# Improved Constraints on Primordial Gravitational Waves using Planck, WMAP, and BICEP/Keck Observations through the 2018 Observing Season

P.A.R. Ade, Z. Ahmed, M. Amiri, D. Barkats, R. Basu Thakur, D. Beck, C.A. Bischoff, J.J. Bock, H. Boenish, E. Bullock, et al.

# BICEP / *Keck* XIII: Improved Constraints on Primordial Gravitational Waves using *Planck*, WMAP, and BICEP/*Keck* Observations through the 2018 Observing Season

BICEP/*Keck* Collaboration: P. A. R. Ade,[1] Z. Ahmed,[2] M. Amiri,[3] D. Barkats,[4] R. Basu Thakur,[5] C. A. Bischoff,[6] D. Beck,[2,7] J. J. Bock,[5,8] H. Boenish,[4] E. Bullock,[9] V. Buza,[10] J. R. Cheshire IV,[9] J. Connors,[4] J. Cornelison,[4] M. Crumrine,[11] A. Cukierman,[7,2] E. V. Denison,[12] M. Dierickx,[4] L. Duband,[13] M. Eiben,[4] S. Fatigoni,[3] J. P. Filippini,[14,15] S. Fliescher,[11] N. Goeckner-Wald,[7] D. C. Goldfinger,[4] J. Grayson,[7] P. Grimes,[4] G. Hall,[11] G. Halal,[7] M. Halpern,[3] E. Hand,[6] S. Harrison,[4] S. Henderson,[2] S. R. Hildebrandt,[5,8] G. C. Hilton,[12] J. Hubmayr,[12] H. Hui,[5] K. D. Irwin,[7,2,12] J. Kang,[7,5] K. S. Karkare,[4,10] E. Karpel,[7] S. Kefeli,[5] S. A. Kernasovskiy,[7] J. M. Kovac,[4,16] C. L. Kuo,[7,2] K. Lau,[11] E. M. Leitch,[10] A. Lennox,[14] K. G. Megerian,[8] L. Minutolo,[5] L. Moncelsi,[5] Y. Nakato,[7] T. Namikawa,[17] H. T. Nguyen,[8] R. O'Brient,[5,8] R. W. Ogburn IV,[7,2] S. Palladino,[6] T. Prouve,[13] C. Pryke,[11,9,*] B. Racine,[4,18] C. D. Reintsema,[12] S. Richter,[4] A. Schillaci,[5] R. Schwarz,[11] B. L. Schmitt,[4] C. D. Sheehy,[19] A. Soliman,[5] T. St. Germaine,[4,16] B. Steinbach,[5] R. V. Sudiwala,[1] G. P. Teply,[5] K. L. Thompson,[7,2] J. E. Tolan,[7] C. Tucker,[1] A. D. Turner,[8] C. Umiltà,[6,14] C. Vergès,[4] A. G. Vieregg,[20,10] A. Wandui,[5] A. C. Weber,[8] D. V. Wiebe,[3] J. Willmert,[11] C. L. Wong,[4,16] W. L. K. Wu,[2] H. Yang,[7] K. W. Yoon,[7,2] E. Young,[7,2] C. Yu,[7] L. Zeng,[4] C. Zhang,[5] and S. Zhang[5]

[1] *School of Physics and Astronomy, Cardiff University, Cardiff, CF24 3AA, United Kingdom*
[2] *Kavli Institute for Particle Astrophysics and Cosmology, SLAC National Accelerator Laboratory, 2575 Sand Hill Rd, Menlo Park, California 94025, USA*
[3] *Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, V6T 1Z1, Canada*
[4] *Center for Astrophysics, Harvard & Smithsonian, Cambridge, MA 02138, U.S.A*
[5] *Department of Physics, California Institute of Technology, Pasadena, California 91125, USA*
[6] *Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221, USA*
[7] *Department of Physics, Stanford University, Stanford, California 94305, USA*
[8] *Jet Propulsion Laboratory, Pasadena, California 91109, USA*
[9] *Minnesota Institute for Astrophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[10] *Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA*
[11] *School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[12] *National Institute of Standards and Technology, Boulder, Colorado 80305, USA*
[13] *Service des Basses Températures, Commissariat à l'Energie Atomique, 38054 Grenoble, France*
[14] *Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[15] *Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*
[16] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*
[17] *Kavli Institute for the Physics and Mathematics of the Universe (WPI), UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[18] *Aix-Marseille Université, CNRS/IN2P3, CPPM, 13288 Marseille, France*
[19] *Physics Department, Brookhaven National Laboratory, Upton, NY 11973*
[20] *Department of Physics, Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, USA*

