

# Fermilab

Cosmology From CMB Lensing and Delensed EE Power Spectra Using  
2019-2020 SPT-3G Polarization Data

FERMILAB-PUB-24-0840-PPD

arXiv:2411.06000

This manuscript has been authored by Fermi Research Alliance, LLC  
under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy,  
Office of Science, Office of High Energy Physics.

# Cosmology From CMB Lensing and Delensed EE Power Spectra Using 2019-2020 SPT-3G Polarization Data

F. Ge<sup>1,\*</sup> M. Millea<sup>1</sup> E. Camphuis<sup>2</sup> C. Daley<sup>3,4</sup> N. Huang<sup>5</sup> Y. Omori<sup>6,7</sup> W. Quan<sup>8,7</sup>  
E. Anderes<sup>9</sup> A. J. Anderson<sup>10,7,6</sup> B. Ansarinejad<sup>11</sup> M. Archipley<sup>7,6</sup> L. Balkenhol<sup>2</sup> K. Benabed<sup>2</sup>  
A. N. Bender<sup>12,7,6</sup> B. A. Benson<sup>10,7,6</sup> F. Bianchini<sup>13,14,15</sup> L. E. Bleem<sup>12,7</sup> F. R. Bouchet<sup>2</sup>  
L. Bryant<sup>16</sup> J. E. Carlstrom<sup>7,16,8,12,6</sup> C. L. Chang<sup>12,7,6</sup> P. Chaubal<sup>11</sup> G. Chen<sup>17</sup> P. M. Chichura<sup>8,7</sup>  
A. Chokshi<sup>17</sup> T.-L. Chou<sup>6,7</sup> A. Coerver<sup>5</sup> T. M. Crawford<sup>7,6</sup> T. de Haan<sup>18</sup> K. R. Dibert<sup>6,7</sup>  
M. A. Dobbs<sup>19,20</sup> M. Doohan<sup>11</sup> A. Doussot<sup>2</sup> D. Dutcher<sup>21</sup> W. Everett<sup>22</sup> C. Feng<sup>23</sup> K. R. Ferguson<sup>24,25</sup>  
K. Fichman<sup>8,7</sup> A. Foster<sup>21</sup> S. Galli<sup>2</sup> A. E. Gambrel<sup>7</sup> R. W. Gardner<sup>16</sup> N. Goeckner-Wald<sup>14,13</sup>  
R. Gualtieri<sup>26</sup> F. Guidi<sup>2</sup> S. Guns<sup>5</sup> N. W. Halverson<sup>27,28</sup> E. Hivon<sup>2</sup> G. P. Holder<sup>23</sup> W. L. Holzapfel<sup>5</sup>  
J. C. Hood<sup>7</sup> D. Howe<sup>17</sup> A. Hryciuk<sup>8,7</sup> F. Kéruzoré<sup>12</sup> A. R. Khalife<sup>2</sup> L. Knox<sup>1</sup> M. Korman<sup>29</sup>  
K. Korneelje<sup>6,7</sup> C.-L. Kuo<sup>13,14,15</sup> A. T. Lee<sup>5,30</sup> K. Levy<sup>11</sup> A. E. Lowitz<sup>7</sup> C. Lu<sup>23</sup> A. Maniyar<sup>13,14,15</sup>  
E. S. Martsen<sup>6,7</sup> F. Menanteau<sup>3,31</sup> J. Montgomery<sup>19</sup> Y. Nakato<sup>14</sup> T. Natoli<sup>7</sup> G. I. Noble<sup>32,33</sup>  
Z. Pan<sup>12,7,8</sup> P. Paschos<sup>16</sup> K. A. Phadke<sup>3,31</sup> A. W. Pollak<sup>17</sup> K. Prabhu<sup>1</sup> M. Rahimi<sup>11</sup> A. Rahlin<sup>6,7</sup>  
C. L. Reichardt<sup>11</sup> D. Riebel<sup>17</sup> M. Rouble<sup>19</sup> J. E. Ruhl<sup>29</sup> E. Schiappucci<sup>11</sup> J. A. Sobrin<sup>10,7</sup> A. A. Stark<sup>34</sup>  
J. Stephen<sup>16</sup> C. Tandoi<sup>3</sup> B. Thorne<sup>1</sup> C. Trendafilova<sup>31</sup> C. Umilta<sup>23</sup> J. D. Vieira<sup>3,23,31</sup> A. Vitrier<sup>2</sup>  
Y. Wan<sup>3,31</sup> N. Whitehorn<sup>25</sup> W. L. K. Wu<sup>13,15</sup> M. R. Young<sup>10,7</sup> and J. A. Zebrowski<sup>7,6,10</sup>

(SPT-3G Collaboration)

<sup>1</sup>Department of Physics & Astronomy, University of California, One Shields Avenue, Davis, CA 95616, USA

<sup>2</sup>Sorbonne Université, CNRS, UMR 7095, Institut d'Astrophysique de Paris, 98 bis bd Arago, 75014 Paris, France

<sup>3</sup>Department of Astronomy, University of Illinois Urbana-Champaign,  
1002 West Green Street, Urbana, IL, 61801, USA

<sup>4</sup>Université Paris-Saclay, Université Paris Cité, CEA, CNRS, AIM, 91191, Gif-sur-Yvette, France

<sup>5</sup>Department of Physics, University of California, Berkeley, CA, 94720, USA

<sup>6</sup>Department of Astronomy and Astrophysics, University of Chicago,  
5640 South Ellis Avenue, Chicago, IL, 60637, USA

<sup>7</sup>Kavli Institute for Cosmological Physics, University of Chicago,  
5640 South Ellis Avenue, Chicago, IL, 60637, USA

<sup>8</sup>Department of Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL, 60637, USA

<sup>9</sup>Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA

<sup>10</sup>Fermi National Accelerator Laboratory, MS209, P.O. Box 500, Batavia, IL, 60510, USA

<sup>11</sup>School of Physics, University of Melbourne, Parkville, VIC 3010, Australia

<sup>12</sup>High-Energy Physics Division, Argonne National Laboratory,  
9700 South Cass Avenue., Lemont, IL, 60439, USA

<sup>13</sup>Kavli Institute for Particle Astrophysics and Cosmology,  
Stanford University, 452 Lomita Mall, Stanford, CA, 94305, USA

<sup>14</sup>Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, CA, 94305, USA

<sup>15</sup>SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA, 94025, USA

<sup>16</sup>Enrico Fermi Institute, University of Chicago, 5640 South Ellis Avenue, Chicago, IL, 60637, USA

<sup>17</sup>University of Chicago, 5640 South Ellis Avenue, Chicago, IL, 60637, USA

<sup>18</sup>High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan

<sup>19</sup>Department of Physics and McGill Space Institute, McGill University,  
3600 Rue University, Montreal, Quebec H3A 2T8, Canada

<sup>20</sup>Canadian Institute for Advanced Research, CIFAR Program in  
Gravity and the Extreme Universe, Toronto, ON, M5G 1Z8, Canada

<sup>21</sup>Joseph Henry Laboratories of Physics, Jadwin Hall,  
Princeton University, Princeton, NJ 08544, USA

<sup>22</sup>Department of Astrophysical and Planetary Sciences,  
University of Colorado, Boulder, CO, 80309, USA

<sup>23</sup>Department of Physics, University of Illinois Urbana-Champaign,  
1110 West Green Street, Urbana, IL, 61801, USA

<sup>24</sup>Department of Physics and Astronomy, University of California, Los Angeles, CA, 90095, USA

<sup>25</sup>Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA

<sup>26</sup>Department of Physics and Astronomy, Northwestern University, 633 Clark St, Evanston, IL, 60208, USA

<sup>27</sup>CASA, Department of Astrophysical and Planetary Sciences,  
University of Colorado, Boulder, CO, 80309, USA

<sup>28</sup>Department of Physics, University of Colorado, Boulder, CO, 80309, USA

<sup>29</sup>Department of Physics, Case Western Reserve University, Cleveland, OH, 44106, USA

<sup>30</sup>Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

<sup>31</sup>Center for AstroPhysical Surveys, National Center for Supercomputing Applications, Urbana, IL, 61801, USA

<sup>32</sup>Dunlap Institute for Astronomy & Astrophysics, University of Toronto,  
50 St. George Street, Toronto, ON, M5S 3H4, Canada

<sup>33</sup>David A. Dunlap Department of Astronomy & Astrophysics,  
University of Toronto, 50 St. George Street, Toronto, ON, M5S 3H4, Canada

<sup>34</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA, 02138, USA

(Dated: November 12, 2024)

From CMB polarization data alone we reconstruct the CMB lensing power spectrum, comparable in overall constraining power to previous temperature-based reconstructions, and an *unlensed*  $E$ -mode power spectrum, with clear detections of the third through tenth acoustic peaks. The observations, taken in 2019 and 2020 with the South Pole Telescope (SPT) and the SPT-3G camera, cover  $1500 \text{ deg}^2$  at 95, 150, and 220 GHz with arcminute resolution and roughly  $4.9 \mu\text{K}$ -arcmin coadded noise in polarization. The power spectrum estimates, together with systematic parameter estimates and a joint covariance matrix, follow from a Bayesian analysis using the Marginal Unbiased Score Expansion (MUSE) method. The  $E$ -mode spectrum at  $\ell > 2000$  and lensing spectrum at  $L > 350$  are the most precise to date. Assuming the  $\Lambda\text{CDM}$  model, and using only these SPT data and priors on  $\tau$  and absolute calibration from Planck, we find  $H_0 = 66.81 \pm 0.81 \text{ km/s/Mpc}$ , comparable in precision to the Planck determination and in  $5.4\sigma$  tension with the most precise  $H_0$  inference derived via the distance ladder. We also find  $S_8 \equiv \sigma_8(\Omega_m/0.3)^{0.5} = 0.850 \pm 0.017$ , providing further independent evidence of a slight tension with low-redshift structure probes. The  $\Lambda\text{CDM}$  model provides a good simultaneous fit to the combined Planck, ACT, and SPT data, and thus passes a powerful test. Combining these CMB datasets with BAO observations, we explore extensions to the  $\Lambda\text{CDM}$  model. We find that the effective number of neutrino species, spatial curvature, and primordial helium fraction are consistent with standard model values, and that the 95% confidence upper limit on the neutrino mass sum is  $0.075 \text{ eV}$ , close to the minimum sum expected from observations of solar and atmospheric neutrino oscillations. The SPT data are consistent with the somewhat weak ( $< 3\sigma$ ) preference for excess lensing power seen in Planck and ACT data relative to predictions of the  $\Lambda\text{CDM}$  model given the combined Planck, ACT, and BAO data sets. We also detect at greater than  $3\sigma$  the influence of non-linear evolution in the CMB lensing power spectrum and discuss it in the context of the  $S_8$  tension. Forthcoming SPT-3G analyses will feature deeper and wider observations in temperature and polarization, providing even tighter constraints and more powerful tests of the  $\Lambda\text{CDM}$  model.

## I. INTRODUCTION

Since the successful completion of the Planck satellite’s sky surveys [1], the frontier in the study of the cosmic microwave background (CMB) anisotropies has been its polarization. Here we present the strongest cosmological constraints to date derived from polarization data alone, constraints that are, by some measures, comparable to those from temperature data alone. Low-noise polarization measurements allow for new tests of the standard cosmological model, as the cosmological parameter inference relies on different signals, and they allow for additional robustness to many sources of potential systematic error.

The data we use come from the SPT-3G camera [2] installed on the South Pole Telescope (SPT) in early 2017. By October of 2024, SPT-3G had been used to survey  $10,000 \text{ deg}^2$ , with the bulk of the Austral winter observing time spent on a  $1500 \text{ deg}^2$  patch. The resulting

data allow for powerful tests of the standard cosmological model,  $\Lambda\text{CDM}$ , via the primary and CMB lensing signals [3], improved constraints on primordial gravitational waves via de-lensing of the degree-scale observations of the BICEP/Keck Collaboration [4], and many other science applications [5–11].

This paper is the first of a series of papers on CMB power spectra, the CMB lensing power spectrum, and cosmological parameter estimates inferred from the 2019 and 2020 Austral winter observations of the  $1500 \text{ deg}^2$  patch. The temperature and polarization maps used in these analyses are the result of four times as much observing time and twice as many detectors as were used for similar analyses that used a half season of data taken in 2018 with a partial SPT-3G focal plane [12–15].

This paper is distinct from the others in the coming series by its focus on polarization data only, and by its use of a novel algorithm for optimally and simultaneously estimating the CMB lensing power spectrum, the *unlensed*<sup>1</sup>

\* fge@ucdavis.edu

<sup>1</sup> We note a subtle distinction between a de-lensed spectrum, which

CMB  $E$ -mode polarization power spectrum ( $EE$ ), and systematics parameters. This algorithm, the Marginal Unbiased Score Expansion (MUSE), was developed by Millea & Seljak [16] and Millea [17], and is applied to real data here for the first time. We present the analysis in some detail and emphasize its advantages, including straightforward incorporation of systematic effects into a statistical model, a lensing potential map with lower noise than if it were produced by a quadratic estimator [e.g. 18], and an unlensed  $EE$  power spectrum with reduced sample variance relative to any inference of the lensed  $EE$  power spectrum [19].

De-lensed power spectra from quadratic estimators have been estimated previously from data by Planck Collaboration et al. [20] and Han et al. [21]. The methodology for optimal lensing reconstruction has been principally developed in these papers: Millea & Seljak [16], Hirata & Seljak [22], Carron & Lewis [23], Millea et al. [24, 25]; and an application to SPTpol data was described in Millea et al. [26]. With the SPT-3G data we analyze here, we are crossing the threshold of polarization noise level below which optimal methods lead to significant improvements in lensing reconstruction.

We focus solely on polarization anisotropies in order to avoid modeling extragalactic foregrounds and their non-Gaussian properties. Such modeling would likely be necessary for a similar analysis which included temperature anisotropies, where the relative impact of foregrounds is larger. Due to the low noise and high angular resolution of the SPT-3G measurements, this initial conservative approach still leaves us with an inferred CMB lensing power spectrum with high signal-to-noise ratio. In the part of sky we observe, over the angular scales used in the analysis, galactic foregrounds are also negligibly small [12, 27].

We compare with other CMB measurements in both temperature and polarization, including those from the Planck Collaboration [28] and the Atacama Cosmology Telescope (ACT) Collaboration [29–32]. We check for consistency of our results in the context of the  $\Lambda$ CDM model. We also determine cosmological parameters assuming the  $\Lambda$ CDM model and extensions from SPT data alone and in combination with these other datasets. Although the SPT observations we use here cover much less sky relative to Planck and ACT, the very low noise ( $4.9 \mu\text{K}$ -arcmin in coadded polarization maps) and arc-minute angular resolution enable measurements of the  $EE$  power spectrum and CMB lensing potential power spectrum ( $\phi\phi$ ) that are the most precise measurements ever made at  $\ell > 2000$  for  $EE$  and  $L > 350$  for  $\phi\phi$ .

We discuss how the signals in our data allow for the determination of cosmological parameters assuming the

$\Lambda$ CDM model. The origins of these determinations have interesting differences compared to those from Planck CMB data. For example, the Hubble constant ( $H_0$ ) inference depends heavily on the lensing information; without it, the error increases by more than 10 times. Consistency between the Planck and SPT inferences of  $H_0$  places further constraints on attempts to solve the  $H_0$  tension [33] with changes to cosmological models [34, 35].

We pay particular attention to the implications of our data for the “negative neutrino mass” problem and its relation to excess lensing power pointed out by Craig et al. [36], and to the suggestion from Amon & Efstathiou [37] that the  $\sigma_8$  tension is associated with non-linear evolution, either due to mismodeling the  $\Lambda$ CDM predictions or by neglect of some new physics that modifies evolution on non-linear scales.

The rest of this paper is constructed as follows. In Sect. II we describe the process for making CMB maps from real and mock observations. We give an overview of MUSE and our data modeling in Sect. III. We present our pipeline and data validation processes, the blinding procedure, and results in Sect. IV. We show the resulting bandpowers and cosmological parameter constraints in Sect. V. We conclude in Sect. VI. Throughout this work we use a series of abbreviations for external datasets tabulated in Tab. III.

## II. DATA AND MAPMAKING

We use data taken with the SPT-3G camera during the 2019 and 2020 winter observing seasons. The  $1500 \text{ deg}^2$  SPT-3G winter footprint extends from  $-42^\circ$  to  $-70^\circ$  in declination and  $-50^\circ$  to  $50^\circ$  in right ascension. The footprint is divided into four subfields with declination centers of  $-44.75^\circ$ ,  $-52.25^\circ$ ,  $-59.75^\circ$ , and  $-67.25^\circ$ , and that allows for fitting different systematics parameters between patches that improve detector response consistency and linearity across the entire field. Observations are performed in two-hour periods and map a single sub-field at a time. During one observation, the telescope performs 72 constant-elevation (and thus constant-declination) scans with a scan speed of  $1 \text{ deg/s}$  in right ascension and an elevation step of  $12.5 \text{ arcmin}$  between scans. The subfields are observed 1034, 902, 772, and 578 times respectively across the two seasons, which produces approximately uniform map noise levels across subfields.

Each of the  $\sim 16,000$  detectors in the SPT-3G camera samples the sky as the telescope scans, producing one-dimensional “timestreams” that record sky brightness as a function of right ascension. The timestreams used in this analysis have on-sky sample rates between  $0.3 \text{ arcmin/sample}$  and  $0.6 \text{ arcmin/sample}$  depending on declination, corresponding to an effective Nyquist frequency well beyond our maximum multipole of interest.

The timestreams are low-pass filtered to reduce aliasing when later binning into maps, and high-pass fil-

---

contains a residual lensing component which must be modeled, and our inference of an unlensed spectrum, which contains no such residual, independent of cosmological model. Although of minor impact, this explains our usage of “unlensed” throughout the text.

tered to remove large-scale instrumental and atmospheric noise. The high-pass filter is performed by deprojecting a set of Legendre polynomials up to 30th-order and a set of sines and cosines up to an equivalent angular scale of  $\ell \leq 300$ .

The filtering described above can produce undesired “ringing” features in the maps around the locations of point sources (a term we will loosely use here to describe both emissive galaxies and the Sunyaev-Zeldovich effect from galaxy clusters, the latter which can be spatially extended). To reduce these features, when fitting the deprojection coefficients, we ignore regions around point sources, specifically around all emissive sources with 150 GHz flux  $> 6$  mJy and all galaxy clusters with a signal-to-noise  $> 10$ , for a total of 2655 objects. This procedure reduces ringing but does not mask the point sources, which instead happens at the map level later in the pipeline.

After filtering, we bin the timestream samples using inverse-variance weights into  $0.5625'$  pixels given a Lambert projection centered on a right ascension of  $0^\circ$  and a declination of  $-59.55^\circ$ . Weights are determined from the detector noise in the range  $320 < \ell < 4000$ . The pixel size was chosen so that the  $1500 \text{ deg}^2$  patch fits into maps with a number of pixels on each side that are powers of two, which maximizes the efficiency of FFTs. For this resolution, this yields  $4096 \times 8192$  maps.

After maps have been made for each observation in the 2019 and 2020 winter observing seasons, we sum these maps into a full-depth “coadd” with a small number of data cuts. In addition to full-depth coadds, we also construct 500 “sign-flip” noise realizations by summing the observations with a random half of them multiplied by  $-1$ , canceling the signal but leaving the statistical properties of the noise unchanged. We use these to quantify the noise covariance, and to directly add to simulated signal maps to produce highly realistic simulations of the full dataset. We estimate that the noise realizations are correlated with each other at the percent level due to the finite number of observations in the dataset, but this is below the level which can significantly impact this analysis.

Next, we apply an isotropic anti-aliasing filter to the  $0.5625'$  maps and rebin them to the final  $2.25'$  resolution,  $1024 \times 2048$  pixel analysis maps. We note that the reason for not binning the timestreams directly into the final  $2.25'$  resolution is exactly to be able to apply this anti-aliasing filter, as otherwise nothing would prevent aliasing in the scan-perpendicular direction. The final Nyquist frequency for these maps corresponds to  $\ell = 4800$ , high enough to accommodate the  $\ell_{\text{max}} = 4000$  of our analysis while at the same time minimizing the pixel count and hence the computational cost of MUSE.

Finally, we apply the pipeline masks described in Sect. III C. These processed and masked  $1024 \times 2048$ -pixel maps of  $Q$  and  $U$  Stokes polarization at 95, 150, and 220 GHz are the basic data,  $d$ , which is input to our analysis pipeline.

### III. METHODOLOGY: BAYESIAN INFERENCE WITH MUSE

#### A. Bandpower and systematics inference

We perform a map-level simultaneous Bayesian inference of the gravitational lensing potential bandpowers, the unlensed CMB  $EE$  bandpowers and systematic parameters. This procedure is described in Millea & Seljak [16] and Millea [17], and summarized below with a few additions specific to this work. The central piece of the analysis is the joint posterior probability function,

$$\mathcal{P}(f, \phi, \theta | d) \quad (1)$$

where  $f$  refers to the unlensed CMB maps (here just polarization),  $\phi$  to the gravitational lensing potential,  $\theta$  to the bandpower and systematics parameter vector, and  $d$  to the maps described in the previous section. The exact details of our modeling and construction of this function are given in the next section. Here, we first discuss how we use it to perform inference.

The goal of our analysis is to marginalize over the “latent”  $f$  and  $\phi$  variables to produce the marginal posterior on just the parameters of interest,  $\theta$ :

$$\mathcal{P}(\theta | d) = \int df d\phi \mathcal{P}(f, \phi, \theta | d) \quad (2)$$

The constraints on  $\theta$  described by this function are formally optimal, satisfying our goal of an optimal analysis. This optimality is achieved because, implicitly, the marginalization sources information from moments of the data at all orders, unlike e.g. quadratic estimators, which only use second-order moments. In particular, this means our results use information from all combinations of the data such as  $EE$ ,  $EB$ ,  $BB$ ,  $EEE$ ,  $EEB$ , etc...

Various techniques exist for performing the difficult high-dimensional integration in Eqn. (2). A popular but computationally expensive choice is to use Hamiltonian Monte Carlo to yield the exact answer, up to sampling errors. MUSE instead performs an approximate but much faster marginalization, which yields a multi-variate Gaussian approximation to the marginal posterior,

$$-2 \log \mathcal{P}(\theta | d) \approx (\theta - \hat{\theta})^\dagger (\Sigma_{\text{MUSE}})^{-1} (\theta - \hat{\theta}) + C \quad (3)$$

for posterior mean  $\hat{\theta}$ , covariance  $\Sigma_{\text{MUSE}}$ , and unimportant constant,  $C$ . The Gaussian approximation is well-motivated due to the central limit theorem as long as many modes of  $f$  and  $\phi$  contribute to the constraint on each  $\theta$ , which is well-satisfied here.

The marginal posterior mean,  $\hat{\theta}$ , is given by solving the following equation for  $\theta$ :

$$s_i^{\text{MAP}}(\theta, d) = \left\langle s_i^{\text{MAP}}(\theta, d') \right\rangle_{d' \sim \mathcal{P}(d' | \theta)}, \quad (4)$$

where  $s_i^{\text{MAP}}(\theta, d)$  is the gradient of the joint posterior with respect to parameter  $\theta_i$ , evaluated at the *maximum a posteriori* (“MAP”) estimate of  $z \equiv (f, \phi)$ , i.e.

$$s_i^{\text{MAP}}(\theta, d) = \frac{d}{d\theta_i} \log \mathcal{P}(\hat{z}(d, \theta), \theta | d) \Big|_{\theta} \quad (5)$$

with

$$\hat{z}(d, \theta) = \arg \max_z \mathcal{P}(z, \theta | d). \quad (6)$$

One way to understand this definition is to consider  $s_i^{\text{MAP}}$  as a data compression, which, for a given value of  $\theta$ , takes as input our  $\sim$ million-dimensional data vector and returns another vector of the same dimensionality as  $\theta$ , which in our case is around a hundred bandpower and systematics parameters. The MUSE algorithm then proceeds by picking a starting guess for  $\theta$  and compressing the data according to this, generating a suite of data simulations given the same  $\theta$  and compressing them as well, and finally iteratively searching for the value of  $\theta$  which make the data and simulations most similar, as compared via this data compression (Eqn. 4). Intuitively, this yields the parameters  $\theta$  which are most likely to have actually generated the real data.

More formally, Millea & Seljak [16] show that following this procedure leads to an asymptotically unbiased estimate of parameters,  $\hat{\theta}$ , for any Gaussian or non-Gaussian latent space,  $z$ . For a Gaussian latent space, the data compression is a sufficient statistic, and therefore  $\hat{\theta}$  gives the exact marginal posterior mean. For non-Gaussian latent spaces such as the one here, the data compression can be considered “lossy”, but this leads only to a possible loss of optimality, with the asymptotic bias remaining zero. For a lensing analysis similar to the one performed here, Millea & Seljak [16] bounds any loss of optimality to less than 10% of the statistical uncertainty.

The covariance,  $\Sigma_{\text{MUSE}}$ , is computed from:

$$\Sigma_{\text{MUSE}}^{-1} = H^\dagger J^{-1} H + \Sigma_{\text{prior}}^{-1} \quad (7)$$

where

$$J_{ij} = \text{cov} \left( s_i^{\text{MAP}}(\hat{\theta}, d), s_j^{\text{MAP}}(\hat{\theta}, d) \right)_{d \sim \mathcal{P}(d | \hat{\theta})} \quad (8)$$

$$H_{ij} = \frac{d}{d\theta_j} \left[ \left\langle s_i^{\text{MAP}}(\hat{\theta}, d) \right\rangle_{d \sim \mathcal{P}(d | \theta)} \right] \Big|_{\theta = \hat{\theta}}, \quad (9)$$

and  $\Sigma_{\text{prior}}$  is the covariance of any assumed prior. Note that this is not a pure Monte Carlo covariance of the MUSE estimate, which would be prohibitive computationally, but rather a cheaper Monte Carlo covariance of just  $s^{\text{MAP}}$ , with analytic propagation to the resulting covariance in parameters. All derivatives in practice are computed with automatic differentiation, and Monte Carlo error stemming from  $J$  is reduced using a method described in App. C.

One aforementioned aspect of MUSE which is key for this analysis is that it is unbiased even for “lossy” data

compressions. This holds as long as the simulation-generating distribution,  $\mathcal{P}(d | \theta)$ , is an accurate representation of the real data. There is not actually *any* requirement on the form of  $s_i^{\text{MAP}}$  for MUSE to remain asymptotically unbiased and for the covariance expressions to be valid. The default choice for  $s_i^{\text{MAP}}$  given in Eqn. (5) is simply the one which makes MUSE exact for Gaussian problems.

Here, we keep this choice largely intact, but modify the joint posterior appearing in Eqns. (5) and (6),  $\mathcal{P}(f, \phi, \theta | d)$ , to be slightly different from the likelihood function used to generate the simulations,  $\mathcal{P}(d | \theta)$ . Normally, the two distributions would be related by:

$$\mathcal{P}(d | \theta) = \frac{\mathcal{P}(d)}{\mathcal{P}(\theta)} \int df d\phi \mathcal{P}(f, \phi, \theta | d). \quad (10)$$

Instead, here we have one statistical model denoted as the *posterior model*,  $\mathcal{P}(f, \phi, \theta | d)$ , and another called the *simulation model*,  $\mathcal{P}(d | \theta)$ .

In the analysis, the posterior model is used in computationally costly high-dimensional maximizations, while the simulation model is only ever used to generate a few hundred samples from  $\mathcal{P}(d | \theta)$ . This hints at the motivation for this choice of separate models, namely to allow defining a simulation model which fully meets our accuracy requirements, while keeping a simpler “surrogate” posterior model which is much faster to compute and makes the analysis more tractable. With an accurate simulation model, the MUSE analysis will remain unbiased, and we find little loss of optimality due to approximations in the posterior model.

We note that our analysis can also be considered a form of “simulation-based inference” (SBI). This term is used to describe a class of algorithms which leverage the ability to generate samples from  $\mathcal{P}(d | \theta)$  directly into the ability to perform inference, and has been of increasing interest in various domains in cosmology [see e.g. 38–44]. In typical high-dimensional SBI applications to-date, an important first step is to obtain a suitable data compression function to reduce the dimensionality of the problem, often by training a neural network on the generated samples themselves [45]. The compressed data is then used to infer parameters with something like Approximate Bayesian Computation (ABC) [46, 47], or by training a neural network surrogate model for the likelihood which can be inverted to obtain the posterior [see 48].

MUSE is similar in that it also involves a data compression function, but rather than learned with a neural network, the generic and semi-analytic  $s_i^{\text{MAP}}$  is used instead. Furthermore, rather than using something like ABC or training surrogate models for inference, we exploit our access to gradients and Jacobians through the posterior and simulation-generating function to derive semi-analytic expressions for the posterior mean and covariance. In this way, MUSE can be considered one of the possible augmentations of SBI suggested by [48] which use additional tractable quantities from the simulator.

