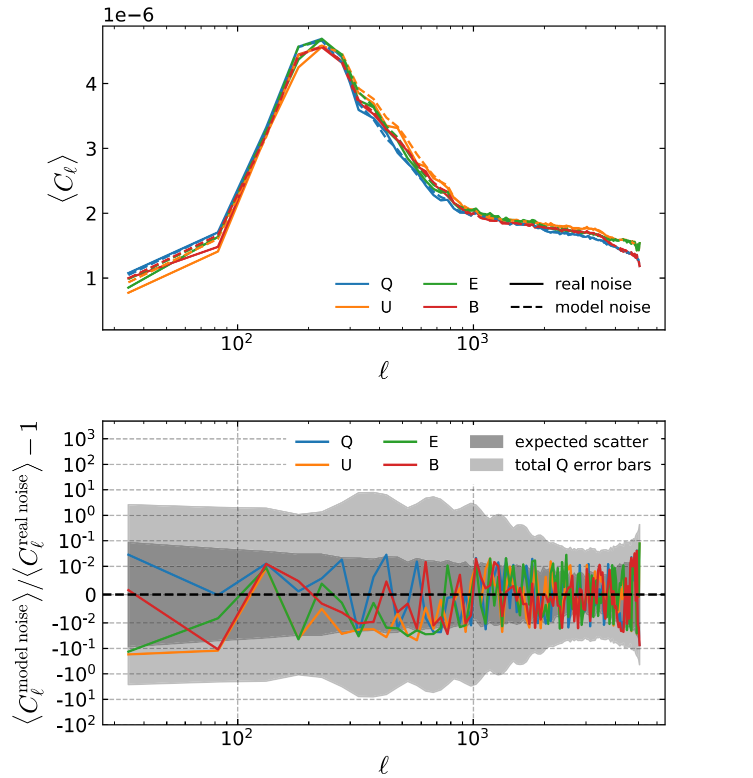

Figure 3. Validation of the approximations underlying our estimate of the noise covariance, $\mathcal{C}_n$ (see Section 3.7). The top panel shows the mean power spectra of 400 real noise realizations and 10^4 model noise realizations that have been masked by $\mathcal{M}$. The bottom is a fractional difference between the two (note the change from linear to log scaling at 10^{-2}). The dark shaded band is the expected scatter due to having only 400 real noise realizations, and the lighter shaded band gives the total CMB + noise error bars in the bins plotted here. The good agreement between the two indicates that our model noise covariance is an accurate representation of the real noise.

et al. 2011; Bleem et al. 2012). SPTpol has employed three different scan strategies for the observations that comprise our final data set.

From 2012 March to 2013 April, SPTpol observed a 100 deg^2 patch of sky ($10° \times 10°$) centered at R.A. 23^{h}30^m and decl. −55°. All observations of this field were made using an azimuthal "lead-trail" scan strategy, where the 100 deg^2 field is split into two equal halves in R.A., a "lead" half-field and a "trail" half-field. The lead half-field is observed first, followed immediately by a trail half-field observation, such that the lead and trail observations occur in the same azimuth-elevation range. Each half-field is observed by scanning the telescope in azimuth right and left across the field and then stepping up in elevation. This lead-trail strategy enables removal of ground pickup. We will refer to these data as the 100d observations.

From 2013 April to 2014 May, SPTpol observed a 500 deg^2 patch of sky, extending from 22^h to 2^h in R.A. and from −65° to −50° in decl. Observations during this time were also made using the "lead-trail" scan strategy, and we will refer to them as the 500d-LT observations.

From 2014 May to 2016 September, while observing the same 500 deg^2 field, SPTpol switched to the "full-field" scan strategy in order to increase sensitivity to larger scales on the sky. In this case, constant-elevation scans are made across the entire range of R.A. of the field. We will refer to these data as the 500d-FULL observations.

Our final data set comprises 6262 100d observations, 858 500d-LT observations, and 3370 500d-FULL observations. Each observation records the time-ordered data (TODs) of each detector, and these TODs are filtered and calibrated before being binned into maps. We highlight that while the lensing reconstruction used in this work is optimal and gauranteed to fully extract the lensing information from the CMB maps, the input maps themselves are not optimal in the same sense. In theory, we could employ a maximum-likelihood mapmaking procedure (for an example, see Aiola et al. 2020); however, because of our fairly simple scan strategy and uniform coverage, this would likely lead to very small improvements in final constraints and is thus not used. The data reduction largely follows previous TE/EE power spectrum analyses, namely Crites et al. (2015) for the 100d observations, and Henning et al. (2018) for the 500d-LT and 500d-FULL observations. Here we only highlight relevant aspects for this analysis.

For the 100d observations, we use slightly different TOD filters compared to previous analysis of these data in Crites et al. (2015). We subtract a fifth-order Legendre polynomial from the TOD of each detector, and then apply a high-pass filter at 0.05 Hz, in order to match the filter choices for

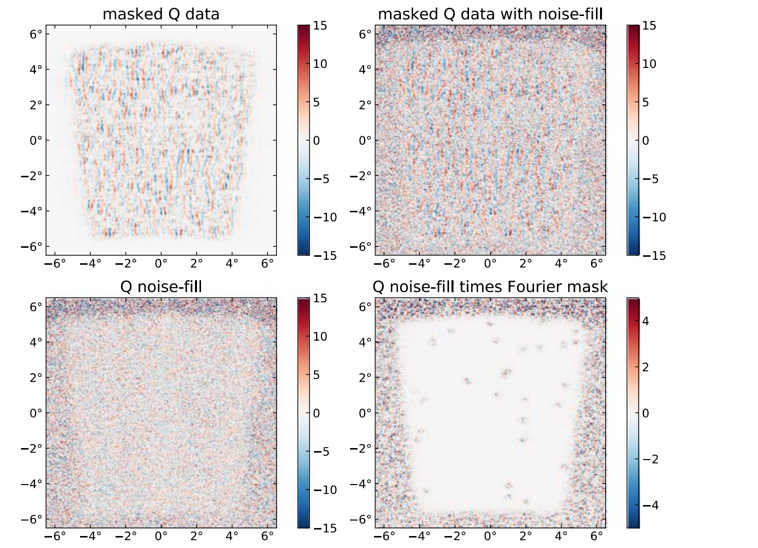

Figure 4. A demonstration of the "noise-fill" procedure described in Section 3.8, which makes it much easier to exactly Wiener filter the data even in the presence of pixel and Fourier-space masking and a noise covariance model that is not diagonal in either space. The top-left panel shows 100D-DEEP data with the mask applied, including Fourier and pixel masks. The top-right panel additionally has the noise-fill, n, added in; this panel is exactly the data, d, which is used in the posterior in Equation (8). The bottom-left panel shows just n, and the bottom-right panel is n multiplied by the Fourier mask. In this last panel, one can see that in the region interior to the mask and in the range of Fourier modes that are not masked by the Fourier mask, no extra noise is added. Here we have plotted just the Q-polarization component; U-polarization behaves qualitatively the same.

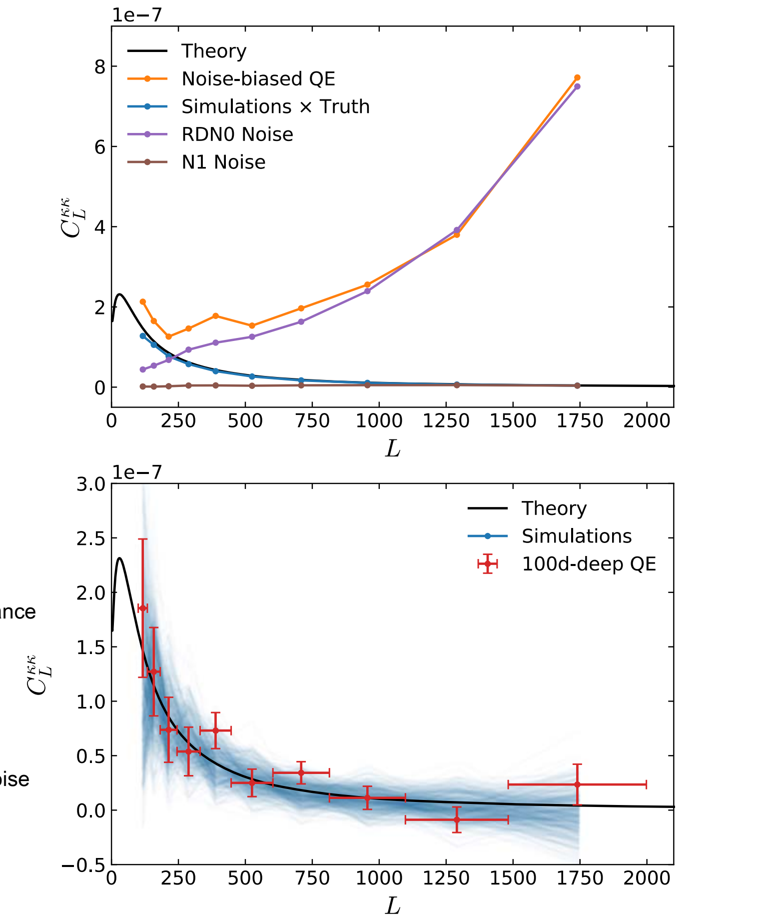

Figure 5. Bandpowers and noise terms from the QE pipeline. The top panel shows the normalized but noise-biased QE power spectrum, along the typical $N_L^{(0),\mathrm{RD}}$ and $N_L^{(1)}$ noise biases that are subtracted. The blue curve is the average cross spectrum between input f maps and $\bar{f}_L^{XY}$ across a suite of simulations, and is used to compute R_L^{MC}. The bottom panel shows the noise-bias-subtracted QE and error bars (from simulations), as well as a cloud of blue lines denoting the noise-debiased simulations used to compute f_{PS}.

500D observations. Based on the size of our map pixels, we apply a low-pass filter at a TOD frequency corresponding to an effective $\ell = 5000$ for anti-aliasing along the scan direction. Electrical cross-talk between detectors could bias our measurement, and in Crites et al. (2015), we applied the cross-talk correction to the power spectra at the end of the analysis. However, in this analysis, we correct cross-talk at the TOD level by measuring a detector-to-detector cross-talk matrix, the same way as described in Henning et al. (2018).

For the 500D-LT observations, we slightly modify the filters as compared to Henning et al. (2018) as well. We subtract a third-order Legendre polynomial from each detector's TOD, and then apply a high-pass filter at $\ell = 100$ to further suppress atmospheric noise. We also apply a low-pass filter at $\ell = 5000$ for anti-aliasing. For the 500D-FULL observations, while using the same high-pass and low-pass filters, we subtract a fifth-order Legendre polynomial instead, due to each scan being twice as long in the scan direction. Electrical cross-talk is corrected as described in Henning et al. (2018).

The TODs of each detector are calibrated relative to one another using an internal thermal source and observations of the Galactic H II region RCW38. The polarization angles of each detector are calibrated by observing an external polarized thermal source, as described in Crites et al. (2015). We bin detector TODs into maps with square 1′ pixels using the oblique Lambert azimuthal equal-area projection, centered at the 100D field center. Because the Bayesian analysis is computationally intensive and scales with the number of pixels, it is advantageous to reduce the number of pixels in the final data map as much as possible. Since our analysis does not use modes above $\ell_{\max} = 3000$, we can, without loss, downgrade the data maps to 3′ arcmin pixels, for which the Nyquist frequency is $\ell_{\mathrm{nyq}} = 3400$. Downgrading is performed by first applying an anti-aliasing isotropic low-pass at ℓ_{nyq}, averaging pixels together, then deconvolving the pixel-window function to match the original 1′ map (the remaining 1′ pixel-window function is accounted for in our forward model for the data). The reason for not making maps directly at 3′ resolution is because the anti-aliasing filter is most easily applied to the intermediate 1′ maps, rather than at the TOD level.

Because we are interested in a low-noise data set where the improvement over the QE is most evident, we only run the analysis on data within the 100D footprint, and only on polarization data. The final data product is a set of co-added 260×260 pixel Q and U maps. The effective noise level of the 100D-DEEP data set inside the mask used in the analysis is 6.0 μK arcmin in polarization over the multipole range $1000 < \ell < 3000$, dipping to 5.8 μK arcmin in the deepest parts of the field.

3. Modeling

To compute the Bayesian posterior for CMB lensing, we require a forward data model and a set of priors. The data, d, which is used as input to the Bayesian pipeline, is a masked and "noise-filled" version of the QU data produced by the

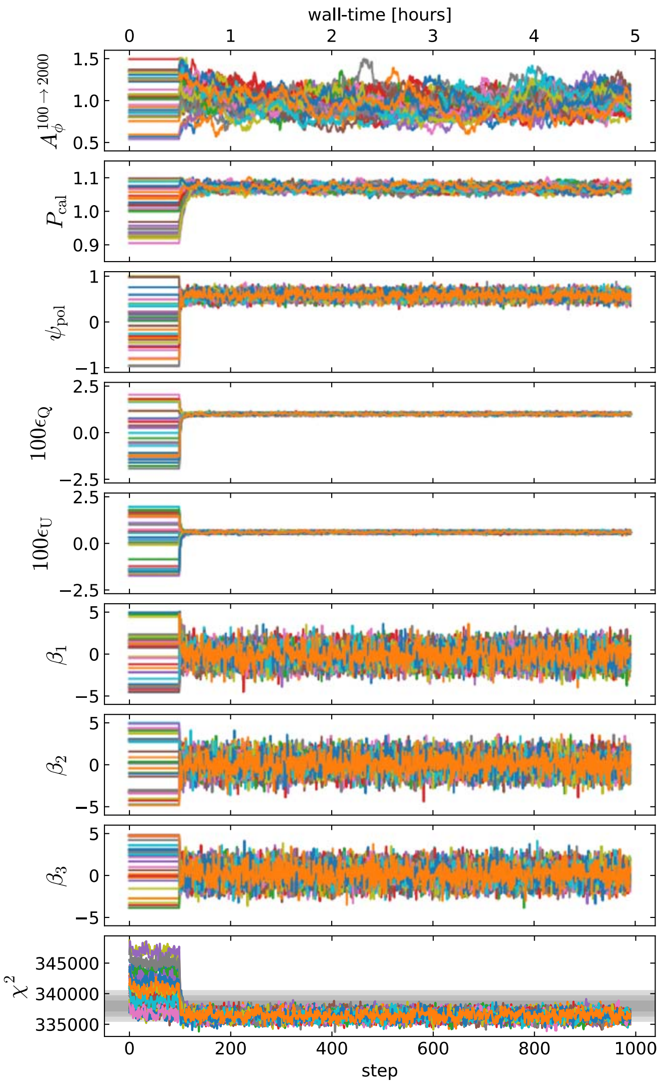

Figure 6. The top eight plots show the trace of the sampled cosmological and systematics parameters, θ, at each step in the Monte Carlo chain. The very bottom plot shows the trace of the χ^2 of the current model point, along with a gray shaded band indicating the expectation based on the number of degrees of freedom. Note that 202,800 other parameters are jointly sampled in this chain (not pictured), corresponding to every pixel or Fourier mode in the CMB polarization and f maps. To aid convergence, the θ are not updated for the first 100 steps in the chain. These 32 independent chains ran across four Tesla V100 GPUs in roughly 5 hr.