(Published in PRL 4 October 2021)

We present results from an analysis of all data taken by the BICEP2, *Keck Array* and BICEP3 CMB polarization experiments up to and including the 2018 observing season. We add additional *Keck Array* observations at 220 GHz and BICEP3 observations at 95 GHz to the previous 95/150/220 GHz data set. The $Q/U$ maps now reach depths of 2.8, 2.8 and 8.8 $\mu\mathrm{K}_{\mathrm{CMB}}$ arcmin at 95, 150 and 220 GHz respectively over an effective area of $\approx 600$ square degrees at 95 GHz and $\approx 400$ square degrees at 150 & 220 GHz. The 220 GHz maps now achieve a signal-to-noise on polarized dust emission exceeding that of *Planck* at 353 GHz. We take auto- and cross-spectra between these maps and publicly available WMAP and *Planck* maps at frequencies from 23 to 353 GHz and evaluate the joint likelihood of the spectra versus a multicomponent model of lensed-$\Lambda$CDM+$r$+dust+synchrotron+noise. The foreground model has seven parameters, and no longer requires a prior on the frequency spectral index of the dust emission taken from measurements on other regions of the sky. This model is an adequate description of the data at the current noise levels. The likelihood analysis yields the constraint $r_{0.05} < 0.036$ at 95% confidence. Running maximum likelihood search on simulations we obtain unbiased results and find that $\sigma(r) = 0.009$. These are the strongest constraints to date on primordial gravitational waves.

PACS numbers: 98.70.Vc, 04.80.Nn, 95.85.Bh, 98.80.Es

*Introduction.*—The ΛCDM standard model of cosmology is able to describe the observable universe in a statistical manner using only six free parameters. Measurements of the cosmic microwave background (CMB) [1] are one the key pillars of this model and now constrain its parameters with percent-level precision (see most recently Ref. [2]).

The ΛCDM model describes how the universe evolved from an initial high energy state ($T \gg 10^{12}$ K), and the conditions at that time can be inferred from observations: fractionally small, Gaussian, adiabatic perturbations with a slightly red power law spectrum ($n_s \lesssim 1$). Inflationary theories naturally explain such conditions as the outcome of a pre-phase of exponential expansion during which the scale of the proto-universe increased by a factor of $\sim e^{60}$. Inflation makes an additional prediction which has not yet been observed—a background of tensor perturbations, also known as gravitational waves (see Ref. [3] for a review and citations to the original literature). There are many specific inflationary models and classes thereof. If we can detect or set limits on primordial gravitational waves we can set limits on these models [4], and probe physics at energy scales far higher than can ever be accessed in laboratory experiments.

A polarization pattern can be decomposed into $E$-mode (gradient) and $B$-mode (curl) components. Under the ΛCDM standard model the CMB polarization pattern is mostly $E$-mode, with a much smaller $B$-mode component which arises due to gravitational deflections (lensing) of the CMB photons after their last scattering [5]. Since primordial gravitational waves will produce $E$-modes and $B$-modes approximately equally it was realized in the late 1990's that the best way to search for them is to look for an excess $B$-mode signal [6–8]. Additional non-primordial $B$-modes are produced by astrophysical foreground emissions, primarily from our own galaxy, but these have different frequency spectra than the CMB, and can be separated from it using multi frequency measurements.

Our BICEP/*Keck* program first reported detection of an excess over the lensing $B$-mode expectation at 150 GHz in Ref. [9]. In a joint analysis using multi-frequency data from the *Planck* experiment it was shown that most or all of this is due to polarized emission from dust in our own galaxy [10, hereafter BKP]. In Ref. [11, hereafter BK14] we improved the constraint using *Keck Array* data at 95 GHz taken during the 2014 season, and in Ref. [12, hereafter BK15] we improved again adding *Keck Array* data at 95 GHz and 220 GHz taken during the 2015 season. In this letter [hereafter BK18] we add large amounts of new data taken by *Keck Array* at 220 GHz and BICEP3 at 95 GHz during the 2016, 2017 and 2018 observing seasons. This paper follows BK15 very closely in the methods, structure, and, in places, even the wording, mainly just adding additional experimental data. This improves the constraint on primordial gravitational waves parameterized by the tensor-to-scalar ratio $r$ by more than a factor of two over our previous result to $r_{0.05} < 0.036$ at 95% confidence, setting important additional limits on inflationary models.