## B. Cosmological parameter inference

The output of MUSE is a parameter mean,  $\hat{\theta}$ , and its covariance,  $\Sigma_{\text{MUSE}}$ . The inferred parameters for this analysis include the  $\phi\phi$  and unlensed  $EE$  spectra binned into bandpowers, and systematics parameters. The systematics-marginalized CMB bandpowers and covariance can be obtained by simply dropping the systematics from the estimate and dropping the corresponding block from the covariance (which is equivalent to marginalizing them). That is, part of the output of MUSE is automatically a set of ‘‘CMB-compressed’’ bandpowers such as those which have been derived for other experiments [49–51]. We can then estimate cosmological parameters,  $\gamma$ , as usual, by forming a cosmological parameter likelihood,

$$-2\log \mathcal{P}(\hat{\theta} | \gamma) = \left[ \theta(\gamma) - \hat{\theta} \right]^\dagger \Sigma_{\text{MUSE}}^{-1} \left[ \theta(\gamma) - \hat{\theta} \right], \quad (11)$$

Simulation model :

$$f \sim \mathcal{N}(0, \mathbb{C}_f^{\text{curv sky}}(A_b^{\text{EE}})) \quad (12)$$

$$\phi \sim \mathcal{N}(0, \mathbb{C}_\phi^{\text{curv sky}}(A_b^{\phi\phi})) \quad (13)$$

$$n^\nu \sim \{n_{\text{signflips}}^\nu\} \quad (14)$$

$$d^{\nu,i} = \mathbb{M}_{\text{fourier}} \cdot \mathbb{M}_{\text{trough}} \cdot \mathbb{M}_{\text{pix}} \cdot (\text{PWF} \cdot \text{TF}^\nu \cdot \mathbb{R}(\psi_{\text{pol}}^\nu) \cdot A_{\text{cal}}^{\nu,i} \cdot \mathbb{B}(\beta_n, \beta_{\text{pol}}^\nu) \cdot \mathbb{G} \cdot \mathbb{P} \cdot \mathbb{L}(\phi) \cdot f + \epsilon_{\text{Q}}^{\nu,i} \cdot t_{\text{Q}}^\nu + \epsilon_{\text{U}}^{\nu,i} \cdot t_{\text{U}}^\nu + n^\nu) \quad (15)$$

Posterior model :

$$f \sim \mathcal{N}(0, \mathbb{C}_f^{\text{flat sky}}(A_b^{\text{EE}})) \quad (16)$$

$$\phi \sim \mathcal{N}(0, \mathbb{C}_\phi^{\text{flat sky}}(A_b^{\phi\phi})) \quad (17)$$

$$\mu^{\nu,i} = \mathbb{M}_{\text{fourier}} \cdot \mathbb{M}_{\text{trough}} \cdot \mathbb{M}_{\text{pix}} \cdot (\text{PWF} \cdot \text{TF}^\nu \cdot \mathbb{R}(\psi_{\text{pol}}^\nu) \cdot A_{\text{cal}}^{\nu,i} \cdot \mathbb{B}(\beta_n, \beta_{\text{pol}}^\nu) \cdot \mathbb{L}(\phi) \cdot f + \epsilon_{\text{Q}}^{\nu,i} \cdot t_{\text{Q}}^\nu + \epsilon_{\text{U}}^{\nu,i} \cdot t_{\text{U}}^\nu) \quad (18)$$

$$d^\nu \sim \mathcal{N}(\mu^\nu, \mathbb{C}_n^\nu) \quad (19)$$

$$\Rightarrow -2\log \mathcal{P}(f, \phi, \theta | d) = f^\dagger \mathbb{C}_f^{-1} f + \phi^\dagger \mathbb{C}_\phi^{-1} \phi + \sum_\nu (d^\nu - \mu^\nu)^\dagger (\mathbb{C}_n^\nu)^{-1} (d^\nu - \mu^\nu) - 2\log \mathcal{P}(\phi) \quad (20)$$

$$\text{where } -2\log \mathcal{P}(\phi) = \|\mathbb{M}_{\text{pix}} \nabla^2 \phi\|^2 / 10^{-8} \quad (21)$$

Reading these models top-to-bottom describes how to produce a random sample from the model, with ‘‘ $\sim$ ’’ used to represent generation of a random sample from the distribution on the RHS and  $\mathcal{N}(\mu, \Sigma)$  a multi-variate Gaussian with mean and covariance  $\mu$  and  $\Sigma$ .

where  $\theta(\gamma)$  is our theory  $\phi\phi$  and unlensed  $EE$  spectra binned according to our specified bandpower window functions (App. A1). For our main results we sample  $\mathcal{P}(\hat{\theta} | \gamma)$  with `Cobaya`<sup>2</sup> [52] and use `CAMB`<sup>3</sup> [53] to compute theory spectra. To get the accurate theory CMB spectra at large  $\ell$ , we set `lens_potential_accuracy=4` and `lens_margin=1250` in `CAMB`. We also set `halofit_version=mead2020` for the non-linear correction in `CAMB`. The convergence criterion of MCMC is set to  $R - 1 < 0.01$  of the Gelman-Rubin statistic.

## C. Map-level model summary

The posterior and simulation models are specified in Eqns. (12-19), with a summary given below and full details given in Appendix A.

To summarize, in the simulation model, we generate curved-sky unlensed CMB maps ( $f$ ) and lensing potential maps ( $\phi$ ) from appropriate covariance operators ( $\mathbb{C}$ ), which are parameterized by the bandpower amplitudes which we will infer ( $A_b^{\text{EE}}$  and  $A_b^{\phi\phi}$ ). We apply the lensing operation ( $\mathbb{L}$ ) to produce lensed CMB maps and project to the flat-sky Lambert pixelization of the data ( $\mathbb{P}$ ). We apply an additional small deflection ( $\mathbb{G}$ ) to account for relativistic aberration due to our proper motion relative to the CMB rest frame. We apply the beam convolu-

<sup>2</sup> <https://cobaya.readthedocs.io/en/latest/>

<sup>3</sup> <https://camb.info/>

tion operator ( $\mathbb{B}(\beta_n, \beta_{\text{pol}}^\nu)$ ), which is parameterized by a set of beam eigenmode amplitudes ( $\beta_n$ ) and the polarization fraction of the beam sidelobes at each frequency ( $\beta_{\text{pol}}^\nu$ ). We rotate the global polarization angle of the maps ( $\mathbb{R}(\psi_{\text{pol}})$ ) by an angle  $\psi_{\text{pol}}$  to model systematic errors in the overall polarization angle of our detectors. In each subfield,  $i$ , we scale the maps by  $A_{\text{cal}}^{\nu,i}$ , to model the uncertain absolute calibration of the instrument. We apply the model transfer function (TF), and apply the pixel window function (PWF). We add in temperature to polarization monopole leakage  $Q$  and  $U$  templates determined directly from our temperature observations ( $t_Q$  and  $t_U$ ) with subfield-dependent amplitudes ( $\epsilon_Q^i$  and  $\epsilon_U^i$ ), as well as add in a sample of the noise ( $n^\nu$ ) from one of the sign-flip noise realizations. Finally, we apply our chosen pixel, “trough”, and Fourier masks (see App. A 9) to produce a simulated data vector,  $d^\nu$ .

The posterior model is nearly the same, with the only differences being that the covariances,  $\mathbb{C}$ , assume flat-sky statistics, the noise is generated from a covariance and is unmasked, and aberration is ignored. For the posterior model, it is possible to calculate the posterior probability function, given in Eqn. (20). Here, we also choose to include a “super-sample” prior,  $\mathcal{P}(\phi)$ , in order to reduce the lensing mean-field (before its impact is automatically removed entirely in the process of obtaining the MUSE estimate).

The final inferred parameter vector is

$$\theta = \{A_b^{\phi,\phi}, A_b^{\text{EE}}, \beta_n, \beta_{\text{pol}}^\nu, A_{\text{cal}}^{\nu,i}, \psi_{\text{pol}}, \epsilon_Q^i, \epsilon_U^i\}, \quad (22)$$

containing 77 bandpower parameters and 47 systematics parameters. The MUSE estimate is a joint best-fit and a covariance between all of them.

We note that the approximation of leaving the noise unmasked, while seemingly benign, is actually critical to obtaining a sufficiently fast posterior function since it avoids an additional large matrix inversion, while having little impact on the final results (see also Sec. 3.7 of [26]). We also note that the ability to directly use the sign-flip noise realizations in the simulation model is extremely advantageous. Without this, a Bayesian analysis would depend on an accurate statistical model of all moments of the noise (power spectrum, bispectrum, etc...), and would require both sampling and evaluating the likelihood of a noise map given such a model. Here, noise modeling is essentially a non-issue since MUSE requires only sampling the noise, and the sign-flip realizations provide these samples and empirically match the true noise distribution.

There are a few effects which we considered but ultimately chose not to include the model. The first effect is polarized foregrounds. The dominant expected foreground comes from polarized point sources. For the same point source flux cut used here, Dutcher et al. [12] construct a set of priors on the foreground amplitudes based on measurements of temperature foregrounds scaled to polarization with very conservative priors on the polar-

ization fraction. Based on this, we find that the maximum possible contamination has power two orders of magnitude lower than our noise spectra at all frequencies. That the foreground contribution is small is also consistent with [54], in which it is found that even for much deeper observations than ours, the bias to lensing from polarized foregrounds is fairly marginal. As mentioned, this finding that the foreground contribution is negligible in polarization for this analysis was key in our decision to focus only on polarization in this paper.

The second effect we ignore is the bias due to spatial correlation between the lensing potential and the location of our masked point source holes [55]. Using the AGORA simulations [56], which include a model for these spatial correlations, we find that for our point source masking threshold, the maximum bias to  $C_L^{\phi\phi}$  would be  $\sim 0.5\%$ . However, Lembo et al. [55] also show that this maximum bias is highly suppressed in an optimally filtered  $\phi$  map such as the MAP  $\phi$  maps which enter our analysis. We thus conclude this effect is also negligible here.

#### IV. VALIDATION OF PIPELINE AND DATA

Before computing our main results, we conducted a set of validation tests which we present in this section. To mitigate and quantify confirmation bias, we performed these tests “blind”, that is, without comparison to data from other experiments or to theory models derived from these experiments. After unblinding, we allowed ourselves to make further changes to the pipeline before producing our final baseline results, but committed to reporting in this paper the parameters that we initially found before any such changes. This motivated and focused the validation work, and allows us to quantify any level of confirmation bias; the initial unblinded results could not be impacted by confirmation bias, only the shifts after unblinding could be. A full discussion of changes made after unblinding is at the end of this section (Sec. IV C).

Our chosen set of rules for blinding were: 1) do not plot our bandpowers on the same plot as bandpowers or theory curves from any other previous analysis or experiment (including from SPT itself), or show bandpowers relative to any theory curves, 2) if we do plot bandpowers, remove all tick marks and tick labels and do not look at numerical values of bandpowers, 3) do not look at absolute cosmological parameters derived from bandpowers (internal null parameter differences, e.g. the parameters from 95 GHz minus parameters from 150 GHz were allowed).

This procedure left us blind to all but very gross departures from  $\Lambda$ CDM and did not allow us to know if bandpowers were consistent with any other experiment before unblinding, while not being overly burdensome.

We now describe the set of validation tests of our pipeline and data which we performed while blinded.



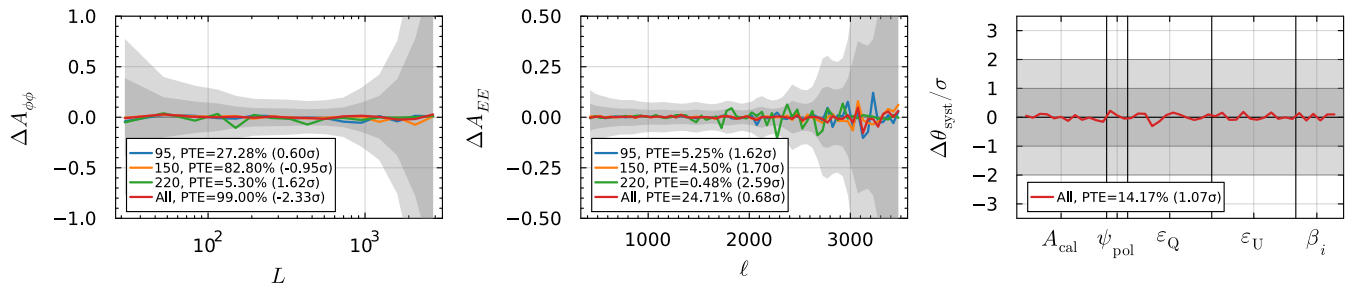


FIG. 1. Verification that our pipeline recovers unbiased bandpower and systematics parameters using 100 mock simulations. The colored lines are the average MUSE parameter estimates over simulations after subtracting the input simulation truth (in the final panel we also divide by the errorbars for one simulation due to the otherwise large dynamic range). The first two panels show  $\phi\phi$  and unlensed  $EE$  bandpower amplitude parameters (and also consider single-frequency runs) and the last panel groups together all the systematics parameters (see Sec. III C and App. A for description of the systematics considered). The PTEs of the colored lines relative to the expected scatter and 2-tailed numbers of  $\sigma$  are shown in the legends. To get the PTEs, we Monte Carlo sample the expected distribution of  $\chi^2$  values, accounting both for the look-elsewhere effect of 4 tests (3 single-frequency runs and 1 all-frequency run) and the fact that we compute the expected scatter itself from 100 mocks. For reference, gray bands are the all-frequency 1 and 2  $\sigma$  errors (note that this is not the expected scatter of any of the lines, rather is a way to judge the size of any potential bias).

## A. Pipeline tests

### 1. Simulated Data

To validate our full bandpower and cosmological parameter estimation pipeline, we use an ensemble of highly realistic simulations of the full dataset which we call “mock simulations”. The signal part of the mock simulations consists of a lensed CMB realization generated assuming a fiducial Planck  $\Lambda$ CDM cosmology and a Gaussian foreground model based on AGORA simulations [56] that is calibrated against data at very high- $\ell$ . These maps are processed through a procedure called mock observation, which generates simulated timestreams and simulates the observation and processing of the timestreams into maps at each frequency, using the identical code which processes the real data.<sup>4</sup> The noise part of the mock simulations comes from sign-flipped noise realizations.

### 2. End-to-end pipeline tests

To check that our pipeline is unbiased, we start with an end-to-end test of the pipeline on a set of 100 mock simulations. First, we compute MUSE estimates of bandpowers and systematic parameters on each mock simulation using all three 95+150+220 GHz polarization maps (hereafter, “all-frequency”). Then we repeat the same procedures on single-frequency polarization maps with

the corresponding systematic parameters fixed to the all-frequency results (this is how we handle systematics in the single-frequency runs on real data as well).

In the left two panels of Fig. 1, we show the mean bandpowers over 100 mock simulations compared to the input truth. The PTE of the difference between the mean and input truth and its conversion to a number of  $\sigma$  is given in the legend.<sup>5</sup> No bias is detected at 3  $\sigma$  significance from the 100 mock simulations. The scatter of the mean bandpowers is at the level of 10% of the statistical uncertainty, as expected from a test with 100 realizations. In the right panel of Fig. 1, we also show the mean systematics parameters over 100 all-frequency mock simulations, here scaled by their 1  $\sigma$  uncertainty. These are also recovered with no significant bias. Single-frequency runs do not vary systematics parameters, hence those estimates are not shown.

We next compare the empirical covariance of the 100 all-frequency estimates with the MUSE covariance given in Eqn. (7). The left column of Fig. 2 compares the square root of the diagonal covariance entries. We see good consistency between the two except for  $A_{\text{cal}}$ . This is because the mock simulations were made without varying systematic parameters within their prior, while the MUSE covariance takes the prior uncertainty of all parameters into account. This difference is particularly evident for prior-dominated parameters such as  $A_{\text{cal}}$ .

In the right column of Fig. 2, we compare the first and second off-diagonal elements in the bandpower covariance matrices, this time fixing the systematics in  $\Sigma_{\text{MUSE}}$

<sup>4</sup> To reduce computational cost and because many observations have identical filtering, we use only 5% of the total observations in this step.

<sup>5</sup> All PTEs we quote are corrected for look-elsewhere in the case that there are multiple tests. We will specify if the conversion to number of  $\sigma$  is 2-tailed, in which case negative values correspond to a PTE better than the mean expectation; otherwise the quoted  $\sigma$  number is 1-tailed.

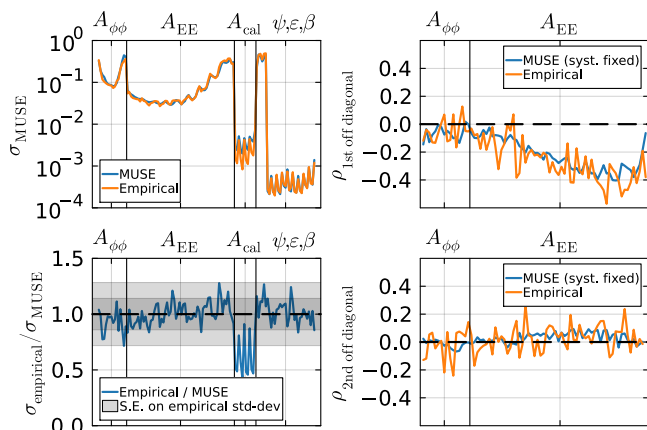


FIG. 2. Verification of the accuracy of the all-frequency MUSE covariance,  $\Sigma_{\text{MUSE}}$ . (*Top left*) Comparison of the square root of the diagonal elements between the MUSE covariance,  $\sigma_{\text{MUSE}} \equiv \sqrt{\text{diag}(\Sigma_{\text{MUSE}})}$ , and the empirical covariance from running the estimate on mock simulations. (*Bottom left*) The ratio of the two curves in the upper panel. The shaded band shows the 1 and 2  $\sigma$  standard error on  $\sigma_{\text{empirical}}$  from the 100 simulations used. These are expected to match for fully likelihood-dominated parameters, and do so for all parameters except the partially prior-dominated  $A_{\text{cal}}$  parameters. (*Top and bottom right*) Comparisons of the first and second off-diagonal elements of the MUSE and empirical correlation matrices. In this case, systematics are assumed fixed in the MUSE covariance, since they are fixed in the simulations.

for a more like-to-like comparison. We see good agreement between the empirical covariance and the MUSE covariance. The anti-correlation in the first off-diagonal elements is caused by projection effects, masking, delensing and the transfer function, where the projection effects explain about half of the anti-correlation. Here the projection effects refers to our use of the flat-sky approximation in the posterior model. This effect has been found in the previous SPT-3G analysis [27].

In Fig. 3, we show the MUSE covariance with its diagonal entries rescaled to unity (i.e. the correlation matrix between all bandpowers and systematics parameters). We find no significant correlation between lensing bandpowers and any other parameter, suggesting a robust determination of the lensing spectrum with little uncertainty from systematic effects. The correlations within the unlensed  $EE$  block are due to masking, delensing, projection and correlation with systematics parameters. The  $EE$  bandpowers are significantly correlated with both calibration uncertainty and the sidelobe polarization fraction, both of which add appreciable systematic uncertainty to our final  $EE$  estimates.

We next obtain  $\Lambda$ CDM cosmological parameter estimates from the 100 all-frequency mock simulations. Unlike the main results which use *Cobaya* and *CAMB* to perform parameter estimation, here, for speed, we have trained and verified the accuracy of an emulator us-

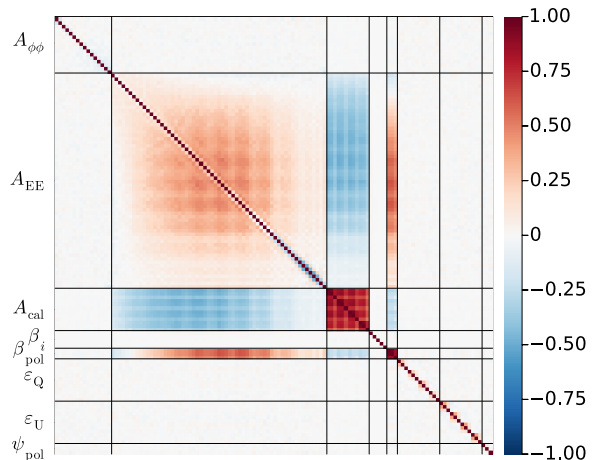


FIG. 3. The posterior correlation matrix of  $\phi\phi$  bandpowers, unlensed  $EE$  bandpowers, and systematics parameters, given all-frequency data.

ing *Capse.jl* [57], and sample using Hamiltonian Monte Carlo.

The one-dimensional cosmological parameter posterior distributions inferred from each mock simulation and the product over all of these posteriors is shown in Fig. 4. It is expected that the product over all posteriors should cover the input truth, which is found to be the case. The non-detection of any bias with 100 sims suggests that, at 68% confidence, we have bounded any possible pipeline-induced systematic error to be less than  $\lesssim 0.1\sigma$  for each cosmological parameter.

To summarize, we have validated the pipeline using 100 highly realistic mock simulations. With 100 simulations, we are able to detect biases at the level of 10% of the statistical uncertainty. We do not detect any such biases to within  $3\sigma$  of the expected scatter on the mean in either our estimates of  $\phi\phi$  lensing spectra, unlensed  $EE$  bandpowers, systematics, or in the cosmological parameters derived from them. We find good agreement of the bandpower covariance matrices between the one derived from mock simulations and the one calculated in MUSE using Eqn. (7).

## B. Data tests

Next, we turn to tests of the data itself. These include intra-frequency null tests using different splits of the data, inter-frequency comparison of  $\phi\phi$  and unlensed  $EE$  bandpower results, pixel-level  $\chi^2$  tests of our model fits, and internal comparison of our systematics estimates with those derived with more traditional methods. We report the results of these tests in this section.

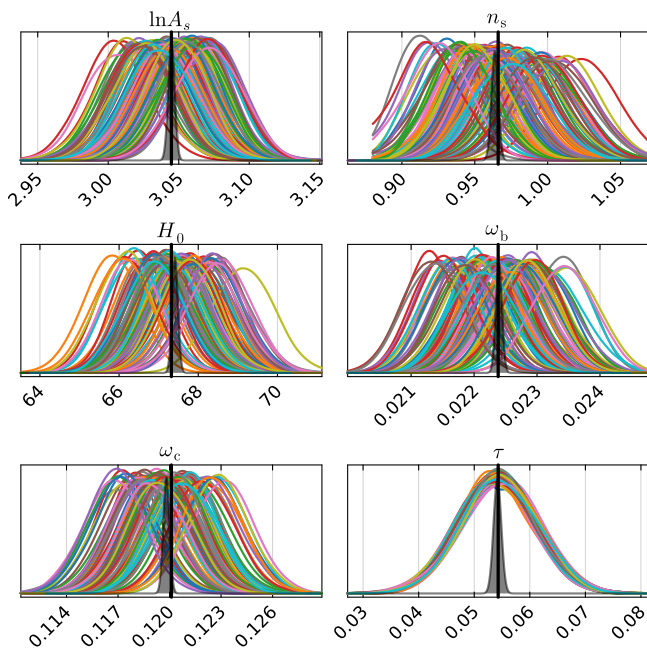


FIG. 4. Verification that our pipeline recovers unbiased cosmological parameters using 100 mock simulations. We infer  $\Lambda$ CDM parameters from each simulated all-frequency dataset, and plot the one-dimensional marginalized posteriors above as different colored lines for each simulation. The shaded gray region in each panel is the product over all posteriors, which should cover the simulation truth, denoted as vertical black lines. Note that the smaller scatter for  $\tau$  is because this parameter is mainly constrained by our prior. Across the other 5 parameters, the shaded region is consistent with the input truth with significance equivalent to  $1.1\sigma$ , suggesting no detection of any bias at the level afforded by our 100 simulations.

### 1. Intra-frequency null tests

Null tests involve splitting the full dataset into two parts and differencing them such that the signal cancels, then checking whether the remaining portion is consistent with the expected statistical properties of noise. Finding something outside of this expectation could indicate the presence of a systematic, which might then also be present in the full dataset coadd at some level. By canceling the signal and thus the sample variance from the signal, these tests can probe systematics to levels well below even what could impact the final results, which are themselves limited by sample variance. Since temperature data does enter our analysis via leakage templates, we also present null TE spectra to more thoroughly vet the full temperature and polarization dataset. Further details of the null tests will be presented in a future work.

We consider 6 types of data splits: sun, moon, azimuth, year, scan, and wafer. The sun and moon tests are designed to check whether a portion of the dataset is significantly contaminated by the radiation from the sun or moon. In these tests, one part of the dataset comprises the data we took when the sun or moon was above

the horizon, and the other part below the horizon. The azimuth test is designed to check whether a portion of the dataset is contaminated by thermal radiation from a building near the telescope. One part of the dataset comprises the data we took when the telescope’s azimuth angle was within 90 degrees of that of the building, and the other part azimuth angles outside that range. The year test is designed to check whether the data we took during the 2019 season are significantly different from the data we took in 2020. The scan test is designed to check whether the data we took when the telescope was scanning in one direction are significantly different from the data we took when the scanning was in the opposite direction. Lastly, the wafer test is designed to check whether the data from one half of the full detector array are significantly different from the data from the other half.

For the scan and wafer tests, we do not expect the null spectra to perfectly cancel the signal due to details of our mapmaking pipeline. For the wafer test, this is because we filter time-ordered data from different detector wafers differently. For the scan test, this is mainly because we do not correct for each individual detector’s time constant effect when converting its time-ordered data into map pixel values. The time constant effects refers to the fact that each detector does not respond to changing radiation from the sky instantaneously. As a result, a map made from left-going scans contain a version of the CMB shifted to the left, and vice-versa for right-going scans. For these two tests, we calculate the expected null spectra through simulations of the mapmaking pipeline and subtract it. This is computed assuming a fiducial cosmology, but because these expected null spectra are very small compared to the signal, it does not significantly imprint model-dependence to these null tests or degrade their usefulness.

For each null test type, we divide each of the two halves of the data into 25 smaller parts and call them bundles. We create 25 null maps by differencing bundles from each half one-by-one, then compute all 300 cross-spectra from these 25 null maps and bin them into the same bandpowers as the analysis. The average of these is what we call a null spectrum, and the standard error is our estimate of the expected scatter of the null spectrum.

The maps used for these tests are curved-sky projections of the same timestreams which enter the main analysis, with spectra computed with `PolSpice` [58]. For convenience, most of the tests have been computed using slightly different masks than the ones used to produce our main results. These tests, which we do not expect to depend significantly on the masking choice, use a pixel mask which includes slightly more sky, and a harmonic space mask which zeros all  $a_{\ell m}$  for  $\ell \in [500, 680]$  and  $m \in [350, 425]$ . This harmonic mask removes contamination from a narrowband signal around 1.1 Hz, which is similar, but not exactly the same as the “trough” mask used to remove the contamination in the MUSE pipeline. As the 1.1 Hz signal presents itself mainly in detector

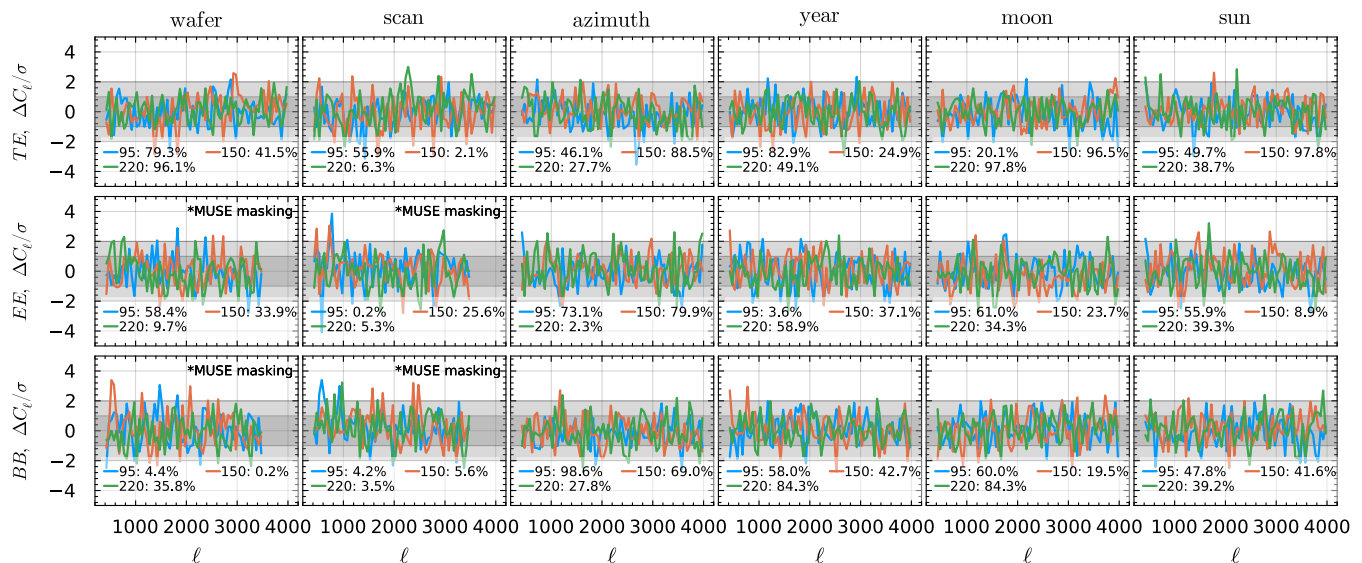


FIG. 5. Spectra of null maps computed from various data splits divided by the expected  $1\sigma$  scatter due to noise. Detector wafer and scan direction tests also have had a small expected contribution from signal leakage subtracted, and in  $EE$  and  $BB$  have been computed with the exact pixel and Fourier space masking used in the main results pipeline (indicated as “MUSE masking”; the other tests are not expected to depend significantly on masking). Legends give the PTE for each test without correcting for the look-elsewhere effect. The  $2\sigma$  threshold when considering the look-elsewhere effect is a PTE  $> 0.07\%$  on any individual test, which is passed in all cases.

wafer and scan direction tests, we have recomputed these with the exact MUSE pixel and Fourier masking to verify the MUSE masking is sufficient to remove it.

After masking, computing null spectra, and subtracting expected contributions as necessary, we calculate a  $\chi^2$  with respect to the expected scatter, assuming no correlation between neighboring bandpowers. To compute a threshold probability-to-exceed (PTE) for the observed  $\chi^2$ , we Monte Carlo sample the expected distribution of  $\chi^2$  values, accounting both for the look-elsewhere effect of 54 tests and the fact that we compute the expected scatter itself from a finite 300 simulations. We consider the test passed if the PTE is  $> 5\%$  (corresponding to a  $< 2\sigma$  deviation) with respect to this Monte Carlo distribution. This corresponds to a PTE  $> 0.07\%$  on any single individual test amongst our suite (this is roughly  $5\%/54$  with an additional correction for the slightly noisy expected scatter).

A summary of all tests and their PTEs is given in Fig. 5. All tests pass at our threshold of  $0.07\%$ , and no obvious features are otherwise visible.