mapmaking described in the previous section (we will describe the masking and what we mean by noise-filled later in this section). The model we assume for d (and later demonstrate is sufficiently accurate) is

$$d = \mathbb{M}_f \mathbb{M}_p \mathbb{R}(\psi_{\rm obs})' \, [P_{\rm cal} \mathbb{R}(\psi_{\rm pol}) \mathbb{T} \mathbb{B}(b_i) \mathbb{L}(\phi) f + \epsilon_Q t_Q + \epsilon_U t_U] + n \tag{1}$$

where

1. f represents the unlensed CMB polarization fields,
2. ϕ is the gravitational lensing potential,
3. n is the instrumental and/or atmospheric noise,
4. $\mathbb{L}(\phi)$ is the lensing operation,
5. $\mathbb{B}(b_i)$ is the beam smoothing operation, controlled by a set of beam eigenmode amplitudes, b_i,
6. $\mathbb{T}$ are the transfer functions,
7. $\mathbb{R}(\psi_{\rm pol})$ is a global Q/U rotation by an angle $\psi_{\rm pol}$, representing the absolute instrumental calibration,
8. $\mathbb{R}_{\rm obs}$ is a fixed but spatially dependent Q/U rotation that aligns the flat-sky Q/U basis vectors to the data observation basis, the inverse of the operation sometimes referred to as "polarization flattening,"
9. $P_{\rm cal}$ is the polarization calibration parameter,
10. $t_{Q/U}$ are temperature-to-polarization monopole leakage templates and $\epsilon_{Q/U}$ are their amplitude coefficients,
11. $\mathbb{M}_p$ and $\mathbb{M}_f$ are pixel-space and Fourier-space masking operations, respectively.

We use the notation that lower-case regular letters represent maps, and double-struck upper-case letters represent linear operators on the $N_{\rm pix}$-dimensional abstract vector space spanned by all possible maps. Later in the paper, we also use the notation that Diagonal(x) refers to a diagonal matrix with the vector x along the diagonal, and diag($\mathbb{X}$) returns the vector along the diagonal of the matrix $\mathbb{X}$.

We adopt Gaussian priors on the fields f, ϕ, and n

$$f \sim \mathcal{N}(0, \mathbb{C}_f(A_f)) \tag{2}$$

$$\phi \sim \mathcal{N}(0, \mathbb{C}_\phi(A_\phi)) \tag{3}$$

$$n \sim \mathcal{N}(0, \mathbb{C}_n), \tag{4}$$

where $\mathbb{C}_f(A_f)$, $\mathbb{C}_\phi(A_\phi)$, and $\mathbb{C}_n$ denote the covariance operators for unlensed CMB polarization, the lensing potential, and the experimental noise, respectively. The first two depend on parameters that control the amplitude of the overall power spectra,

$$\mathbb{C}_f(A_f) = A_f \mathbb{C}_f^0 \tag{5}$$

$$\mathbb{C}_\phi(A_\phi) = \mathbb{C}_\phi^0 + (A_\phi - 1)\mathbb{W} \mathbb{C}_\phi^0, \tag{6}$$

where $\mathbb{C}_f^0$ and $\mathbb{C}_\phi^0$ are evaluated at the best-fit Planck cosmology. The lensing amplitude parameter, A_ϕ, is the main cosmological parameter of interest in this work, and scales the amplitude of the fiducial lensing potential within some window, $\mathbb{W}$. The window allows us to estimate the amplitude just within a given multipole range, which here we take to be $\ell = (100, 2000)$ to match previous SPT lensing analyses. This parameter is sometimes denoted as $A_\phi^{100\to2000}$, but throughout this work, unless otherwise stated or included for clarity, we will drop the superscript and simply refer to

$$A_\phi \equiv A_\phi^{100\to2000}. \tag{7}$$

The unlensed CMB amplitude parameter, A_f, functions as a proxy for the Planck absolute calibration, and allows us to marginalize over the uncertainty in this quantity. Incorporating the $A_{\rm L}$ parameter is slightly less straightforward than either A_ϕ or A_f, and this discussion is delayed until Section 6.1. All other cosmological parameters not explicitly sampled are assumed to be perfectly known and fixed to their true value given the fiducial model.

We assume uniform priors on the cosmological and instrumental parameters A_ϕ, A_f, $P_{\rm cal}$, $\psi_{\rm pol}$, ϵ_Q, and ϵ_U, and unit normal priors on the b_i (discussed in Section 3.4).

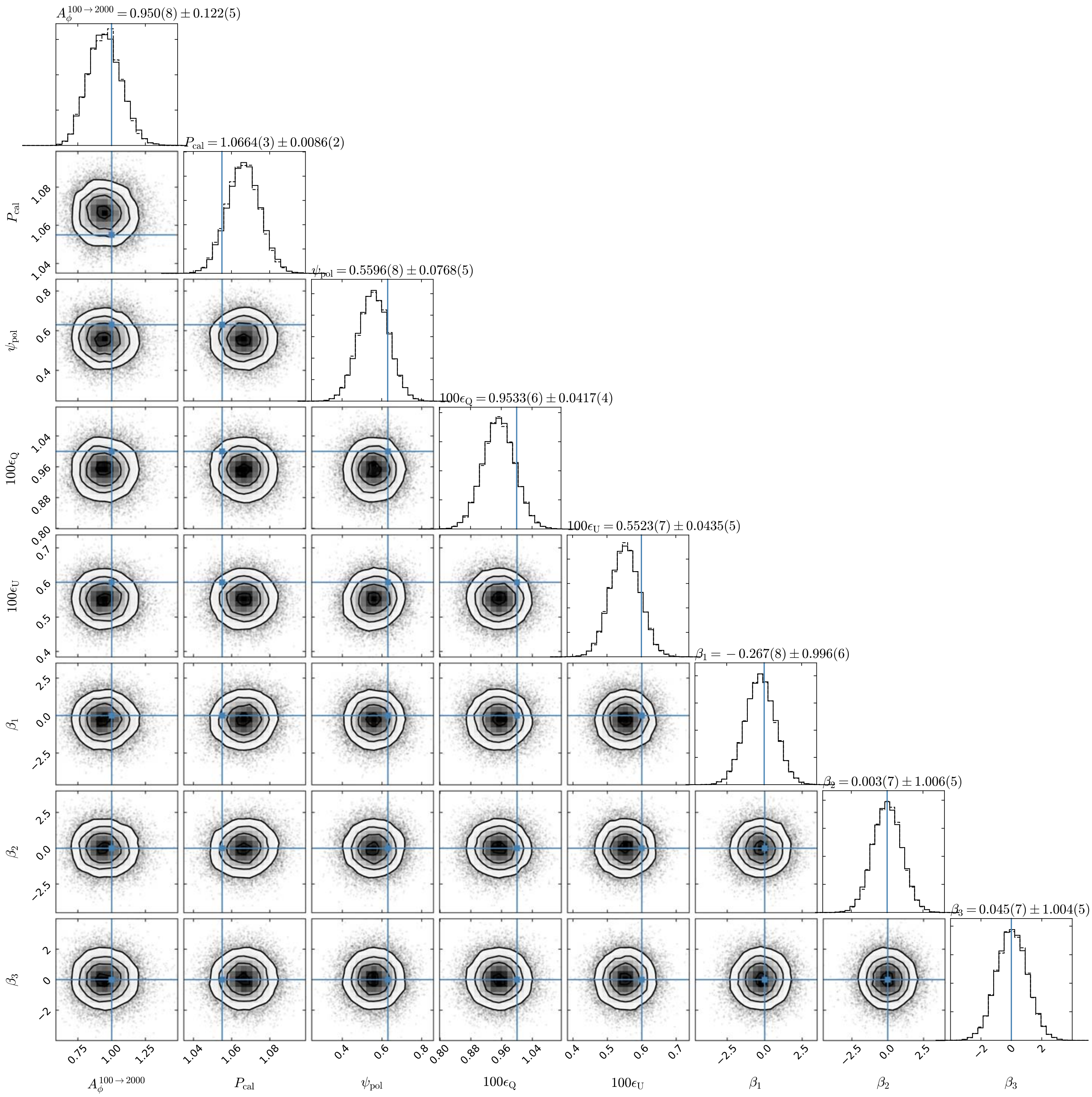

Figure 7. Constraints on sampled parameters, θ, from our baseline chain. The 2D plots show 1σ, 2σ, and 3σ posterior contours as black lines, with binned 2D histograms of the samples shown inside of the 3σ boundary and individual samples shown beyond that. The first column is the main cosmological parameter of interest $A_\phi^{100\to2000}$, and the remaining columns are systematics parameters. The ability to easily and jointly constrain cosmological and systematics parameters in this manner, while implicitly performing optimal lensing reconstruction and delensing, is a unique strength of the Bayesian procedure. Here, we find <5% correlation between $A_\phi^{100\to2000}$ and any systematics, meaning $\sigma(A_\phi^{100\to2000})$ is increased by <2% upon marginalizing over systematic uncertainty. For the systematics parameters, the blue lines denote an estimate from an external procedure, and the agreement in all cases is an important consistency check. The 1D histograms also include the posterior from a separate independent chain as a dashed line, indicating the distributions are sufficiently well converged. More quantitatively, the numbers in parentheses in the titles give an estimate of the standard error on the last digit of the posterior mean and of the posterior standard deviation.

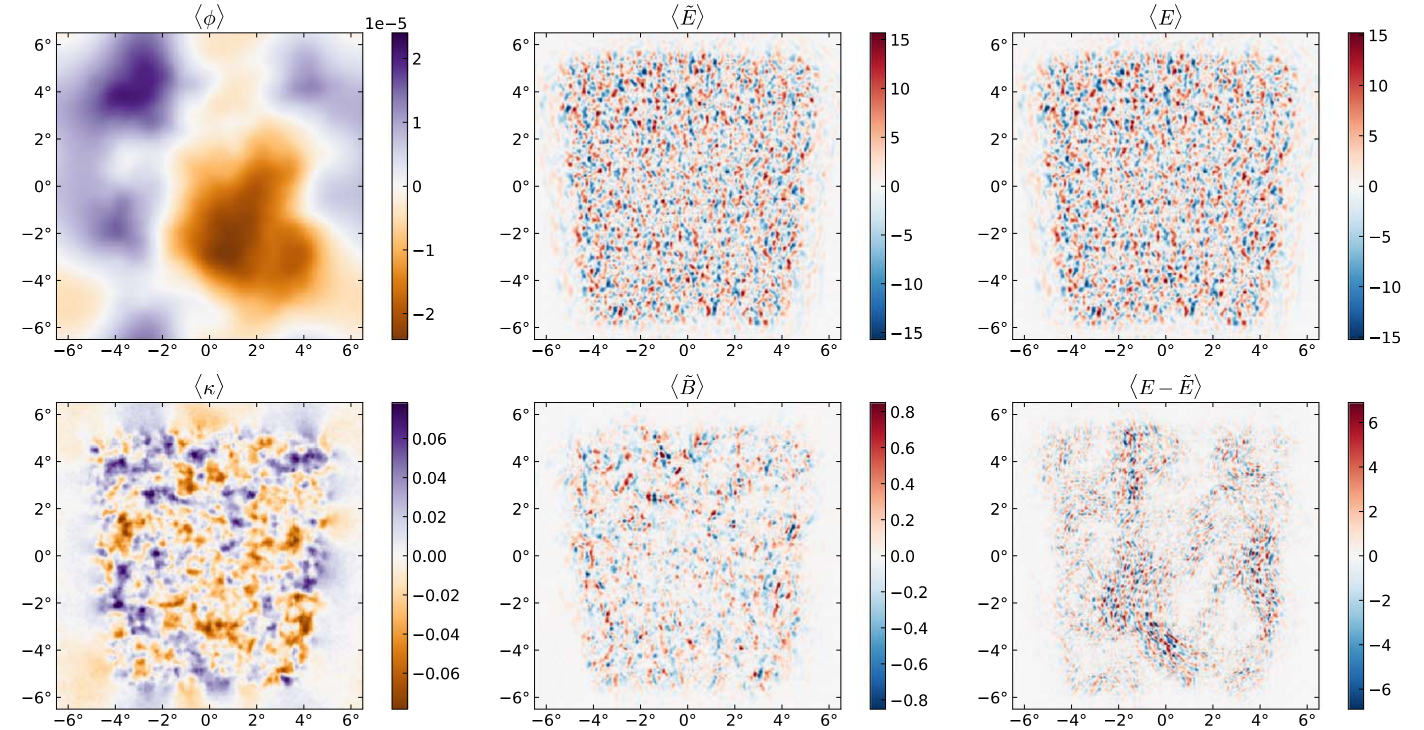

Figure 8. Posterior mean maps, computed by averaging over the Monte Carlo samples in our chains. The quantities ϕ and $\kappa = -\nabla^2\phi/2$ are the lensing potential and convergence maps, and $\tilde{E}$ and $\tilde{B}$ are the lensed E- and B-mode polarization maps. The posterior of any quantity can be computed by post-processing the chain and averaging; for example, the bottom-right panel shows the posterior mean of $(E - \tilde{E})$, i.e., the lensing contribution to the E-mode map. These maps are in some sense only a byproduct of the A_ϕ inference, but if a single point estimate of any of these quantities is required elsewhere, these are the best estimates to use. As expected, these maps qualitatively resemble Wiener filtered data, wherein low signal-to-noise modes are suppressed. The Monte Carlo error in these maps is more quantitat explored in Figure 9.