*Instrument and observations.*—The BICEP2 receiver observed at 150 GHz from 2010–2012 [13]. The *Keck Array* was essentially five copies of BICEP2 running in parallel from 2012–2019, initially at 150 GHz but switching over time to 95 and 220 GHz [14]. BICEP3 is a single similar, but scaled up, receiver which commenced science observations in the 2016 Austral winter season [15]. Whereas the BICEP2 and *Keck* 150 & 220 GHz receivers each contained ≈ 500 bolometric detectors BICEP3 contains ≈ 2500 detectors. The aperture size is also increased from ≈ 0.25 m to ≈ 0.5 m. The *Keck* receivers were mounted on a single telescope mount (movable platform), while BICEP3 occupies a separate mount previously used for BICEP2 on a nearby building. All of these telescopes are located at the South Pole Station in Antarctica. The mounts scan the receivers across the sky, and the cryogenic detectors track the intensity of the incoming microwave radiation. The detectors are arranged as interleaved orthogonally polarized pairs in the focal planes and the pair difference timestreams are thus measures of the polarized emission from the sky [16]. At the South Pole the atmosphere is exceptionally transparent and stable at the observation frequencies [17, Fig. 5].

BICEP2 and *Keck Array* both mapped a region of sky centered at RA 0h, Dec. −57.5° with an effective area of ≈ 400 square degrees. BICEP3 has a larger instantaneous field of view and hence naturally maps a larger sky area with an effective area of ≈ 600 square degrees. We have perturbed the center of the BICEP3 scan region such that most of this additional area falls on the higher declination side of the sky patch in an attempt to stay away from regions where the *Planck* data indicates polarized dust contamination may be higher. The BK15 data set consisted of 4/17/2 receiver-years at 95/150/220 GHz respectively. BICEP3 is equivalent to about eight of the *Keck Array* 95 GHz receivers [15] so the BK18 data set is equivalent to about 28/18/14 *Keck* receiver-years at 95/150/220 GHz respectively.

*Maps and Power Spectra*—We make maps and power spectra using the same procedures as in our previous series of papers. The timestream data are binned into pixels on the sky using knowledge of the pointing direction of the telescope at each moment in time, together with the relative angles from the telescope boresight to each individual detector pair. By taking data with the receivers rotated at a range of angles, maps of the Stokes parameters $Q$ and $U$ can be constructed.

The maps at each observing frequency are subjected to a matrix purification operation [9, 18] such that they contain only structures sourced by $B$-modes of the underlying sky pattern. This allows us to measure the $B$-modes in the presence of the much brighter ΛCDM $E$-modes. The maps are then inverse noise variance apodized, Fourier transformed and rotated from the $Q/U$ to the $E/B$ basis. In this paper we use our own maps at 95, 150 and 220 GHz plus the 23 & 33 GHz

bands of WMAP [19][20] and the 30, 44, 143, 217 and 353 GHz maps from the NPIPE processing of the *Planck* data [21][22]. For illustration purposes we can inverse Fourier transform to form $E/B$ maps. Fig. 1 shows $E$- and $B$-mode maps at 95, 150 and 220 GHz. (See Appendix A for the full set of $T/Q/U$ maps.)

We take the variance within annuli of the Fourier plane to estimate the angular power spectra. Fig. 2 shows the $EE$ and $BB$ auto- and cross-spectra for the BICEP/*Keck* bands plus the *Planck* 353 GHz band which remains important for constraining the polarized dust contribution. Comparing this plot to Fig. 2 of BK15 we can see that the uncertainties are dramatically reduced for the auto- and cross-spectra of the 95 and 220 GHz bands. The model plotted is a "baseline" lensed-ΛCDM+dust model from our previous BK15 analysis, which remains a good description of the data. The $EE$ spectra were not used to derive the model but agree well with it under the assumption that $EE/BB = 2$ for dust, as is known to be close to the case [23, 24].