## 2. Inter-frequency agreement

We next check for inter-frequency agreement of the  $\phi\phi$  and unlensed  $EE$  bandpowers inferred from single-frequency maps. We follow the same procedure for inference as described in Sect. III, except that the data,  $d$ , includes only a single-frequency map and the systematic parameters are fixed to the value derived from the

all-frequency run; only the bandpowers  $A_b^{EE}$  and  $A_b^{\phi\phi}$  are free parameters. To quantify the expected scatter, we run this identical procedure on 100 mock simulations.

In Fig. 6, we show the ratios of bandpowers inferred from maps at different frequencies relative to the expected scatter. The blue lines show our result prior to unblinding, which showed good overall agreement. The legends give 2-tailed effective number of  $\sigma$  for the  $\chi^2$  values, and all are consistent with statistical fluctuations to within  $\pm 2\sigma$ . We also note that with model changes after unblinding, shown as orange lines, agreement improved slightly, particularly in the comparison between 95 GHz and 150 GHz.

## 3. Goodness-of-fit

We next compare the pixel-level goodness-of-fit of our model to the observed data. To do so, we define  $\chi^2 = -2 \log \mathcal{P}(\hat{z}(\hat{\theta}, d), \hat{\theta} | d)$ , that is, minus twice the posterior probability function evaluated at the MUSE estimate of parameters,  $\hat{\theta}$ , and the MAP given that  $\hat{\theta}$ . Because of our decoupling of the posterior and simulation models, we cannot compute the expected  $\chi^2$  simply from counting degrees of freedom. Instead, we use the distribution of  $\chi^2$  obtained from mock simulations to quantify the expected distribution.

Tab. I gives the data  $\chi^2$  values observed for single-frequency and all-frequency runs before unblinding, as compared to this expectation. We see that the values are on the order of  $10^8$ , despite that the number of pixels

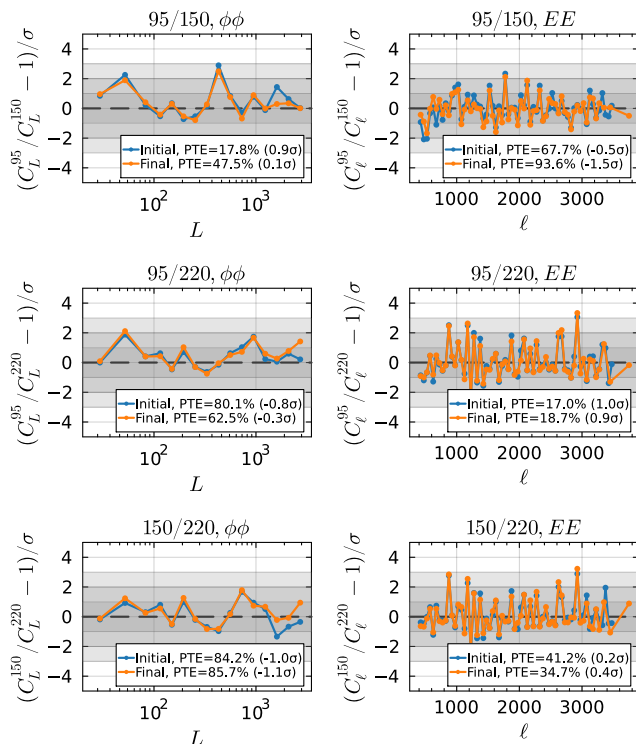


FIG. 6. Tests of the inter-frequency bandpower agreement of our data. Each panel shows the ratio of  $\phi\phi$  or unlensed  $EE$  bandpowers from single-frequency MUSE runs at different frequencies, normalized by the  $1\sigma$  expected scatter as computed from simulations. Legends show the  $\chi^2$  and 2-tailed effective number of  $\sigma$  fluctuation of this  $\chi^2$  given the number of degrees of freedom. The blue line shows the result using our model before unblinding, which yielded acceptable  $\chi^2$  in all cases. The orange line shows the result after the post-unblinding changes to our model, which notably slightly improves the agreement between 95 GHz and 150 GHz.

in  $d$  is close to  $10^7$  for the all-frequency case, confirming the fact that the expectation is not given simply by counting degrees of freedom. That the single-frequency expected  $\chi^2$  are larger than the all-frequencies case is due to the fixing of systematic parameters in those cases. Regardless, with the expected distribution mapped out by simulations, we find that our observed values in all cases are consistent with a statistical fluctuation within  $3\sigma$ . This suggests our model is sufficient to describe the observed data.

#### 4. Agreement of systematics with alternative estimates

Our analysis infers systematic parameters jointly with bandpowers by incorporating systematics into a map-level model for the data. In this section, we compare our systematics estimates to independent determinations from alternate methods which have traditionally been used to estimate systematics in CMB analyses.

	$ \chi_{\text{obs}}^2/10^6 $	$ \chi_{\text{expected}}^2/10^6 $	PTE	$n_\sigma$
95+150+220	93.33	$93.50 \pm 0.12$	4.0%	2.1
95	108.21	$108.28 \pm 0.18$	37.0%	0.9
150	109.32	$109.51 \pm 0.19$	7.0%	1.8
220	100.97	$101.00 \pm 0.04$	24.0%	1.2

TABLE I. The  $\chi^2$  of the best-fit pixel-level model against the data, as compared to the expectation computed from mock simulations, as described in Sec. IV B 3. The number of  $\sigma$  is 1-tailed. We find values within  $3\sigma$  of the expectation for individual frequency fits, as well as for our baseline combination of all frequencies.

In this comparison, we consider the global polarization rotation angle,  $\psi_{\text{pol}}$ , the amplitude of monopole temperature-to-polarization leakage,  $\epsilon_Q^i$  and  $\epsilon_U^i$ , and the calibration at each frequency,  $A_{\text{cal}}^{\nu,i}$ . Details of how these are estimated with alternate methods are given in Appendix A. Briefly here,  $\psi_{\text{pol}}$  can be estimated by assuming the sky  $EB$  spectrum is zero, and searching for the angle correction which achieves this in the data.  $\epsilon_Q^i$  and  $\epsilon_U^i$  can be estimated by assuming that the sky  $TQ$  and  $TU$  spectra are zero and finding the deprojection of leakage templates which achieves this.  $A_{\text{cal}}^{\nu,i}$  can be estimated taking ratios of power spectra to a known calibrated spectrum and averaging over very broad multipole ranges. Here,  $A_{\text{cal}}^{150}$  is fit via ratio to Planck, which performs its own absolute calibration, and  $A_{\text{cal}}^{95}$  and  $A_{\text{cal}}^{220}$  are fit from internal comparison to the Planck-calibrated SPT 150 GHz spectrum.

The source of MUSE estimates of these systematics is generally the same as described above, but happens implicitly and jointly inside the global MUSE fit, driven by our model definition (e.g. the model assumes the CMB  $EB$  spectrum is zero and allows a global rotation controlled by  $\psi_{\text{pol}}$  to absorb any  $EB$  power observed in the data during the fit).

The best-fit systematic parameters inferred with MUSE from the SPT data are summarized in Tab. II. Broadly speaking, we find around 1% temperature to polarization leakage, a little less than half of a degree of global polarization angle miscalibration, and beam side-lobes (see Appendix A 4) which are polarized at around the 50% level.

In Fig. 7, we compare to systematic parameters derived from the alternative methods. Up to small differences expected due to slightly different choices of masking and filtering made in these alternative estimates, we find good agreement between our two sets of estimates. We also note that in general the MUSE systematic estimate is slightly tighter, likely due to simultaneously fitting for all systematics and increased optimality of the Bayesian approach.

Posterior			
	95 GHz	150 GHz	220 GHz
$A_{\text{cal}}^1$	$0.8977 \pm 0.0074$	$0.9276 \pm 0.0074$	$0.8612 \pm 0.0084$
$A_{\text{cal}}^2$	$0.8928 \pm 0.0072$	$0.9133 \pm 0.0071$	$0.864 \pm 0.0079$
$A_{\text{cal}}^3$	$0.8861 \pm 0.0073$	$0.9399 \pm 0.0075$	$0.849 \pm 0.0083$
$A_{\text{cal}}^4$	$0.8775 \pm 0.008$	$0.9272 \pm 0.0081$	$0.8405 \pm 0.0097$
$100 \epsilon_Q^1$	$0.291 \pm 0.025$	$0.319 \pm 0.021$	$0.492 \pm 0.059$
$100 \epsilon_Q^2$	$0.402 \pm 0.024$	$0.39 \pm 0.022$	$0.418 \pm 0.06$
$100 \epsilon_Q^3$	$0.555 \pm 0.025$	$0.838 \pm 0.021$	$2.096 \pm 0.066$
$100 \epsilon_Q^4$	$0.603 \pm 0.033$	$0.912 \pm 0.03$	$2.221 \pm 0.084$
$100 \epsilon_U^1$	$0.584 \pm 0.027$	$0.74 \pm 0.025$	$0.735 \pm 0.064$
$100 \epsilon_U^2$	$0.648 \pm 0.025$	$0.748 \pm 0.023$	$0.642 \pm 0.058$
$100 \epsilon_U^3$	$0.851 \pm 0.027$	$1.238 \pm 0.023$	$1.33 \pm 0.063$
$100 \epsilon_U^4$	$0.83 \pm 0.035$	$1.174 \pm 0.03$	$1.121 \pm 0.092$
$\psi_{\text{pol}} [^\circ]$	$0.393 \pm 0.024$	$0.419 \pm 0.021$	$-0.188 \pm 0.079$
$\beta_{\text{pol}}$	$0.44 \pm 0.20$	$0.60 \pm 0.28$	$0.51 \pm 0.26$
Prior			
	95 GHz	150 GHz	220 GHz
$A_{\text{cal}}^1$		$0.9275 \pm 0.0094$	
$A_{\text{cal}}^2$		$0.9095 \pm 0.0089$	
$A_{\text{cal}}^3$		$0.9386 \pm 0.0096$	
$A_{\text{cal}}^4$		$0.927 \pm 0.011$	
$\beta_{\text{pol}}$	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$

TABLE II. Posteriors on systematics parameters given our all-frequency MUSE run (top section) and priors which were input (bottom section). Blank entries or parameters not listed under priors have uniform priors on  $(-\infty, \infty)$ . Superscripts denote each of the four subfields. Note that no cosmological model is assumed here.

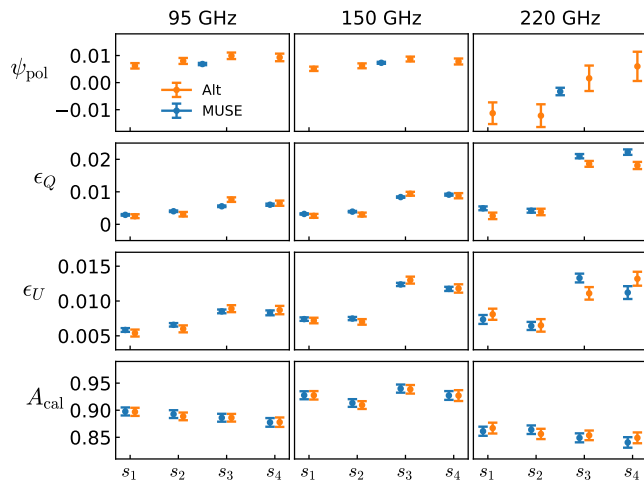


FIG. 7. Agreement of polarization angle rotation angle ( $\psi_{\text{pol}}$ ), monopole T→P leakage ( $\epsilon_Q$  and  $\epsilon_U$ ) and calibration estimates ( $A_{\text{cal}}$ ), between MUSE estimates and alternative methods which have been previously used. The groups of 4 points refer to the 4 different subfields in cases where those are estimated separately. Details of the alternative estimates are given in App. A.

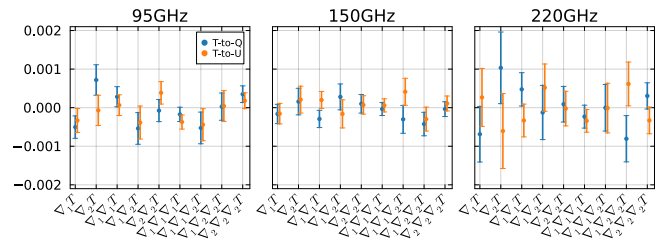


FIG. 8. No evidence for dipole, quadrupole, or octopole T→P leakage. These are MUSE posteriors from a run on all-frequency data that includes these extra leakage templates in addition to the other usual bandpower and systematics parameters. In our baseline run, we thus do not include such higher-order templates.

### 5. Higher order T-to-P leakage

In our baseline model, only monopole T-to-P leakage is modeled. However, it is possible that higher-order leakage exists. To verify our baseline choice, we add leakage templates up to octopole in the MUSE model and infer their leakage amplitudes. These templates correspond to higher-order spatial gradients of the observed temperature maps. The results are shown in Fig. 8, where no higher-order T-to-P leakage amplitude parameters are detected at  $> 3\sigma$ .

### 6. MAPs

We note that prior to unblinding we did look at MAP estimates of the lensing potential and of the lensed and unlensed CMB polarization, and thus present them in this section of the paper. These are shown in Fig. 9.

### C. Post-unblinding changes

With all of these validation tests passed, we unblinded results, fit cosmological models, and compared to data from other experiments. We found  $\Lambda$ CDM was a good fit to our own data and that our  $\phi\phi$  spectrum agreed very well with other experiments. However, the  $EE$  spectrum disagreed with the Planck theory prediction by a smooth slope (shown in the middle panel of Fig. 10), leading to an overall  $3.5\sigma$  disagreement across 5  $\Lambda$ CDM parameters (excluding  $\tau$ ) between Planck and SPT datasets. The slope was most significant at small scales and would have been undetectable by Planck, ACT, or previous SPT measurements, thus could have indicated a newly discovered departure from  $\Lambda$ CDM. We were thus motivated to exhaustively check the robustness of the slope. Doing so, however, we discovered that its likely origin was an overly strong assumption we had been making in our polarization beam modeling.

Prior to unblinding, we had assumed the polarized and temperature beams were the same at the scales used

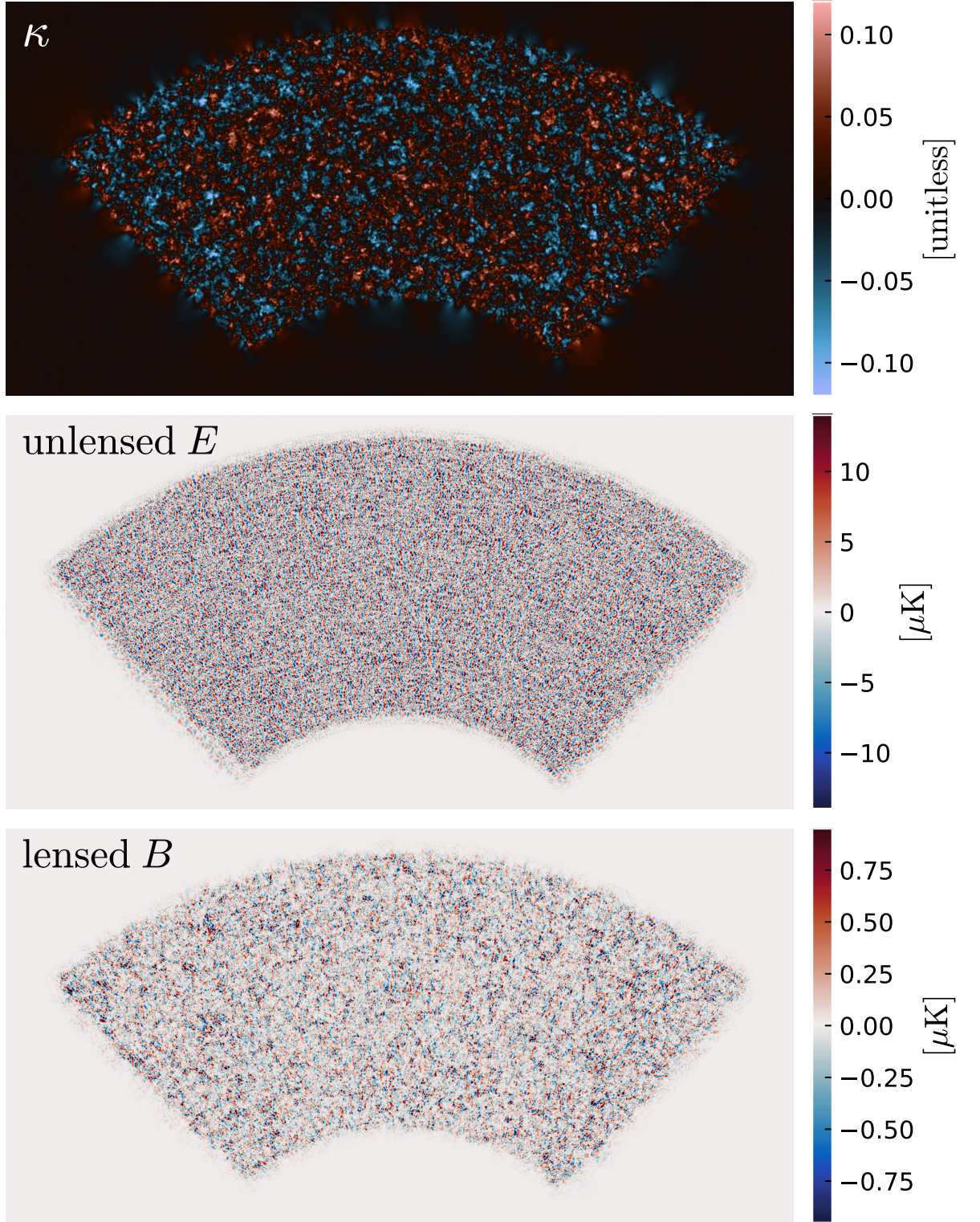


FIG. 9. Maximum a posteriori (MAP) maps from 95+150+220 GHz data at the MUSE estimate of theory and systematics parameters,  $\hat{\theta}$ . MAPs correspond to a filtering of the data which maximizes signal relative to noise (akin to a linear Wiener filter, but in the case of the MAP  $\kappa$ , a non-linear filter). The statistical anisotropy visible particularly in the  $E$  map originates from this filtering. No other additional smoothing has been applied, although we have subtracted the simulation-mean of  $\kappa$ . Except for this subtraction, these MAPs correspond exactly to  $\hat{z}$  which enters the MUSE estimate in Eqn. (5). We have chosen to show only unlensed  $E$  and lensed  $B$ , as lensed and unlensed  $E$  look nearly identical at the scale of this figure, and unlensed  $B$  is effectively zero by assumption.

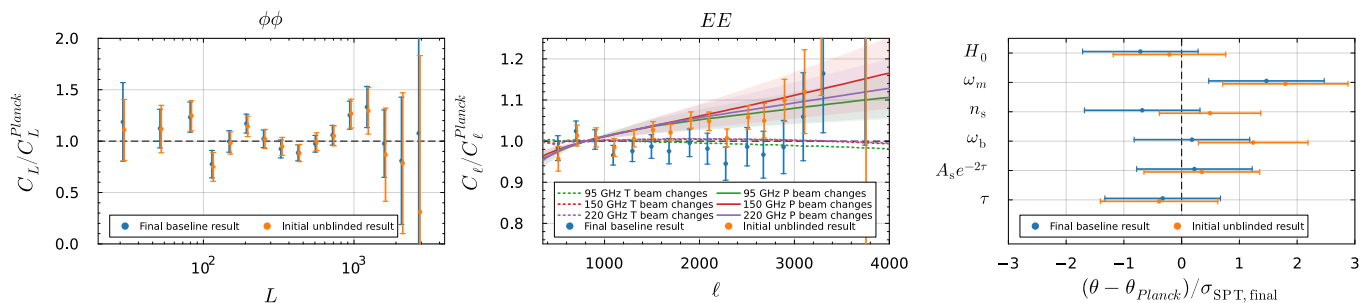


FIG. 10. Changes to our results made after unblinding, and their origin in changes to beams. In the left two panels, blue and orange data points show  $\phi\phi$  and unlensed  $EE$  bandpowers relative to Planck  $\Lambda$ CDM theory for our final result and our initial unblinded result, respectively. Colored lines in the  $EE$  panel show ratios of temperature (dashed lines) and polarization beams (solid lines) between final and initial unblinded versions. For the polarized beam changes, we also show the posterior uncertainty from the final polarized beam fit with shaded bands indicating  $1\sigma$ . In the right panel, the same two colors show the corresponding changes to cosmological parameter results from the SPT dataset.

in the analysis. While the temperature beam sidelobes are mapped with high signal-to-noise via observations of bright point sources and planets, we lack sufficiently bright polarized sources to similarly map the sidelobe response of the polarized beams. To check the robustness of the  $EE$  slope to this assumption, we devised a model of the polarized beams to realistically account for how much these might differ from the temperature beams. Appendix A 4 describes the model in more detail, but in summary, we took the polarized beams to be the sum of two components. The first is a central main lobe which can be modeled physically using the basic optics of the telescope, and is expected from physical modeling to be fully polarized and thus the same in polarization as in temperature. This model is fit to match temperature beams themselves at small scales where the model is a good description of the beam. The second component is a diffuse sidelobe stemming from effect such as scattering or reflection from optical elements in the camera and scattering from panel gaps in the primary mirror, which is not necessarily expected to be fully polarized. We marginalize over the polarization fraction of the sidelobe component with a new parameter at each frequency,  $\beta_{\text{pol}}^\nu$ . In this definition, our earlier assumption had corresponded to  $\beta_{\text{pol}}^\nu = 1$ , which makes the polarized beam identical to the temperature beam.

With these free parameters introduced, we re-ran the MUSE fit, finding  $\beta_{\text{pol}}^\nu = [0.44 \pm 0.20, 0.60 \pm 0.28, 0.51 \pm 0.26]$  at [95, 150, 220] GHz. That is, our CMB data alone is able to internally constrain the sidelobe polarization fractions, and it constrains them away from our previously assumed value of unity with mild significance. The ability to constrain these parameters at all arises from inter-frequency differences, as changing the  $\beta_{\text{pol}}^\nu$  changes the beam at each frequency differently. Indeed, if we re-examine the inter-frequency agreement of our data with the  $\beta_{\text{pol}}^\nu$  marginalized, we find slightly improved agreement. This is shown as the orange line in Fig. 6. Thus, even though our original blind inter-frequency tests did not show any statistically significant evidence for disagreement, there was indeed some small systematic dis-

agreement hidden below the noise, which is alleviated by this improved beam modeling.

Regarding the slope in  $EE$  relative to Planck  $\Lambda$ CDM, the change from  $\beta_{\text{pol}}^\nu = 1$  to the new marginalized posterior values largely removes it. It also increases our  $EE$  uncertainties at  $l > 2000$  by as much as a factor of two for some bins, which, however, is highly correlated between bins, with the effect much like an additional relative calibration between high- $l$  and low- $l$ . The impact to our  $\phi\phi$  spectrum of this post-unblinding change is negligible, as optimal lensing estimation is largely insensitive to smooth changes to beams. Since much of our cosmological information is coming from lensing, it also means that our  $\Lambda$ CDM constraints from SPT data are not that strongly affected; for example we originally found  $H_0 = 67.3 \pm 0.77$  with the sidelobe polarization fraction fixed to unity and  $H_0 = 66.81 \pm 0.81$  with it marginalized. A larger impact occurs in the  $(n_s, \Omega_b h^2)$  plane, with both parameters shifting just over  $1\sigma$  and the posteriors broadening. These changes are all summarized in Fig. 10. Our final agreement with Planck across 5  $\Lambda$ CDM parameters is  $0.5\sigma$ , with the change roughly coming from equal parts central value shifts and posterior broadening in the  $(n_s, \Omega_b h^2)$  plane.

We consider this beam model sufficient to capture our polarized beam uncertainty because we find no significant shifts in results if we change the main lobe model to a simple Gaussian, suggesting the exact details of the main lobe modeling are not relevant. Additionally, allowing for a free linear dependence of the sidelobe polarization fraction as a function of angular distance from the peak response does not significantly shift results either, suggesting the model of a constant sidelobe polarization fraction is sufficient.

With this improvement to the polarized beam modeling, some additional small tweaks mentioned in App. A which lead to insignificant changes, and all validation tests passing at the same or better levels than they were prior to unblinding, we adopted this as our final baseline model. All results we will now present in the following section utilize this baseline model.



Name	Data Set
SPT	This work, i.e. SPT-3G 2019/2020 <sup>a</sup> unlensed $EE$ + optimal $\phi\phi$ from 95+150+220 GHz
SPT $\phi\phi$	Just the $\phi\phi$ spectrum from the SPT dataset, with systematics and unlensed $EE$ still marginalized over
PlanckT&E	Planck 2018 high- $\ell$ $TT, TE, EE$ + low- $\ell$ $TT$ [59]
Planck $\phi\phi$	Planck NPIPE PR4 $\phi\phi$ <sup>b</sup> [60]
Planck	PlanckT&E + Planck $\phi\phi$
ACT $\phi\phi$	ACT DR6 $\phi\phi$ <sup>c</sup> [29, 30]
ACT	ACT DR4 $TT, TE, EE$ <sup>d</sup> [31, 32] + ACT $\phi\phi$
WMAP	WMAP9 <sup>e</sup> ( $\ell_{\min}^{\text{TE}} = 24$ , i.e. no $\tau$ information) [61]
BAO	6dF [62] + SDSS DR7 MGS [63] + SDSS DR12 LRG [64] + SDSS DR16 Ly $\alpha$ [65] + DESI [66–68]

<sup>a</sup> The likelihood will be publicly available after the acceptance of this paper for publication. We refer to this data in the figures as “2yr-main”; the “main” field is the 1500 deg<sup>2</sup> field as defined in [3].

<sup>b</sup> [https://github.com/carronj/planck\\_PR4\\_lensing](https://github.com/carronj/planck_PR4_lensing)

<sup>c</sup> [https://github.com/ACTCollaboration/act\\_dr6\\_lenslike](https://github.com/ACTCollaboration/act_dr6_lenslike)

<sup>d</sup> <https://github.com/ACTCollaboration/pyactlike>

<sup>e</sup> <https://github.com/HTJense/pyWMAP>

TABLE III. Glossary of data set abbreviations used in the text and in figure captions. When combining SPT with WMAP and Planck data, we sum the likelihood directly assuming no correlation, an assumption justified by the small sky area of the SPT main field (1500 deg<sup>2</sup>). When combining ACT DR4 with WMAP and Planck data, we followed the procedure in [31]. When using both Planck NPIPE PR4 lensing and ACT DR6 lensing, we used the built-in likelihood implementation of `act_dr6_lenslike` package with `variant=actplanck_baseline` which accounts for correlations between ACT and Planck lensing due to overlapping sky coverage. For the BAO dataset, we followed the proposal in Sect. 3.3 of Adame et al. [66] to replace the DESI BGS and lowest redshift LRG measurements with SDSS MGS ( $z_{\text{eff}} \sim 0.15$ ) and two SDSS DR12 LRG points ( $z_{\text{eff}} \sim 0.38, 0.51$ ), as well as replacing the DESI Ly $\alpha$  point with joint eBOSS+DESI Ly $\alpha$  BAO. With the exception of Sec. VI, any CMB dataset combinations we use includes a prior on  $\tau$  of  $0.051 \pm 0.006$  [69]. In Sec. VI, we instead drop the  $\tau$  prior and PlanckT&E is to be understood as including the Planck low- $\ell$   $EE$  likelihood.

## V. RESULTS

In this section we present our  $\phi\phi$  and unlensed  $EE$  bandpowers and resulting cosmological parameter constraints in various model spaces, with and without inclusion of some external datasets. Nomenclature for the external datasets, as well as our SPT-3G dataset, is described in Tab. III. We follow the convention of naming cosmological parameters in [70] (see Table. 1 there for detail). The setup for MCMC analysis is described in Sect. III B. For simplicity, we adopt a Gaussian prior,  $\tau \sim \mathcal{N}(0.051, 0.006)$ , using the result from the joint processing of Planck HFI and LFI data [69] as baseline unless otherwise specified, instead of using Planck low- $\ell$   $EE$  likelihood or WMAP TE data at  $\ell < 24$ . Unless otherwise specified, all uncertainties quoted in this paper represent the 68% confidence interval.

### A. Bandpowers

In Fig. 11, we show our inference of  $\phi\phi$  and unlensed  $EE$  bandpowers. The  $\Lambda$ CDM model that best fits these data provides a good fit with  $\chi^2 = 75.8$  given 71 degrees of freedom (76 bandpowers minus 5 cosmological parameters constrained beyond their prior), which corresponds to a PTE of 32.7%. By this test, the SPT data are consistent with the standard cosmological model. We also show as an orange band the 68% confidence level predictions of the  $\Lambda$ CDM model for the CMB lensing and

unlensed  $EE$  power spectra given these SPT data.

We can also check if the SPT data are consistent with the  $\Lambda$ CDM predictions given Planck data. The 68% confidence level predictions of the  $\Lambda$ CDM model given Planck data are shown as a gray band, with the best-fit  $\Lambda$ CDM model given Planck data shown as a black line. The  $\chi^2$  of the SPT bandpowers given the Planck best-fit  $\Lambda$ CDM model is 80.7, which corresponds to a PTE of 20.2%. Thus, the SPT bandpowers agree with the Planck  $\Lambda$ CDM best-fit prediction at the  $1.3\sigma$  level.

In Fig. 12, we compare the SPT bandpowers to prior measurements. The SPT lensing bandpower measurements extend to  $L \sim 3000$  and are more precise than prior published measurements at  $L > 350$ . The unlensed  $EE$  bandpowers extend to  $\ell \sim 3500$  and are more precise than prior published measurements at  $\ell > 2000$  where we have clear detections of the eighth, ninth and tenth acoustic peaks.