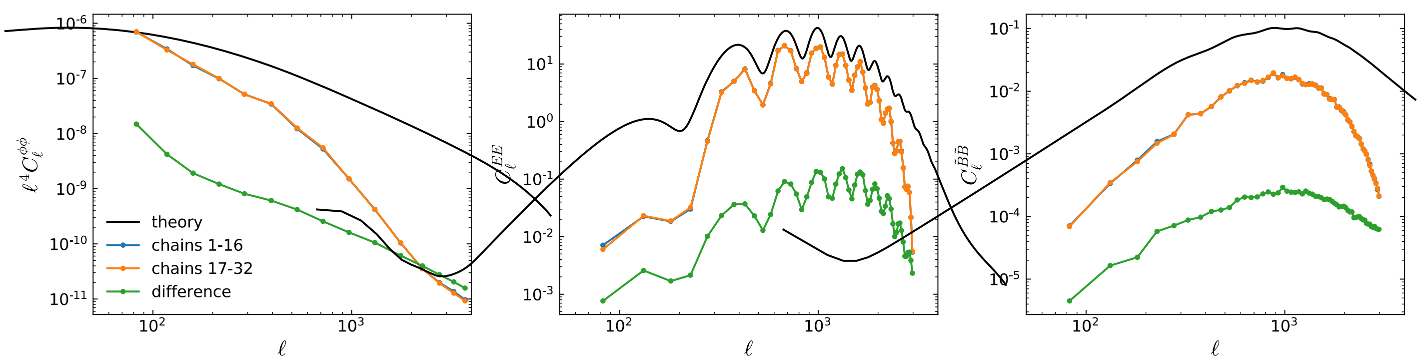

Figure 9. The blue and orange lines (nearly coincident) show the power spectra of (from left to right) posterior mean ϕ, unlensed E, and lensed B maps, as determin from one-half of the 32 independent 100dEEP MCMC chains vs. the other half. The power spectra of posterior mean maps is expected to be suppressed relative to theory, similar to the suppression that arises when Wiener filtering. The green line shows the power spectra of the map differences between these two sets of chai Across almost all scales, these differences are one to two orders of magnitude below the spectra of the mean maps, demonstrating the level of convergence of the chains. The smallest scales in ϕ are the only region where the difference is larger than the mean. An analysis that required better accuracy here could run more cha although we note these scales do not impact the determination of $A_\phi^{100-2000}$.

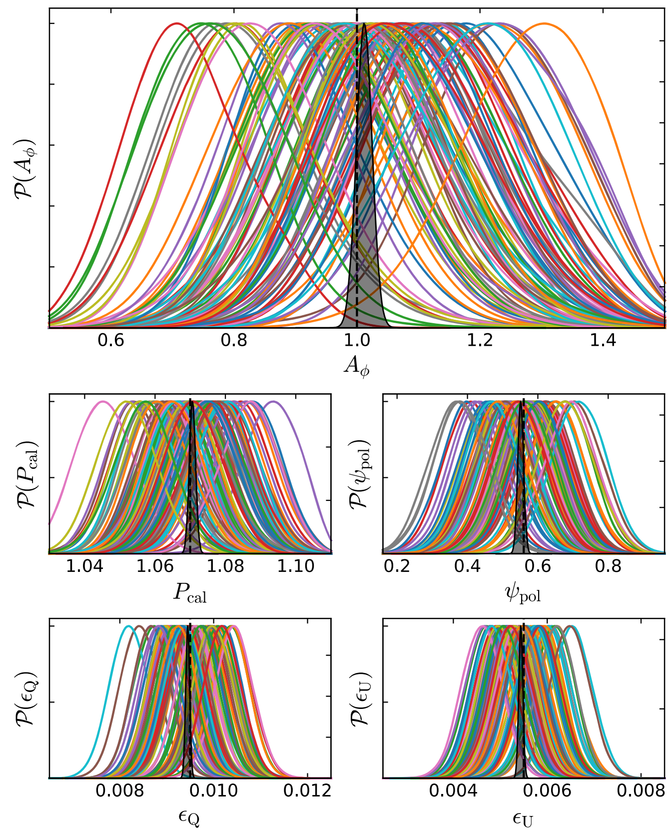

Figure 10. Validation of the Bayesian pipeline on simulations. The colored lines in each panel denote the posterior distributions from each of 100 simulated 100D-DEEP data sets (these include real noise realizations). The shaded black curve is the product of all of these probability distributions. Note that, for clarity, all distributions have been normalized to their maximum value. The true value of the systematics parameters in these simulations comes from the best-fit 100D-DEEP results, and is denoted by the vertical dashed line in each plot. The shaded black curve bounds possible systematic errors in the Bayesian pipeline due to mismodeling of the instrumental noise or pipeline errors, and we find no evidence for either to within the 10% of the statistical error afforded by the 100 simulations.

This set of choices fully specifies the posterior distribution over all variables, given in Equation (8):

$$
\begin{aligned}
\mathcal{P}(f, \phi, A_f, A_\phi, P_{\rm cal}, \psi_{\rm pol}, \epsilon_Q, \epsilon_U, b_i \,|\, d)
&\propto \frac{\exp\left\{-\frac{[d - \mathbb{M}_f \mathbb{M}_p \mathbb{B}_{\rm obs}(P_{\rm cal}\,\mathbb{R}(\psi_{\rm pol}))\,\mathbb{B}\,\mathbb{T}(b_i)\,\mathbb{L}(\phi)\, f + \epsilon_Q t_Q + \epsilon_U t_U]^2}{2\mathbb{C}_n}\right\}}{\det \mathbb{C}_n^{\frac{1}{2}}} \\
&\times \frac{\exp\left\{-\frac{f^2}{2\mathbb{C}_f(A_f)}\right\}}{\det \mathbb{C}_f(A_f)^{\frac{1}{2}}} \\
&\times \frac{\exp\left\{-\frac{\phi^2}{2\mathbb{C}_\phi(A_\phi)}\right\}}{\det \mathbb{C}_\phi(A_\phi)^{\frac{1}{2}}}\, \mathcal{P}(b_i)
\end{aligned}
\tag{8}
$$

where we use the shorthand $x^2/\Sigma \equiv x^\dagger \Sigma^{-1} x$ here and throughout the paper.

Following the terminology of MAW20, we refer to this as the "joint posterior," in contrast to the "marginal posterior," which would analytically marginalize out f.

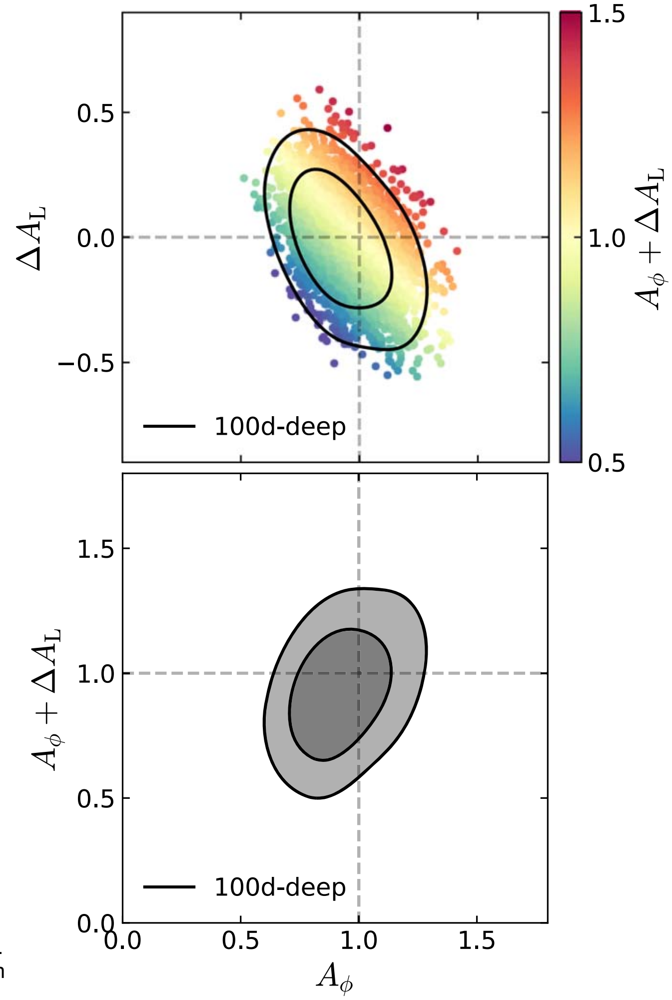

Figure 11. (Top panel) Joint constraints from the 100D-DEEP data set on the amplitude of the lensing potential, $A_\phi^{100\to2000}$, and the residual lensing-like power, ΔA_L. The correlation coefficient between the two is $r = -0.40(5)$, demonstrating only about 9(3)% of the $A_\phi^{100\to2000}$ constraint originates from the power spectrum of the data. (Bottom panel) The same posterior as in the top panel but in terms of the $A_L = A_\phi^{100\to2000} + \Delta A_L$ parameter, which controls the total lensing-like power in the data model. These results demonstrate the unique ability of the Bayesian lensing procedure to infer parameters from an optimal lensing reconstruction and from delensed bandpowers while easily and exactly accounting for correlations between the two.

3.1. Calibration

Performing a change-of-variables from $f \to f/\sqrt{A_f}$ in Equation (8) makes it clear that the posterior constrains only the product $P_{\rm cal}\sqrt{A_f}$. Thus, without loss of generality, we fix $A_f = 1$ in our sampling and only explicitly sample the $P_{\rm cal}$ parameter. The resulting constraints on $P_{\rm cal}$ can be interpreted as a constraint on $P_{\rm cal}\sqrt{A_f}$, or equivalently as a constraint on the SPT polarization calibration when calibrating to a perfectly known theory unlensed CMB spectrum given by the Planck best fit.

An estimate of $P_{\rm cal}$ can be obtained by comparing SPTpol E maps with those made by Planck. For the 500D data, Henning et al. (2018) measured $P_{\rm cal} = 1.06$, and for the 100D data, Crites et al. (2015) measured $P_{\rm cal} = 1.048$. A weighted combination of the two predicts $P_{\rm cal} \sim 1.055$ for the 100D-DEEP data.

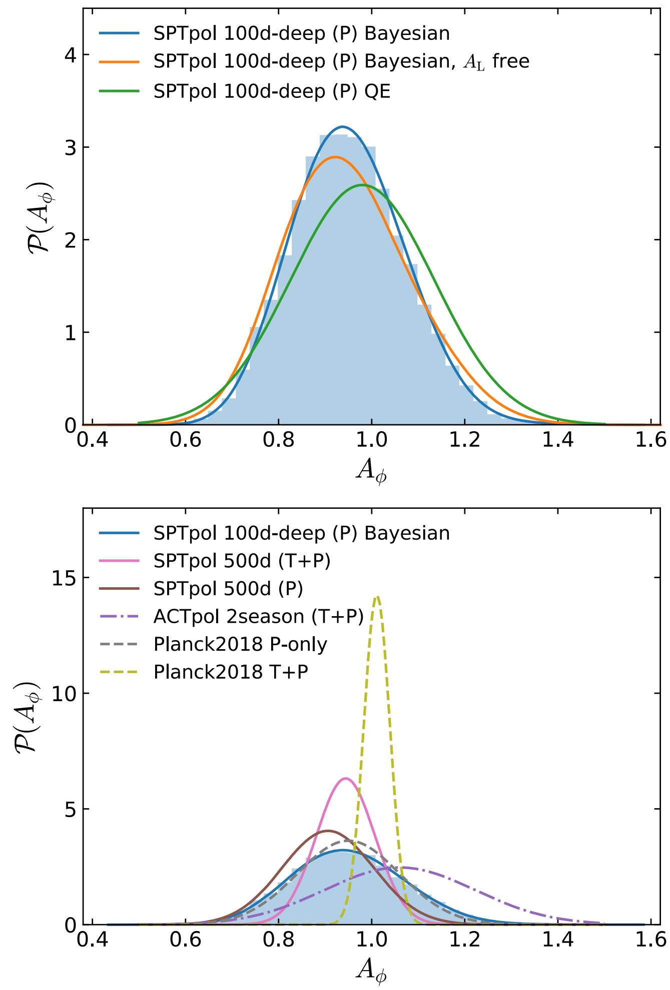

Figure 12. (Top panel) Posterior distribution of $A_\phi^{100\to2000}$ as determined by the Bayesian and QE procedures. The blue bars are a histogram of the samples in the chain from the Bayesian procedure, and the solid blue line is the Blackwell–Rao posterior. The orange curve removes information from the power spectrum of the data by marginalizing over $A_{\rm L}$, and the green curve is the Gaussian estimate from fitting the QE bandpowers. The 17% improvement in error bar in the $A_{\rm L}$-marginalized Bayesian case over the QE is a main result of this work. (Bottom panel) Comparison of the Bayesian result with other measurements of A_ϕ in the literature. The result here achieves the lowest-yet effective noise level on ϕ, although other results achieve better A_ϕ constraints with a larger observation region.

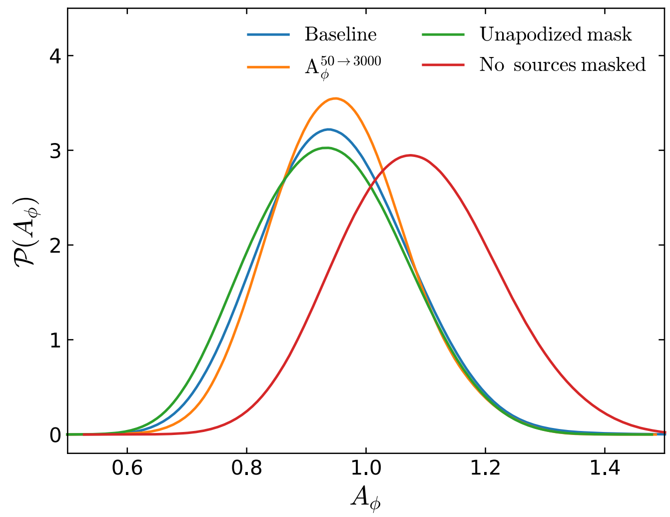

Figure 13. Constraints on A_ϕ given various changes to the analysis as compared to the baseline result, as described in Section 6.4.

This external estimate of $P_{\rm cal}$, however, is not directly used, because we do not correct the raw data by a best-fit $P_{\rm cal}$. Instead, we include $P_{\rm cal}$ in the forward model for the data and sample its value in our MCMC chains. Note that this approach is unique for a lensing analysis, because it means that the calibration is jointly estimated at the same time as other systematics, at the same time as cosmological parameters, and even at the same time as the reconstructed ϕ maps themselves. We will see in Section 6.3 that this has concrete benefits, mainly that it reduces the impact of the uncertainty on $P_{\rm cal}$ on the final cosmological uncertainty. As a consistency check, we will also show that the range of $P_{\rm cal}$ values allowed by the MCMC chain is consistent with $P_{\rm cal} \sim 1.055$.