To test for systematic contamination we carry out our usual "jackknife" internal consistency (null) tests on the new 95 GHz and 220 GHz data as described in Appendices B and C. Fig. 3 upper shows the noise spectra for the three main BK18 bands after correction for the filter and beam suppression. In an auto-spectrum the quantity which determines the ability to constrain $r$ is the fluctuation of the noise bandpowers rather than their mean. The lower panel therefore shows the effective sky fraction as inferred from the fractional noise fluctuation. Together, these panels provide a useful synoptic measure of the loss of information due to noise, filtering, and $EE/BB$ separation in the lowest bandpowers (and we are glad to see taken up by others as e.g. Fig. 6 of Ref. [25]).

*Likelihood Analysis.*—We perform likelihood analysis using the methods introduced in BKP and refined in BK14 & BK15. We use the Hamimeche-Lewis approximation [26] to the joint likelihood of the ensemble of 66 $BB$ auto- and cross-spectra taken between the BICEP/*Keck*, WMAP and *Planck* maps. We compare the observed bandpower values for $20 < \ell < 330$ (9 bandpowers per spectrum) to an eight parameter model of lensed-ΛCDM+$r$+dust+synchrotron+noise and explore the parameter space using `COSMOMC` [27] (which implements a Markov chain Monte Carlo method). As in our previous analyses the bandpower covariance matrix is derived from 499 simulations of signal and noise, explicitly setting to zero terms such as the covariance of signal-only bandpowers with noise-only bandpowers or covariance of BICEP/*Keck* noise bandpowers with WMAP/*Planck* noise bandpowers (see Appendix H of BK15 and Appendix B of Ref. [28] for details). We deal with the differing sky coverage of the BICEP3 and BICEP2/*Keck* maps as described in Appendix D. The tensor/scalar power ratio $r$ is evaluated at a pivot scale of 0.05 Mpc$^{-1}$, and we fix the tensor spectral index $n_t = 0$. A `COSMOMC` module containing the data and model is available for download at `http://bicepkeck.org`. The following paragraphs briefly summarize the foreground model.

We include dust with amplitude $A_{\rm d,353}$ evaluated at 353 GHz and $\ell = 80$. The frequency spectral behavior is taken as a modified black body spectrum with $T_{\rm d} = 19.6$ K and frequency spectral index $\beta_{\rm d}$. In a significant change from the baseline analysis choices of BK15, we remove the prior on the dust frequency spectral index which was previously applied based on *Planck* data in other regions of sky—with the improvement in the *Keck* 220 GHz sensitivity this prior is no longer needed. The spatial power spectrum is taken as a power law $\mathcal{D}_\ell \propto \ell^{\alpha_{\rm d}}$ marginalizing uniformly over the (generous) range $-1 < \alpha_{\rm d} < 0$ (where $\mathcal{D}_\ell \equiv \ell\,(\ell+1)\,C_\ell/2\pi$). *Planck* analysis consistently finds approximate power law behavior of both the $EE$ and $BB$ dust spectra with exponents $\approx -0.4$ [23, 24].

We include synchrotron with amplitude $A_{\rm sync,23}$ evaluated at 23 GHz (the lowest WMAP band) and $\ell = 80$, assuming a simple power law for the frequency spectral behavior $A_{\rm sync} \propto \nu^{\beta_{\rm s}}$, and using a Gaussian prior $\beta_{\rm s} = -3.1 \pm 0.3$ taken from the analysis of WMAP 23 and 33 GHz data in Ref [29]. We note that analysis of 2.3 GHz data from S-PASS in conjunction with WMAP and *Planck* finds $\beta_{\rm s} = -3.2$ with no detected trends with galactic latitude or angular scale [30], and that Ref. [31] analyzed the S-PASS and WMAP 23 GHz data and found $\beta_{\rm s} = -3.22 \pm 0.06$ in the BICEP2 sky patch. The spatial power spectrum is taken as a power law $\mathcal{D}_\ell \propto \ell^{\alpha_{\rm s}}$ marginalizing over the range $-1 < \alpha_{\rm s} < 0$ [32]. Ref. [30] finds a value at the bottom end of this range ($\approx -1$) from the S-PASS data for $BB$ at high galactic latitude.