### B. Origins of $\Lambda$ CDM parameter constraints

The SPT bandpowers represent a significant advance in our capacity to constrain cosmological parameters with polarization data alone. In Fig. 13 we show constraints on four  $\Lambda$ CDM parameters from SPT, ACT, and Planck, assuming the  $\Lambda$ CDM model, and limiting to only  $EE$  spectra and polarization-based lensing reconstruction, or in the case of ACT, also including temperature in the lensing reconstruction (since no polarization-only lensing likelihood is available). We see significantly tighter con-

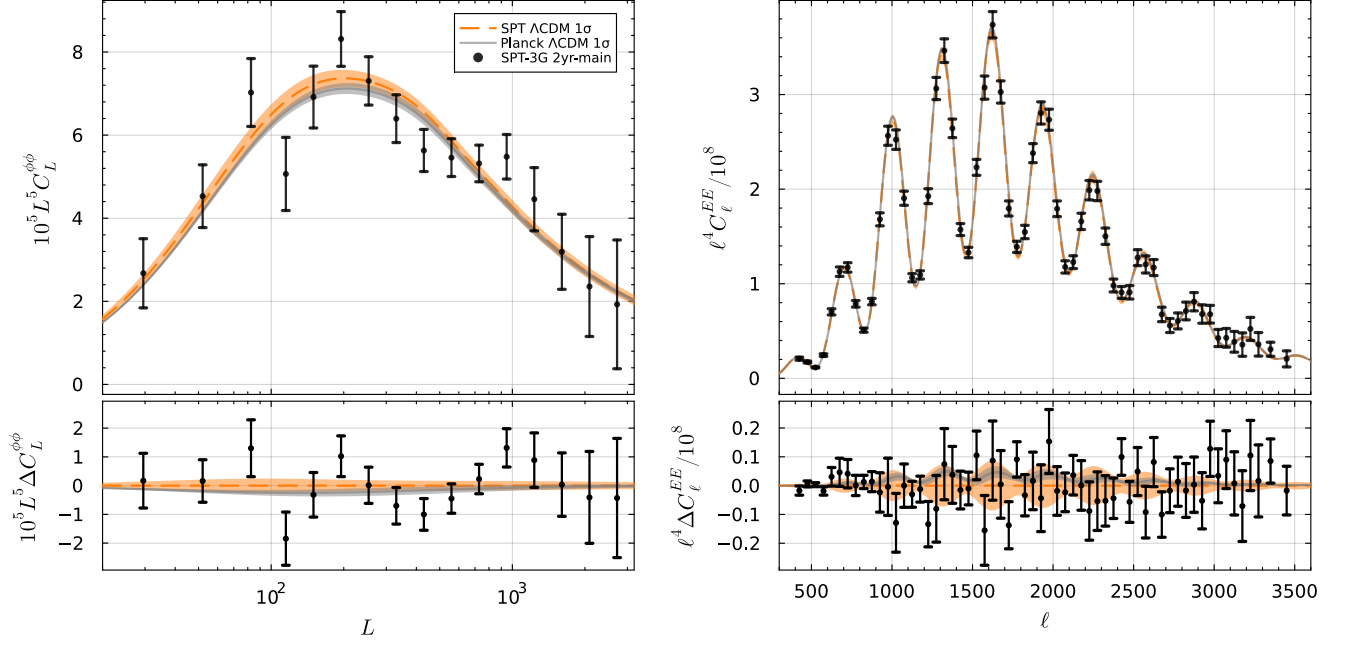


FIG. 11. *Top panels:* MUSE measurement of the CMB lensing  $\phi\phi$  bandpowers and the unlensed  $EE$  bandpowers using SPT-3G 2019/20 CMB polarization data. The lensing bandpowers represent a  $38\sigma$  detection of lensing power. These are the most precise measurements to date at  $\ell > 2000$  for  $EE$  and  $L > 350$  for  $\phi\phi$ . *Bottom panels:* Residual bandpowers with respect to the SPT best-fit cosmology. The gray and orange bands show the 68% confidence region of  $\Lambda$ CDM model spectra fit to Planck and SPT data, respectively.

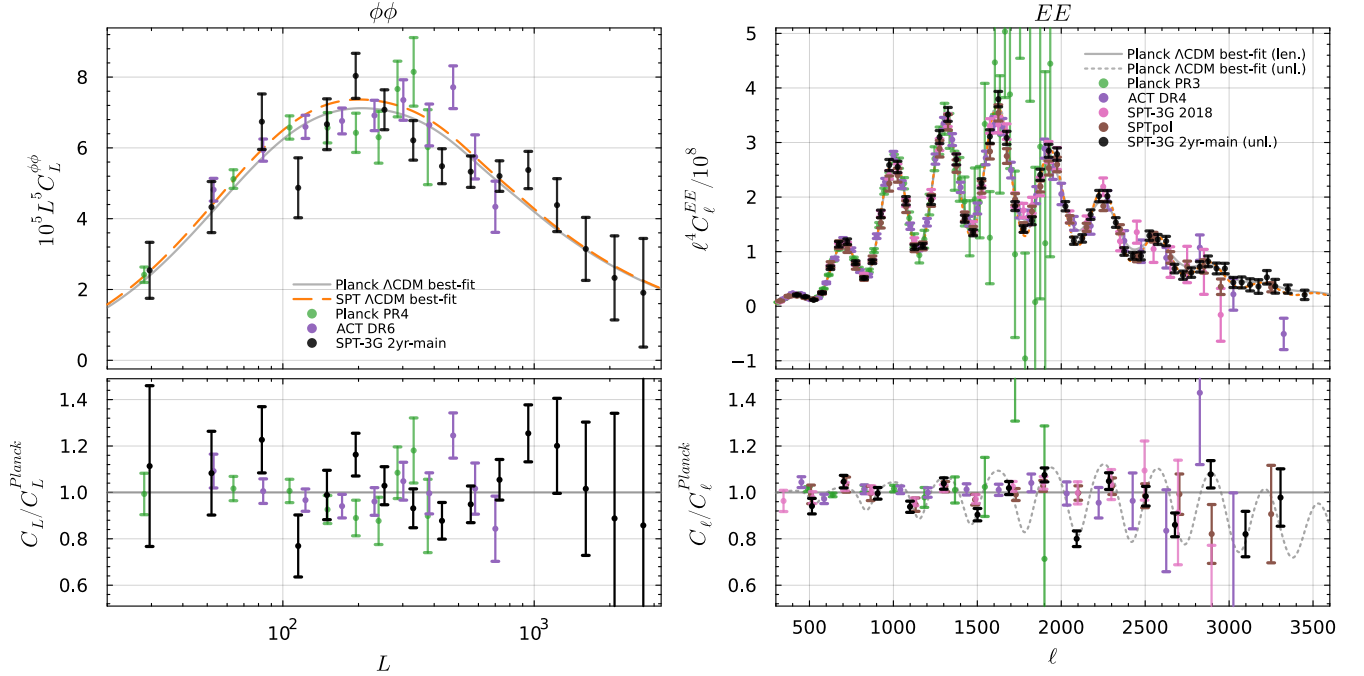


FIG. 12. Comparison of SPT bandpowers with existing measurements. *Top panels:* CMB  $\phi\phi$  and  $EE$  bandpowers. *Bottom panels:* Bandpowers divided by the Planck best-fit cosmology predictions. Note that the  $EE$  bandpowers have been further rebinned as compared to the top panel for clarity. As a reminder, the SPT  $EE$  points are inferences of unlensed  $EE$ . The dotted gray line shows the unlensed  $EE$  model prediction. The orange dashed and the solid gray lines are the same as Fig. 11.

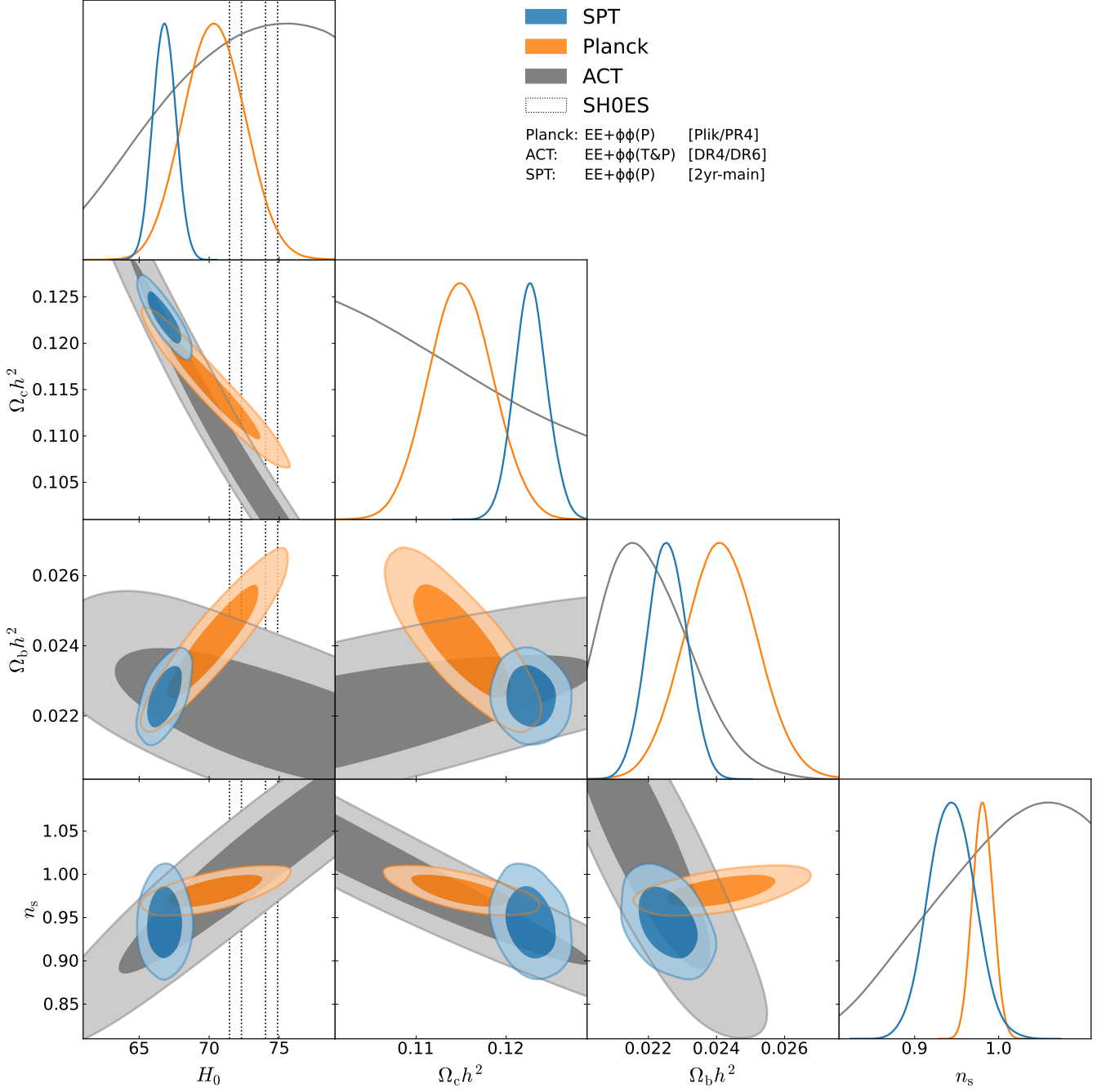


FIG. 13. Posterior distributions of  $\Lambda$ CDM cosmological parameters derived only from polarization  $EE$  spectra and polarization-based lensing reconstruction, or in the case of ACT, also including temperature in the lensing reconstruction (for technical reasons). We see tighter constraints on  $\Omega_b h^2$ ,  $\Omega_c h^2$ , and  $H_0$  from the SPT data. We show the SH0ES constraint on  $H_0$  for comparison. See also Tab. III for more details on dataset definitions.

straints on three of the parameters from SPT, while the higher dynamic range of the Planck  $EE$  data (which extends to lower  $\ell$  than shown in Fig. 12) leads to a tighter constraint on  $n_s$  for Planck.

Constraints on  $\Lambda$ CDM parameters from SPT data and other dataset combinations, now including both temperature and polarization as detailed in Tab. III, are summarized in Tab. IV. The SPT constraints are comparable to full Planck results in the matter density,  $\Omega_m h^2$ , and angular scale of the sound-horizon,  $\theta_*$ . The constraining power on  $\Omega_m h^2$  from SPT is mainly due to the CMB lensing bandpowers, which are sensitive to the amount of matter in the Universe. This is evidenced by the fact that adding just SPT lensing to WMAP data reduces the error on  $\Omega_m h^2$  by over a factor of 3 from 0.0045 [61] to 0.0014, and that further adding SPT  $EE$  to this combination leaves the error unchanged.

That the sensitivity to  $\Omega_m h^2$  primarily comes from the lensing reconstruction rather than primary power spectra (in this case unlensed  $EE$ ) is a unique feature of these data compared to prior CMB datasets. The sensitivity of primary power spectra to matter density is weakened due to its weighting toward small scales that are sensitive to conditions at horizon crossing; at such early times the matter density is negligible compared to the radiation density. Conversely, the lensing power spectrum, if probed precisely enough, has high sensitivity to  $\Omega_m h^2$ . This comes mainly (but not solely) from how CMB lensing power depends on the scale of matter-radiation equality. Additionally, the sensitivity increases at the small scales probed here, with  $C_L^{\phi\phi} \propto (\Omega_m h^2)^2$  at  $L \sim 200$  and asymptoting to  $(\Omega_m h^2)^3$  by  $L \sim 2000$  [71].

The SPT data are also sensitive to the angular scale of the sound horizon,  $\theta_*$ . This sensitivity comes from the  $EE$  bandpowers, where the peak and trough locations of the  $EE$  spectrum are highly sensitive to  $\theta_*$ . The importance of the  $EE$  bandpowers for this constraint, as opposed to the lensing bandpowers, can be seen in the comparison of  $\theta_*$  constraints from SPT, WMAP+SPT and WMAP+SPT $\phi\phi$ ; the exclusion of SPT  $EE$  significantly degrades the  $\theta_*$  determination. We note that the constraint on  $\theta_*$  is known to be improved by delensing [72], a result which applies to our unlensed spectra inferences here as well, although we do not explicitly attempt to quantify exactly how much improvement we have achieved.

The constraints on  $\tau$  from the SPT dataset are driven almost entirely by Planck low- $\ell$  polarization data, included here as a  $\tau$  prior. Although lensing can, in theory, help constrain  $\tau$  via degeneracy breaking, the contribution at current noise levels is negligible. The constraint on  $\log(10^{10} A_s)$  is similarly largely driven by Planck, since it is degenerate both with  $\tau$  and the absolute calibration, the latter which also comes from Planck.

The SPT constraints on  $\Omega_b h^2$  and  $n_s$  are weaker than those inferred from Planck. Both  $n_s$  and  $\Omega_b h^2$  cause tilts of the CMB  $EE$  spectrum, since  $\Omega_b h^2$  affects the photon diffusion damping at small angular scales and  $n_s$  affects

the shape of the primordial spectrum. Our observations lack the very largest scales probed by Planck, which, logarithmically, contain a significant range of scales useful for constraining a tilt. Additionally, at small scales our uncertainties are increased by polarized beam uncertainty, preventing what could otherwise be a slightly better constraint. Our final result is a 2.5% determination of the the baryon density.

With the baryon density, matter density, and  $\theta_*$  determined, one has everything necessary, in  $\Lambda$ CDM, for calculating both  $H_0$  and  $\Omega_m$  [e.g. 73]. Adding in  $A_s$  and  $n_s$  one can also calculate  $\sigma_8$  and  $S_8$ . We show constraints on these four derived parameters in Tab. IV.

### C. Agreement with other CMB datasets in $\Lambda$ CDM

Before providing interpretations, it is useful to establish the consistency level of our  $\Lambda$ CDM parameter results with other CMB measurements. Such comparison provides tests of the  $\Lambda$ CDM model that are complementary to and stronger than the bandpower-level tests above.

We perform a simple consistency check applicable to uncorrelated datasets by calculating the  $\chi^2$  between the two sets of parameters,

$$\chi^2 = (\vec{x}_1 - \vec{x}_2)^T (C_1 + C_2)^{-1} (\vec{x}_1 - \vec{x}_2), \quad (23)$$

where  $\vec{x}$  is a vector of parameter estimates,  $C$  are the parameter covariances from each, and the subscripts 1 & 2 distinguish datasets. We then compute the PTE for the given  $\chi^2$  and quote a 1-tailed conversion to number of  $\sigma$ . Since, as discussed, our estimates of  $\tau$  and  $A_s$  are highly correlated to Planck, we consider only  $\vec{x} = (\theta_*, \Omega_b h^2, \Omega_c h^2, n_s)$ . For these parameters, the SPT results are very close to independent of both Planck and ACT due to non-overlapping sky area and multipole ranges. We find  $\Lambda$ CDM parameters from SPT and Planck datasets are consistent at  $0.8\sigma$ , and from SPT and ACT datasets at  $0.8\sigma$ .

A triangle plot comparing marginal posterior distributions from SPT+WMAP, ACT+WMAP, and Planck is shown in Fig. 14.

We now turn to estimates of three parameters of particular contemporary interest in cosmology:  $H_0$ ,  $S_8$ , and  $\Omega_m$ .

### D. The Hubble constant, $H_0$

The Hubble constant inferred from Planck data, assuming the  $\Lambda$ CDM model, has been found to be in tension with the most precise “distance ladder” measurement [74, 75], which uses Cepheid-calibrated SNe Ia, at the  $5.8\sigma$  level.<sup>6</sup> In Fig. 15, we show the Hubble constant

<sup>6</sup> See also, e.g. Freedman & Madore [76], Freedman et al. [77], Riess et al. [78] for comparison of alternative “rung 2” methods

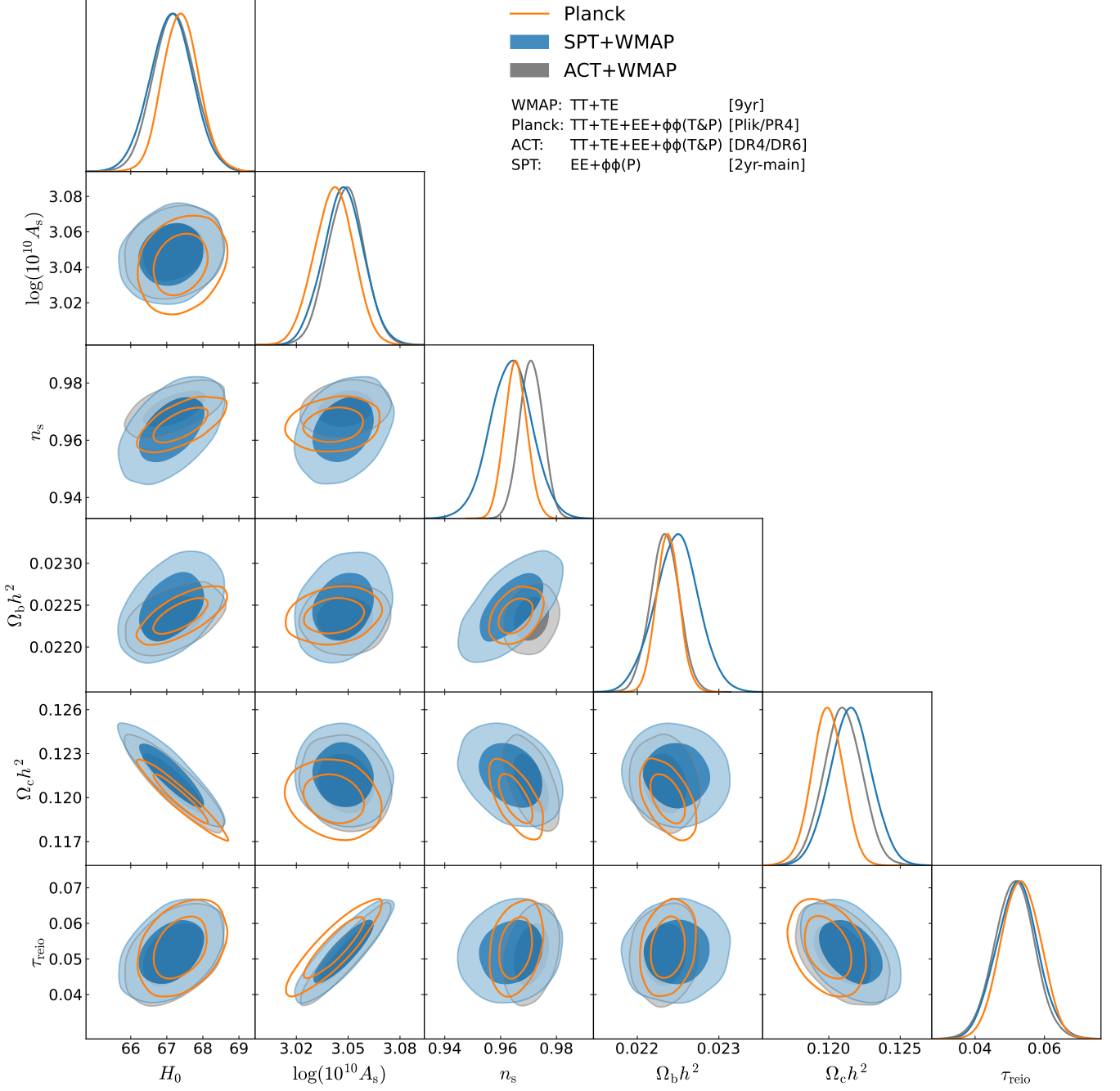


FIG. 14. Posterior distributions of  $\Lambda$ CDM cosmological parameters derived from SPT+WMAP, ACT+WMAP, and Planck. The details of the datasets are listed in Tab. III. We see that either ACT or SPT, when combined with WMAP for constraints on larger angular scales, achieve constraints on cosmological parameters with similar constraining power as the constraints from Planck.

Parameters	SPT	Planck	Planck+SPT	WMAP+SPT	WMAP+SPT $\phi\phi$
$\log(10^{10} A_s)$	$3.045 \pm 0.024$	$3.042 \pm 0.011$	$3.046 \pm 0.011$	$3.047 \pm 0.012$	$3.042 \pm 0.013$
$n_s$	$0.944 \pm 0.027$	$0.965 \pm 0.004$	$0.9647 \pm 0.0037$	$0.9637 \pm 0.0076$	$0.961 \pm 0.011$
$100\theta_{MC}$	$1.04162 \pm 0.00066$	$1.04092 \pm 0.00031$	$1.04104 \pm 0.00028$	$1.04154 \pm 0.00057$	$1.0394 \pm 0.0022$
$\Omega_b h^2$	$0.02255 \pm 0.00057$	$0.02237 \pm 0.00014$	$0.02239 \pm 0.00013$	$0.02248 \pm 0.00027$	$0.02226 \pm 0.00045$
$\Omega_c h^2$	$0.1227 \pm 0.0018$	$0.1199 \pm 0.0011$	$0.12039 \pm 0.00094$	$0.1215 \pm 0.0014$	$0.1210 \pm 0.0014$
$\tau_{reio}$	$0.052 \pm 0.006$	$0.0532 \pm 0.0056$	$0.0547 \pm 0.0058$	$0.052 \pm 0.006$	$0.0515 \pm 0.0059$
$H_0$ [km/s/Mpc]	$66.81 \pm 0.81$	$67.4 \pm 0.5$	$67.28 \pm 0.42$	$67.1 \pm 0.6$	$66.4 \pm 1.1$
$100\theta_*$	$1.04181 \pm 0.00069$	$1.0411 \pm 0.0003$	$1.04125 \pm 0.00027$	$1.04173 \pm 0.00057$	$1.0396 \pm 0.0022$
$\Omega_m h^2$	$0.1459 \pm 0.0019$	$0.143 \pm 0.001$	$0.14340 \pm 0.00089$	$0.1446 \pm 0.0015$	$0.1439 \pm 0.0015$
$\Omega_m$	$0.327 \pm 0.011$	$0.3147 \pm 0.0068$	$0.3169 \pm 0.0058$	$0.3211 \pm 0.0085$	$0.326 \pm 0.011$
$\sigma_8$	$0.8143 \pm 0.0088$	$0.8104 \pm 0.0051$	$0.8137 \pm 0.0046$	$0.8177 \pm 0.0065$	$0.8118 \pm 0.0086$
$S_8 = \sqrt{\Omega_m}/0.3$	$0.850 \pm 0.017$	$0.830 \pm 0.012$	$0.8363 \pm 0.0097$	$0.846 \pm 0.016$	$0.847 \pm 0.016$

TABLE IV. Parameter estimates for various datasets assuming  $\Lambda$ CDM. The datasets are defined in Tab. III.

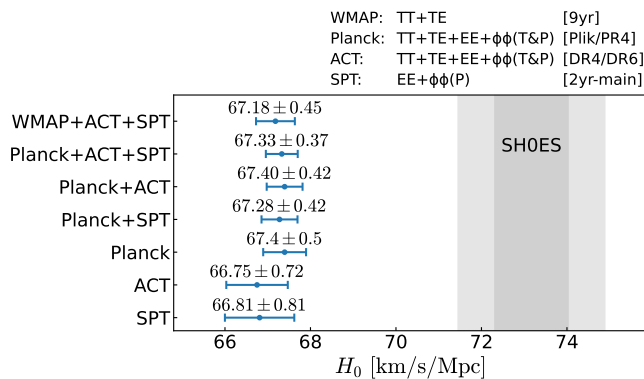


FIG. 15. Comparison of the  $\Lambda$ CDM constraints on  $H_0$  inferred from different CMB observations with the Cepheid-calibrated SNe Ia observations of the SH0ES collaboration [74, 75]. All CMB  $\Lambda$ CDM constraints are significantly lower.

inferred from various observations. Using the SPT data alone, we report

$$H_0 = 66.81 \pm 0.81 \text{ km/s/Mpc.} \quad (24)$$

This is consistent with the Hubble constant inferred from Planck, and is in  $5.4\sigma$  tension with [74]. This confirms previous hints of a tension given only CMB polarization data, where Planck, ACTpol, and SPTpol yielded  $H_0 = 68.7 \pm 1.3$  [79], but almost halves the error bar and increases the significance of the tension from less than  $3\sigma$  to greater than  $5\sigma$ .

The agreement with previous CMB constraints, that rely heavily on temperature data, is also significant as a passed test of  $\Lambda$ CDM. Of course, the continued disagreement with distance ladder methods remains a mystery. We expect that these new SPT data will be a significant hurdle for many of the proposed models [33], and models yet to come, that address the Hubble tension.

Combining Planck, ACT, and SPT, we report

$$H_0 = 67.33 \pm 0.37 \text{ km/s/Mpc,} \quad (25)$$

which results in tension with the  $H_0$  reported in [74] at  $6.2\sigma$ . This is the tightest constraint on  $H_0$  inferred from CMB observations assuming the  $\Lambda$ CDM model. We also note that the combination WMAP+ACT+SPT achieves  $H_0 = 67.18 \pm 0.45$  km/s/Mpc, a result with precision nearly the same as that of the Planck result and nearly independent of Planck data. Overall, inferences of the Hubble constant from a wide range of CMB observations are in good agreement. While the potential for confirmation bias to drive this agreement could be a concern, we highlight that both the unlensed  $EE$  and  $\phi\phi$  spectra which enter our SPT result here were obtained blindly, and while there were some changes to  $H_0$  after we unblinded, these changes were small compared to the magnitude of the tension.

### E. The matter density, $\Omega_m$

There has recently emerged some degree of tension between  $\Omega_m$  as inferred from BAO, and values inferred from supernova surveys [66]. We summarize the situation in Fig. 16. The biggest discrepancy is between DESI BAO and the DES-Y5 determination of  $\Omega_m$  which yields a  $2.5\sigma$  difference. Inferences from CMB data, with consistency among different observations, are more precise than those from the supernovae and BAO, lying in between the low BAO value and the higher supernova values. From SPT data alone we infer

$$\Omega_m = 0.327 \pm 0.011. \quad (26)$$

This is the highest central value of the inferences from different choices of CMB dataset, though consistent with all of them, and with an uncertainty only 65% greater than the Planck uncertainty.

### F. The structure amplitude parameter, $S_8$

The amplitude of structure measured from galaxy weak lensing surveys also shows mild tension with the  $\Lambda$ CDM

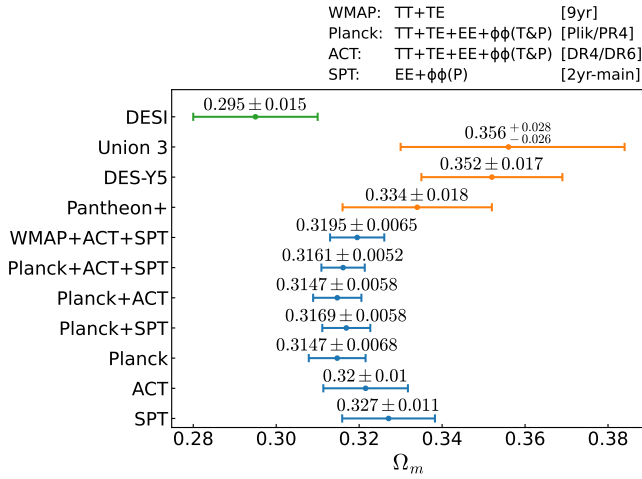


FIG. 16. Comparison of the  $\Lambda$ CDM constraints on  $\Omega_m$  inferred from different CMB observations (blue) with those from SNe observations (orange) and DESI BAO (green). The SNe observations include Pantheon+ [80], DES-Y5 [81] and Union 3 [82]. The biggest discrepancy is between DESI BAO and the DES-Y5 determination of  $\Omega_m$ , a  $2.5\sigma$  difference.

inference using CMB data. In Fig. 17, we show the value of  $S_8 \equiv \sigma_8 \sqrt{\Omega_m}/0.3$  inferred from various observations. From SPT, we find

$$S_8 = 0.850 \pm 0.017, \quad (27)$$

which is in agreement with the result from Planck at a similar level of precision. Our result is in  $2\sigma$ ,  $3.2\sigma$ , and  $3\sigma$  disagreement with HSC-Y3 [83], KiDs-1000 [84] and DES-Y3 [85], respectively. From the joint fit of Planck+ACT+SPT, we get

$$S_8 = 0.8380 \pm 0.0084, \quad (28)$$

which is at  $1.8\sigma$ ,  $3.3\sigma$ , and  $3.3\sigma$  tension with HSC-Y3, KiDs and DES-Y3 results. We also note that the constraint on  $S_8$  from a combination of WMAP+ACT+SPT is consistent with the  $S_8$  inferred from Planck alone, which provides a nearly-independent check of the Planck result.