For the QE pipeline where there is no analogous approach, we do correct the data; however, we correct by the best-fit value from the Bayesian pipeline for easier comparison between the two. All of the systematics parameters described in the following subsections are handled in the same way as $P_{\rm cal}$, by sampling in the Bayesian case and by applying a best-fit correction in the QE case.

3.2. Global Polarization Angle

Assuming negligible foregrounds and a non-parity-violating cosmological model, we expect the cross-spectra between TB and EB to be consistent with zero. A systematic error in the global polarization angle calibration of the instrument, $\psi_{\rm pol}$, can also create a signal in these channels. A typical approach is to determine $\psi_{\rm pol}$ by finding the value that nulls the TB and EB channels (Keating et al. 2012). This was the approach taken in Wu et al. (2019b) for a subset of the same data used here, which found $\psi_{\rm pol} = 0°.63 \pm 0°.04$.

We include the global polarization rotation in the forward data model in the form of the operator $\mathbb{R}(\psi_{\rm pol})$, and jointly infer $\psi_{\rm pol}$ along with the other systematics and cosmological parameters. Because the prior on ϕ assumes no correlation between EB (i.e., $\mathbb{C}_\phi$ is diagonal in EB Fourier space), the MCMC chain will implicitly try to find the $\psi_{\rm pol}$ that nulls the EB channel. As we will see in Section 6.4, the value we find is consistent with the determination from Wu et al. (2019b).

3.3. Temperature-to-polarization Leakage

Because the measured polarization signal effectively comes from differencing the measured intensity along two different polarization axes, any systematic mismatch affecting just one of the axes can leak the CMB temperature signal into polarization. Depending on the nature of the mismatch, different functions of the temperature map can be leaked into Q and U. For example, a gain variation between detectors will leak a copy of the T map directly, whereas pointing errors, errors in the beamwidth, or beam ellipticity will leak higher-order gradients of the T map (Ade et al. 2015). Because the temperature map is measured with very high signal-to-noise, the presence of leakage can be detected by cross correlating temperature and Q or U maps (this correlation should be zero on average for the true CMB, given a Fourier mask with appropriate symmetries). Additionally, if any correlation is detected, it can simply be subtracted given an appropriate amplitude.

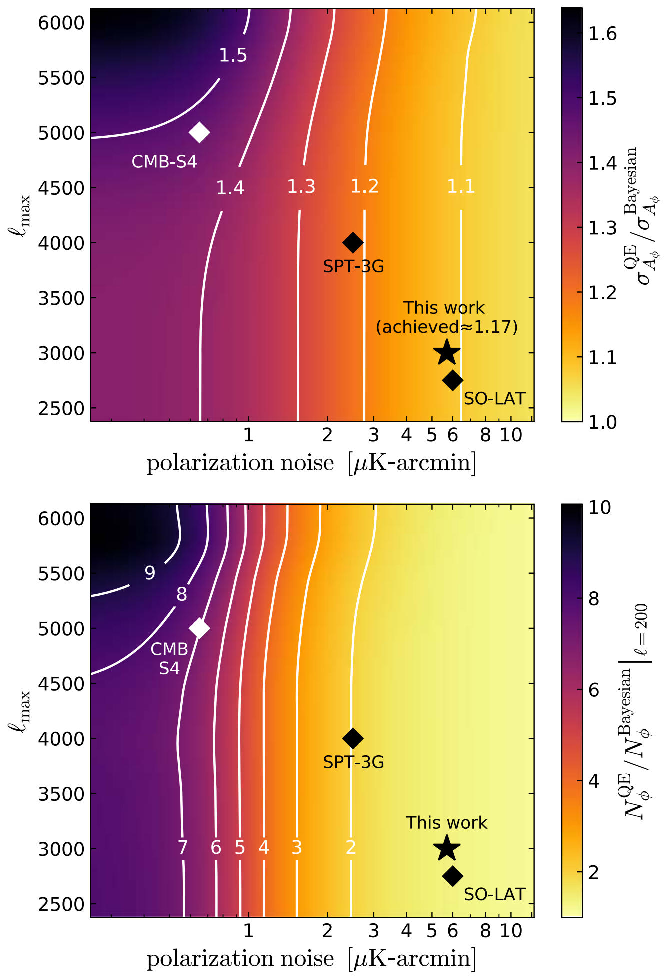

Figure 14. Forecasted improvement of Bayesian lensing reconstruction over the quadratic estimate, computed from a suite of map-level mask-free simulations. The x-axis gives the noise level in polarization, and the y-axis gives the largest ℓ used in the reconstruction. The top panel shows the improvement in the error bar on $A_\phi^{100\rightarrow2000}$. The bottom panel shows the improvement in the effective noise in the lensing reconstruction, $N_\ell^{\phi\phi}$, at $\ell = 200$. This work achieves a slightly better improvement in $A_\phi^{100\rightarrow2000}$ than predicted from these simulations due to minor sub-optimalities present in our (and typical) QE pipelines when masking and other analysis complexities exist. Forecasts for the deep CMB-S4 survey, SPT-3G, and Simons Observatory LATs are shown as diamonds. The latter lies almost directly on top of the star denoting the current work, but is offset only for visual clarity. These simulations cover roughly 100 deg^2, although the relative improvements are not expected to scale appreciably with $f_{\rm sky}$.

For the 100d-DEEP data, cross correlating with the appropriate templates demonstrates that only gain-type leakage exists at appreciable levels in the maps. This type leads to a leakage of the form,

$$\begin{pmatrix} Q \\ U \end{pmatrix} \rightarrow \begin{pmatrix} Q \\ U \end{pmatrix} + \begin{pmatrix} \epsilon_Q T \\ \epsilon_U T \end{pmatrix} \tag{9}$$

where ϵ_Q and ϵ_U are coefficients that capture the total leakage to each channel. Minimizing the TQ and TU cross-correlation yields best-fit values of

$$\epsilon_Q = 0.010 \quad \epsilon_U = 0.006. \tag{10}$$

As for the other systematics, these values are only used as a consistency check, and instead the leakage templates are included in the forward model, and ϵ_Q and ϵ_U are sampled. For convenience, we also define the spin-2 polarization fields, $t_Q \equiv (T, 0)$ and $t_U \equiv (0, T)$, which allow writing the leakage contribution in the form seen in Equation (8). Finally, we note that the coefficients are small enough that no T noise is introduced in the deprojection or marginalization over the leakage templates, thus the T field can be taken as a fixed truth given by the measurement and does not need to be additionally sampled. As we will see in Section 6.4, the values preferred by the chain are in agreement with Equation (10).

3.4. Beams

For the 100d field, the beam window function and error covariance are measured using eight independent observations of Mars. The beam in the field observations is further broadened by pointing jitter, which we estimate by making a second beam measurement using bright point sources in the 100d field, and convolving it with the Mars-derived beam. Full details can be found in Crites et al. (2015). For the 500d field, the beam is measured using seven independent Venus observations, and pointing jitter is convolved in the same way as above. Full details can be found in Henning et al. (2018), where a cross-check is also performed by comparing with Planck beams and maps. The 100d-DEEP beam is computed by averaging over beam-convolved simulations of the 100d and 500d fields, combined given the appropriate weights.

The forward data model includes the beam uncertainty in the form of a beam operator parameterized by free beam eigenmode amplitudes:

$$\mathcal{B}(\beta_i) = \mathcal{B}_0 + \beta_1\mathcal{B}_1 + \beta_2\mathcal{B}_2 + \ldots \tag{11}$$

where $\mathcal{B}_0$ is the best-fit beam, the β_i are beam eigenmode amplitudes, and the $\mathcal{B}_i$ are the perturbations to the beam operator determined from an eigenmode decomposition of the beam covariance matrix. An image of $\mathcal{B}_0$ is shown in the top-left panel of Figure 1. We normalize the $\mathcal{B}_i$ such that the β_i have unit normal priors, which are included in the sampling. We keep three eigenmodes in the chain. As we will see in Section 6.3, none are appreciably constrained beyond their prior, indicating that the data is consistent with the fiducial beam determination.

3.5. Masking

Our analysis applies a pixel mask, $\mathcal{M}_p$, which selects the 100d-DEEP field and masks bright discrete sources. The mask border is built by thresholding the noise pixel variance at five times its minimum value, straightening the resulting edge with a smoothing filter, and finally applying a 1 deg^2 cosine apodization window. The source mask is composed of known galaxy clusters (Vanderlinde et al. 2010), and point sources detected in temperature with fluxes greater than 50 mJy (Everett et al. 2020). In total, the effective sky fraction left unmasked is 99.9 deg^2. This pixel mask is shown in the top-right panel of Figure 1.

We note that neither Bayesian nor QE pipelines require that the mask be apodized. However, while the Bayesian pipeline remains optimal for any mask, hard mask edges can lead to larger Monte Carlo corrections and slight sub-optimalities in the QE pipeline. To facilitate a fairer comparison, we have chosen to use apodization in the baseline case, but also present results with an unapodized mask in Section 6.4.

In the Fourier domain, we apply a Fourier-space mask, $\mathbb{M}_f$, shown in the bottom-right panel of Figure 1. The center part of the mask is built by thresholding the 2D transfer function at 0.9 to remove modes, mainly in the ℓ_x direction, which are significantly affected by the TOD filtering and for which the approximation that $\mathbb{T}$ is diagonal in QU Fourier space breaks down. We additionally apply an $\ell_{\rm max} = 3000$ upper bound to limit the possible contamination from polarized extragalactic point sources. Although there is not much information beyond $\ell = 3000$ at these noise levels, we note that this choice is likely quite conservative and can probably be significantly relaxed in the future.

The total masking operator is chosen as $\mathbb{M} = \mathbb{M}_f \mathbb{M}_p$, i.e., pixel masking happens first. To produce the data that is input to the Bayesian pipeline, d, we apply $\mathbb{M}$ to the raw data map that is output by the mapmaking procedure. We then also self-consistently include $\mathbb{M}$ in the data model itself. Because $\mathbb{M}_f$ and $\mathbb{M}_p$ do not commute exactly, there is some small leakage of masked Fourier modes into d. Our analysis features a fairly conservative $\mathbb{M}_f$, and it is not a problem that the effective Fourier mask leaks slightly into the region that is formally masked by $\mathbb{M}_f$, specifically by around $\Delta\ell \sim 10$ (set by the width of the mask kernel window function). For future analyses where a more precise cut might be desired, one could fully remove any leakage by directly deprojecting the undesired modes from the data and including the deprojection operator in the data model.

3.6. Transfer Functions

The filters applied to the TOD during mapmaking imprint an effective transfer function on the data maps, dependent on the scanning strategy and filtering choices made for each type of observation. We approximate these transfer functions, $\mathbb{T}$, as diagonal in QU Fourier space, and estimate them, as well as validating the approximation, with a set of full pipeline simulations. The full pipeline simulations are fairly computationally costly, and we take two steps to reduce the cost of this step of the analysis: (1) we simplify each simulation by reducing the number of individual observations that are included, and (2) we reduce the total number of simulations needed from ~ 400 to only 20 using a variance canceling technique.

The full pipeline simulations start with a Gaussian realization of the CMB given the best-fit 2015 Planck plikHM_T-T_lowTEB_lensing lensed power spectra (Planck Collaboration et al. 2016). A small expected galactic and extragalactic Gaussian foreground contribution is also added, and then a smoothed version of the SPTpol beam window function is convolved. Note that because the TOD filtering is linear by construction and approximately diagonal in QU Fourier space, it is not crucial that these simulations exactly match the true sky power, nor that they contain the right level of lensing or foreground non-Gaussianity.

From these, we generate mock TOD by virtually scanning the sky using the recorded pointing information from actual observations. For each scan strategy (100D, 500D-LT, and 500D-FULL), we mock-observe the simulated sky into TOD, process TOD into maps, and then co-add these maps in the same way as the real data. The first of the two improvements mentioned above is that we only use a subset of the actual observations (in practice, 20), since many observations have identical scan strategies and would have effectively identical transfer functions. In parallel to these full pipeline simulations, we also perform a simple projection of the beam-convolved CMB+foregrounds to the flat-sky, with no other filtering applied.

We can achieve sufficient accuracy on $\mathbb{T}$ with only 20 simulations by using a new variance canceling technique. This method computes the transfer function as,

$$\mathbb{T} = \mathrm{Diagonal}\left\langle \mathrm{Re}\left[\frac{(\mathbb{M}_p f_{\rm full-pipeline})_{QU,\ell}}{(\mathbb{M}_p f_{\rm projected})_{QU,\ell}}\right]\right\rangle_{20\ {\rm sims}} \quad (12)$$

where the f variables in the numerator and denominator are the mock-observed and projected maps, respectively, and $\mathbb{M}_p$ is the pixel mask. The presence of the projected map in the denominator cancels sample variance in the estimate, leading to much quicker Monte Carlo convergence. However, this comes at the cost that Equation (12) is actually a biased estimate of the true effective transfer function.

With a simple test, we can verify (1) that this bias is small, (2) that our approximation that $\mathbb{T}$ is diagonal in QU Fourier space is sufficient, (3) that there is negligible Monte Carlo error due to using only 20 pipeline simulations, and (4) that our usage of only 20 observations per simulation is valid. For a set of simulations separate than those used to estimate $\mathbb{T}$ and using a different set of 20 observations within each simulation, we compare the result of the full pipeline simulation versus simply applying $\mathbb{T}$ to the projected map for the same realization. In the top panel of Figure 2, we show these difference maps, and in the bottom panel, we show their power spectrum averaged over a few realizations. In both top and bottom panels, we multiply by the full mask, $\mathbb{M}$, so as to pick out only modes relevant for the analysis. We see that the difference is one to four orders of magnitude below the noise spectrum; hence, $\mathbb{T}$ is a very accurate representation of the true transfer function, particularly at smaller scales, which drive the lensing constraint. The final estimate of $\mathbb{T}$ used in the analysis is shown in the bottom-left panel of Figure 1.

We note that the variance canceling technique employed here may be of wider use, but only if full pipeline simulations are not required to quantify uncertainty; otherwise, a larger set of simulations is needed anyway. Here we did not need such a larger set because the Bayesian pipeline does not use simulations to quantify uncertainty at all, and because for the QE pipeline, we have used simulations from the forward data model, as this model is demonstrated sufficiently accurate for our purposes.