Finally we include sync/dust correlation parameter $\epsilon$ (called $\rho$ in some other papers [24, 30, 33]). As in BK15 we marginalize over the full possible range $-1 < \epsilon < 1$.

We hold the lensing $B$-mode spectrum fixed at that predicted for the *Planck* 2018 cosmological parameters [2, Table 2]. Results of our baseline analysis are shown in Fig. 4 and yield the following statistics: $r_{0.05} = 0.014^{+0.010}_{-0.011}$ ($r_{0.05} < 0.036$ at 95% confidence), $A_{\rm d,353} = 4.4^{+0.8}_{-0.7}\,\mu{\rm K}^2$, $A_{\rm sync,23} < 1.4\,\mu{\rm K}^2$ at 95% confidence, and $\beta_{\rm d} = 1.49^{+0.13}_{-0.12}$. For $r$, the zero-to-peak likelihood ratio is 0.46. Taking $\frac{1}{2}\,(1 - f\,(-2\log L_0/L_{\rm peak}))$, where $f$ is the $\chi^2$ CDF (for one degree of freedom), we estimate that the probability to get a likelihood ratio smaller than this is 11% if, in fact, $r = 0$. As compared to the previous BK15 analysis, the likelihood curve for $r$ tightens considerably with the peak position shifting down slightly, and the $A_{\rm d}$ curve tightens slightly. In addition the $A_{\rm sync}$ curve now peaks at zero—the weak evidence for synchrotron we saw in BK15 is no longer present. (Using the S-PASS data [30] we estimate that the expectation is $A_{\rm sync,23} \approx 0.4\,\mu{\rm K}^2$ in the BICEP/*Keck* field, which is consistent with our $A_{\rm sync}$ likelihood curve.) In the BK15 analysis the constraint on $\beta_{\rm d}$ was prior dominated, but for BK18 we see that the data is able to constrain this parameter almost as well as the prior previously did. Interestingly the peak value selected is very close to the mean value from *Planck* 2018 analysis of larger regions