### G. Amplitude of nonlinear structure growth

One proposed phenomenological resolution to the  $S_8$  tension is a suppression of the matter power spectrum at non-linear scales, potentially due to baryonic feedback or dark matter structure growth which differs from the standard picture [37, 86]. The CMB lensing reconstruction we present in this paper is the most precise measurement of the CMB lensing spectrum at the smaller angular scales where non-linear effects are more important, and can be used to further constrain the space of possible models which resolve this tension. To do so, we introduce a new parameter,  $A_{\text{mod}}^{\text{CMB}}$ , equivalent to the  $A_{\text{mod}}$  parameter of

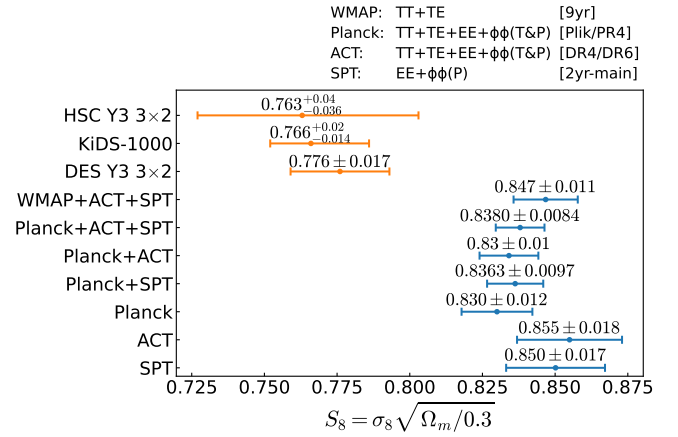


FIG. 17. Comparison of the  $\Lambda$ CDM constraints on  $S_8$  inferred from different CMB observations (blue) with those from galaxy surveys of HSC Y3, KiDs-1000 and DES-Y3 (orange). The estimate from Planck+ACT+SPT differs with the one from HSC-Y3 at  $2\sigma$ , the one from KiDs at  $3.3\sigma$ , and with the one from DES-Y3 at  $3\sigma$ .

[37] under the Limber approximation, but here measured at CMB lensing scales and redshifts. We implement this by directly scaling the non-linear correction to the linear-theory CMB lensing spectrum,

$$C_L^{\phi\phi} = C_{L,\text{lin}}^{\phi\phi} + A_{\text{mod}}^{\text{CMB}} (C_{L,\text{nonlin}}^{\phi\phi} - C_{L,\text{lin}}^{\phi\phi}), \quad (29)$$

where  $C_{L,\text{lin}}^{\phi\phi}$  and  $C_{L,\text{nonlin}}^{\phi\phi}$  are the linear and non-linear CMB lensing spectra, the latter computed via our default *Halofit*. This modified lensing spectrum is then also propagated to changes in the lensed primary spectra.

As a small aside, we have verified that constraints on  $A_{\text{mod}}^{\text{CMB}}$  are insensitive to the choice of nonlinear and baryonic feedback models. Given that  $A_{\text{mod}}^{\text{CMB}}$  is the parameter most sensitive to the non-linear prediction, our constraints on other cosmological parameters in this paper are insensitive as well.

Fig. 18 shows results of fitting the  $\Lambda$ CDM+ $A_{\text{mod}}^{\text{CMB}}$  model to various CMB datasets. We take the Planck primary CMB measurements as a baseline, which tightly limit the linear contribution to lensing, then explore the impact to constraints on the non-linear contribution,  $A_{\text{mod}}^{\text{CMB}}$ , from adding lensing measurements from Planck, ACT, or the SPT results from this work. As expected from the spectra shown in Fig. 12, there are significant improvements in constraining the non-linear contribution from Planck to ACT, and again from ACT to SPT. The Planck primary CMB together with SPT detect a non-zero  $A_{\text{mod}}^{\text{CMB}}$  at  $> 3\sigma$  for the first time.

All measurements are consistent with the standard value  $A_{\text{mod}}^{\text{CMB}} = 1$ , and the tightest combination of all lensing measurements yields  $A_{\text{mod}}^{\text{CMB}} = 1.60 \pm 0.39$ . The  $A_{\text{mod}}$  inferred from DES-Y3 and KiDs is  $A_{\text{mod}} = 0.820 \pm 0.042$  [86]. This latter result measures the amplitude of non-linear structure growth at redshift  $z \sim 0.3$  and scales of  $k \sim 1 \text{ Mpc}^{-1}$ , whereas our CMB lensing measurements

are sensitive to higher redshifts and larger scales. Using a Fisher calculation based on our actual achieved error-bars, in the top panel of Fig. 18 we have computed the range of  $(k, z)$  which contribute most to our constraints on total lensing power, as well as to just the non-linear contribution to the lensing power, i.e. to  $A_{\text{mod}}$ .

This plot, taken together with our  $A_{\text{mod}}$  constraints, suggests (at  $\sim 2\sigma$ ) that if the  $S_8$  tension is to be explained by some non-standard physics suppressing the non-linear matter power spectrum, these effects have not started at the higher redshift and larger scales of CMB lensing as compared to galaxy lensing, and evolve in the decade of separation in  $(k, z)$  before suppressing the galaxy lensing power by  $\sim 20\%$  at later times and smaller scales. We note that with even better future measurements of small-scale CMB lensing, sensitivity to  $A_{\text{mod}}^{\text{CMB}}$  will push to smaller scales; e.g., significantly tightening up measurement of lensing power in a broad band centered on  $L = 3000$  will lead to sensitivity to galaxy weak lensing scales, although at  $z \simeq 1$  to 2 rather than  $z = 0.3$ .

## H. Massive neutrinos

The  $\Lambda$ CDM model, as usually understood, includes a strong assumption about the spectrum of neutrino masses, namely that the sum is 0.058 eV. This particular value follows from assuming that the lightest neutrino mass is effectively zero, that normal ordering applies, and that the usual interpretation of atmospheric and solar neutrino oscillations is correct. Here we relax this assumption and study constraints on the sum of neutrino masses,  $\Sigma m_\nu$ . For a review of neutrino physics and its cosmological significance see [87].

Relative to the case of massless neutrinos, neutrino rest masses increase the mean energy density and therefore the expansion rate. Increased expansion rate reduces growth on scales below the neutrino free-streaming length. Increasing mass also decreases the neutrino free-streaming length. On scales above the free-streaming length, neutrinos can cluster, approximately canceling the impact on growth of the increased expansion rate. For a recent discussion of the cosmological impact of massive neutrinos see [88].

With the release of the DESI BAO data, cosmological constraints on neutrino mass have become an especially interesting topic. In the context of the  $\Lambda$ CDM +  $\Sigma m_\nu$  model space, the DESI data in combination with Planck high- $\ell$   $TT$ ,  $TE$ ,  $EE$  and the Planck  $\tau$  determination serve to tighten up the predictions for CMB lensing power as a function of assumed  $\Sigma m_\nu$ . As has been shown in [36] and [89], these predictions, assuming a minimal value of  $\Sigma m_\nu = 0.058$  eV, indicate somewhat less lensing power than observed in either Planck lensing or ACT DR6 lensing. This is puzzling, since boosting the expected lensing power by changing the neutrino mass sum would require a *decrease* to below the minimum of 0.058 eV.

In Table V, we show the constraints on  $\Sigma m_\nu$  from

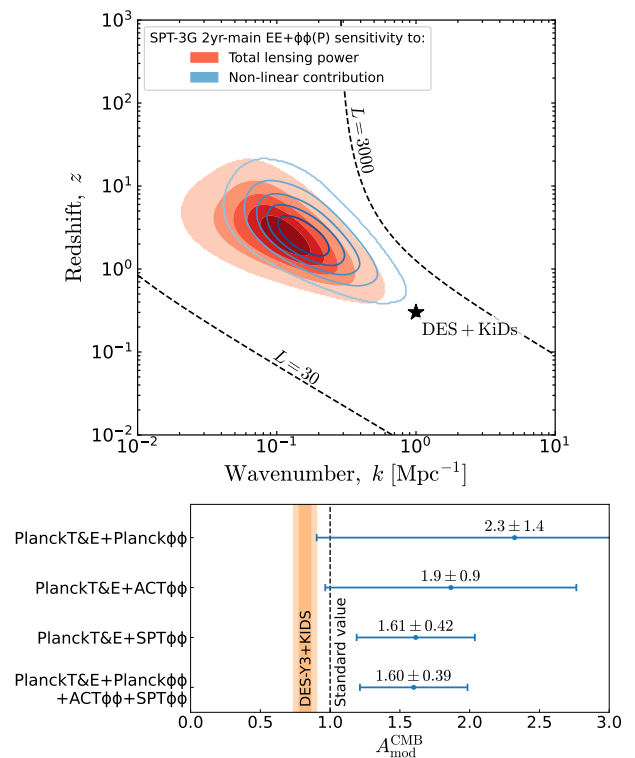


FIG. 18. *Top panel:* Redshifts and scales to which our lensing measurements are sensitive. Colored regions are iso-contours in the contribution to the square root of the Fisher information from logarithmic bins in  $k$  and  $z$ . The outermost contour corresponds to 10% of the peak. Red is for the Fisher information in the total lensing power, whereas blue is for the Fisher information just in  $A_{\text{mod}}^{\text{CMB}}$  (and is weighted more towards higher  $L$ ). Under the Limber approximation, a given  $(k, z)$  contributes to a single CMB lensing multipole,  $L$ , and dashed lines denote this mapping for scales of  $L = 30$  and  $L = 3000$ . *Bottom panel:* Constraints on  $A_{\text{mod}}^{\text{CMB}}$ , defined as the amplitude of non-linear contributions to the CMB lensing spectrum (Eqn. 29), from Planck primary CMB measurements (labeled PlanckT&E) in combination with lensing measurements from Planck, ACT DR6, and the SPT dataset presented here. The orange band shows the 68% and 95% constraints of  $A_{\text{mod}}$  inferred from DES-Y3 and KiDs datasets in [86].

various datasets. Combining the CMB observations of Planck, SPT, and ACT, we report the following constraint on neutrino mass

$$\Sigma m_\nu < 0.20 \text{ eV (95\% C.L.)}. \quad (30)$$

Additionally including BAO measurements yields

$$\Sigma m_\nu < 0.075 \text{ eV (95\% C.L.)}, \quad (31)$$

which puts extreme pressure on the inverted hierarchy minimal mass sum of about 0.098 eV.



### I. Excess lensing power?

As mentioned, Craig et al. [36] pointed out that the lensing power inferred from the Planck lensing reconstruction and from the ACT DR6 lensing reconstruction are somewhat in excess of  $\Lambda$ CDM predictions given DESI BAO and primary CMB data. Here we investigate this claim, first with  $L$ -independent lensing template parameters and then with the phenomenological “negative neutrino mass” parameter used in [36],  $\Sigma\tilde{m}_\nu$ , which can be thought of as an  $L$ -dependent lensing template parameter.

To conduct our analysis, we define three  $L$ -independent lensing template parameters:

1.  $A_{\text{recon}}$  scales the  $\Lambda$ CDM model lensing power used to predict the reconstructed lensing power,
2.  $A_{2\text{pt}}$  scales the model lensing power used to compute any lensed power spectra (2-point correlation functions), and
3.  $A_{\text{lens}}$  scales both of these model power spectra; i.e., it is what we call  $A_{\text{recon}}$  and  $A_{2\text{pt}}$  when we force them to be equal to each other.

Note that what we call  $A_{2\text{pt}}$  has often been called  $A_L$ .

We consider four model spaces, each an extension of  $\Lambda$ CDM. They are  $\Lambda$ CDM+ $A_{\text{recon}}$ ,  $\Lambda$ CDM+ $A_{2\text{pt}}$ ,  $\Lambda$ CDM+ $A_{\text{lens}}$ , and  $\Lambda$ CDM+ $A_{\text{recon}}$ + $A_{2\text{pt}}$ . Taking  $\theta$  to be the  $\Lambda$ CDM parameters, the extensions to the model spaces that include these template parameters are done via

$$C_L^{\phi\phi}(A_X, \theta) = A_X C_L^{\phi\phi}(\theta), \quad (32)$$

where  $X = \text{recon}, 2\text{pt}, \text{or lens}$  and  $C_L^{\phi\phi}(\theta)$  on the right hand side is the  $\Lambda$ CDM model evaluated at  $\theta$ .<sup>7</sup> To assist in comparison with [36, 89], in this subsection only we take “PlanckT&E” to include the Planck low- $\ell$   $EE$  likelihood and remove the  $\tau$  prior, and we restrict our use of ACT data to ACT DR6 lensing only. Our choice of BAO data can be found in Table III.

In the dash-dotted curves of the top panel of Fig. 19 we see the well-known “ $A_{\text{lens}}$ ” or “ $A_L$ ” anomaly, which in our language is an  $A_{2\text{pt}}$  anomaly driven by Planck PR3  $TT$  data [90]. This is a preference, in the lensed  $TT$ ,  $TE$ , and  $EE$  spectra, for excess lensing power beyond what one gets from the  $\Lambda$ CDM model. In the lower left panel we can see that if both  $A_{\text{recon}}$  and  $A_{2\text{pt}}$  are allowed to vary freely, the 95% confidence regions for the four considered data combinations have central values greater than 1 in both dimensions and no overlap with the  $\Lambda$ CDM point of  $A_{\text{recon}} = A_{2\text{pt}} = 1$ .

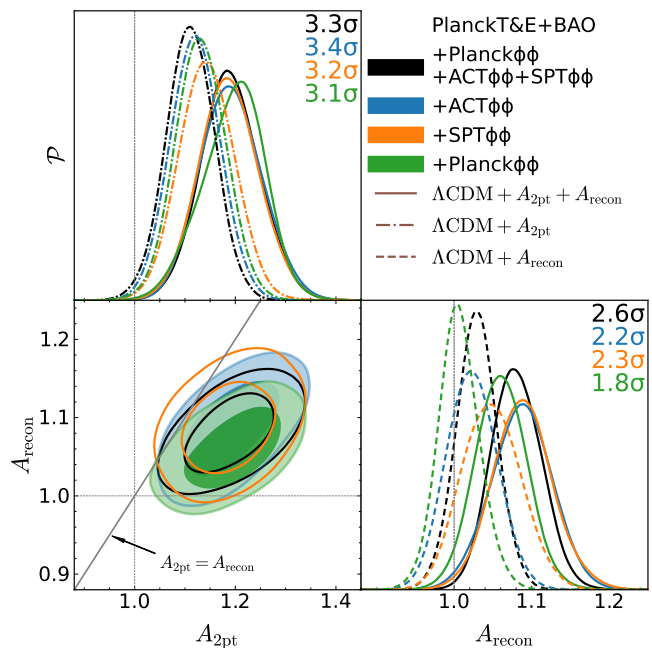


FIG. 19. Posteriors of the lensing power phenomenological scaling parameters  $A_{2\text{pt}}$  and  $A_{\text{recon}}$ . Different combinations of CMB lensing data added to the Planck primary CMB and BAO are shown in different colors. The 2D contours and solid lines show the marginal posterior from the  $\Lambda$ CDM+ $A_{2\text{pt}}$ + $A_{\text{recon}}$  model; the dot-dashed and dashed lines show the results from  $\Lambda$ CDM+ $A_{2\text{pt}}$  and  $\Lambda$ CDM+ $A_{\text{recon}}$  respectively. The  $\sigma$  values give the level at which  $A_{\text{recon}}$  ( $A_{2\text{pt}}$ ) values less than 1 are excluded, for the case of marginalization over  $A_{2\text{pt}}$  ( $A_{\text{recon}}$ ); i.e., for the solid curves.

Turning to the question of an excess of reconstructed lensing power, we see in the solid lines of the right panel that if we marginalize over  $A_{2\text{pt}}$  (solid lines) then there is some preference for  $A_{\text{recon}} > 1$ . The significances of these preferences are shown as ranging from  $1.8\sigma$  to  $2.6\sigma$ . From these significance levels and from the blue, orange, and green curves we see that we get a very similar result whether we use Planck, ACT DR6, or SPT lensing reconstructions, with the ACT DR6 and SPT curves nearly overlapping. If we remove DESI BAO data<sup>8</sup> (a case not shown in the figure), the preferences for  $A_{\text{recon}} > 1$  persist, although with somewhat reduced significance ranging from  $1.4\sigma$  to  $2.2\sigma$ . The preference for  $A_{\text{recon}} > 1$  is weakened much more if we fix  $A_{2\text{pt}}$  to 1 rather than marginalize over it (dashed curves); any evidence for excess reconstructed lensing power is then quite weak.

This dependence of the  $A_{\text{recon}}$  posteriors on the treatment of  $A_{2\text{pt}}$  (fixing to 1 or marginalizing) makes sense given the correlation we see in the lower left panel. The

<sup>7</sup> Recall that in this paper the  $\Lambda$ CDM model has  $\Sigma m_\nu$  fixed to 0.058 eV.

<sup>8</sup> When removing DESI we also add back in BAO data we had removed (see Tab. III), namely the SDSS DR16 LRG, ELG, and QSO BAO [65] and the DR12 BAO bin with  $z_{\text{eff}} = 0.61$  [64].

lower the assumed  $A_{2\text{pt}}$ , the lower the distribution of  $A_{\text{recon}}$  values. Physically, this correlation emerges because as  $A_{2\text{pt}}$  is increased, the  $\Lambda\text{CDM}$  parameters used to fit the Planck T&E data adjust in a way that decreases lensing power (both  $\Omega_{\text{m}}h^2$  and  $A_s$  decrease). With the  $\Lambda\text{CDM}$  lensing power decreased, a given level of reconstructed CMB lensing power needs a larger  $A_{\text{recon}}$  to fit it.

Readers who are familiar with [89] may be surprised at this sensitivity of  $A_{\text{recon}}$  to choice of treatment of  $A_{2\text{pt}}$  (marginalizing it or fixing it to 1), given arguments made in that paper about the *un*importance of the 2-point lensing anomaly to their conclusion of excess lensing power. But these two different results do not contradict each other. Green & Meyers [89] define a slightly  $L$ -dependent lensing template parameter,  $A_L(\Sigma\tilde{m}_\nu)$ , and an  $L$ -independent template parameter,  $B_{\text{lens}}$ , which are approximately related to our parameters by  $A_{\text{recon}} = A_L(\Sigma\tilde{m}_\nu)$  and  $A_{2\text{pt}} = A_L(\Sigma\tilde{m}_\nu)B_{\text{lens}}$ .<sup>9</sup> The resulting correlation between  $B_{\text{lens}}$  and  $\Sigma\tilde{m}_\nu$  is much smaller than the correlation we see in Fig. 19, which explains why they see the posterior for  $\Sigma\tilde{m}_\nu$  not change much if they switch from fixing  $B_{\text{lens}} = 1$  to marginalizing over it.

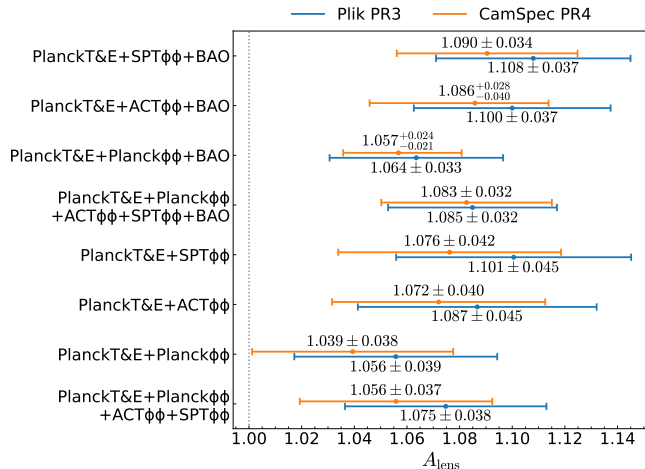


FIG. 20. Constraints on the lensing power phenomenological scaling parameter  $A_{\text{lens}}$  (68% C.L.) using Planck PR3 Plik T&E likelihood (in blue) and CamSpec PR4 T&E likelihood (in orange).

We have a strong theoretical prior that the lensing power one would infer from lensed CMB spectra (given the right cosmological model) should be equal, within the expected uncertainties, to what one would infer in a CMB lensing reconstruction.<sup>10</sup> So, even though the bulk

of the probability lies to one side of the  $A_{\text{recon}} = A_{2\text{pt}}$  line, it is worth exploring what we find if we force these two parameters to be equal. This is the model space we explore with an even greater variety of dataset combinations in Fig. 20, and it is also very similar to the model space chiefly studied by Green & Meyers [89] (with their  $B_{\text{lens}} = 1$ ).

One of the dataset variations we consider throughout Fig. 20 is the replacement of the Planck PR3 Plik T&E likelihood with the CamSpec PR4 T&E likelihood [91].<sup>11</sup> The Planck PR4 maps use more timestream data than does PR3 and the CamSpec PR4 T&E likelihood uses more sky than does the Plik PR3 T&E likelihood. If the  $A_{2\text{pt}}$  anomaly is driven by some fluctuations that drive the spectra away from the mean (in a manner similar to enhanced lensing) then we would expect the support for it to go down as we add in more data. For Plik PR3 we have, from  $TT, TE, EE + \text{lowE}$ ,  $A_L = 1.180 \pm 0.065$  [90]. With the additional PR4 data, we instead have  $A_L = 1.095 \pm 0.056$  [91]. Note that there is one other PR4 T&E likelihood [92], in which they find an even lower  $A_L = 1.039 \pm 0.052$ .

The lower half of the entries in Fig. 20 do not include any BAO data. We see that without the BAO data any conclusion of  $A_{\text{lens}} > 1$  by more than  $2\sigma$  is not very robust, occurring only for one of the dataset combinations. With the addition of BAO data (top half of the figure), such a conclusion becomes much more robust. The combinations that include Planck T&E, BAO, and all three lensing reconstructions become robust to the difference between PR3 and PR4. With PR3 specifically we find PlanckT&E+BAO+Planck $\phi\phi$ +ACT $\phi\phi$ +SPT $\phi\phi$  yields

$$A_{\text{lens}} = 1.085 \pm 0.032 \quad (33)$$

and a  $2.7\sigma$  exclusion of  $A_{\text{lens}} < 1$ . Note that the constraints are weakened by removing DESI from the BAO data set. Also, if one does so, the results become dependent again on the choice of Planck T&E. Removing DESI (in the same manner as done above) and switching from Planck PR3 T&E to Planck PR4 T&E we find  $A_{\text{lens}} = 1.059 \pm 0.029$ .

We have checked that switching to the  $L$ -dependent lensing template used in [36, 89] in place of  $A_{\text{lens}}$  delivers a similar result. Their template depends on their neutrino-mass-like parameter  $\Sigma\tilde{m}_\nu$ . For PlanckT&E(PR3)+BAO+Planck $\phi\phi$ +ACT $\phi\phi$ +SPT $\phi\phi$ , we find

$$\Sigma\tilde{m}_\nu = -0.122 \pm 0.072 \text{ eV} \quad (34)$$

and a  $2.7\sigma$  exclusion of  $\Sigma\tilde{m}_\nu > 0.058 \text{ eV}$ .

We note that this evidence for excess lensing power arises from 3 types of constraints: 1) lensing power measurements, 2) inferences of  $\Lambda\text{CDM}$  parameters (other

<sup>9</sup> This is approximate because 1) we have ignored the slight  $L$ -dependence of  $A_L(\Sigma\tilde{m}_\nu)$  and 2) their template is defined with respect to  $\Lambda\text{CDM}$  models with  $\Sigma m_\nu = 0 \text{ eV}$ .

<sup>10</sup> Although there are models in which this is violated too; see Craig et al. [36].

<sup>11</sup> This replacement does not change the low- $\ell$   $TT$  and  $EE$  likelihood, which remain PR3.

than  $\tau$ ) from the combination of CMB and BAO data, and 3) inference of  $\tau$  from large angular-scale CMB polarization. We have seen that the preference for  $A_{\text{lens}} > 1$  is robust to selection of different datasets for (1). Of particular relevance for this paper, the preference is present when the Planck and ACT lensing measurements are replaced with the SPT one. The SPT lensing is inferred entirely from polarization, making it much less susceptible to possible biases from extragalactic foregrounds. Regarding (2), we see that a more than  $2\sigma$  preference for  $A_{\text{lens}} > 1$  requires either BAO data (including DESI) or PR3 T&E. The next release of BAO data from DESI should be informative, as will the forthcoming lensed  $TT$ ,  $TE$ , and  $EE$  spectra from SPT-3G. Note that  $\tau$  (item 3 above) is inferred from Planck low- $\ell$   $EE$  data. The importance of  $\tau$  comes from its implications for  $A_s$  as the combination  $A_s e^{-2\tau}$  is very well determined by CMB data: a higher  $\tau$  would lead to higher  $A_s$  and hence a higher lensing power prediction. Independent and competitive constraints on  $\tau$  may come via the kinematic SZ effect [11, 93].

The conclusion that there is an excess of lensing power also relies on assumption of the  $\Lambda$ CDM model. Craig et al. [36] introduced a number of changes to the model that could possibly restore concordance. Lynch et al. [94] noted that non-standard recombination can simultaneously reduce the  $H_0$  tension and lead to a higher predicted lensing power. The DESI collaboration points out that with free  $w_0$  and  $w_a$  the constraints on  $\Sigma m_\nu$  are significantly relaxed [66].

## J. Other $\Lambda$ CDM extensions

We also explore  $\Lambda$ CDM extensions with curvature, the density of light relics, primordial abundance of helium, and a time-varying dark energy equation of state. Constraints on additional parameters are summarized in Table V.

For the  $\Lambda$ CDM extension with free curvature and for the ones with changing dark energy equation of state, we find that all CMB datasets are consistent within  $2\sigma$  with the fiducial values of the  $\Lambda$ CDM model. The constraints from different CMB datasets are consistent. Adding the BAO data significantly tightens the constraints. This is due to the BAO sensitivity to the expansion rate and the distance-redshift relation over redshifts when the dark energy begins to dominate the density.

The energy density of additional relativistic particles other than photons is parameterized by the effective number of neutrino species,  $N_{\text{eff}}$ , defined via

$$\rho_{\text{rad}} = \rho_\gamma \left[ 1 + N_{\text{eff}} \frac{7}{8} \left( \frac{4}{11} \right)^{4/3} \right], \quad (35)$$

where  $\rho_\gamma$  is the photon energy density. Assuming the standard cosmological model,  $N_{\text{eff}} = 3.044$  [96–98]. The CMB power spectrum is sensitive to  $N_{\text{eff}}$  [99–103].

By increasing the density of light relics, one increases the expansion rate during the radiation-dominated era. To keep the precisely measured angular scale of the sound horizon unchanged, the expansion rate after last scattering also increases [e.g. 104]. Therefore, there is a positive correlation between  $H_0$  and  $N_{\text{eff}}$ . The marginal posterior distribution of  $N_{\text{eff}}$  and  $H_0$  is shown in the left panel of Fig. 21. Combining Planck, SPT, ACT, and BAO, we get

$$N_{\text{eff}} = 2.83 \pm 0.13 \quad (36)$$

$$H_0 = 66.52 \pm 0.87 \text{ km/s/Mpc} \quad (37)$$

where  $N_{\text{eff}}$  is consistent with 3.044 from the standard model prediction, but  $H_0$  is still in  $5.4\sigma$  tension with the SH0ES [74] measurement.

One can further extend the model to allow for a free  $\Sigma m_\nu$  as well. The marginal posterior of  $N_{\text{eff}}$  and  $\Sigma m_\nu$  is shown in the middle panel of Fig. 21. There is nearly no degeneracy between  $N_{\text{eff}}$  and  $\Sigma m_\nu$ . Adding SPT data to Planck improves the constraints on  $N_{\text{eff}}$  slightly and  $\Sigma m_\nu$  by about a factor of 2 as discussed in Sect. V H. The lensing information in SPT data is useful for constraining the neutrino mass sum, as discussed in Sect. V H, and the improvement of the  $N_{\text{eff}}$  constraint is mostly from the more precisely measured  $EE$  bandpowers.

The density of light relics is also partially degenerate with the primordial fraction of baryonic mass in helium,  $Y_{\text{P}}$ . Both  $N_{\text{eff}}$  and  $Y_{\text{P}}$  affect the photon diffusion damping of the CMB power spectra at small angular scales. In the right panel of Fig. 21, we show the marginal posterior of  $N_{\text{eff}}$  and  $Y_{\text{P}}$ . Compared to the constraints using only Planck data, the uncertainty is reduced by  $\sim 41\%$  with the addition of SPT data. When we fit jointly for  $N_{\text{eff}}$  and  $Y_{\text{P}}$  using Planck, ACT, SPT, and BAO measurements, we find values of  $N_{\text{eff}}$  that are consistent with the standard model prediction of 3.044,  $N_{\text{eff}} - Y_{\text{P}}$  contours that are consistent with BBN predictions, and  $Y_{\text{P}}$  constraints that are consistent with direct measurements of HII regions in metal-poor galaxies [95].

## VI. CONCLUSIONS

In this paper, we presented an optimal joint inference of the CMB lensing potential power spectrum and the unlensed CMB  $EE$  power spectrum, derived entirely from CMB polarization maps. These maps were made from observations of  $1500 \text{ deg}^2$  taken during the SPT-3G 2019 and 2020 Austral winter observing seasons. The inference was performed using a Bayesian framework called the Marginal Unbiased Score Expansion (MUSE) method, which also enables propagation of systematic uncertainties into the estimated bandpower uncertainties. The inferred bandpowers have smaller uncertainties than previously published results at  $\ell > 2000$  for the  $EE$  spectrum and at  $L > 350$  for the lensing spectrum, with a  $38\sigma$  detection of lensing power.