3.7. Noise Covariance

The noise covariance is inferred from noise realizations that come directly from the real data using the "sign-flipping" method also used by previous SPT and BICEP analyses (e.g., BICEP2 Collaboration et al. 2014; Wu et al. 2019b). This method works by multiplying a random half of the N = 10,490 observations that enter the final data co-add by −1 before summing them. This cancels the signal but leaves the statistical properties of the noise unchanged, as long as no observation-to-observation correlations exist (which is expected to be the case). This is repeated $M = 400$ times, yielding M nearly independent noise realizations. We will refer to these as real noise realizations and to the distribution from which they are drawn as the real noise.

As we will describe in Section 4.2, the QE pipeline only requires the average 2D power spectrum of the noise as well as an approximate white-noise level. This is sufficient because the noise only enters the QE pipeline for the purposes of Wiener filtering the data, where an approximate Wiener filter is computed, and the impact of this approximation is captured in a Monte Carlo correction applied at the end of the pipeline. This does not lead to any bias, only a small sub-optimality of the final result. The Bayesian pipeline does not apply any Monte Carlo corrections, and thus needs to perform the Wiener filter (which also arises in the Bayesian case) more exactly. This in turn necessitates a full model for a noise covariance operator, $\mathcal{N}_n$, which needs to be as accurate as possible. We will refer to this as the model noise, and samples from this covariance as model noise realizations.

The real SPT noise is non-white, as instrumental and atmospheric 1/f noise dominates at large scales. It is anisotropic, as spatial modes in the scan-parallel and scan-perpendicular directions map onto different temporal modes, and are affected differently by TOD filtering. Finally, it is inhomogeneous, as some spatial regions are observed slightly deeper than others; in particular, the lead-trail scanning strategy used in the 100D and 500D-LT observations causes some regions near the center and right edges of the final 100D-DEEP field to have noise levels a few tens of percent lower than the rest of the field.

With only $M = 400$ real noise realizations, but the most generic $\mathcal{N}_n$ corresponding to an $N_{\rm pix} \times N_{\rm pix}$ matrix where $N_{\rm pix} = 2 \cdot 260^2$, some form of regularization is needed to choose a unique $\mathcal{N}_n$. The choice we make here is motivated by retaining the flexibility to model the complexity of the real noise just described while keeping $\mathcal{N}_n$ fast to invert and to square-root,[41] as both are needed to sample Equation (8). Specifically, we define the model noise covariance, $\mathcal{N}_n$, as

$$\mathcal{N}_n \equiv \square\,\square\,\square\,\square^\dagger \qquad (13)$$

where $\square$ is diagonal in QU pixel space and $\square$ is diagonal in QU Fourier space. That is to say, we model the noise as having an arbitrary non-white anisotropic power spectrum that is spatially modulated in pixel space. With this choice, we have that

$$\mathcal{N}_n^{-1} = \square\,\square^{\dagger\,-1}\,\square^{-1} \qquad (14)$$

$$\sqrt{\mathcal{N}_n} = \sqrt{\square}\,\square^\dagger, \qquad (15)$$

where both operators can be easily applied to vectors with only a few fast Fourier transforms (FFTs). We solve for $\square$ and $\square$ by requiring that the variance in each individual 2D Fourier mode and the variance in each individual pixel be identical for noise realizations drawn from $\mathcal{N}_n$ and for the real noise realizations. These are $2N_{\rm pix}$ constraints for the $2N_{\rm pix}$ combined degrees of freedom in $\square$ and $\square$, yielding the following solution for the diagonal entries of these matrices

$$\square = {\rm Diagonal}({\rm std}(\{n\})_{{\rm QU},x}) \qquad (16)$$

$$\square = {\rm Diagonal}({\rm var}(\{\square\,\square^{-1} n\})_{{\rm QU},l}) \qquad (17)$$

where the standard deviation and variance are taken across the M noise realizations.

We note that the noise realizations used in these averages are the raw sign-flipped combinations of the actual data, with no extra operators deconvolved or masks applied. Hence, the noise term, n, is not multiplied by any extra factors in Equation (1). Additionally, we smooth both $\square$ and $\square$ with small Gaussian kernels, since the uniform scan strategy and large number of observations employed should average away any significant across neighboring pixels or across neighboring Fourier modes.

We plot $\square$ and $\square$ in the middle two panels of Figure 1. The top panel shows the spatially varying pixel variance pattern in $\square$, and the bottom panel shows the non-white anisotropic Fourier noise pattern. To verify that model noise realizations drawn from $\mathcal{N}_n$ are largely indistinguishable from real noise realizations, we show in Figure 3 the mean Q, U, E, and B power spectra of the 400 real noise realizations along with the mean power spectra of 10^4 model noise realizations. We find excellent agreement, the difference between the two completely explained by the scatter expected due to having only 400 real noise realizations (dark shaded band). Additionally, any systematic difference between them is less than 1% of the total Q sample variance error bars (lighter shaded band; note the switch from linear to log scaling at 10^{-2}). As a further check, in Section 5.2 we will use the model noise covariance to analyze simulated data, which includes real noise realizations, finding no evidence for biases to A_f due to any difference between these two.

3.8. The Noise-fill Procedure

The fact that $\mathcal{N}_n$ is not diagonal in either Fourier or map bases presents a challenge for exactly Wiener filtering the data in the presence of a masking operation that is also not diagonal in either space. Whether explicitly stated or not, computing such Wiener filters usually involves approximating the noise as diagonal in one of the two bases, with the impact of the approximation difficult to quantify for a Bayesian analysis. Here, we develop an alternate procedure, which involves artificially adding noise to the data in a particular way so as to make the Wiener filter problem easier to solve, and then demonstrating that the resulting degradation in constraints is negligible. To our knowledge, this has not been described before, and could be of general use.

The challenge can be understood by considering the following toy problem. Suppose we observe some map that is the sum of some signal s and noise n, both defined on the full pixel/Fourier plane, then apply a mask, $\square$, which is a rectangular matrix mapping the full set of pixels/Fourier modes to a smaller subset of just the unmasked ones. The data model is thus given by $d = \square(s + n)$. The residual between data and signal model is $(d - \square s)$, and the covariance of this quantity is $\square\,\square\,\square^\dagger$, where $\square$ is the noise covariance. Defining the signal covariance as $\square$, the log-posterior for this problem is thus

$$\log \square(s \mid d) \propto -\frac{(d - \square s)^2}{2\,\square\,\square\,\square^\dagger} - \frac{s^2}{2\,\square}. \qquad (18)$$

Evaluating the posterior or its gradients with respect to s requires inverting $\square\,\square\,\square^\dagger$. Maximizing the posterior (i.e., Wiener filtering) requires this as well, as the solution is given

[41] We note that for our purposes, the matrix square-root is any $\square$ for which $\mathcal{N}_n = \square\,\square^\dagger$.

by

$$s = [\mathcal{S}^{-1} + \mathcal{M}^{\dagger}(\mathcal{M}\mathcal{N}\mathcal{M}^{\dagger})^{-1}\mathcal{M}]^{-1}\mathcal{M}^{\dagger}(\mathcal{M}\mathcal{N}\mathcal{M}^{\dagger})^{-1}d. \tag{19}$$

However, since $\mathcal{M}$ is not a square matrix, these inverses cannot be simplified away or trivially computed. Sometimes, as a simplifying assumption, $\mathcal{M}$ and $\mathcal{N}$ are taken to be diagonal in the same basis (e.g., $\mathcal{N}$ is assumed to be white noise). In this case, the inverse can be computed explicitly (often in practice by setting the noise to infinity or to a very large floating point number). Since in our case we wish to not make this simplification, we cannot take this approach.

The more general solution we use instead involves artificially filling in the masked data with extra noise, $\bar{n}$, such that the new data model is

$$d' = d + \bar{n} = \mathcal{M}(s + n) + \bar{n}, \tag{20}$$

where we are now considering $\mathcal{M}$ as a square operator but with some rows that are zero. Note that the extra noise does not shift the mean of the data. However, the covariance of the data residual becomes

$$\mathcal{M}\mathcal{N}\mathcal{M}^{\dagger} + \bar{\mathcal{N}}, \tag{21}$$

where $\bar{\mathcal{N}}$ denotes the covariance for $\bar{n}$. Since we are free to choose $\bar{\mathcal{N}}$, we can choose it such that the new data residual covariance is easy to invert, in particular such that it is equal to $\alpha^2\mathcal{N}$ for some constant α (explained below). This happens when

$$\bar{\mathcal{N}} = \alpha^2\mathcal{N} - \mathcal{M}\mathcal{N}\mathcal{M}^{\dagger}. \tag{22}$$

We can draw a realization of $\bar{n}$ from $\mathcal{N}(0, \bar{\mathcal{N}})$ by computing $\bar{\mathcal{N}}^{\frac{1}{2}}\xi$ where ξ is a unit random normal vector. This can in turn be computed by evolving the following ordinary differential equation (ODE) from t = 0 to t = 1,

$$\frac{dy}{dt} = -\frac{1}{2}[\bar{\mathcal{N}}t + (1 - t)\mathbb{1}]^{-1}(\mathbb{1} - \bar{\mathcal{N}})y(t) \tag{23}$$

starting from $y(0) = \xi$ (Allen et al. 2000). The quantity in brackets in Equation (23) can be inverted with the conjugate gradient method. The ordinary differential equation (ODE) itself is requires a stiff solver (we use CVODE_BDF from the Sundials.jl package; Hindmarsh et al. 2005; Rackauckas & Nie 2017).

Ideally, α can be set to unity; however, noise correlations between the masked and unmasked regions may force us to chose some $\alpha > 1$ to ensure the rest of Equation (22) is a positive-definite operator and thus a valid covariance. If this were the case, we would be adding noise to unmasked regions of the data, ultimately degrading the final result. The necessary value of α can be found by direct search, as the ODE will be singular if α is not large enough. One would not use the noise-fill method if α much higher than unity was required (or perhaps one would promote α to some scale-dependent quantity if only certain scales needed a larger value), but here we find $\alpha = 1$ is sufficient to keep Equation (22) numerically positive-definite, confirming that we have not introduced any appreciable degradation of our constraints.

Overall, the ODE solution and α search are not particularly costly, and only need to be done once at the beginning of any analysis. Once d′ is computed, the new posterior is given by the much simpler

$$\log \mathcal{P}(s \,|\, d') \propto -\frac{(d' - \mathcal{M}s)^2}{2\alpha^2\mathcal{N}} - \frac{s^2}{2\mathcal{S}}. \tag{24}$$

Note that, when generating simulated data, it is not necessary to actually perform this procedure. Instead, it is equivalent to simply generate data from a model $d = \mathcal{M}s + \alpha n$, i.e., to leave the noise unmasked and scale it by α. This is very convenient for the simulation pipeline, and it is only on the real data, where one does not have access to s and n separately, that one needs to explicitly perform the noise-fill. An added benefit of this approach is that the likelihood term in the posterior becomes a full $N_{\rm pix}$-dimensional χ^2; thus, its expectation value and scatter are easy to compute. We use this in the later sections to ascertain goodness-of-fit. As a final sanity check, we have verified that using different realizations of the noise-fill yield no shift in the resulting constraints on A_ϕ. In Figure 4, we plot example data and noise-fills for the 100D-DEEP data set. The top-right panel shows the noise-filled data, the equivalent to d' from Equation (20), but for the actual 100D-DEEP case (we drop the prime on d in the rest of this paper for brevity). The bottom-right panel shows $\bar{n}$ after applying the Fourier mask, which gives intuition for why no degradation occurs due to the noise-fill, since the added noise tends to zero in the unmasked pixels and for the unmasked Fourier modes. The toy problem discussed in this section otherwise maps directly onto the real case, with the straightforward inclusion of the additional lensing and systematics operators.

3.9. Negligible Effects

To conclude this section, we mention a few effects that are expected to be negligible for this data set and are thus not modeled. Both Bayesian and QE pipelines ignore sky curvature, instead working in the flat-sky approximation, which is very accurate for the modestly sized 100 deg^2 patch considered here. The lensing operation is implemented with LenseFlow (MAW20), which assumes the Born approximation. Post-Born effects are not detectable until much lower noise levels and are thus ignored (Pratten & Lewis 2016; Beck et al. 2018; Böhm et al. 2018; Fabbian et al. 2018). Finally, we do not model galactic or extragalactic foregrounds. The 100D-DEEP field is in a region of sky particularly free of galactic contamination, and we conservatively mask modes below $\ell \sim 500$; thus, we expect negligible polarized galactic dust foregrounds (Planck Collaboration et al. 2020b). Extragalactic foregrounds are expected to be much smaller in polarization than in temperature, and here we only use polarization. Given that we also conservatively mask modes above $\ell = 3000$, we follow Wu et al. (2019b) in concluding extragalactic foregrounds can be ignored in this analysis.

4. Lensing Analysis

4.1. Bayesian Lensing

The Bayesian sampling pipeline very closely follows the methodology described in MAW20, and uses the same code, CMBLensing.jl on Github. Conceptually it is extremely straightforward: it is simply a Monte Carlo sampler of the full posterior given in Equation (8). Beyond this, there are a few practical details that we describe in this section.

First, we perform the standard change-of-variables from $(f, \phi) \to (f', \phi')$ and sample the posterior in terms of (f', ϕ')

instead. In this parameterization, the posterior is less degenerate and better conditioned, yielding much better performance of the sampling algorithm. This was extensively discussed in MAW20, and we apply the same re-parameterization as described there almost without change. Specifically, we take

$$\phi^\circ \equiv \mathcal{G}(A_\phi)\,\phi \tag{25}$$

$$f^\circ \equiv \mathbb{L}(\phi)\,\mathcal{D}\,f. \tag{26}$$

The operator $\mathcal{D}$ is defined to be diagonal in EB Fourier space, and $\mathcal{G}(A_\phi)$ is diagonal in Fourier space, with

$$\mathcal{D} \equiv \left[\frac{\mathcal{C}_f + 2\mathcal{N}_f}{\mathcal{C}_f}\right]^{\frac{1}{2}} \tag{27}$$

$$\mathcal{G}(A_\phi) \equiv \left[\frac{\mathcal{C}_\phi(A_\phi) + 2\mathcal{N}_\phi}{\mathcal{C}_\phi(A_\phi)}\right]^{\frac{1}{2}} \tag{28}$$

where $\mathcal{N}_f$ should approximate the sum of instrumental noise and lensing-induced excess CMB power, and $\mathcal{N}_\phi$ should approximate noise in the ϕ reconstruction. Here, we find a sufficient choice is to set $\mathcal{N}_f$ to isotropic 12 μK arcmin white noise, and $\mathcal{N}_\phi$ to the 2D QE $N^{(0)}$ bias. We note that the optimal choice of these operators is not precisely defined, and poor choices do not affect results, instead only lead to slower convergence.