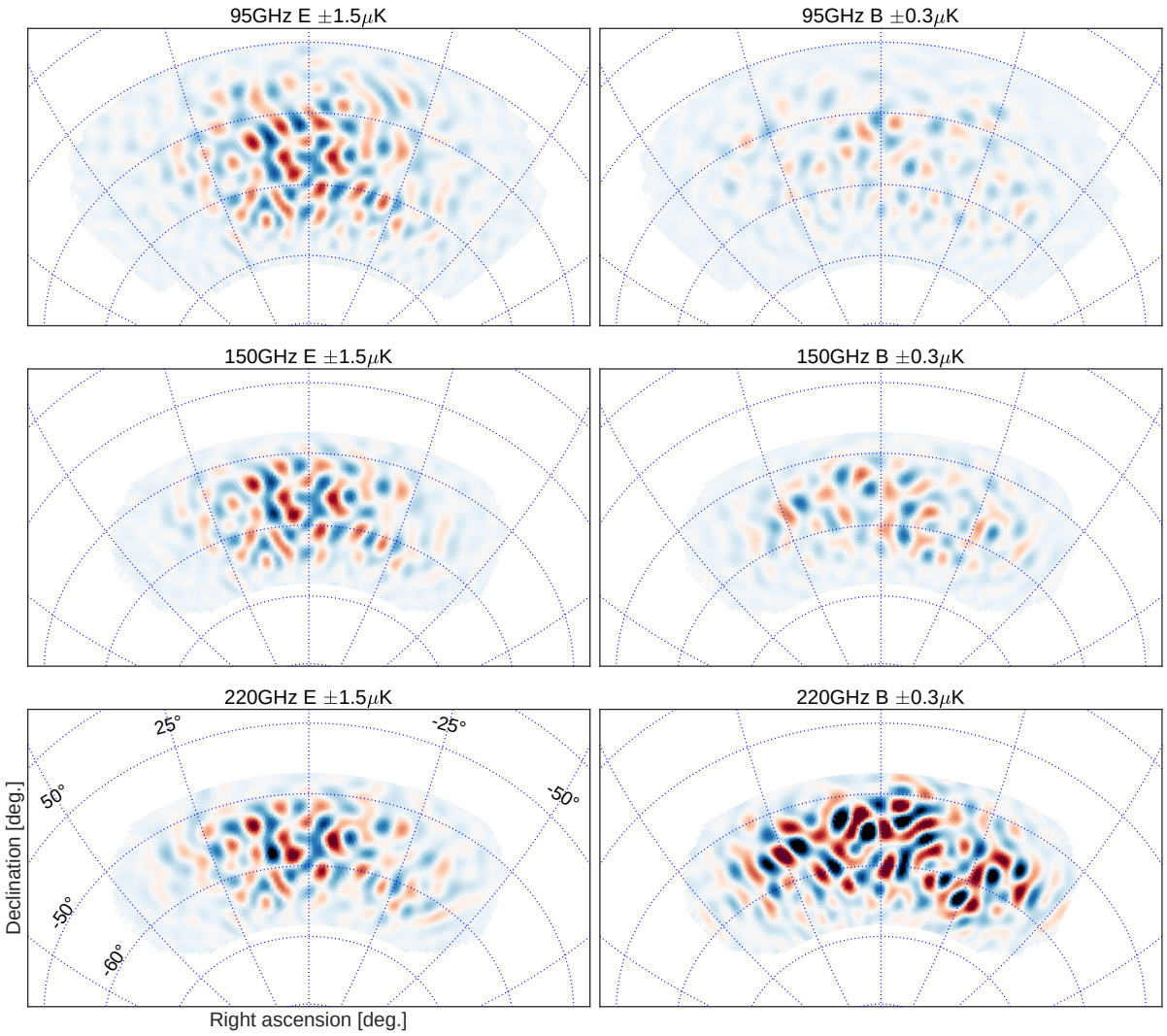


FIG. 1. $E$-mode (left column) and $B$-mode (right column) maps at 95, 150 and 220 GHz in CMB units, and filtered to degree angular scales ($50 < \ell < 120$). Note the differing color ranges left and right. The $E$ maps are dominated by ΛCDM signal, and hence are highly correlated across all three bands. The 95 GHz $B$ map is approximately equal parts lensed-ΛCDM signal and noise. At 150 and 220 GHz the $B$ maps are dominated by polarized dust emission.

of sky $\beta_{\rm d} = 1.53$ [24].

The maximum likelihood model has parameters $r_{0.05} = 0.011$, $A_{\rm d,353} = 4.4\,\mu{\rm K}^2$, $A_{\rm sync,23} = 0.6\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.5$, $\beta_{\rm s} = -3.0$, $\alpha_{\rm d} = -0.66$, $\alpha_{\rm s} = 0.00$, and $\epsilon = -0.11$. This model is an acceptable fit to the data with the probability to exceed (PTE) the observed value of $\chi^2$ being 0.94. Thus, while the dust spectrum might in general be expected to exhibit fluctuations about power law spatial spectral behavior greater than that expected for a Gaussian random field, for the present the model continues to be an adequate description of the data—see Appendix D for further details.

In Appendix E we explore variation and validation of the likelihood. In Appendix E 2 we vary the baseline analysis choices and data selection, finding that these do not significantly alter the results, and that the data do not prefer allowing decorrelation of the dust pattern in the model. We also find that the value of $A_{\rm d}$ is very similar when evaluated over the larger BICEP3 sky coverage region and the smaller BICEP2/*Keck* sky region. Freeing the amplitude of the lensing power we obtain $A_{\rm L}^{\rm BB} = 1.03^{+0.08}_{-0.09}$, and the $r$ constraint hardly changes. In Appendix E 3 we verify that the likelihood analysis is unbiased, and in Appendix E 4 we explore a suite of alternate foreground models. As part of our standard data reduction we "deproject" leading order temperature to polarization leakage [9, 34]—in Appendix F we quantify possible residual leakage and some other possible systematics.