Ext. Parameter	WMAP+SPT	Planck	Planck+SPT	Planck+ACT+SPT	Planck+ACT+SPT+BAO
$\Omega_k$	$-0.021^{+0.014}_{-0.011}$	$-0.0070 \pm 0.0056$	$-0.0098 \pm 0.0052$	$-0.0076^{+0.0053}_{-0.0047}$	$0.0014 \pm 0.0014$
$w_0$	$-1.42^{+0.29}_{-0.47}$	$-1.48^{+0.20}_{-0.37}$	$-1.52^{+0.18}_{-0.35}$	$-1.52^{+0.19}_{-0.43}$	$-1.075 \pm 0.046$
$w_0$	$-1.10^{+0.30}_{-0.69}$	$-1.18^{+0.46}_{-0.62}$	$-1.16^{+0.43}_{-0.60}$	$-1.18^{+0.48}_{-0.58}$	$-0.65 \pm 0.23$
$w_a$	$-1.1^{+0.6}_{-1.9}$	$-1.06^{+0.61}_{-1.94}$	$-1.18^{+0.52}_{-1.82}$	$-1.21^{+0.57}_{-1.79}$	$-1.15^{+0.71}_{-0.59}$
$N_{\text{eff}}$	$2.95 \pm 0.33$	$2.89 \pm 0.18$	$2.85 \pm 0.16$	$2.66 \pm 0.14$	$2.86 \pm 0.13$
$\Sigma m_\nu$	$< 0.38 \text{ eV}$	$< 0.31 \text{ eV}$	$< 0.17 \text{ eV}$	$< 0.20 \text{ eV}$	$< 0.075 \text{ eV}$
$N_{\text{eff}}$	$2.95 \pm 0.32$	$2.89 \pm 0.19$	$2.84 \pm 0.17$	$2.67 \pm 0.14$	$2.83 \pm 0.13$
$\Sigma m_\nu$	$< 0.38 \text{ eV}$	$< 0.31 \text{ eV}$	$< 0.17 \text{ eV}$	$< 0.21 \text{ eV}$	$< 0.061 \text{ eV}$
$N_{\text{eff}}$	$2.79 \pm 0.52$	$2.8 \pm 0.3$	$2.77 \pm 0.26$	$2.60 \pm 0.23$	$2.89 \pm 0.23$
$Y_{\text{P}}$	$0.256 \pm 0.028$	$0.248 \pm 0.018$	$0.250 \pm 0.016$	$0.246 \pm 0.014$	$0.241 \pm 0.015$

TABLE V. Summary of constraints on selected extensions to the  $\Lambda$ CDM model. The errors indicate the 68% confidence region, and the upper limits on the sum of neutrino masses are 95% upper limits.

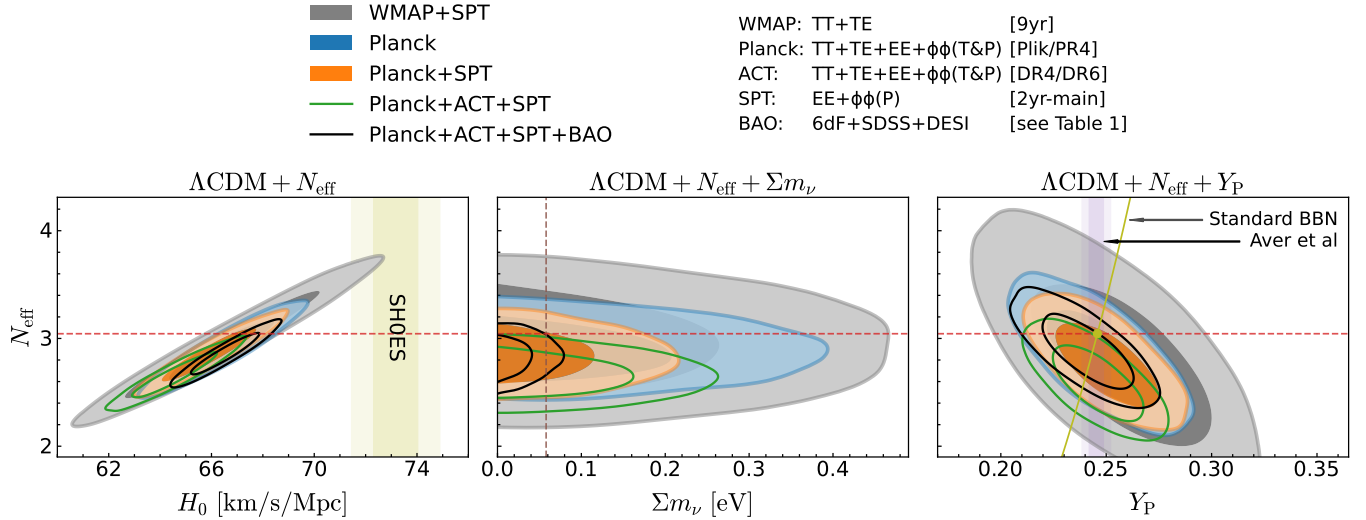


FIG. 21. Marginal posterior distributions of the  $\Lambda$ CDM extension models involving  $N_{\text{eff}}$ . The horizontal red dashed line shows  $N_{\text{eff}} = 3.044$ . The yellow vertical band in the left panel shows the constraint on  $H_0$  from Breuval et al. [74]. The vertical brown dashed line shows  $\Sigma m_\nu = 0.058 \text{ eV}$ . The purple vertical band shows the inference of the primordial helium fraction from Aver et al. [95].

After checking that the  $\Lambda$ CDM model provides a good fit to the SPT data (combined with a prior on  $\tau$ ), we reported constraints on the  $\Lambda$ CDM parameters. We found  $H_0 = 66.81 \pm 0.81 \text{ km/s/Mpc}$ , a significantly tighter constraint than from prior polarization-only measurements. It is consistent with the Planck  $TT/TE/EE$  plus Planck lensing constraints assuming  $\Lambda$ CDM, and in  $5.4\sigma$  tension with the most precise distance ladder measurement [74]. The SPT constraints on  $\Omega_c h^2$  and  $\Omega_b h^2$  are also tighter than from prior polarization-only or polarization plus CMB lensing measurements. We also found  $S_8 = 0.850 \pm 0.017$ , consistent with the result from Planck data and with comparable uncertainty.

We estimated  $\Lambda$ CDM parameters for several other CMB datasets and checked for consistency between these estimates. These consistency tests are stringent tests of the  $\Lambda$ CDM model. All differences investigated were within expectations given the measurement uncertainties and the  $\Lambda$ CDM model. The comparison between

SPT and Planck provides a particularly interesting test of the  $\Lambda$ CDM model, because the SPT constraints are more heavily weighted toward small-scale polarization power, and especially lensing power, than is the case for Planck.

We estimated  $\Lambda$ CDM parameters from SPT data in combination with other CMB datasets. When combined with primary CMB and lensing from Planck and ACT, we find  $H_0 = 67.33 \pm 0.37 \text{ km/s/Mpc}$ , a  $6.2\sigma$  tension with the distance ladder [74], and  $S_8 = 0.8380 \pm 0.0084$ , in  $3.3\sigma$  tension with  $S_8$  from the  $3 \times 2$ pt analysis of KiDS-1000 [84] and also from the  $3 \times 2$ pt analysis of DES-Y3 [85].

Motivated by the work of Amon & Efstathiou [37] and Preston et al. [86] regarding the  $S_8$  tension, we investigated the sensitivity of our lensing measurement to non-linear corrections. By introducing the  $A_{\text{mod}}^{\text{CMB}}$  parameter that scales the amplitude of non-linear corrections to the CMB lensing spectrum, we found  $A_{\text{mod}}^{\text{CMB}} = 1.60 \pm 0.39$ . This is the first  $> 3\sigma$  “detection” of the non-linear cor-

rection for CMB lensing, and is  $\sim 2\sigma$  higher than the value needed for resolving the  $S_8$  tension between galaxy lensing and CMB measurements. Our data clearly do not favor any suppression of the non-linear contribution, though the uncertainties are still large. A suppressed  $S_8$  could arise at lower redshifts and from slightly smaller scales than the ones we are probing.

The CMB lensing potential is also sensitive to the neutrino mass sum,  $\Sigma m_\nu$ . We investigated constraints on  $\Sigma m_\nu$  with the SPT lensing spectrum. In the  $\Lambda$ CDM+ $\Sigma m_\nu$  model, we find  $\Sigma m_\nu < 0.20$  eV (95% C.L.) from the joint fit to Planck, ACT, and SPT data. Adding BAO data, the constraint is further tightened to  $\Sigma m_\nu < 0.075$  eV (95% C.L.), placing pressure on the inverted hierarchy minimal mass sum of 0.098 eV.

With the addition of BAO data to the CMB datasets, the strongest challenge to the standard model we see in our results comes from an investigation of lensing power amplitudes. We introduced  $A_{2\text{pt}}$  to scale the  $\Lambda$ CDM model CMB lensing spectrum used to calculate lensed model  $TT/TE/EE$  spectra and  $A_{\text{recon}}$  to scale that same  $\Lambda$ CDM model CMB lensing spectrum, but for comparison to the reconstructed CMB lensing spectrum instead. With both of these parameters free, the  $\Lambda$ CDM point in the  $A_{\text{recon}}, A_{2\text{pt}}$  plane is outside the 95% confidence region for a variety of dataset combinations. After marginalization over  $A_{2\text{pt}}$ , Planck lensing, ACT DR6 lensing, and SPT lensing all favor  $A_{\text{recon}} > 1$  at  $\sim 2\sigma$ .

We also set  $A_{\text{recon}} = A_{2\text{pt}}$  and called the single template parameter  $A_{\text{lens}}$ . We investigated the posterior distribution of  $A_{\text{lens}}$  for a variety of datasets. For our most comprehensive combination of datasets we find  $A_{\text{lens}} = 1.085 \pm 0.032$  and a  $2.7\sigma$  exclusion of  $A_{\text{lens}} < 1$ . Following Craig et al. [36] and Green & Meyers [89], we found a similar constraint on their neutrino-mass-like  $L$ -dependent lensing template parameter,  $\Sigma \tilde{m}_\nu$ . For the same dataset we found a  $2.7\sigma$  exclusion of  $\Sigma \tilde{m}_\nu < 0.058$  eV. We also saw that these constraints are weakened by dropping DESI BAO, or by switching from Planck PR3 T&E data to Planck PR4 T&E data. Doing both we find  $A_{\text{lens}} = 1.059 \pm 0.029$ .

We also discussed the cosmological inference with SPT CMB lensing potential bandpowers and the unlensed  $EE$  bandpowers on other one- and two-parameter extensions of the  $\Lambda$ CDM model. These model extensions include varying the amount of spatial curvature, the dark energy equation of state and its evolution, the density of light relics, and the primordial fraction of baryonic mass in helium. The constraints on the extended parameters are summarized in Tab. V. Combining the SPT results with Planck, ACT, and BAO, we saw no significant preference for any of these extensions.

Although the 2019-2020 SPT-3G CMB polarization data analyzed here primarily constrain  $H_0$  and  $S_8$ , SPT-3G constraints on other  $\Lambda$ CDM parameters and extensions will be improved significantly in the near future. The addition of temperature information to both the power spectrum and lensing analyses will significantly

improve the constraints on the inferred cosmological parameters, in particular the effective number of neutrino species,  $\Sigma m_\nu$ , and curvature [3]. With polarized beam uncertainties identified as an important contribution to our systematic error budget, we have planned additional deep observations of polarized point sources to better map out the polarized beam response and reduce this source of uncertainty in future analyses. The 1500 deg<sup>2</sup> field has since been observed for three more seasons. The noise level of the full 2019-2023 1500 deg<sup>2</sup> data is expected to be 1.7 times lower than the 2019-20 observations, and the figure of merit for  $\Lambda$ CDM parameters is expected to be 2 times better [3]. The addition of  $\sim 8500$  deg<sup>2</sup> more sky mapped with SPT-3G by the end of 2024 will strengthen constraints even further, eventually surpassing those from Planck [3]. The analysis of these data, particularly the ultra-deep 1500 deg<sup>2</sup> data, will benefit even more from optimal analyses such as the MUSE method demonstrated here. This work has demonstrated the strength of joint Bayesian CMB lensing potential bandpower reconstruction and unlensed CMB  $EE$  bandpower estimation, in particular its ability to propagate systematic uncertainties, and paves the way for future analyses of SPT-3G observations and similarly deep datasets in the future.

## ACKNOWLEDGMENTS

The South Pole Telescope program is supported by the National Science Foundation (NSF) through awards OPP-1852617 and OPP-2332483. Partial support is also provided by the Kavli Institute of Cosmological Physics at the University of Chicago. 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, the Infinity Cluster hosted by Institut d’Astrophysique de Paris and the Peloton Cluster of the High Performance Computing Core Facility at the University of California, Davis. We acknowledge the computing resources provided on Swing and Crossover, high-performance computing clusters operated by the Laboratory Computing Resource Center at Argonne National Laboratory. The UC Davis group acknowledges support from Michael and Ester Vaida. Argonne National Laboratory’s work was supported by the U.S. Department of Energy, Office of High Energy Physics, under contract DE-AC02-06CH11357. The Paris group has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (grant agreement No 101001897), and

funding from the Centre National d'Etudes Spatiales. The Melbourne authors acknowledge support from the Australian Research Council's Discovery Project scheme (No. DP210102386). Work at Fermi National Accelerator Laboratory, a DOE-OS, HEP User Facility managed

by the Fermi Research Alliance, LLC, was supported under Contract No. DE-AC02-07CH11359. The SLAC group is supported in part by the Department of Energy at SLAC National Accelerator Laboratory, under contract DE-AC02-76SF00515.

- 
- [1] Aghanim, N., et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. 2020, *Astron. Astrophys.*, 641, A1, doi: [10.1051/0004-6361/201833880](https://doi.org/10.1051/0004-6361/201833880)
- [2] Sobrin, J. A., Anderson, A. J., Bender, A. N., et al. The Design and Integrated Performance of SPT-3G. 2022, *Ap. J. Suppl.*, 258, 42, doi: [10.3847/1538-4365/ac374f](https://doi.org/10.3847/1538-4365/ac374f)
- [3] Prabhu, K., et al. Testing the  $\Lambda$ CDM Cosmological Model with Forthcoming Measurements of the Cosmic Microwave Background with SPT-3G. 2024. <https://arxiv.org/abs/2403.17925>
- [4] Ade, P. A. R., Ahmed, Z., Amiri, M., et al. Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season. 2021, *Phys. Rev. Lett.*, 127, 151301, doi: [10.1103/PhysRevLett.127.151301](https://doi.org/10.1103/PhysRevLett.127.151301)
- [5] Guns, S., Foster, A., Daley, C., et al. Detection of Galactic and Extragalactic Millimeter-wavelength Transient Sources with SPT-3G. 2021, *Astrophys. J.*, 916, 98, doi: [10.3847/1538-4357/ac06a3](https://doi.org/10.3847/1538-4357/ac06a3)
- [6] Chaubal, P. S., Reichardt, C. L., Gupta, N., et al. Improving Cosmological Constraints from Galaxy Cluster Number Counts with CMB-cluster-lensing Data: Results from the SPT-SZ Survey and Forecasts for the Future. 2022, *Astrophys. J.*, 931, 139, doi: [10.3847/1538-4357/ac6a55](https://doi.org/10.3847/1538-4357/ac6a55)
- [7] Ferguson, K. R., Anderson, A. J., Whitehorn, N., et al. Searching for axionlike time-dependent cosmic birefringence with data from SPT-3G. 2022, *Phys. Rev. D*, 106, 042011, doi: [10.1103/PhysRevD.106.042011](https://doi.org/10.1103/PhysRevD.106.042011)
- [8] Schiappucci, E., Bianchini, F., Aguena, M., et al. Measurement of the mean central optical depth of galaxy clusters via the pairwise kinematic Sunyaev-Zel'dovich effect with SPT-3G and DES. 2023, *Phys. Rev. D*, 107, 042004, doi: [10.1103/PhysRevD.107.042004](https://doi.org/10.1103/PhysRevD.107.042004)
- [9] Ansarinejad, B., Raghunathan, S., Abbott, T. M. C., et al. Mass calibration of DES Year-3 clusters via SPT-3G CMB cluster lensing. 2024, *J. of Cosm. & Astropart. Phys.*, 2024, 024, doi: [10.1088/1475-7516/2024/07/024](https://doi.org/10.1088/1475-7516/2024/07/024)
- [10] Tandoi, C., Guns, S., Foster, A., et al. Flaring Stars in a Nontargeted Millimeter-wave Survey with SPT-3G. 2024, *Astrophys. J.*, 972, 6, doi: [10.3847/1538-4357/ad58db](https://doi.org/10.3847/1538-4357/ad58db)
- [11] Raghunathan, S., Ade, P. A. R., Anderson, A. J., et al. 2024, First Constraints on the Epoch of Reionization Using the Non-Gaussianity of the Kinematic Sunyaev-Zel'dovich Effect from the South Pole Telescope and  $\{it\}$  Herschel $\}$ -SPIRE Observations, arXiv, doi: [10.48550/arXiv.2403.02337](https://doi.org/10.48550/arXiv.2403.02337)
- [12] Dutcher, D., et al. Measurements of the E-mode polarization and temperature-E-mode correlation of the CMB from SPT-3G 2018 data. 2021, *Phys. Rev. D*, 104, 022003, doi: [10.1103/PhysRevD.104.022003](https://doi.org/10.1103/PhysRevD.104.022003)
- [13] Balkenhol, L., et al. Constraints on  $\Lambda$ CDM extensions from the SPT-3G 2018 EE and TE power spectra. 2021, *Phys. Rev. D*, 104, 083509, doi: [10.1103/PhysRevD.104.083509](https://doi.org/10.1103/PhysRevD.104.083509)
- [14] —. Measurement of the CMB temperature power spectrum and constraints on cosmology from the SPT-3G 2018 TT, TE, and EE dataset. 2023, *Phys. Rev. D*, 108, 023510, doi: [10.1103/PhysRevD.108.023510](https://doi.org/10.1103/PhysRevD.108.023510)
- [15] Pan, Z., et al. Measurement of gravitational lensing of the cosmic microwave background using SPT-3G 2018 data. 2023, *Phys. Rev. D*, 108, 122005, doi: [10.1103/PhysRevD.108.122005](https://doi.org/10.1103/PhysRevD.108.122005)
- [16] Millea, M., & Seljak, U. Marginal unbiased score expansion and application to CMB lensing. 2022, *Phys. Rev. D*, 105, 103531, doi: [10.1103/PhysRevD.105.103531](https://doi.org/10.1103/PhysRevD.105.103531)
- [17] Millea, M. 2022, Improved Marginal Unbiased Score Expansion (MUSE) via Implicit Differentiation, arXiv, doi: [10.48550/arXiv.2209.10512](https://doi.org/10.48550/arXiv.2209.10512)
- [18] Hu, W., & Okamoto, T. Mass reconstruction with cmb polarization. 2002, *Astrophys. J.*, 574, 566, doi: [10.1086/341110](https://doi.org/10.1086/341110)
- [19] Hotinli, S. C., Meyers, J., Trendafilova, C., Green, D., & van Engelen, A. The benefits of CMB delensing. 2022, *JCAP*, 04, 020, doi: [10.1088/1475-7516/2022/04/020](https://doi.org/10.1088/1475-7516/2022/04/020)
- [20] Planck Collaboration, Aghanim, N., Akrami, Y., et al. Planck 2018 Results. VIII. Gravitational Lensing. 2018, arXiv:1807.06210 [astro-ph]. <https://arxiv.org/abs/1807.06210>
- [21] Han, D., et al. The Atacama Cosmology Telescope: delensed power spectra and parameters. 2021, *JCAP*, 01, 031, doi: [10.1088/1475-7516/2021/01/031](https://doi.org/10.1088/1475-7516/2021/01/031)
- [22] Hirata, C. M., & Seljak, U. Analyzing Weak Lensing of the Cosmic Microwave Background Using the Likelihood Function. 2003, *Physical Review D*, 67, doi: [10.1103/PhysRevD.67.043001](https://doi.org/10.1103/PhysRevD.67.043001)
- [23] Carron, J., & Lewis, A. Maximum a Posteriori CMB Lensing Reconstruction. 2017, *Physical Review D*, 96, 063510, doi: [10.1103/PhysRevD.96.063510](https://doi.org/10.1103/PhysRevD.96.063510)
- [24] Millea, M., Anderes, E., & Wandelt, B. D. Bayesian Delensing of CMB Temperature and Polarization. 2019, *Physical Review D*, 100, 023509, doi: [10.1103/PhysRevD.100.023509](https://doi.org/10.1103/PhysRevD.100.023509)
- [25] —. Sampling-based inference of the primordial CMB and gravitational lensing. 2020, *Phys. Rev. D*, 102, 123542, doi: [10.1103/PhysRevD.102.123542](https://doi.org/10.1103/PhysRevD.102.123542)
- [26] Millea, M., Daley, C. M., Chou, T. L., et al. Optimal Cosmic Microwave Background Lensing Reconstruction and Parameter Estimation with SPTpol Data. 2021, *Astrophys. J.*, 922, 259, doi: [10.3847/1538-4357/ac02bb](https://doi.org/10.3847/1538-4357/ac02bb)
- [27] Balkenhol, L., et al. Measurement of the CMB temperature power spectrum and constraints on cosmology from the SPT-3G 2018 TT, TE, and EE dataset. 2023, *Phys. Rev. D*, 108, 023510, doi: [10.1103/PhysRevD.108.023510](https://doi.org/10.1103/PhysRevD.108.023510)

- [28] Planck Collaboration, Aghanim, N., Akrami, Y., et al. Planck 2018 results. I. Overview and the cosmological legacy of Planck. 2020, *A&A*, 641, A1, doi: [10.1051/0004-6361/201833880](https://doi.org/10.1051/0004-6361/201833880)
- [29] Madhavacheril, M. S., et al. The Atacama Cosmology Telescope: DR6 Gravitational Lensing Map and Cosmological Parameters. 2024, *Astrophys. J.*, 962, 113, doi: [10.3847/1538-4357/acff5f](https://doi.org/10.3847/1538-4357/acff5f)
- [30] Qu, F. J., et al. The Atacama Cosmology Telescope: A Measurement of the DR6 CMB Lensing Power Spectrum and Its Implications for Structure Growth. 2024, *Astrophys. J.*, 962, 112, doi: [10.3847/1538-4357/acfe06](https://doi.org/10.3847/1538-4357/acfe06)
- [31] Aiola, S., Calabrese, E., Maurin, L., et al. The Atacama Cosmology Telescope: DR4 Maps and Cosmological Parameters. 2020, arXiv e-prints, 2007, arXiv:2007.07288. <http://adsabs.harvard.edu/abs/2020arXiv200707288A>
- [32] Choi, S. K., et al. The Atacama Cosmology Telescope: a measurement of the Cosmic Microwave Background power spectra at 98 and 150 GHz. 2020, *JCAP*, 12, 045, doi: [10.1088/1475-7516/2020/12/045](https://doi.org/10.1088/1475-7516/2020/12/045)
- [33] Di Valentino, E., Mena, O., Pan, S., Visinelli, L., Yang, W., Melchiorri, A., Mota, D. F., Riess, A. G., & Silk, J. In the realm of the Hubble tension—a review of solutions. 2021, *Class. Quant. Grav.*, 38, 153001, doi: [10.1088/1361-6382/ac086d](https://doi.org/10.1088/1361-6382/ac086d)
- [34] Schöneberg, N., Franco Abellán, G., Pérez Sánchez, A., Witte, S. J., Poulin, V., & Lesgourgues, J. The H0 Olympics: A fair ranking of proposed models. 2022, *Phys. Rept.*, 984, 1, doi: [10.1016/j.physrep.2022.07.001](https://doi.org/10.1016/j.physrep.2022.07.001)
- [35] Khalife, A. R., Zanjani, M. B., Galli, S., Günther, S., Lesgourgues, J., & Benabed, K. Review of Hubble tension solutions with new SH0ES and SPT-3G data. 2024, *JCAP*, 04, 059, doi: [10.1088/1475-7516/2024/04/059](https://doi.org/10.1088/1475-7516/2024/04/059)
- [36] Craig, N., Green, D., Meyers, J., & Rajendran, S. No  $\nu_s$  is Good News. 2024. <https://arxiv.org/abs/2405.00836>
- [37] Amon, A., & Efstathiou, G. A non-linear solution to the  $S_8$  tension? 2022, *Mon. Not. Roy. Astron. Soc.*, 516, 5355, doi: [10.1093/mnras/stac2429](https://doi.org/10.1093/mnras/stac2429)
- [38] Hahn, C., Vakili, M., Walsh, K., Hearin, A. P., Hogg, D. W., & Campbell, D. 2017, Approximate Bayesian Computation in Large Scale Structure: Constraining the Galaxy-Halo Connection, doi: [10.1093/mnras/stx894](https://doi.org/10.1093/mnras/stx894)
- [39] Levasseur, L. P., Hezaveh, Y. D., & Wechsler, R. H. 2017, Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing, doi: [10.3847/2041-8213/aa9704](https://doi.org/10.3847/2041-8213/aa9704)
- [40] Coogan, A., Karchev, K., & Weniger, C. 2020, Targeted Likelihood-Free Inference of Dark Matter Substructure in Strongly-Lensed Galaxies, arXiv, doi: [10.48550/arXiv.2010.07032](https://doi.org/10.48550/arXiv.2010.07032)
- [41] Legin, R., Hezaveh, Y., Levasseur, L. P., & Wandelt, B. 2022, Simulation-Based Inference of Strong Gravitational Lensing Parameters, arXiv, doi: [10.48550/arXiv.2112.05278](https://doi.org/10.48550/arXiv.2112.05278)
- [42] Lin, K., von wietersheim-Kramsta, M., Joachimi, B., & Feeney, S. A Simulation-Based Inference Pipeline for Cosmic Shear with the Kilo-Degree Survey. 2023, *Monthly Notices of the Royal Astronomical Society*, 524, 6167, doi: [10.1093/mnras/stad2262](https://doi.org/10.1093/mnras/stad2262)
- [43] Lemos, P., Parker, L. H., Hahn, C., Ho, S., Eickenberg, M., Hou, J., Massara, E., Modi, C., Moradinezhad Dizgah, A., Régalo-Saint Blancard, B., & Spergel, D. 2023, SimBIG: Field-level Simulation-based Inference of Large-scale Structure (eprint: arXiv:2310.15256), 18, doi: [10.48550/arXiv.2310.15256](https://doi.org/10.48550/arXiv.2310.15256)
- [44] Karchev, K., Trotta, R., & Weniger, C. SECRET: Supernova Ia Cosmology with Truncated Marginal Neural Ratio Estimation. 2023, *Monthly Notices of the Royal Astronomical Society*, 520, 1056, doi: [10.1093/mnras/stac3785](https://doi.org/10.1093/mnras/stac3785)
- [45] Charnock, T., Lavaux, G., & Wandelt, B. D. Automatic Physical Inference with Information Maximizing Neural Networks. 2018, *Physical Review D*, 97, 083004, doi: [10.1103/PhysRevD.97.083004](https://doi.org/10.1103/PhysRevD.97.083004)
- [46] Rubin, D. B. Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician. 1984, *The Annals of Statistics*, 12, 1151, doi: [10.1214/aos/1176346785](https://doi.org/10.1214/aos/1176346785)
- [47] Beaumont, M. A., Zhang, W., & Balding, D. J. Approximate Bayesian Computation in Population Genetics. 2002, *Genetics*, 162, 2025
- [48] Cranmer, K., Brehmer, J., & Louppe, G. The frontier of simulation-based inference. 2020, *Proceedings of the National Academy of Science*, 117, 30055, doi: [10.1073/pnas.1912789117](https://doi.org/10.1073/pnas.1912789117)
- [49] Dunkley, J., Calabrese, E., Sievers, J., et al. The Atacama Cosmology Telescope: Likelihood for Small-Scale CMB Data. 2013, *Journal of Cosmology and Astroparticle Physics*, 2013, 025, doi: [10.1088/1475-7516/2013/07/025](https://doi.org/10.1088/1475-7516/2013/07/025)
- [50] Prince, H., & Dunkley, J. Data Compression in Cosmology: A Compressed Likelihood for Planck Data. 2019, *Physical Review D*, 100, 083502, doi: [10.1103/PhysRevD.100.083502](https://doi.org/10.1103/PhysRevD.100.083502)
- [51] Prince, H., Calabrese, E., & Dunkley, J. 2024, A Foreground-Marginalized 'BK-lite' Likelihood for the Tensor-to-Scalar Ratio, arXiv, doi: [10.48550/arXiv.2403.00085](https://doi.org/10.48550/arXiv.2403.00085)
- [52] Torrado, J., & Lewis, A. Cobaya: Code for Bayesian Analysis of hierarchical physical models. 2021, *JCAP*, 05, 057, doi: [10.1088/1475-7516/2021/05/057](https://doi.org/10.1088/1475-7516/2021/05/057)
- [53] Lewis, A., Challinor, A., & Lasenby, A. Efficient computation of CMB anisotropies in closed FRW models. 2000, *Astrophys. J.*, 538, 473, doi: [10.1086/309179](https://doi.org/10.1086/309179)
- [54] Qu, F. J., Millea, M., & Schaan, E. 2024, Impact & Mitigation of Polarized Extragalactic Foregrounds on Bayesian Cosmic Microwave Background Lensing, doi: [10.48550/arXiv.2406.15351](https://doi.org/10.48550/arXiv.2406.15351)
- [55] Lembo, M., Fabbian, G., Carron, J., & Lewis, A. CMB Lensing Reconstruction Biases from Masking Extragalactic Sources. 2022, *Physical Review D*, 106, 023525, doi: [10.1103/PhysRevD.106.023525](https://doi.org/10.1103/PhysRevD.106.023525)
- [56] Omori, Y. AGORA: Multicomponent Simulation for Cross-Survey Science. 2024, *Monthly Notices of the Royal Astronomical Society*, 530, 5030, doi: [10.1093/mnras/stae1031](https://doi.org/10.1093/mnras/stae1031)
- [57] Bonici, M., Bianchini, F., & Ruiz-Zapatero, J. Capse.jl: efficient and auto-differentiable CMB power spectra emulation. 2023. <https://arxiv.org/abs/2307.14339>
- [58] Chon, G., Challinor, A., Prunet, S., Hivon, E., & Szapudi, I. Fast estimation of polarization power spectra using correlation functions. 2004, *Mon. Not. Roy. Astron. Soc.*, 350, 914, doi: [10.1111/j.1365-2966.2004.](https://doi.org/10.1111/j.1365-2966.2004.)