With the re-parameterized target posterior in hand, we now describe the sampler. For both convenience and efficiency, the sampling is broken up into separate Gibbs steps where we sample different conditional slices of Equation (8). The Gibbs procedure ensures that after a sufficiently long time, the chain of conditional samples asymptotes to draws from the joint distribution.

The first Gibbs step samples the conditional distribution of f given the other variables. The advantage of splitting this off as its own Gibbs step is that this conditional is Gaussian and can be sampled exactly by running one conjugate gradient solver. This solver involves inverting the operator shown below in Equation (29), where we have left out instrumental parameters and beam and transfer functions for clarity.[42] We use a nested pre-conditioner, wherein we precondition Equation (29) with Equation (30), which itself involves a conjugate gradient solution using Equation (31) as a pre-conditioner. In Equation (31) we use a noise operator, $\hat{\mathcal{C}}_n$, which is an approximate EB Fourier-diagonal version of $\mathcal{C}_n$, making the final pre-conditioner explicitly invertible.

$$\mathcal{C}_f^{-1} + \mathbb{L}(\phi)^\dagger \mathsf{M}_p^\dagger \mathsf{M}_f^\dagger \mathcal{C}_n^{-1} \mathsf{M}_f \mathsf{M}_p\, \mathbb{L}(\phi) \tag{29}$$

$$\mathcal{C}_f^{-1} + \mathsf{M}_p^\dagger \mathsf{M}_f^\dagger \mathcal{C}_n^{-1} \mathsf{M}_f \mathsf{M}_p \tag{30}$$

$$\mathcal{C}_f^{-1} + \mathsf{M}_f^\dagger \hat{\mathcal{C}}_n^{-1} \mathsf{M}_f. \tag{31}$$

The advantage of this scheme is that it minimizes the number of times we need to compute the action of Equation (29), which involves two lensing operations and hence is much costlier than the others. With the nested preconditioning, only a few applications of Equation (29) are necessary per solution.

The second Gibbs step samples the conditional distribution of ϕ given the other variables. This sample is drawn via Hamiltonian Monte Carlo (Betancourt 2017), which involves sampling a random momentum, p, from a chosen mass matrix, and then performing a symplectic integration to evolve the Hamiltonian for the system. Poor choices of mass matrix or large symplectic integration errors yield a slower converging chain, but do not bias the result asymptotically. We find that 25 leap-frog symplectic integration steps with step size $\epsilon = 0.02$ per Gibbs pass yield nearly optimal convergence efficiency. We note that to control symplectic integration error, we also need at least a 10-step fourth-order Runge–Kutta ODE integration as part of the LenseFlow solver (in MAW20, only seven steps were needed, likely due to simpler masking). Finally, the mass matrix should ideally approximate the Hessian of the log-posterior; here we use,

$$\Lambda_\phi(A_\phi) = \mathcal{G}(A_\phi)^{-2}\left[\mathcal{N}_\phi^{-1} + \mathcal{C}_\phi(A_\phi)^{-1}\right]. \tag{32}$$

The final Gibbs passes sample the conditionals of each of the remaining scalar parameters in turn, A_ϕ, $P_{\rm cal}$, $\psi_{\rm pol}$, ϵ_Q, ϵ_U, and the β_i. Since these are 1D distributions, we sample by evaluating the log-posterior along a grid of values, interpolating it, then using inverse-transform sampling to get an exact sample. Importantly, in all cases except A_ϕ, these parameters are "fast" parameters because $\mathbb{L}(\phi)f$ remains constant along the conditional slice and can be computed just once at the beginning of the pass. Indeed, sampling these parameters accounts for <5% of the total runtime of a chain, and one could imagine adding many other instrumental parameters like these at almost no computational cost. Sampling A_ϕ is somewhat costlier because Equation (25) couples A_ϕ and f, meaning that each grid point of A_ϕ requires lensing a new map (however, the decorrelating effect of the re-parameterization far outweighs this increased computational cost).

4.2. Quadratic Estimate

The QE analysis closely follows those of the 100 deg^2 and 500 deg^2 SPTpol analyses (Story et al. 2015; Wu et al. 2019a). It uses the standard SPT QE pipeline, and so is completely independent from the Bayesian code. We give a brief review of the QE pipeline here and take note of aspects particular to this analysis, referring the reader to the previous works for a more comprehensive treatment.

The QE uses correlations between Fourier modes in pairs of CMB maps to estimate the lensing potential. Here we use the same modified form of the Hu & Okamoto (2002) estimator as in Wu et al. (2019a),

$$\bar{\phi}_L^{XY} = \int d^2\ell\, \bar{X}_\ell \bar{Y}^*_{\ell-L} W^{XY}_{\ell,\ell-L}, \tag{33}$$

where $\bar{X}$ and $\bar{Y}$ are inverse-variance filtered data maps and W^{XY} is a weighting function with $XY \in$ {EE, EB}.

The inverse-variance filtering used for the QE does not employ the noise-fill procedure outlined in Section 3.8, opting instead to leave the existing pipeline unmodified. Here, the noise is approximated as the sum of two components. The first is a pixel-space diagonal component, $\mathcal{N}_{n,p} = \mathsf{M}_p^{-1}\mathcal{N}\mathsf{M}_p^{-1}$, where M_p is the pixel mask and $\mathcal{N}$ is a homogeneous white noise covariance specified by the noise levels at the end of Section 2. The second is a Fourier-space diagonal component, $\mathcal{N}_{n,f}$, which includes the power spectrum of atmospheric foregrounds and excess instrumental 1/f noise not captured in the first component, and is determined empirically from the real noise realizations. Inverse-variance filtering can then be

[42] The exact operator to be inverted can be derived by taking the derivative d/df of Equation (8), setting it equal to zero, and solving for f.

performed by solving the following equation for $\bar{X}$ with conjugate gradient:

$$[\mathbb{C}^{-1} + \mathbb{M}^{\dagger}\mathbb{C}_{n,p}^{-1}\mathbb{M}]\,\bar{X} = \mathbb{M}^{\dagger}\mathbb{C}_{n,p}^{-1}\,d_{\rm QE}, \tag{34}$$

where $\mathbb{C} = \mathbb{C}_f + \mathbb{C}_{n,f}$ and $\mathbb{M} = \mathbb{M}\,\mathbb{B}$.

We then correct each estimator, $\bar{\phi}_L^{XY}$, by (1) subtracting a mean-field bias, $\bar{\phi}_L^{XY,\rm MF}$, computed from an average over simulations, (2) normalizing by the analytic response, $R_L^{XY,\rm Analytic}$, and (3) summing the debiased and normalized estimates. We account for the impact of the pixel mask, not captured by the analytic response, with an isotropic Monte Carlo correction, $R_L^{\rm MC}$. This is computed by fitting a smooth curve to the ratio $C_\ell^{\bar{\phi}\times\phi_{\rm true}}/C_\ell^{\phi\phi,\rm theory}$, averaged over simulations. This gives a normalized unbiased estimate

$$\hat{\phi}_L = \frac{1}{R_L^{\rm MC}}\frac{\sum_{XY}\bar{\phi}_L^{XY} - \bar{\phi}_L^{XY,\rm MF}}{\sum_{XY}R_L^{XY,\rm Analytic}}. \tag{35}$$

To obtain constraints on A_ϕ, we take the auto-spectrum of $\hat{\phi}_L$ to form biased lensing power spectra, $\hat{C}_\ell^{\phi\phi}$. We then estimate the typical $N_L^{(0),\rm RD}$ and $N_L^{(1)}$ biases using simulations, and apply a final multiplicative Monte Carlo correction $f_{\rm PS}$ as in Wu et al. (2019a). No foreground correction is applied, so the final expression for the debiased bandpowers is

$$\hat{C}_\ell^{\phi\phi} = f_{\rm PS}[C_\ell^{\hat{\phi}\hat{\phi}} - N_L^{(0),\rm RD} - N_L^{(1)}]. \tag{36}$$

We calculate the covariance between the bandpowers, Σ, by running a Monte Carlo over the entire procedure. Figure 5 shows the bandpowers of $C_L^{\kappa\kappa} \equiv L^4 C_L^{\phi\phi}/4$, along with error bars computed from the diagonal of Σ.

Since the bandpower errors are assumed Gaussian, the resulting A_ϕ constraints are also Gaussian and are given by

$$A_\phi^{\rm QE} = \frac{\hat{C}_\ell^{\phi\phi}\,(\Sigma^{-1})_{\ell\ell'}\,C_{\ell'}^{\phi\phi}}{C_\ell^{\phi\phi}\,(\Sigma^{-1})_{\ell\ell'}\,C_{\ell'}^{\phi\phi}} \tag{37}$$

$$\sigma(A_\phi^{\rm QE}) = \frac{1}{\sqrt{C_\ell^{\phi\phi}\,(\Sigma^{-1})_{\ell\ell'}\,C_{\ell'}^{\phi\phi}}}, \tag{38}$$

where the summation over $\ell\ell'$ is implied. For this calculation, we truncate Σ at the third off-diagonal, beyond which we do not resolve any nonzero covariance to within Monte Carlo error, consistent with the expectation that the correlation should be small for very distant bins. We note, however, that correlation between neighboring bins can be as large as 10% and has a significant impact on the final uncertainties.

5. Validation

5.1. Chain Convergence

One of the main challenges of the Bayesian procedure is ensuring the Monte Carlo chains are sufficiently converged and are thus yielding stationary samples from the true posterior distribution. A large body of work exists on verifying chain convergence, and many methods of varying sophistication exist. Our experience has been that the most robust and accurate check is actually the simplest, namely just running multiple independent chains in parallel starting from different initial points, and ensuring that the quantities of interest have identical statistics between the different chains. Here, we are in a fortunate position where this is possible, largely because: (1) it is computationally feasible to run many chains and to run existing chains for longer if there is any doubt, and (2) we find no evidence for complicated multimodal distributions, so convergence is not about finding multiple maxima but rather simply a matter of getting enough samples to smoothly map out the (mildly) non-Gaussian posteriors of interest.

Checking for convergence usually begins by visually inspecting the samples from a chain. For the baseline 100D-DEEP chain, we show the sampled values of the cosmological and systematics parameters comprising θ in Figure 6. Our default runs evolve 32 chains in parallel (batches of eight chains per Tesla V100 GPU) and hold θ fixed for the first 100 steps to give the f and φ maps a chance to find the bulk of the posterior first, which reduces the needed burn-in time. Note that the starting point for our chains is a sample from the prior, not just for θ but also for the f and φ maps themselves.[43] Despite this, Figure 6 shows that all θ converge to the same regions in parameter space, and no "long wavelength" drift is seen in the samples.

We also check convergence by splitting the 32 chains into two sets of 16 and estimating parameter constraints from each set. The 1D posteriors from two sets of the baseline 100D-DEEP case are shown in Figure 7. Here we remove a burn-in period of 200 samples from the beginning of each chain. We find that all contours overlap closely, and no conclusions would be reasonably changed by picking one half over the other.

To make the convergence diagnostics more quantitative, we use the following procedure throughout this paper whenever quoting any number derived from a Monte Carlo chain. We first compute the effective sample size (ESS) of the quantity of interest given the observed chain auto-correlation (Goodman & Weare 2010). We then use bootstrap resampling to estimate the Monte Carlo error, wherein (1) we draw N random samples with replacement from the chain where N is the ESS, (2) we compute the quantity in question using these samples, then (3) we repeat this thousands of times and measure the scatter. The scatter gives a 1σ Monte Carlo error, which we report using the typical notation that M digits in parentheses indicate an error in the last M digits of the quantity, i.e., 1.23(4) is shorthand for 1.23 ± 0.04. We use this not only for the posterior mean, but also standard deviations, correlation coefficients, or any other quantity estimated from the chain.

For example, skipping ahead to the results presented in the next section, the constraint on A_ϕ from the 100D-DEEP chain is

$$A_\phi = 0.949(8) \pm 0.122(5). \tag{39}$$

This is to say, the standard error on the mean is 0.008, which is an acceptable 6% of the 1σ posterior uncertainty of 0.122(5), and could be reduced further by running the chain longer if desired.

If we are interested only in constraints on A_ϕ, then Equation (39) gives us what we need to know about how accurate our posterior inference on this quantity is. It is the case, however, that not all modes in the corresponding φ samples in the chain are necessarily converged to this same level. This will not affect A_ϕ since not all modes are informative for A_ϕ, and the errors in Equation (39) tell us about the convergence of the sum total of all modes that are

[43] Note that due to the "curse of dimensionality," these random starting points are much farther apart in the high-dimensional parameter space than might seem from looking at any 1D projection.

informative. In other applications, however, we might care about other modes, for example for delensing external data sets or for cross correlating with other tracers of large-scale structure. We can check the convergence for all modes at the field level by computing posterior mean maps and comparing the power spectrum of the difference when estimated again from two independent sets of 16 chains. Figure 8 shows posterior mean maps and Figure 9 shows the power spectrum differences from the two independent sets. Across a wide range of scales in ϕ, E, and B, the power of the difference maps is one to two orders of magnitude below the signal. The only exception is very small scales in ϕ; indeed, this is an example of modes for which the standard error is larger than the mean, but which are not informative for A_ϕ. If one uses these samples for a downstream analysis, one could use the bootstrap resampling procedure with the maps themselves to estimate the Monte Carlo error in whatever final quantity was computed from these samples.

5.2. Simulations

Having verified in the previous section that Monte Carlo errors in our chains are sufficiently small, we now verify the pipeline itself, as well as our noise covariance approximation. This is done by running chains on simulated data and checking that, on average, we recover the input truth. Crucially, the simulations we use include real noise realizations, while the posterior itself uses the model noise covariance. If the statistics of the real noise were different in a way not captured by the model noise covariance, we would expect to see some bias against the input truth in these simulations.