Fig. 5 shows the constraints in the $r$ vs. $n_s$ plane for

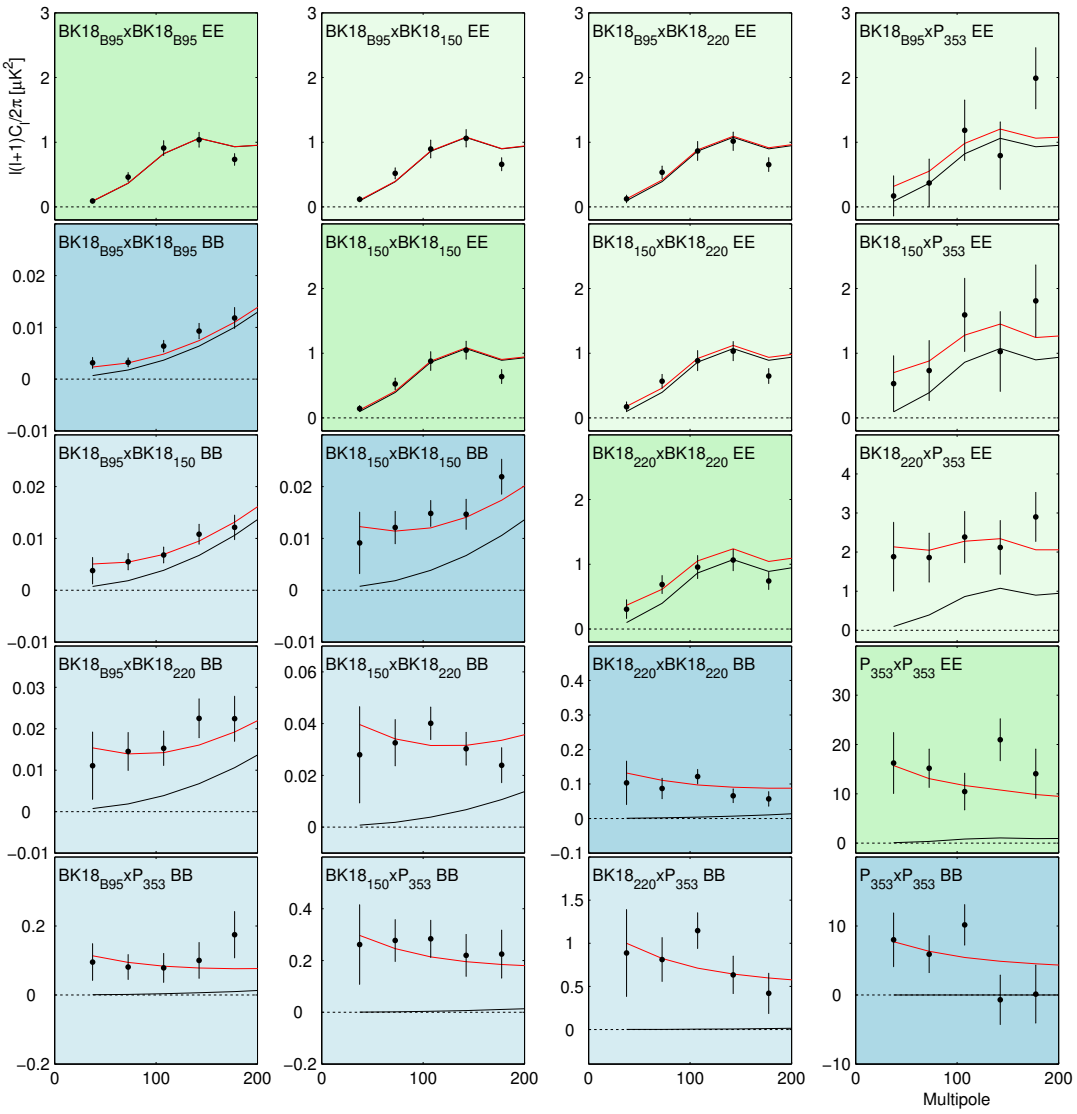


FIG. 2. *EE* (green) and *BB* (blue) auto- and cross-spectra calculated using the BICEP3 95 GHz map, the BICEP2/*Keck* 150 GHz map, the *Keck* 220 GHz map, and the *Planck* 353 GHz map (with the auto-spectra in darker colors). The BICEP/*Keck* maps use all data taken up to and including the 2018 observing season—we refer to these as BK18. The black lines show the model expectation values for lensed-ΛCDM, while the red lines show the expectation values of a baseline lensed-ΛCDM+dust model from our previous BK15 analysis ($r = 0$, $A_{\rm d,353} = 4.7\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.6$, $\alpha_{\rm d} = -0.4$). Note that the model shown was fit to *BB* only and did not use the BICEP3 95 GHz points shown (which are entirely new). The agreement with the spectra involving 95 GHz and all the *EE* spectra (under the assumption that $EE/BB = 2$ for dust) is therefore a validation of the model.

the *Planck* 2018 baseline analysis [2] and when adding in BK18 & BAO. The BK18 data shrinks the contours in the vertical ($r$) direction while the BAO data shrinks the contours in the horizontal ($n_s$) direction and shifts the centroid slightly to the right. The $\phi^{2/3}$ model now lies outside the 95% contour as does the band of natural inflation models.