07737. x
- [59] Aghanim, N., et al. Planck 2018 results. V. CMB power spectra and likelihoods. 2020, *Astron. Astrophys.*, 641, A5, doi: [10.1051/0004-6361/201936386](https://doi.org/10.1051/0004-6361/201936386)
- [60] Carron, J., Mirmelstein, M., & Lewis, A. CMB lensing from Planck PR4 maps. 2022, *JCAP*, 09, 039, doi: [10.1088/1475-7516/2022/09/039](https://doi.org/10.1088/1475-7516/2022/09/039)
- [61] Hinshaw, G., Larson, D., Komatsu, E., et al. Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. 2013, *Ap. J. Suppl.*, 208, 19, doi: [10.1088/0067-0049/208/2/19](https://doi.org/10.1088/0067-0049/208/2/19)
- [62] Beutler, F., Blake, C., Colless, M., Jones, D. H., Staveley-Smith, L., Campbell, L., Parker, Q., Saunders, W., & Watson, F. The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. 2011, *Monthly Notices of the Royal Astronomical Society*, 416, 3017
- [63] Ross, A. J., Samushia, L., Howlett, C., Percival, W. J., Burden, A., & Manera, M. The clustering of the SDSS DR7 main Galaxy sample – I. A 4 per cent distance measure at  $z = 0.15$ . 2015, *Mon. Not. Roy. Astron. Soc.*, 449, 835, doi: [10.1093/mnras/stv154](https://doi.org/10.1093/mnras/stv154)
- [64] Alam, S., et al. The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: cosmological analysis of the DR12 galaxy sample. 2017, *Mon. Not. Roy. Astron. Soc.*, 470, 2617, doi: [10.1093/mnras/stx721](https://doi.org/10.1093/mnras/stx721)
- [65] —. Completed SDSS-IV extended Baryon Oscillation Spectroscopic Survey: Cosmological implications from two decades of spectroscopic surveys at the Apache Point Observatory. 2021, *Phys. Rev. D*, 103, 083533, doi: [10.1103/PhysRevD.103.083533](https://doi.org/10.1103/PhysRevD.103.083533)
- [66] Adame, A. G., et al. DESI 2024 VI: Cosmological Constraints from the Measurements of Baryon Acoustic Oscillations. 2024. <https://arxiv.org/abs/2404.03002>
- [67] —. DESI 2024 IV: Baryon Acoustic Oscillations from the Lyman Alpha Forest. 2024. <https://arxiv.org/abs/2404.03001>
- [68] —. DESI 2024 III: Baryon Acoustic Oscillations from Galaxies and Quasars. 2024. <https://arxiv.org/abs/2404.03000>
- [69] Akrami, Y., et al. *Planck* intermediate results. LVII. Joint Planck LFI and HFI data processing. 2020, *Astron. Astrophys.*, 643, A42, doi: [10.1051/0004-6361/202038073](https://doi.org/10.1051/0004-6361/202038073)
- [70] Ade, P. A. R., et al. Planck 2013 results. XVI. Cosmological parameters. 2014, *Astron. Astrophys.*, 571, A16, doi: [10.1051/0004-6361/201321591](https://doi.org/10.1051/0004-6361/201321591)
- [71] Pan, Z., Knox, L., & White, M. Dependence of the Cosmic Microwave Background Lensing Power Spectrum on the Matter Density. 2014, *Mon. Not. Roy. Astron. Soc.*, 445, 2941, doi: [10.1093/mnras/stu1971](https://doi.org/10.1093/mnras/stu1971)
- [72] Green, D., Meyers, J., & van Engelen, A. CMB Delensing Beyond the B Modes. 2016, arXiv:1609.08143 [astro-ph, physics:hep-ph, physics:hep-th]. <https://arxiv.org/abs/1609.08143>
- [73] Knox, L., & Millea, M. Hubble constant hunter’s guide. 2020, *Phys. Rev. D*, 101, 043533, doi: [10.1103/PhysRevD.101.043533](https://doi.org/10.1103/PhysRevD.101.043533)
- [74] Breuval, L., Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., Romaniello, M., Murakami, Y. S., Scolnic, D., Anand, G. S., & Soszyński, I. Small Magellanic Cloud Cepheids Observed with the Hubble Space Telescope Provide a New Anchor for the SH0ES Distance Ladder. 2024, *Astrophys. J.*, 973, 30, doi: [10.3847/1538-4357/ad630e](https://doi.org/10.3847/1538-4357/ad630e)
- [75] Riess, A. G., et al. A Comprehensive Measurement of the Local Value of the Hubble Constant with  $1 \text{ km s}^{-1} \text{ Mpc}^{-1}$  Uncertainty from the Hubble Space Telescope and the SH0ES Team. 2022, *Astrophys. J. Lett.*, 934, L7, doi: [10.3847/2041-8213/ac5c5b](https://doi.org/10.3847/2041-8213/ac5c5b)
- [76] Freedman, W. L., & Madore, B. F. Progress in direct measurements of the Hubble constant. 2023, *JCAP*, 11, 050, doi: [10.1088/1475-7516/2023/11/050](https://doi.org/10.1088/1475-7516/2023/11/050)
- [77] Freedman, W. L., Madore, B. F., Jang, I. S., Hoyt, T. J., Lee, A. J., & Owens, K. A. Status Report on the Chicago-Carnegie Hubble Program (CCHP): Three Independent Astrophysical Determinations of the Hubble Constant Using the James Webb Space Telescope. 2024. <https://arxiv.org/abs/2408.06153>
- [78] Riess, A. G., et al. JWST Validates HST Distance Measurements: Selection of Supernova Subsample Explains Differences in JWST Estimates of Local  $H_0$ . 2024. <https://arxiv.org/abs/2408.11770>
- [79] Addison, G. E. 2021, High  $H_0$  Values from CMB E-mode Data: A Clue for Resolving the Hubble Tension?, arXiv, doi: [10.48550/arXiv.2102.00028](https://arxiv.org/abs/10.48550/arXiv.2102.00028)
- [80] Brout, D., et al. The Pantheon+ Analysis: Cosmological Constraints. 2022, *Astrophys. J.*, 938, 110, doi: [10.3847/1538-4357/ac8e04](https://doi.org/10.3847/1538-4357/ac8e04)
- [81] Abbott, T. M. C., et al. The Dark Energy Survey: Cosmology Results with  $\sim 1500$  New High-redshift Type Ia Supernovae Using the Full 5 yr Data Set. 2024, *Astrophys. J. Lett.*, 973, L14, doi: [10.3847/2041-8213/ad6f9f](https://doi.org/10.3847/2041-8213/ad6f9f)
- [82] Rubin, D., et al. Union Through UNITY: Cosmology with 2,000 SNe Using a Unified Bayesian Framework. 2023. <https://arxiv.org/abs/2311.12098>
- [83] Miyatake, H., et al. Hyper Suprime-Cam Year 3 results: Cosmology from galaxy clustering and weak lensing with HSC and SDSS using the emulator based halo model. 2023, *Phys. Rev. D*, 108, 123517, doi: [10.1103/PhysRevD.108.123517](https://doi.org/10.1103/PhysRevD.108.123517)
- [84] Heymans, C., et al. KiDS-1000 Cosmology: Multi-probe weak gravitational lensing and spectroscopic galaxy clustering constraints. 2021, *Astron. Astrophys.*, 646, A140, doi: [10.1051/0004-6361/202039063](https://doi.org/10.1051/0004-6361/202039063)
- [85] Abbott, T. M. C., et al. Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing. 2022, *Phys. Rev. D*, 105, 023520, doi: [10.1103/PhysRevD.105.023520](https://doi.org/10.1103/PhysRevD.105.023520)
- [86] Preston, C., Amon, A., & Efstathiou, G. A non-linear solution to the S8 tension – II. Analysis of DES Year 3 cosmic shear. 2023, *Mon. Not. Roy. Astron. Soc.*, 525, 5554, doi: [10.1093/mnras/stad2573](https://doi.org/10.1093/mnras/stad2573)
- [87] Lesgourgues, J., Mangano, G., Miele, G., & Pastor, S. 2013, *Neutrino Cosmology* (Cambridge University Press)
- [88] Loverde, M., & Weiner, Z. J. Massive neutrinos and cosmic composition. 2024. <https://arxiv.org/abs/2410.00090>
- [89] Green, D., & Meyers, J. The Cosmological Preference for Negative Neutrino Mass. 2024. <https://arxiv.org/abs/2407.07878>
- [90] Aghanim, N., et al. Planck 2018 results. VI. Cosmological parameters. 2020, *Astron. Astrophys.*, 641, A6, doi: [10.1051/0004-6361/201833910](https://doi.org/10.1051/0004-6361/201833910)



- [91] Rosenberg, E., Gratton, S., & Efstathiou, G. CMB power spectra and cosmological parameters from Planck PR4 with CamSpec. 2022, Mon. Not. Roy. Astron. Soc., 517, 4620, doi: [10.1093/mnras/stac2744](https://doi.org/10.1093/mnras/stac2744)
- [92] Tristram, M., et al. Cosmological parameters derived from the final Planck data release (PR4). 2024, Astron. Astrophys., 682, A37, doi: [10.1051/0004-6361/202348015](https://doi.org/10.1051/0004-6361/202348015)
- [93] Alvarez, M. A., Ferraro, S., Hill, J. C., Hložek, R., & Ikape, M. Mitigating the optical depth degeneracy using the kinematic Sunyaev-Zel'dovich effect with CMB-S4. 2021, Phys. Rev. D, 103, 063518, doi: [10.1103/PhysRevD.103.063518](https://doi.org/10.1103/PhysRevD.103.063518)
- [94] Lynch, G. P., Knox, L., & Chluba, J. DESI and the Hubble tension in light of modified recombination. 2024. <https://arxiv.org/abs/2406.10202>
- [95] Aver, E., Berg, D. A., Olive, K. A., Pogge, R. W., Salzer, J. J., & Skillman, E. D. Improving helium abundance determinations with Leo P as a case study. 2021, JCAP, 03, 027, doi: [10.1088/1475-7516/2021/03/027](https://doi.org/10.1088/1475-7516/2021/03/027)
- [96] Bennett, J. J., Buldgen, G., De Salas, P. F., Drewes, M., Gariazzo, S., Pastor, S., & Wong, Y. Y. Y. Towards a precision calculation of  $N_{\text{eff}}$  in the Standard Model II: Neutrino decoupling in the presence of flavour oscillations and finite-temperature QED. 2021, JCAP, 04, 073, doi: [10.1088/1475-7516/2021/04/073](https://doi.org/10.1088/1475-7516/2021/04/073)
- [97] Akita, K., & Yamaguchi, M. A precision calculation of relic neutrino decoupling. 2020, JCAP, 08, 012, doi: [10.1088/1475-7516/2020/08/012](https://doi.org/10.1088/1475-7516/2020/08/012)
- [98] Froustey, J., Pitrou, C., & Volpe, M. C. Neutrino decoupling including flavour oscillations and primordial nucleosynthesis. 2020, JCAP, 12, 015, doi: [10.1088/1475-7516/2020/12/015](https://doi.org/10.1088/1475-7516/2020/12/015)
- [99] Bashinsky, S., & Seljak, U. Neutrino perturbations in CMB anisotropy and matter clustering. 2004, Phys. Rev. D, 69, 083002, doi: [10.1103/PhysRevD.69.083002](https://doi.org/10.1103/PhysRevD.69.083002)
- [100] Baumann, D., Green, D., Meyers, J., & Wallisch, B. Phases of New Physics in the CMB. 2016, JCAP, 01, 007, doi: [10.1088/1475-7516/2016/01/007](https://doi.org/10.1088/1475-7516/2016/01/007)
- [101] Follin, B., Knox, L., Millea, M., & Pan, Z. First Detection of the Acoustic Oscillation Phase Shift Expected from the Cosmic Neutrino Background. 2015, Phys. Rev. Lett., 115, 091301, doi: [10.1103/PhysRevLett.115.091301](https://doi.org/10.1103/PhysRevLett.115.091301)
- [102] Ge, F., Cyr-Racine, F.-Y., & Knox, L. Scaling transformations and the origins of light relics constraints from cosmic microwave background observations. 2023, Phys. Rev. D, 107, 023517, doi: [10.1103/PhysRevD.107.023517](https://doi.org/10.1103/PhysRevD.107.023517)
- [103] Shao, H., Givans, J. J., Dunkley, J., Madhavacheril, M., Qu, F., Farren, G., & Sherwin, B. Cosmological limits on the neutrino mass sum for beyond- $\Lambda$ CDM models. 2024. <https://arxiv.org/abs/2409.02295>
- [104] Hou, Z., Keisler, R., Knox, L., Millea, M., & Reichardt, C. How Massless Neutrinos Affect the Cosmic Microwave Background Damping Tail. 2013, Phys. Rev. D, 87, 083008, doi: [10.1103/PhysRevD.87.083008](https://doi.org/10.1103/PhysRevD.87.083008)
- [105] Reinecke, M., Belkner, S., & Carron, J. Improved cosmic microwave background (de-)lensing using general spherical harmonic transforms. 2023, Astron. Astrophys., 678, A165, doi: [10.1051/0004-6361/202346717](https://doi.org/10.1051/0004-6361/202346717)
- [106] Millea, M., Anderes, E., & Wandelt, B. D. Bayesian delensing of CMB temperature and polarization. 2019, Phys. Rev. D, 100, 023509, doi: [10.1103/PhysRevD.100.023509](https://doi.org/10.1103/PhysRevD.100.023509)
- [107] Knopp, T., Boberg, M., & Grosser, M. NFFT.jl: Generic and Fast Julia Implementation of the Nonequidistant Fast Fourier Transform. 2023, SIAM Journal on Scientific Computing, 45, C179
- [108] Hu, W., Hedman, M. M., & Zaldarriaga, M. Benchmark parameters for CMB polarization experiments. 2003, Phys. Rev. D, 67, 043004, doi: [10.1103/PhysRevD.67.043004](https://doi.org/10.1103/PhysRevD.67.043004)
- [109] Henning, J. W., Sayre, J. T., Reichardt, C. L., et al. Measurements of the Temperature and E-mode Polarization of the CMB from 500 Square Degrees of SPT-pol Data. 2018, Astrophys. J., 852, 97, doi: [10.3847/1538-4357/aa9ff4](https://doi.org/10.3847/1538-4357/aa9ff4)
- [110] Planck Collaboration, Aghanim, N., Akrami, Y., et al. Planck 2018 results. V. CMB power spectra and likelihoods. 2019, A&A, 641, A5, doi: [10.1051/0004-6361/201936386](https://doi.org/10.1051/0004-6361/201936386)
- [111] Dutcher, D., et al. Measurements of the E-mode polarization and temperature-E-mode correlation of the CMB from SPT-3G 2018 data. 2021, Phys. Rev. D, 104, 022003, doi: [10.1103/PhysRevD.104.022003](https://doi.org/10.1103/PhysRevD.104.022003)
- [112] Mancini, A. S., Piras, D., Alsing, J., Joachimi, B., & Hobson, M. P. COSMOPOWER: emulating cosmological power spectra for accelerated Bayesian inference from next-generation surveys. 2021, MNRAS, 511, 1771, doi: [10.1093/mnras/stac064](https://doi.org/10.1093/mnras/stac064)

## Appendix A: Modeling details

In this appendix, we give full details about our simulation and posterior models, proceeding in the order in which each piece is used to generate a simulation. In cases where the posterior model is different from the simulation model, each is described individually.

### 1. Lensed CMB

*Simulation model* In the simulation model, the lensed CMB is computed on the curved sky. Spherical harmonic coefficients for the unlensed CMB and for the lensing potential are sampled from a normal distribution with variance given by the appropriate  $C_\ell$  values. We use the DUCCLIBRARY to perform an inverse spherical harmonic transform of the coefficients to obtain simulated HEALPIX  $N_{\text{side}} = 2048$  maps. The lensing operation is computed with LENSPYX [105], taking as input the unlensed CMB maps,  $f$ , and the gravitational lensing potential maps,  $\phi$ . The lensed map

is then projected to the Lambert projection using Non-Uniform FFTs (NFFTs), as described in the next section.

For the MUSE covariance calculation, we also require the Jacobian of the lensed map with respect to bandpower parameters. Since here we use external libraries for SHTs and curved-sky lensing, we must define the Jacobians for these operations for the autodifferentiation (AD) system. The Jacobian of a spherical harmonic transformation (SHT),  $f = \mathbb{Y}a$  and the Jacobian with respect to  $f$  of the lensing operation,  $\tilde{f} = \mathbb{L}(\phi)f$ , are trivially just  $\mathbb{Y}$  and  $\mathbb{L}$  themselves since they are linear operators. The Jacobian of lensing with respect to  $\phi$  is derived from

$$f(x + \nabla(\phi + \epsilon)) = f(x + \nabla\phi) + \nabla f(x + \nabla\phi) \cdot \nabla\epsilon + O(\epsilon^2), \quad (\text{A1})$$

where  $x$  is a pixel location and  $\epsilon$  is a small perturbation to the lensing potential. The second term demonstrates that the Jacobian is the operator  $\nabla f(x + \nabla\phi) \cdot \nabla$ , and involves the spatial gradient of a lensed map, which itself is computed with `LENSPYX`. With these AD rules implemented, our curved-sky lensing model is fully compatible with the MUSE covariance calculation.

*Posterior model* In the posterior model, we assume the flat-sky approximation, so simulated maps are generated from white-noise maps FFT-convolved with appropriate kernels. We implement lensing with `LENSEFLOW` [106], which is GPU-compatible and automatically differentiable up to second-order as needed by the MUSE covariance calculation.

*Bandpower parameters* Both simulation and posterior models take as input a theory unlensed CMB spectrum and gravitational lensing potential spectrum. We introduce free parameters which we will infer,  $A_b^{\text{EE}}$  and  $A_b^{\phi\phi}$ , which control the amplitude of these spectra relative to a fiducial model within each bin. Our  $EE$  bandpowers have linearly spaced bins with  $\Delta\ell = 50$  between 350 and 3500, and our  $\phi\phi$  bandpowers have 12 logarithmically spaced bins between 20 and 3000. We take the fiducial model to be the Planck best-fit spectrum, but this introduces no dependence on Planck, rather is just a choice of definition of the bandpower amplitude parameters. Within a bin, the parameters scale the theory spectrum uniformly across  $\ell$ 's. One can show that with this choice, our inferred  $A_b^{\text{EE}}$  and  $A_b^{\phi\phi}$  parameters correspond to a constraint on an inverse-variance weighted average of the theory spectrum across the bin. Since this inverse-variance weighting arises only implicitly, we also construct explicit bandpower window functions (used in parameter estimation) from an interpolation of the diagonal of the MUSE covariance,  $\Sigma_{\text{MUSE}}$ , to every  $\ell$ . The tests on simulations in Sec. IV A 2, which demonstrate no bias and correct scatter of cosmological parameters derived from simulations, provide verification of this procedure.

## 2. NFFT-based projections

Within our model, we perform interpolation between maps in different projections in four different places: 1) as mentioned, in the simulation model, the lensed CMB is projected from `HEALPIX`  $N_{\text{side}} = 2048$  to a Lambert pixelization, 2) in both simulation and posterior models, the transfer function model involves projecting from a Lambert to an Equirectangular pixelization, applying some weights, then projecting back to Lambert, 3) when computing the transfer function, a set of end-to-end pipeline simulations are compared to a direct projection of the `HEALPIX`  $N_{\text{side}} = 8192$  maps which are input to the end-to-end simulations, and 4) to model deflections due to relativistic aberration.

To ensure easy differentiability, GPU compatibility, and the best possible accuracy, we perform these interpolations using Non-Uniform FFTs (NFFTs). NFFT interpolation implicitly uses information from all pixels to compute the interpolated value at a given location, as opposed to other typically-used interpolations such as bilinear interpolation, which uses only the nearest pixel neighbors. It is exact for a band-limited signal (i.e. when there is no power beyond the Nyquist frequency), and remains very accurate for nearly band-limited signals such as ours, where power beyond the Nyquist frequency is highly suppressed by the beam.

We use the CPU and GPU compatible NFFT library `NFFT.jl` [107]. An interpolation is performed by a pair of forward and adjoint NFFTs,  $\mathcal{F}$ , with appropriate grid-points,  $\gamma$ , which describe the relative location of the pixel centers in the map projections being interpolated between,

$$f_1 = \mathcal{F}(\gamma_1)\mathcal{F}(\gamma_2)^\dagger f_2. \quad (\text{A2})$$

The adjoint of the projection operator, which arises in automatic differentiation, is then simply

$$(\mathcal{F}(\gamma_1)\mathcal{F}(\gamma_2)^\dagger)^\dagger = \mathcal{F}(\gamma_2)\mathcal{F}(\gamma_1)^\dagger. \quad (\text{A3})$$

We verify these projections are accurate enough in Fig. 22. For use cases (1) and (2), in the left panel, we generate a simulated lensed CMB map in the Lambert pixelization. We then project to equirectangular and back to Lambert, as well as to `HEALPIX`  $N_{\text{side}} = 2048$  and back to Lambert. We then examine the change to the spectrum of the original map. Fig. 22 demonstrates this change is  $\lesssim 0.3\%$  up to the maximum multipole used in the analysis,  $\ell < 3500$ . For use case (3), in the right panel, we compare the spectrum of the projected `HEALPIX`  $N_{\text{side}} = 8192$  maps to the input theory spectrum, again finding  $\lesssim 0.5\%$  changes up to the maximum multipole. These differences are well below our uncertainties; for example typical  $EE$  bandpower errors are at the 5 to 10% level.

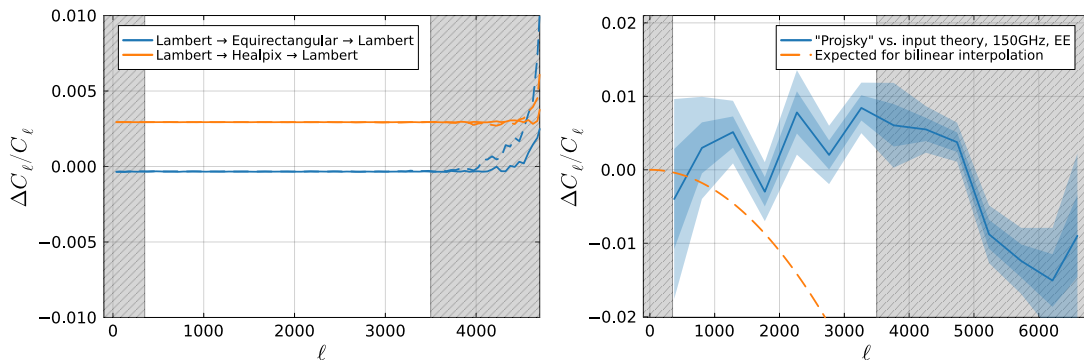


FIG. 22. (Left) Change to the  $EE$  (solid) and  $BB$  (dashed) power spectrum of a lensed CMB simulation when using NFFTs to project a pair of Q/U maps to Equirectangular or HEALPIX  $N_{\text{side}} = 2048$  pixelizations and back. This demonstrates negligible modeling error in our simulation and posterior models due to these projections. (Right) Comparison of the pseudo-powerspectra of NFFT Lambert 2.25' projections of the HEALPIX  $N_{\text{side}} = 8192$  maps which are input to mock observation (labeled “projskies”) vs. the input theory which generated those HEALPIX maps. Blue bands represent 1 and 2  $\sigma$  Monte Carlo error from the finite number of simulations used to compute the blue curve. This confirms the “projskies” sufficiently recover the input theory so as not to bias the transfer function calculation, where they are used. For reference, this plot also shows the expected change to the power spectra if bilinear interpolation was instead used, which shows much worse performance than NFFT interpolation.

### 3. Relativistic aberration

Relativistic aberration distorts the CMB maps due to our proper motion relative to the CMB rest frame  $\square$ . This is a small but non-negligible effect, which we expect could bias  $\theta_s$  by as much as  $\sim 0.3\sigma_{\theta_s}$  if unaccounted for. One method for modeling aberration is to apply a first-order correction to the power spectrum  $\square$ ,

$$C_\ell \rightarrow C_\ell + C_\ell \frac{d \log C_\ell}{d \log \ell} \beta \langle \cos \psi \rangle, \quad (\text{A4})$$

where  $\beta$  is our peculiar Lorentz factor and  $\langle \cos \psi \rangle$  measures the average angle to the direction of proper motion for pixels in the field. We note that this correction applies to a lensed spectrum, and does not apply to the unlensed spectra inferred here. In particular, some a priori unknown part of the aberration deflection is interpreted in the posterior as very long wavelength gravitational lensing deflection, and then is automatically removed when producing inferences of the unlensed spectrum. Eqn. (A4) is thus an upper bound on the impact aberration might have on our particular analysis, but it could be much smaller. Without knowing the exact magnitude, we choose to simply include the aberration effect directly in the map-level model. Since the effect is small, it is sufficient to only include it in the simulation model to prevent biases.

The deflection angle,  $\Delta\psi$ , is given by

$$\Delta\psi = \psi - \cos^{-1} \left( \frac{\cos \psi - \beta}{1 - \beta \cos \psi} \right) \quad (\text{A5})$$

where  $\psi$  is the angle on the sphere to the direction of proper motion. We assume a fixed velocity of  $\beta = 0.00128$  and a direction of  $(l, b) = (264^\circ, 48^\circ)$  [1], ignoring next-order effects from uncertainties on these values. The deflection is performed via NFFTs.

We note that we do not model the CMB dipole also induced by our proper motion, as this is filtered out by timestream and map-level high-pass filters.

### 4. Beams

Next, the CMB maps are convolved with a model for the polarized instrument beams at each frequency,  $\mathbb{B}_P^\nu(\beta_n, \beta_{\text{pol}}^\nu)$ . We assume a rotationally symmetric convolution kernel, performed on the Lambert maps with FFTs in both simulation and posterior models. The shape of the convolution kernel in real space,  $B(\theta)$ , or in multipole space,  $B_\ell$ , is controlled by free parameters which are marginalized over in the analysis. These parameters include a set of temperature beam eigenmode amplitudes,  $\beta_n$ , and the polarization fraction of the beam sidelobes at each frequency,  $\beta_{\text{pol}}^\nu$ . We first

describe the temperature beams, then how we derive the polarization beams from them. More detail on the beam determination will be presented in Huang et al., (in prep).

#### a. Temperature beams

The first step in determining our polarized beams is to determine the temperature beams. At low radii ( $\lesssim 2$  arcmin), temperature beams are determined from observations of several bright AGN in our observing field. The AGN, however, do not have sufficient brightness to fully map out the beam at larger scales (referred to as the beam “sidelobes”), where instead we use observations of Saturn. In turn, observations of Saturn cannot be used at small scales, since its intensity saturates detectors and changes the response as compared to typical CMB observations. Therefore, we stitch the AGN and Saturn maps into a composite beam map, which uses the AGN measurements at low radii, and Saturn for the sidelobes.

To create the composite map, we must account for several systematic effects. First, Saturn is very slightly extended. Before stitching, we convolve the AGN maps with a disk of 8.55 arcsec (the average angular diameter of Saturn during our observations). We later deconvolve the same disk from the final beam. The SPT-3G detectors have a finite response time, which spreads signals along the scan direction. While the response time is constant, the on-sky scan speed varies with declination. Therefore, we deconvolve the map-space effect of the time constants from each map (both for individual AGN and Saturn), and then reconvolve with the mean effect of the time constant over the SPT-3G winter field footprint. We also attempt to remove background signals (primarily the CMB) from both the Saturn maps and AGN maps. Since Saturn moves across the sky, we are able to subtract direct observations of the backgrounds from the Saturn maps. To do this, we mock observe Planck maps for each of our Saturn observations, and subtract the mock-observed maps from our Saturn maps. For the AGN maps, we are unable to do this. However, we are only concerned with small scale ( $\lesssim 7$  arcmin) information in the AGN maps. On these scales, the CMB varies very little, and can be modeled with only low-order modes. Therefore, we fit a simple background model to each AGN consisting of a slope in radius and an offset. We fit these parameters (using the Saturn maps as a reference for the beam) at 95 and 150 GHz, since the CMB is expected to be common between bands, and the 220 GHz maps are significantly noisier (both from instrumental noise, and higher astrophysical backgrounds). We subtract the background model from all three bands.

Finally, we apply a very basic model of intra-band frequency dependence to the beam. We model the beam at each band as a Gaussian made of two components, one which is fixed across the band, and one which scales with frequency. We model the frequency scaling as  $\sigma(\nu) \propto 1/\nu$ , which is the expected behavior for a diffraction-limited beam. We use the same constant of proportionality across all three bands. In order to fit these parameters, we assume that the 95 GHz beam is diffraction limited (i.e. the frequency independent component is non-existent). We approximate the “effective beam frequency” for each band using the SPT-3G mean bandpass functions, and the source spectrum. Then, we fit the three remaining parameters (the constant of proportionality in the frequency dependent portion, and the two frequency independent portions of the 150 and 220 GHz beams) using observations of the brightest AGN. This simple model allows us to approximate the change in the beam by calculating the effective beam frequency for different source spectra, and convolving the beam maps (or multiplying their spectra) with the appropriate Gaussian. We apply this model to transform the final beam from the mean AGN spectrum to the CMB spectrum, and to transform the Saturn maps to match the AGN spectrum before stitching.