Figure 10 shows these posterior distributions. The simulation truth uses the same fiducial Planck cosmology used in the baseline model (Section 3). Additionally, we include simulated systematics at a level given by the best-fit values of the 100d-DEEP analysis itself, to confirm that we recover nonzero values of the systematics parameters. The colored lines are the posteriors from each of the N = 100 simulations performed, and the shaded black curve is the product of all N. Because the simulated data are independent (ignoring the very small correlations between our sign-flipped noise realizations) and because the θ shown in this figure have a uniform prior, the product can also be interpreted as a single posterior given N data,

$$\mathcal{P}(\theta|d_1)\mathcal{P}(\theta|d_2)\ldots\mathcal{P}(\theta|d_N) = \mathcal{P}(\theta|d_1, d_2, \ldots, d_N). \quad (40)$$

This indicates that the black shaded contour should also, on average, cover the input truth. If there were any systematic biases affecting the inference of θ, either from noise mismodeling or from errors in the pipeline, we would expect to find a noticeable bias, which we do not. With N = 100 simulations, we have formally checked against biases at the level of $1/\sqrt{N} \approx 10\%$ of the 1σ error bar for any single realization.

6. Results

6.1. Joint A_ϕ and A_L Constraints

The A_ϕ constraint obtained from the QE explicitly does not use information from the power spectrum of the data because the weights $W^{XY}_{\ell,\ell-L}$ in Equation (33) are zero when $L = 0$. The Bayesian constraint, however, extracts all information, including whatever may be contained in the power spectrum, as well as in all higher-order moments (bispectra, trispectra, etc.). To facilitate a fairer comparison between the two, and as a consistency check, it is useful to separate out the power spectrum information in the Bayesian case.

A natural way to do so is by adding a correction to the noise covariance operator such that,

$$\mathcal{C}_n \rightarrow \mathcal{C}_n + \Delta A_L\, \mathbb{L}\, \mathcal{C}_{\rm len}\, \mathbb{L}^\dagger, \quad (41)$$

where ΔA_L is a new free parameter, $\mathbb{L} \equiv$ [illegible], and

$$\mathcal{C}_{\rm len} = {\rm Diagonal}(C_\ell(A_\phi^{100\to2000} = 1) - C_\ell(A_\phi^{100\to2000} = 0)). \quad (42)$$

This is similar to the effect of marginalizing over an extra data component that is Gaussian and has a lensing-like power spectrum with amplitude controlled by ΔA_L, but that does not have the non-Gaussian imprint of real lensing. The similarly is only partial, however, because the correction is sometimes negative (lensing reduces power at the top of peaks in the E-mode power spectrum), while an extra component could only have a positive power contribution. Directly modifying the noise covariance remedies this, and can add or subtract power as long as the sum of the noise and lensing-like contributions still yields a positive-definite total covariance (which is the case for the range of ΔA_L explored by the MCMC chains here).

With this modification, both nonzero ΔA_L and nonzero A_ϕ can generate lensing-like power in the data. The sum of the two parameters thus gives the total lensing-like effect on the data power spectrum, and most closely matches the typical definition of the A_L or $A_{\rm lens}$ parameter, which in our case is a "derived" parameter,

$$A_L = A_\phi + \Delta A_L. \quad (43)$$

If no residual lensing-like power beyond the actual lensing generated by A_ϕ is needed to explain the data, one expects to find $\Delta A_L = 0$ and $A_L = 1$.

Because the power spectrum of the data could be just as well explained by $\Delta A_L = 1$ and $A_\phi = 0$, the extent to which we infer nonzero A_ϕ when ΔA_L is a free parameter confirms that not just power spectrum information is contributing to the constraint, but also quadratic $L \neq 0$ modes and higher-order moments. Correspondingly, marginalizing over ΔA_L is equivalent to removing power spectrum information from the A_ϕ constraint, giving us the tool needed to separate out this information.

A consequence of the modification to the $\mathcal{C}_n$ operator in Equation (41) is that it is no longer easily factorizable in any simple basis. This presents three new numerical challenges for our MCMC chains: (1) applying the inverse of $\mathcal{C}_n$, (2) drawing Gaussian samples with covariance $\mathcal{C}_n$, and (3) computing the determinant of $\mathcal{C}_n$. Inversion turns out to be fairly easily performed with a negligible $\mathcal{O}(10)$ iterations conjugate gradient. Sampling is performed by computing $\mathcal{C}_n^{\frac{1}{2}}\, x$ with the same ODE-based solution used in Equation (23). The determinant (as a function of ΔA_L) is the most difficult piece, but can be computed utilizing the method described in Fitzsimons et al. (2017). This involves swapping the log determinant for a trace

$$\begin{aligned}\log\det[\mathcal{N}_n + \Delta A_{\rm L}\, \mathbb{L}\, \mathcal{C}_{\rm len}\, \mathbb{L}^\dagger] \\ = \sum_{k=1}^{\infty} \mathrm{tr}\{[-\Delta A_{\rm L}\, \mathbb{L}\, \mathcal{C}_{\rm len}\, \mathbb{L}^\dagger \mathcal{N}_n^{-1}]^k\} + C,\end{aligned} \tag{44}$$

where C is a constant that is independent of $\Delta A_{\rm L}$ and can thus be ignored. The trace is then evaluated stochastically using a generalization of Hutchinson's method (Hutchinson 1990) to complex vectors (Iitaka & Ebisuzaki 2004), which evaluates the trace of some matrix $\mathbb{M}$ as $\langle z^\dagger \mathbb{M} z \rangle$ where z are vectors of unit-amplitude random-phase complex numbers, here in the EB Fourier domain. The summation in Equation (44) converges since our matrix is positive-definite, and only 20 terms are needed to give sufficient accuracy in the $\Delta A_{\rm L}$ region explored by the chain. Note also that because the powers of $\Delta A_{\rm L}$ factor out of the trace, the traces can be precomputed once at the beginning of the chain. In terms of sampling, $\Delta A_{\rm L}$ is a "fast" parameter and does not significantly impact chain runtime.

In the top panel of Figure 11, we show joint constraints on $\Delta A_{\rm L}$ and A_f from the 100D-DEEP data. Here we find,

$$\Delta A_{\rm L} = 0.024(9) \pm 0.170(7) \tag{45}$$

$$A_f^{100\to 2000} = 0.955(14) \pm 0.135(10). \tag{46}$$

The two parameters are visibly degenerate, with cross-correlation coefficient $\rho = -0.40(5)$. One can calculate by how much $\sigma(A_f)$ is degraded due to marginalizing over $\Delta A_{\rm L}$ as $1/\sqrt{1-r^2}$, which here gives a 9(3)% degradation. Thus, relatively little information on A_f comes from the power spectrum of the data; instead, most of the constraining power originates from lensing non-Gaussianity. Because of this small impact and for simplicity, we fix $\Delta A_{\rm L} = 0$ for the remaining results in this paper. However, we note that the 9(3)% contribution from the power spectrum is important to keep in mind when comparing to the QE result in the next section.

Correlations between A_f and $A_{\rm L}$ have been negligible in all previous lensing results from data, but are of considerable interest moving forward as it is likely that they will need to be accurately quantified in the future. Previous work on this topic includes Schmittfull et al. (2013), who computed the correlation between A_f estimated via the QE and $A_{\rm L}$ estimated via a traditional power spectrum analysis, finding at most a 10% correlation for temperature maps at Planck-like noise levels. Peloton et al. (2017) extended similar calculations to polarization, finding correlations in the 5%–70% range for CMB-S4-like polarization maps, depending on the exact multipole ranges considered, if a realization-dependent noise subtraction is performed, and whether T, E, and/or B are used to estimate $A_{\rm L}$. The correlation is largest when using B, since B is entirely sourced by lensing and thus contains much of the same information as f. For the 100D-DEEP data, there is twice the Fisher information for $A_{\rm L}$ in B as compared to E, which means our observed correlation should be on the higher end. This is counteracted by the fact that our data is noisier than the CMB-S4 noise levels assumed in Peloton et al. (2017), meaning we should see a lower correlation. Ultimately, although we have not repeated their calculation for our exact noise levels, our observed correlation has the same sign and reasonably agrees in amplitude with their prediction, despite the fairly different analysis.

It is useful to consider what it would take for frequentist methods such as the ones used in these previous works to reach equivalence with the Bayesian approach in terms of quantifying A_f-$A_{\rm L}$ correlations, or more generally, quantifying correlations between the reconstructed lensing potential and the CMB. First, they would need to be extended beyond the QE, which would introduce computational cost and conceptual complexity. Second, they would need to be extended to compute not just correlations of the lensing reconstruction with the raw (lensed) data, but also with delensed data as well. Although not immediately obvious, this is automatically handled in the Bayesian approach. This is because, despite that the Bayesian procedure does not constrain $A_{\rm L}$ by way of explicitly forming a delensed power spectrum, it exactly accounts for the actual posterior distribution of the lensed data maps. For example, if f were perfectly known such that there were no scatter in the MCMC f samples, this would yield no excess lensing variance when estimating $A_{\rm L}$, simply an anisotropic but perfectly known lensed CMB covariance, corresponding to perfect delensing. Whether it is as easy to estimate such correlations in the frequentist approach is unclear, but we highlight the relative simplicity with which it was attained here. It required no additional costly simulations or complex analytic calculations, only the introduction of $\Delta A_{\rm L}$ into the posterior.

Although outside of the scope of this paper, this approach can be used not just for $\Delta A_{\rm L}$ but any other cosmological parameter that controls the unlensed power spectra. It thus serves as a Bayesian analog to existing frequentist methods for parameter estimation from delensed power spectra (Han et al. 2021), immediately allowing inclusion of lensing reconstruction data, and giving a path to the type of joint constraints from both that will be important for optimally inferring cosmological parameters from future data (Green et al. 2017).

6.2. Improvement over Quadratic Estimate

One of the main goals of this work is to demonstrate an improvement in the Bayesian pipeline when compared to the QE result. This improvement arises because the QE ceases to be approximately minimum-variance around 5 μK arcmin, close to the noise levels of the 100D-DEEP observations.

The baseline 100D-DEEP Bayesian constraint is

$$A_f = 0.949(8) \pm 0.122(5)\ (\text{Bayesian}). \tag{47}$$

For the exact same data set, the QE constraint yields

$$A_f = 0.995 \pm 0.154\ (\text{QE}). \tag{48}$$

This represents an improvement in the 1σ error bar of 26(5)%, summarized in Figure 12.

The shift in the central value between the two results is $\Delta A_f = 0.046(8)$. Note that these results are "nested" because the QE uses only quadratic combinations of the data while the Bayesian result implicitly uses all-order moments. Because of this, one can follow Gratton & Challinor (2020; hereafter GC20) to calculate the standard deviation of the expected shift as $\sigma_{\Delta A_f} = (\sigma_{\rm QE}^2 - \sigma_{\rm Bayesian}^2)^{\frac{1}{2}} = 0.10(6)$. The observed shift therefore falls within the 1σ expectation.

Of this improvement, we have ascertained in the previous section that 9(3)% stems from the power spectrum of the data, which is not used by the QE, but could be included if we combined with traditional power spectrum constraints on $A_{\rm L}$. This leaves a 17(6)% improvement as the fairest comparison between Bayesian and QE results. To ascertain whether this is in line with expectations, we have performed a suite of generic mask-free 100 deg^2 simulations with varying noise levels and $\ell_{\max}$ cutoffs for the reconstruction. For each of these simulations, we compute the QE or joint MAP ϕ estimate, compute the cross-correlation coefficient, r_L, with the true ϕ map, then compute the effective Gaussian noise, given by $N_L^{\phi\phi} = C_L^{\phi\phi}(1/r_L^2 - 1)$. From this noise, we compute Gaussian constraints on A_ϕ without including the power spectrum of the data, such that these should be compared to the 17(6)% result. Improvements in A_ϕ and in $N_{L=200}^{\phi\phi}$ are shown in Figure 14. Near the noise levels of the 100d DEEP field, we find around a 10% expected improvement on A_ϕ.

6.3. Joint Systematics and Cosmological Constraints

A unique feature of the Bayesian approach is the ability to jointly estimate cosmological and systematics parameters by simply adding free parameters to the posterior and sampling them in the chain. Here, we have added parameters for the polarization calibration, $P_{\rm cal}$, the global polarization angle calibration, $\psi_{\rm pol}$, temperature-to-polarization monopole leakage template coefficients, ϵ_Q and ϵ_U, and three beam eigenmode amplitudes, b_1, b_2, and b_3.

Figure 7 shows constraints on all of these parameters jointly with the main cosmological parameter of interest, A_ϕ. For $P_{\rm cal}$, $\psi_{\rm pol}$, ϵ_Q and ϵ_U, the blue lines indicate the best-fit value obtained from the external estimation procedures described in Sections 3.1–3.3. The chain results agree with these in all cases, which is an important consistency check. The beam amplitude parameters, b_i, are sampled with unit Gaussian priors centered at zero. If the data is not sensitive to them, we expect the posterior is also a unit Gaussian centered at exactly zero, which is indeed what we find.

If our main cosmological result significantly depends on knowledge of any of these systematics, we would find a correlation between these parameters and A_ϕ. Instead, we find that no parameter is correlated at more than the 5% level. Using the measured covariance across all parameters, Σ_{ij}, we can calculate the fractional amount by which $\sigma(A_\phi)$ decreases if the systematics were fixed to their best fit in the 100d DEEP chain as[44]

$$\sqrt{\Sigma_{11}/(\Sigma^{-1})_{11}} \approx 0.01, \tag{49}$$

where $i = j = 1$ is the entry corresponding to A_ϕ. Thus, the systematic error contribution to the Bayesian A_ϕ measurement is less than 1% of the statistical error.

Although in this paper we do not propagate any systematic errors through the QE pipeline, for some of the same data used here, this has already been done by Story et al. (2015) and Wu et al. (2019b). The approach there is to modify the input data, for example, multiply it by $1 + s(P_{\rm cal})$ to mimic a 1σ error in the $P_{\rm cal}$ parameter, where $s(P_{\rm cal})$ is determined from some external calibration procedure. The resulting change to A_ϕ is then taken as the 1σ systematic error on A_ϕ due to $P_{\rm cal}$, and the errors from several systematics are added in quadrature (hence assuming that they are all Gaussian and uncorrelated). For $P_{\rm cal}$, because the quadratically estimated lensing potential power spectrum depends on the fourth power of the data, the systematic error on A_ϕ scales as $4 \times s(P_{\rm cal})$ to linear order, and can become significant even for modest calibration error. Indeed, using the above procedure, Wu et al. 2019b found that the systematic error on A_ϕ from polarization was nearly half of the statistical uncertainty.

6.4. Consistency Checks

Having presented our baseline results in the previous subsections, we now perform a number of consistency checks to see if various analysis choices have any impact on the final results. The corresponding constraints on A_ϕ for each case discussed here are pictured in Figure 13.