With the systematic effects accounted for, the stitching process is relatively simple. For each pair of AGN and Saturn maps, we fit for an amplitude to account for the difference in intensity between the sources. As discussed above, we cannot use the innermost portion of the Saturn maps. The AGN maps also become noise-dominated at large distances from the source, so we perform this fit over an annulus with inner radii of (2.25, 1.5, 1.0) arcmin, and outer radii of (3.5, 3.0, 2.5) arcmin at (95, 150, 220) GHz. Finally, each composite map is constructed by linearly transitioning from the AGN map to the Saturn map over the same annulus.

We make composite maps from all pairs of AGN and Saturn maps. We also vary several systematic parameters: the choice of stitching radii, uncertainty in the time constants, the choice of the background fitting region. For each of these variations, we create the full set of composite maps. Then, we construct cross spectra of composite maps such that none of the component maps are the same. If we included the same component map more than once, we would incur a noise bias in the beam. To measure the final beam, we take the mean of these cross spectra, and apply the two final effects described above (specifically, deconvolving the Saturn disk, and converting from the AGN spectrum to the CMB spectrum).

The beam uncertainty is measured from the cross spectra. By including both different Saturn and AGN maps, as well as the systematic variations, we capture both the map noise, systematic uncertainty from our analysis choices, and systematic differences between the AGN. Using the Monte-Carlo method, this uncertainty is propagated to a covariance between temperature  $B_\ell^\nu$ 's at each frequency,  $\nu$ , and multipole,  $\ell$ . Finally, we add one additional uncertainty directly

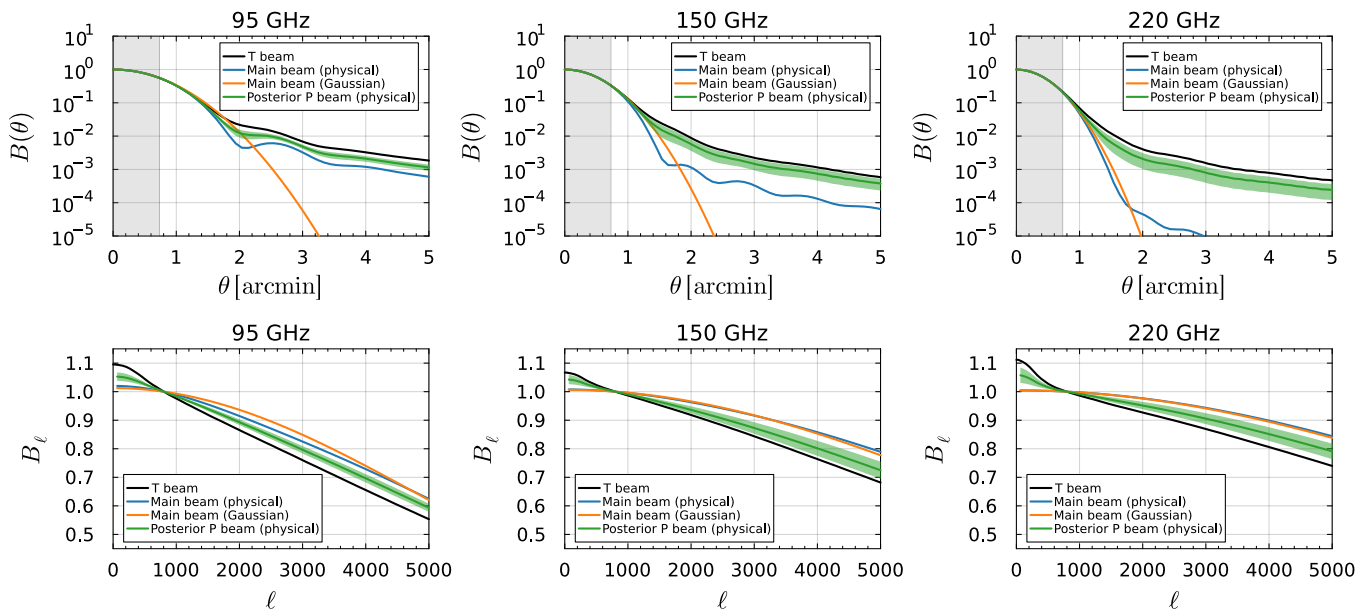


FIG. 23. Beams used in this analysis, in real space (top row) and multipole space (bottom row), at each frequency (each column). Black lines show the temperature beams. Blue and orange lines show our best fits to the main beam, where the fitting region is denoted by the gray shaded region. We used the physical main beam for our final result, but we also verified that using the Gaussian main beam leads to no significant shifts in results (see Sect. IV C). Green lines and green bands represent the posterior mean and  $1\sigma$  uncertainty on the polarized beams from the MUSE analysis, which includes marginalization over  $\beta_n$ ,  $\beta_{\text{pol}}$ , and all other systematics and bandpower parameters.

to the covariance matrix, which accounts for uncertainty in the detector bandpasses and the AGN spectral index. This covariance is decomposed into eigenmodes, and in the analysis the amplitude of 5 eigenmodes,  $\beta_n$  is free to vary jointly with the bandpower and other systematics parameters. We note that these are eigenmodes of the joint covariance between frequencies and multipoles, so each  $\beta_n$  changes the beam simultaneously at each frequency (in different ways at each frequency). The total uncertainty in the determination of the temperature beams is 0.2% at  $\ell = 3000$ , and is sub-dominant to the uncertainty on polarized beams which we now describe.

#### b. Polarization beams

While the temperature beams are mapped out to very high signal-to-noise at angular scales relevant to this analysis via observations of bright point sources, we lack sufficiently bright polarized sources to similarly map the sidelobe response of the polarized beams. Prior to unblinding, we had assumed the polarized and temperature beams were identical at the scales used in the analysis. However, the temperature beam has significant diffuse sidelobes due to diffraction and scattering that may not be fully polarized. To account for this, we devised a model of the beams which allows the polarization of the beam sidelobes to vary. This model assumes the beams are made up of a fully polarized main central beam, and a diffuse beam sidelobe which may be partially unpolarized.

We consider four components in the model for the main beam 1) Gaussian illumination pattern of the primary by the detector lenslets, 2) truncation of that illumination pattern by the Lyot stop, 3) averaging the frequency dependent beam over the detector bandpass weighted by a CMB spectrum, and finally 4) Gaussian broadening of the diffracted beam due to a frequency independent geometric aberration.

We model the electric field illumination pattern on the primary mirror as a truncated Gaussian:

$$V(r, \nu) = \exp\left\{\left(\frac{-r^2}{2(\sigma_0 \frac{\nu_0}{\nu})^2}\right)\right\} \Pi(R) \quad (\text{A6})$$

where  $\sigma_0$  is the width of the Gaussian beam from the lenslets at some arbitrary fiducial frequency  $\nu_0$ , and  $\Pi(R)$  is a

disk of radius  $R$ :

$$\Pi(R) = \begin{cases} 1 & r \leq R \\ 0 & r > R \end{cases} \quad (\text{A7})$$

For SPT-3G,  $R \approx 3.75$  m. The illumination on the sky can then be computed from the Fraunhofer diffraction equation,

$$U(\rho, z, \nu) = 2\pi \int_0^\infty V(r, \nu) J_0\left(\frac{2\pi r \rho}{c/\nu z}\right) r dr \quad (\text{A8})$$

$$U(\theta, \nu) = 2\pi \int_0^\infty V(r, \nu) J_0\left(\frac{2\pi r \theta}{c/\nu}\right) r dr \quad (\text{A9})$$

where  $z$  is the distance to the image plane,  $\rho$  is the distance from the origin on the image plane, and  $J_0$  is the zero-th order Bessel function. In the second line, we have taken the small angle approximation  $\rho/z = \theta$  to put this in terms of angle on the sky. The beam on the sky is the intensity which is the square of the electric field pattern,

$$B(\theta, \nu)_{\text{ideal}} = U(\theta, \nu)^2 \quad (\text{A10})$$

The beam for our wide frequency bands is then integrated over the bandpass and source spectrum. We integrate over the beam intensity, since the electric field phases will decohere at any significant frequency separation. Adjacent frequencies where phase coherence matters have very nearly the same illumination pattern, because both the spectra of our sources and bandpasses vary slowly. Thus we have,

$$B(\theta)_{\text{ideal}} = \frac{\int_{\text{bandpass}} I(\nu) t(\nu) B(\theta, \nu) d\nu}{\int_{\text{bandpass}} I(\nu) t(\nu) d\nu}, \quad (\text{A11})$$

where  $t(\nu)$  is the bandpass function, and  $I(\nu) = \frac{dB}{dT}|_{T=T_{\text{CMB}}}$  is the observed spectrum.

Finally, we need to account for the effect of geometric aberrations. This is simply a matter of convolving  $B(\theta)_{\text{ideal}}$  with a Gaussian, a representation of geometric aberrations, so the final beam is,

$$B(\theta) = B(\theta)_{\text{ideal}} \otimes \exp\left\{\left(-\frac{\theta^2}{2\sigma_{\text{geom}}^2}\right)\right\} \quad (\text{A12})$$

In principle, we expect  $\sigma_{\text{geom}}$  to be the same for all three bands.

This model for the main beam has five free parameters,  $\sigma_0$ ,  $\sigma_{\text{geom}}$ , and an overall amplitude at each frequency,  $A_\nu$ , and predicts the main beam simultaneously in each of our bands. We fit this model to the observed temperature beams as determined in the previous section, limiting to an inner radius of  $\theta < 0.75$  arcmin where we observe a good fit, suggesting the main beam dominates and the fit is not biased by the presence of temperature beam sidelobes. For this fit, we construct a  $\chi^2$  of the form  $\sum_\nu \int d\theta (B_{\text{model}}^\nu(\theta, A, \sigma_0, \sigma_{\text{geom}}) - B_{\text{observed}}^\nu(\theta))^2$  which ignores uncertainties in the temperature beams, since we expect the systematic uncertainty from our choice of physical model to dominate to the total uncertainties in fit.

With this fiducial main beam determined, we denote the difference between this and the temperature beam as the sidelobes, and model the polarized beams as the sum of the fixed main beam and the sidelobes, the latter scaled by some unknown scale-independent polarization fractions,  $\beta_{\text{pol}}^\nu$ . That is, the polarized beams used in our analysis are,

$$\mathbb{B}_P^\nu(\beta_n, \beta_{\text{pol}}^\nu) = \mathbb{B}_{\text{main}}^\nu + \beta_{\text{pol}}^\nu (\mathbb{B}_T^\nu(\beta_n) - \mathbb{B}_{\text{main}}^\nu) \quad (\text{A13})$$

where the temperature beam at each frequency is

$$\mathbb{B}_T^\nu(\beta_n) = \mathbb{B}_0^\nu + \beta_1 \mathbb{B}_1^\nu + \beta_2 \mathbb{B}_2^\nu + \dots, \quad (\text{A14})$$

where  $\mathbb{B}_0$  is the best-fit beam and  $\mathbb{B}_n$  is the  $n$ -th eigenmode of the beam covariance. The  $\beta_n$  is the amplitude of the  $n$ -th eigenmode, which is the free parameter. The  $\mathbb{B}_n$  is normalized such that  $\beta_i$  has a unit normal prior.

## 5. Absolute calibration

In the analysis, we apply calibration factors,  $A_{\text{cal}}^{\nu, i}$ , to each frequency map,  $\nu$ , and subfield,  $i$ , independently. We assume a Gaussian prior on the  $A_{\text{cal}}^{150, i}$  based on calibrating our maps to Planck, and leave uniform priors for  $A_{\text{cal}}^{95, i}$  and  $A_{\text{cal}}^{220, i}$ . App. B describes our choice of prior for the 150 GHz calibration.

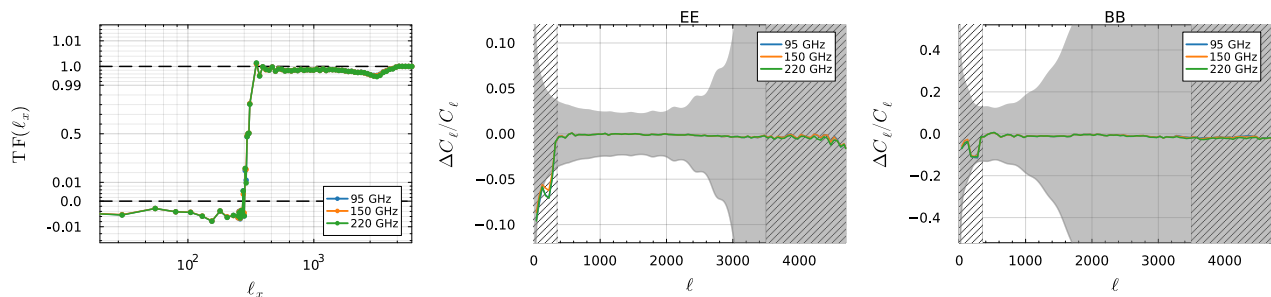


FIG. 24. *Left panel:* The filtering applied in the scan direction as part of the transfer function model. This function is fit on simulations at the spline interpolation points denoted with markers. Note the non-linear y-axis scaling. *Right two panels:* Comparison of the  $EE$  and  $BB$  power spectra of an end-to-end pipeline simulation of the CMB as compared to the model transfer function applied to the same input sky which generated the simulation. Across the multipole range used for the analysis, the model is accurate to  $\lesssim 0.2\%$  in  $EE$  and  $\lesssim 2\%$  in  $BB$ , both less than 10% of our uncertainties (denoted as gray bands).

## 6. Global polarization angle

To model a systematic error on the global polarization angle calibration, we rotate the global polarization angle of the entire field by an angle,  $\psi_{\text{pol}}^\nu$ , at each frequency,  $\nu$ . The rotation is given by,

$$\mathbb{R}(\psi) \begin{pmatrix} Q \\ U \end{pmatrix} = \begin{pmatrix} \cos(2\psi) & -\sin(2\psi) \\ \sin(2\psi) & \cos(2\psi) \end{pmatrix} \begin{pmatrix} Q \\ U \end{pmatrix}. \quad (\text{A15})$$

## 7. Transfer function

The transfer function accounts for the filtering of the timestreams during mapmaking described in Sec. II. In our transfer function model  $\text{TF}$ , we must approximate the impact of this filtering directly to a Lambert-projected map, since our model is not at the level of timestreams. The non-trivial procedure of sequentially deprojecting templates while masking point sources still corresponds to a linear operator, but not one which is diagonal in any easily accessible basis.

Our model begins by first reprojecting the map to an Equirectangular projection using NFFTs, such that individual scans are straight lines in the horizontal direction. We then create a single map-level deprojection template by multiplying the point source mask assumed in mapmaking and applying a 1D filtering in the horizontal direction using FFTs where the 1D filter is a free function which we will fit for using simulations. This template is subtracted from the entire map (including the masked regions), before projecting back to Lambert. The operation can be represented as:

$$\text{TF} = \mathbb{P}^{-1} \cdot [\mathbb{1} - (\mathbb{1} - \text{TF}(\ell_x)) \cdot \mathbb{M}_{\text{ptsrc}}] \cdot \mathbb{P} \quad (\text{A16})$$

where  $\mathbb{P}$  is the NFFT projection from Lambert to Equirectangular,  $\mathbb{M}_{\text{ptsrc}}$  is the point source mask and  $\text{TF}(\ell_x)$  is the fitted FFT kernel in the horizontal direction.

To fit for the appropriate function,  $\text{TF}(\ell_x)$ , we compare 1) mock observations of the signal,  $m$ , which simulate the entire timestream filtering and mapmaking process, and 2) direct projections of the simulated signal maps which were fed into the mock observations,  $p$ . The relation between these maps are

$$m = \text{PWF}_{2.25'} \cdot \text{TF} \cdot \text{PWF}_{8192}^2 \cdot p. \quad (\text{A17})$$

The factor of the HEALPIX  $N_{\text{side}} = 8192$  pixel window function is due to the interpolation of  $p$  onto mock timestreams during the simulation process, and does not arise in the real data. The factor of Lambert  $2.25'$  pixel window function,  $\text{PWF}_{2.25'}$  arises from the binning of these timestreams into  $2.25'$  resolution maps. We then minimize the function  $\|m_{\text{true}} - m_{\text{model}}\|^2$  by fitting  $\text{TF}(\ell_x)$  at a set of spline points. We find that a single pair of  $m$  and  $p$  simulations is sufficient to obtain a stable estimate, and have verified the result changes negligibly between a few different simulations.

The left panel of Fig. 24 shows the fitted  $\text{TF}(\ell_x)$ . As expected, it is roughly zero until around  $\ell_x = 300$ , then transitions to unity at higher multipoles. We choose to allow some small negative values as it gives a better overall

fit with no other observed impact. The right two panels of the figure then show the comparison of  $EE$  and  $BB$  spectra of the mock observations with those computed by applying the best-fit model transfer function to  $p$ . Across the multipole range used for the analysis, the model is accurate to  $\lesssim 0.2\%$  in  $EE$  and  $\lesssim 2\%$  in  $BB$ , both less than 10% of our uncertainties (which are denoted as gray bands).

## 8. Temperature-to-polarization leakage

T-to-P leakage can be caused by a variety of sources. Gain mismatch between detector pairs leaks a scaled copy of the temperature map into polarization (“monopole” leakage), whereas differential detector pointing and beam ellipticity introduce copies of the first and second derivatives of the temperature map [“dipole” and “quadrupole” leakage; 108, 109]. We find no evidence of dipole or quadrupole leakage in the SPT-3G data (Figure 8), so we only correct for monopole leakage.

In the MUSE simulation and posterior model, we construct the T-to-P leakage as  $t_Q^\nu = (T^\nu, 0)$  and  $t_U^\nu = (0, T^\nu)$ , where  $T^\nu$  is the temperature map at frequency  $\nu$ . The temperature maps are made following the same map-making procedure in Sect. II. The amplitudes of the T-to-P templates,  $\epsilon_Q^\nu$  and  $\epsilon_U^\nu$ , are free parameters. Note that we used independent leakage parameters per frequency per subfield to be consistent with the choice of calibration parameters.

## 9. Masking

We apply a pixel mask ( $M_{\text{pix}}$ ) in the model, which includes a border mask and a point source mask. The border mask is set by zeroing out the areas with weights less than 30% of the median, padding the border of the mask with zero within the radius of 120 arcmin and apodizing with a cosine apodization function of 60 arcmin radius. The point source mask contains emissive sources with brightness greater than 50 mJy and clusters with signal-to-noise ratio greater than 15 at 150GHz.

We also apply a trough mask ( $M_{\text{trough}}$ ) to exclude specific modes from analysis. As discussed in Sect. A 7, the pixel mask aliases the power to higher  $\ell_x$  modes in Equirectangular projection due to the TOD filtering, and smears out the power when the maps are projected to Lambert. Therefore, we cut out the modes with  $\ell_x < 400$  in Equirectangular projection to avoid this aliasing power. We also masked out the region that has been contaminated in the observation by a narrow-band signal around 1.1 Hz.

The mid-pass filter in the Fourier space ( $M_{\text{fourier}}$ ) constrains the modes used in the analysis. We include modes between  $350 \leq \ell \leq 4000$ .

The total mask used in the analysis is  $M = M_{\text{fourier}} \cdot M_{\text{trough}} \cdot M_{\text{pix}}$ , that is,  $M_{\text{pix}}$  is applied first. The mask is also applied to the real data map before passed down into the analysis.

## 10. Monte-carlo correction

With the aforementioned modeling ingredients specified, we can compare the pseudo power spectrum of the signal part of  $d^\nu$  between simulation-model simulations, as in Eqn. (15), and mock observations. We find a remaining 0.3% discrepancy at  $\ell \lesssim 1250$  in the  $EE$  spectrum between mock observations and MUSE simulations. We believe this is due to residual transfer function modeling errors. Before unblinding, we implement this as an empirical multiplicative Monte Carlo (MC) correction to the  $A_b^{\text{EE}}$  parameters, with a shape fit from these simulations. We note that this correction is small relative to our uncertainties and corresponds to a  $0.01 \sigma$  shift in  $H_0$  or a  $0.07 \sigma$  shift in  $\Omega_b h^2$ .

## 11. Noise modeling

*Simulation model* In the simulation model, we use sign-flip noise realizations generated empirically from the real data. This guarantees the noise in the simulation model has the same statistical properties at the true data noise, up to the small correlation between sign-flip realizations, which is not large enough to appreciably impact our final results.

*Posterior model* In the posterior model, we must be able to compute the posterior probability function, thus must explicitly specify a random distribution for the noise. We chose a Gaussian model with a covariance derived from the sign-flip noise realizations themselves. We model the noise as diagonal in Fourier space with isotropic noise spectra,

$$N_\ell = A(\ell/3000)^\alpha, \quad (\text{A18})$$



where  $A$  and  $\alpha$  are parameters to fit to 20 sign-flip noise realizations across the region  $500 \leq \ell \leq 3000$ , separately at each frequency. We find this gives an acceptable fit, and note that any deviations of this model from true noise distribution will only lead to slight loss of optimality, not any bias.

### Appendix B: Calibration priors and alternative systematics estimates

Here, we describe the priors on  $A_{\text{cal}}^{150,i}$  which enter the MUSE analysis, and the alternative estimates of  $A_{\text{cal}}^{95,i}$ ,  $A_{\text{cal}}^{220,i}$ ,  $\epsilon_Q$ ,  $\epsilon_U$ , and  $\psi_{\text{pol}}$  which are compared against the MUSE results.

In practice, calibrating to Planck is performed as two steps. First, we determine a calibration factor for temperature maps in each subfield,  $T_{\text{cal}}^i$ . Then, we determine a correction on top of this for polarization maps,  $P_{\text{cal}}$ , which is assumed the same across all subfields. By our convention, these factors multiply the SPT data, rather than multiplying the theory model like the MUSE definition of  $A_{\text{cal}}^{\nu,i}$ . The relation is thus

$$A_{\text{cal}}^{150,i} = 1/(T_{\text{cal}}^{150,i} P_{\text{cal}}^{150}), \quad (\text{B1})$$

The temperature calibration is determined by computing the cross-spectrum of the full-depth 150 GHz SPT-3G data with the full-depth 143 GHz *Planck* map passed through the SPT-3G observing pipeline normalized by the cross-spectrum of two half-depth SPT-3G maps,

$$T_{\text{cal, external}}^{150} = \frac{\text{SPT}_{\text{full}}^{150} \times \text{Planck}_{\text{full}}^{143}}{\text{SPT}_{\text{half1}}^{150} \times \text{SPT}_{\text{half2}}^{150}}. \quad (\text{B2})$$

The uncertainty on  $T_{\text{cal, external}}^{150}$  is computed from the scatter in the quantity when it is computed with different subsets of SPT observations entering the SPT maps. After unblinding, we also realized we had neglected to fold in a 0.25% uncertainty on the *Planck* absolute calibration [110], which leads to insignificant changes, but is fixed in our final results.

Although not used in MUSE, we also obtain temperature calibrations of 95 and 220GHz SPT-3G maps by cross-correlating against the 150GHz map (this yields tighter constraints than calibrating these against Planck). The calibrations are

$$T_{\text{cal, internal}}^{95} = \frac{\text{SPT}_{\text{half1}}^{150} \times \text{SPT}_{\text{half2}}^{150}}{\text{SPT}_{\text{half1}}^{95} \times \text{SPT}_{\text{half2}}^{150}} T_{\text{cal, external}}^{150},$$

$$T_{\text{cal, internal}}^{220} = \frac{\text{SPT}_{\text{half1}}^{150} \times \text{SPT}_{\text{half2}}^{150}}{\text{SPT}_{\text{half1}}^{220} \times \text{SPT}_{\text{half2}}^{150}} T_{\text{cal, external}}^{150}. \quad (\text{B3})$$

We recalibrate the  $Q$  and  $U$  maps by this factor, then correct them for temperature-to-polarization leakage and polarization angle calibration as described below. Then, we determine  $P_{\text{cal}}^{150}$  by cross-correlating the SPT 150 GHz polarization maps to Planck 143 GHz polarization maps in a manner analogous to the absolute temperature calibration in Eqn. (B3). The estimated  $P_{\text{cal}}^{150}$  is then corrected by an additional factor of  $\sqrt{0.966}$  to match the final recalibration factor applied to the 143 GHz *Planck* polarization bandpowers [Eqn. 45, 110], and additionally has a 0.5% uncertainty folded in corresponding to the uncertainty on the *Planck* polarization calibration. Similarly used only for comparison, we then determine  $P_{\text{cal}}^{95}$  and  $P_{\text{cal}}^{220}$  by internally cross-correlating the SPT maps.

Next, we describe the alternative estimate of temperature-to-polarization leakage and global polarization angle correction which is used above in determining the calibration factors and is compared against the MUSE results in Fig. 7. The main difference from the MUSE estimate is that these are performed at the power spectrum rather than at the map level, and are done with an iterative procedure which assumes a theory cosmological model rather than simultaneously and jointly with all systematics and bandpower parameters.

Independent estimates of the monopole leakage coefficients,  $\epsilon^{Q,TT}$ ,  $\epsilon^{Q,TQ}$ ,  $\epsilon^{U,TT}$ , are formed by fitting a template to the  $C_{\ell}^{TQ}$  and  $C_{\ell}^{TU}$  cross-spectra using cross-spectra of the half-depth data maps and cross-spectra from simulations that have no monopole leakage:

$$C_{\ell, \text{template}}^{TQ} = \epsilon^{Q,TT} C_{\ell, \text{data}}^{TT} + \epsilon^{Q,TQ} C_{\ell, \text{sim}}^{TQ}, \quad (\text{B4})$$

$$C_{\ell, \text{template}}^{TU} = \epsilon^{U,TT} C_{\ell, \text{data}}^{TT} + \epsilon^{U,TU} C_{\ell, \text{sim}}^{TU}. \quad (\text{B5})$$

In practice,  $C_{\ell, \text{sim}}^{TU}$  scatters around null, and therefore  $\epsilon^{U,TU}$  can take arbitrary values. For this reason, the second term is assumed to be 0 and only fits to  $\epsilon^{Q,TT}$ ,  $\epsilon^{Q,TQ}$ ,  $\epsilon^{U,TT}$  are carried out [111].

The observed (and possibly rotated)  $Q$  and  $U$  are related to the primordial  $\tilde{Q}$  and  $\tilde{U}$  by:

$$(Q + iU) = (\tilde{Q} + i\tilde{U})e^{i2\psi_{\text{pol}}}, \quad (\text{B6})$$

which introduces artificial  $TB$  and  $EB$  correlations:

$$C_\ell^{TB} = -\sin(2\psi_{\text{pol}})\tilde{C}_\ell^{TE} \quad (\text{B7})$$

$$C_\ell^{EB} = \frac{1}{2}\sin(4\psi_{\text{pol}})(\tilde{C}_\ell^{BB} - \tilde{C}_\ell^{EE}), \quad (\text{B8})$$

where  $\psi_{\text{pol}}$  is an rotation angle between 0 and  $2\pi$ .

A joint fit to the T-to-P and polarization rotation parameters is performed by generating 100 ( $\epsilon^{Q,TT}$ ,  $\epsilon^{Q,TQ}$ ,  $\epsilon^{U,TT}$ ,  $\psi_{\text{pol}}$ ) Latin hypercube training points, which are applied to the data maps (first T-to-P deprojection, then polarization rotation), before computing the resulting  $TB$  and  $EB$  data spectra. An emulator is then trained using the **CosmoPower** [112] package, which allows us to predict  $C_{\ell,\text{template}}^{TQ}$ ,  $C_{\ell,\text{template}}^{TU}$ ,  $C_\ell^{TB}$ , and  $C_\ell^{EB}$  at arbitrary points in the full parameter space. An MCMC chain is then run to find the best-fit parameters that produces  $C_\ell^{TQ}$  and  $C_\ell^{TU}$  spectra that match with data as well as  $\psi_{\text{pol}}$  that results in the  $C_\ell^{TE}$  and  $C_\ell^{EB}$  that is consistent with null.

### Appendix C: Covariance shrinkage

We develop and use a new procedure to reduce Monte Carlo error in the MUSE covariance. As a reminder, the covariance is given by

$$\Sigma = H^{-1}J(H^{-1})^\dagger = H^{-1}\langle s s^\dagger \rangle (H^{-1})^\dagger, \quad (\text{C1})$$

where  $s \equiv s^{\text{MAP}} - \langle s^{\text{MAP}} \rangle$  [see also Eqns. (7-9)]. The  $H$  matrix originates from second order derivatives of the posterior function and hence contains very little MC error. The  $J$  matrix, however, is a pure covariance of some finite set of samples of  $s$  and can contain significant MC error.

Under a singular value decomposition of  $H = USV^\dagger$ , the covariance,  $\Sigma$ , can be rewritten as,

$$VS^{-1}U^\dagger\langle s s^\dagger \rangle US^{-1}V^\dagger = VS^{-1/2}\langle (S^{-1/2}U^\dagger s)(s^\dagger US^{-1/2}) \rangle S^{-1/2}V^\dagger \quad (\text{C2})$$

Note that in the case of a Gaussian latent space,  $H = J$ , hence the quantity in angle brackets in Eqn. (C2) is the identity matrix. Even for the non-Gaussian latent space of the lensing problem, we have verified on simplified cases with thousands of samples of  $s$ , and for our exact problem with hundreds of samples, that this matrix remains extremely close to diagonal. This motivates a method of reducing MC error by applying a shrinkage estimator to compute the covariance in angle brackets in Eqn. (C2). Specifically, we shrink towards a diagonal target covariance with arbitrary diagonal entries. We find that this drastically reduces MC error of the covariance, and importantly, of its inverse, which is much more prone to MC error. Note that naively applying a diagonal shrinkage estimator to  $J = \langle s s^\dagger \rangle$  directly would not work, since nothing guarantees the  $J$  matrix is close to diagonal nor would there be any *a priori* known shrinkage target. Instead, we are essentially applying the shrinkage estimator to a preconditioned version of  $J$ , where the preconditioning is based on  $H$ , which itself contains almost no MC error and provides excellent preconditioning.