Our baseline case constrains $A_\phi^{100\to2000}$. As a first check, we extend this range to encompass $A_\phi^{50\to3000}$. Here, we find

$$A_\phi^{50\to3000} = 0.957(8) \pm 0.114(5), \tag{50}$$

which is an additional 7(7)% tighter than the baseline result, and consistent with the shift expected from GC20.

We next check if mask apodization has significant impact. Although the QE produces an unbiased answer regardless of mask, hard mask edges lead to larger Monte Carlo corrections and slightly larger sub-optimality of the final estimator. Conversely, the Bayesian pipeline, in theory, always produces both an unbiased and optimal result. This can be an advantage because, depending on the point-source flux cut, adding a large number of apodized holes to the map can reduce the effective sky area of the observations by a non-negligible amount. One solution sometimes used in the QE case is to inpaint point-source holes rather than leave them masked, and then demonstrate on simulations that negligible bias is introduced due to the inpainting (Benoit-Lévy et al. 2013; Raghunathan et al. 2019). The inpainting is often performed by sampling a constrained Gaussian realization of the CMB within the masked region, given the data just outside of the masked region. The Bayesian pipeline corresponds to simultaneously inpainting all point-source holes with a different realization at each step in the MCMC chain, while accounting for the non-Gaussian statistics of the lensed CMB given the ϕ map at that chain step. In practice, one could imagine that the ringing created by hard mask edges induces large degeneracies in the posterior and leads to poor chain convergence. It is thus useful to verify that the Bayesian pipeline works with an unapodized mask, meaning point sources can simply be masked without apodization, and the pipeline can be used as is without extra steps.

To keep the apodized and unapodized cases nested, we take the original mask and set it to zero everywhere in the apodization taper. The result is the green curve in Figure 13, which gives

$$A_\phi^{100\to2000} = 0.937(15) \pm 0.124(9), \tag{51}$$

consistent with the GC20 expected shift. The slightly looser constraint is consistent with the unapodized case not using the data within the apodization taper, although longer chains would be needed to exactly confirm this. We do not observe a significantly worse auto-correlation length for this chain as compared to the apodized case, demonstrating that mask apodization has little effect on the Bayesian analysis.

The point-source mask serves to reduce foreground contamination. Here, we have used a mask built from point sources detected in temperature, but have not attempted to cross-check if these same point sources are bright in polarization. As a

[44] We could also calculate this by running a separate chain with these explicitly fixed, which we have done as a consistency check, but using Σ_{ij} directly is easier and is less affected by Monte Carlo error.

simple check, we consider leaving point sources completely unmasked. In this case, we find the red curve in Figure 13. This result and the baseline case are also nested. However, this time the shift in central value is inconsistent at 2.8σ given GC20. Visually inspecting the reconstructed κ map (not pictured here) reveals obvious residuals at the locations of a few of the brightest previously masked sources. Evidently, some level of point-source masking is necessary to mitigate foreground biases even in polarization. Our mask is based on a 50 mJy flux cut in temperature. For future analyses, it will be important to determine the flux cut that is a good trade-off between reducing foreground biases but not excising too much data.

7. Conclusion

We conclude with a summary of the main results along with some remarks about the Bayesian procedure and future prospects for this type of analysis. One of the main goals of this work was to apply, for the first time, a full Bayesian reconstruction to very deep CMB polarization data, and observe an improvement over the QE. This work is the second optimal lensing reconstruction ever applied to data, and the first to actually infer cosmological parameters that control the lensing potential itself. Doing so is particularly natural in the Bayesian framework, as extra parameters can always be added (sometimes trivially) and sampled over. We found a 26% improved error bar on A_f in the Bayesian case as compared to the QE, and a 17% improvement after removing power spectrum information.

As instrumental noise levels continue to improve in the future, we expect this relative improvement will increase. In Figure 14, we forecast the relative improvement in A_f, as well more generically the relative improvement in the effective noise level of the f reconstruction at $L = 200$ (the choice of particular L here is arbitrary, and we note that the result is only moderately sensitive to scale). By the time noise levels of the deep CMB-S4 survey are reached, the relative improvement will be around 50% for A_f. The full story is even more optimistic, however, as A_f is not the best parameter to reflect the lower-noise reconstruction possible in the Bayesian case. This is because once a mode becomes signal dominated, A_f is no longer improved by further reducing the noise for that mode (only more sky can help). If we instead consider directly the effective noise level itself, which will be more indicative of the types of improvements one can achieve on parameters that are determined from noise-dominated regions of the spectra, we see that improvements of up to factors of seven are possible.

Looking toward the future, the main challenges we foresee for the Bayesian approach are twofold. The first is related to a fundamental difference between the Bayesian and QE (or any frequentist) methods. In the frequentist case, one is free to use various approximations in the process of computing an estimator, or to null various data modes, as long as the final result is debiased (usually via Monte Carlo simulations), and this bias can be demonstrated to be sufficiently cosmology-independent. The Bayesian approach does not have any notion of debiasing; instead, a forward model for the full data must be provided and guaranteed to be sufficiently accurate so as to ensure biases in the final answer are small. The solution we have employed here is to build the forward model with approximations to things like the transfer function, $\mathbb{T}$, or the noise covariance, $\mathbb{C}_n$, which are as accurate as more sophisticated full pipeline simulations, but not prohibitive to compute at each step in the MCMC chain. Pushing to larger scales, larger sky fractions, and more complex scanning strategies will require upgrading these approximations, while maintaining high computational speed. The toolbox for these types of improvements includes things like machine-learning models (e.g., Münchmeyer & Smith 2019, for a CMB application), sparse operators such as the BICEP observation matrix (Ade et al. 2015), or other physically motivated analytic approximations.

The second challenge of the Bayesian approach is computational. For reference, the Monte Carlo simulations needed to compute the QE here take around 10 minutes across a few hundred CPU cores. Conversely, the Bayesian MCMC chains take about 5 hr on four GPUs, with interpretable results returned within around an hour.[45] Ignoring the mild total allocation cost of these calculations, the main difference is the longer wall-time of the MCMC chain. Since the computation is roughly dominated by FFTs, a naive scaling to, e.g., the full SPT-3G 1500 deg^2 footprint along with an upgraded 2′ pixel resolution (to reach scales of $\ell \sim 5000$) gives around one week for a chain. Because the MCMC chains do not appear to require a long burn-in time, the total runtime can be reduced fairly efficiently by running more chains in parallel on more GPUs, or potentially on TPUs. Along with some planned code optimizations, we expect it will be possible to obtain results for a full SPT-3G data set in under a day. Additionally, much of the runtime will be dominated by Wiener filtering, where our current algorithm can likely be improved, making scaling to even larger data sets possible. It may be noteworthy to highlight that the computational tools in play here, GPUs, linear algebra, and automatic differentiation, are the identical building blocks of machine learning, and are the subject of rapid technological improvements.

The overall experience of Bayesian lensing in this work is encouraging, solving and side-stepping many difficulties that arise in other procedures. While some development is needed to extend beyond the data set considered here, this approach appears to be a viable option for future CMB probes that will depend on methods such as these for the next generation of lensing analyses.

M.M. thanks Uros Seljak for useful discussions. The South Pole Telescope (SPT) is supported by the National Science Foundation through grants PLR-1248097 and OPP-1852617. Partial support is also provided by the NSF Physics Frontier Center grant PHY-1125897 to the Kavli Institute of Cosmological Physics at the University of Chicago, the Kavli Foundation, and the Gordon and Betty Moore Foundation grant GBMF 947. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. This research also used the Savio computational cluster resource provided by the Berkeley Research Computing program at the University of California, Berkeley (supported by the UC Berkeley Chancellor, Vice Chancellor for Research, and Chief Information Officer). The Melbourne group acknowledges support from the University of Melbourne and an Australian Research Council's Future Fellowship (FT150100074). Argonne National Laboratory's

[45] The same code can run on CPUs by switching a flag, although it is factors of several slower and mainly useful for debugging.

work was supported by the U.S. Department of Energy, Office of High Energy Physics, under contract DE-AC02-06CH11357. We also acknowledge support from the Argonne Center for Nanoscale Materials.

ORCID iDs

M. Millea https://orcid.org/0000-0001-7317-0551
P. A. R. Ade https://orcid.org/0000-0002-5127-0401
B. A. Benson https://orcid.org/0000-0002-5108-6823
F. Bianchini https://orcid.org/0000-0003-4847-3483
L. E. Bleem https://orcid.org/0000-0001-7665-5079
T. M. Crawford https://orcid.org/0000-0001-9000-5013
M. A. Dobbs https://orcid.org/0000-0001-7166-6422
W. Everett https://orcid.org/0000-0002-5370-6651
S. Guns https://orcid.org/0000-0001-7143-2853
N. Gupta https://orcid.org/0000-0001-7652-9451
G. P. Holder https://orcid.org/0000-0002-0463-6394
S. S. Meyer https://orcid.org/0000-0003-3315-4332
L. M. Mocanu https://orcid.org/0000-0002-2416-2552
S. Patil https://orcid.org/0000-0001-5871-7520
C. L. Reichardt https://orcid.org/0000-0003-2226-9169
B. R. Saliwanchik https://orcid.org/0000-0002-5089-7472
G. Smecher https://orcid.org/0000-0002-5560-187X
A. A. Stark https://orcid.org/0000-0002-2718-9996
J. D. Vieira https://orcid.org/0000-0001-7192-3871
N. Whitehorn https://orcid.org/0000-0002-3157-0407
W. L. K. Wu https://orcid.org/0000-0001-5411-6920

References

Abazajian, K. N., Adshead, P., Ahmed, Z., et al. 2016, arXiv:1610.02743
Adachi, S., Aguilar Faúndez, M. A. O., Akiba, Y., et al. 2020, PhRvL, 124, 131301
Ade, P. A. R., Aikin, R. W., Barkats, D., et al. 2015, ApJ, 814, 110
Aiola, S., Calabrese, E., Maurin, L., et al. 2020, JCAP, 2020, 047
Allen, E. J., Baglama, J., & Boyd, S. K. 2000, Algebra Appl., 310, 167
Anderes, E., Wandelt, B. D., & Lavaux, G. 2015, ApJ, 808, 152
Beck, D., Fabbian, G., & Errard, J. 2018, PhRvD, 98, 043512
Benoit-Lévy, A., Déchelette, T., Benabed, K., et al. 2013, A&A, 555, A37
Benson, B. A., Ade, P. A. R., Ahmed, Z., et al. 2014, Proc. SPIE, 9153, 91531P
Betancourt, M. 2017, arXiv:1701.02434
BICEP2 Collaboration, Ade, P. A. R., Aikin, R. W., et al. 2014, PhRvL, 112, 241101
Bleem, L., Ade, P., Aird, K., et al. 2012, JLTP, 167, 859
Böhm, V., Sherwin, B. D., Liu, J., et al. 2018, PhRvD, 98, 123510
Carlstrom, J. E., Ade, P. A. R., Aird, K. A., et al. 2011, PASP, 123, 568
Carron, J., & Lewis, A. 2017, PhRvD, 96, 063510
Crites, A. T., Henning, J. W., Ade, P. A. R., et al. 2015, ApJ, 805, 36
Everett, W. B., Zhang, L., Crawford, T. M., et al. 2020, ApJ, 900, 55
Fabbian, G., Calabrese, M., & Carbone, C. 2018, JCAP, 02, 050
Fitzsimons, J., Cutajar, K., Osborne, M., Roberts, S., & Filippone, M. 2017, arXiv:1704.01445
Goodman, J., & Weare, J. 2010, CAMCS, 5, 65
Gratton, S., & Challinor, A. 2020, MNRAS, 499, 3410
Green, D., Meyers, J., & van Engelen, A. 2017, JCAP, 2017, 005
Han, D., Sehgal, N., MacInnis, A., et al. 2021, JCAP, 2021, 031
Henning, J. W., Sayre, J. T., Reichardt, C. L., et al. 2018, ApJ, 852, 97
Hindmarsh, A. C., Brown, P. N., Grant, K. E., et al. 2005, ACM Trans. Math. Softw., 31, 363
Hirata, C. M., & Seljak, U. 2003a, PhRvD, 68, 083002
Hirata, C. M., & Seljak, U. 2003b, PhRvD, 67, 043001
Hu, W., & Okamoto, T. 2002, ApJ, 574, 566
Hutchinson, M. F. 1990, Communications in Statistics—Simulation and Computation, 19, 433
Iitaka, T., & Ebisuzaki, T. 2004, PhRvE, 69, 057701
Keating, B. G., Shimon, M., & Yadav, A. P. S. 2012, ApJL, 762, L23
Lewis, A., & Challinor, A. 2006, PhR, 429, 1
Millea, M., Anderes, E., & Wandelt, B. D. 2019, PhRvD, 100, 023509
Millea, M., Anderes, E., & Wandelt, B. D. 2020, PhRvD, 102, 123542
Münchmeyer, M., & Smith, K. M. 2019, arXiv:1905.05846
Padin, S., Staniszewski, Z., Keisler, R., et al. 2008, ApOpt, 47, 4418
Peloton, J., Schmittfull, M., Lewis, A., Carron, J., & Zahn, O. 2017, PhRvD, 95, 043508
Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A15
Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020a, A&A, 641, A8
Planck Collaboration, Akrami, Y., Ashdown, M., et al. 2020b, A&A, 641, A11
Pratten, G., & Lewis, A. 2016, JCAP, 2016, 047
Rackauckas, C., & Nie, Q. 2017, JORS, 5, 15
Raghunathan, S., Holder, G. P., Bartlett, J. G., et al. 2019, JCAP, 11, 037
Schmittfull, M. M., Challinor, A., Hanson, D., & Lewis, A. 2013, PhRvD, 88, 063012
Seljak, U., & Hirata, C. M. 2004, PhRvD, 69, 043005
Story, K. T., Hanson, D., Ade, P. A. R., et al. 2015, ApJ, 810, 50
The Simons Observatory Collaboration, Ade, P., Aguirre, J., et al. 2019, JCAP, 2019, 056
Vanderlinde, K., Crawford, T. M., de Haan, T., et al. 2010, ApJ, 722, 1180
Wu, W. L. K., Mocanu, L. M., Ade, P. A. R., et al. 2019a, ApJ, 884, 70
Wu, W. L. K., Mocanu, L. M., Ade, P. A. R., et al. 2019b, ApJ, 884, 70
Zaldarriaga, M., & Seljak, U. 1999, PhRvD, 59, 123507