*Conclusions.*—The BKP analysis yielded a 95% confidence constraint $r_{0.05} < 0.12$, which BK14 improved to $r_{0.05} < 0.09$, and BK15 improved to $r_{0.05} < 0.07$. The BK18 result described in this letter, $r_{0.05} < 0.036$, represents a fractional improvement equivalent to the two previous steps combined. The BK18 simulations have a median 95% upper limit of $r_{0.05} < 0.019$.

The distributions of maximum likelihood $r$ values in simulations where the true value of $r$ is zero gave $\sigma(r_{0.05}) = 0.020$ for BK15 which is reduced to $\sigma(r_{0.05}) = 0.009$ for BK18 (see Appendix E 3 for details). Such simulations can also be used to investigate the degree to which the analysis is limited by foregrounds and lensing. Running the baseline BK18 analysis on simulations which contain no lensing $B$-modes gives $\sigma(r_{0.05}) = 0.004$, while running without foreground parameters on simulations which contain no dust gives $\sigma(r_{0.05}) = 0.007$.

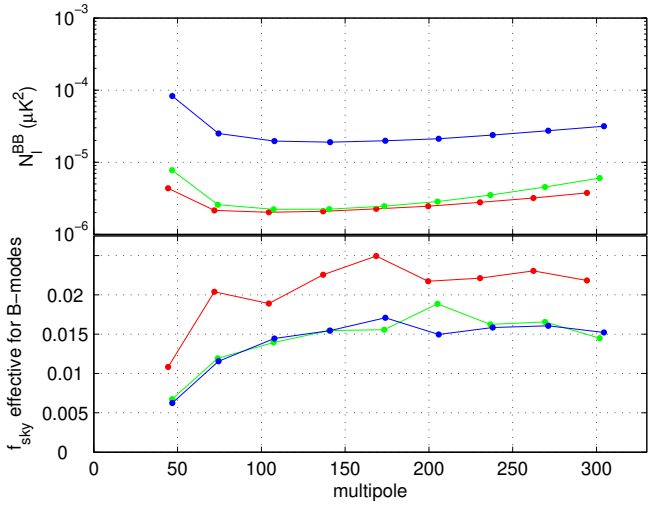


FIG. 3. *Upper:* The noise spectra of the BICEP3 95 GHz map (red), the BICEP2/*Keck* 150 GHz map (green) and the *Keck* 220 GHz maps (blue). The spectra are shown after correction for the filtering of signal which occurs due to the beam roll-off, timestream filtering, and $B$-mode purification. (Note that no $\ell^2$ scaling is applied.) *Lower:* The effective sky fraction as calculated from the ratio of the mean noise realization bandpowers to their fluctuation $f_{\rm sky}(\ell) = \frac{1}{2\ell\Delta\ell}\left(\frac{\sqrt{2}\bar{N}_b}{\sigma(N_b)}\right)^2$, i.e. the observed number of $B$-mode degrees of freedom divided by the nominal full-sky number. The turn-down at low $\ell$ is due to mode loss to the timestream filtering and matrix purification.

Running without foreground parameters on simulations which contain neither lensing or dust gives $\sigma(r_{0.05}) = 0.002$.

Fig. 6 shows the BK18 noise uncertainties in the $\ell \approx 80$ bandpowers as compared to the signal levels. The signal-to-noise on polarized dust emission of our 220 GHz band is now considerably higher than that of the *Planck* 353 GHz band—i.e. the 220 × 220 noise point is much further below the dust band than the P353×P353 point. Additional BICEP3 data taken during 2019–2021 will reduce the noise by a factor greater than 2 and $\sqrt{2}$ for 95 × 95 and 95×W23 respectively, and we have also recorded additional data at 220 and 270 GHz.

Fig. 7 shows the estimated CMB-only component of the BK18 $B$-mode bandpowers versus measurements from other experiments. See Appendix D for a description of how the CMB-only power spectrum estimate is calculated.

Figure 2 shows that the BK18 data is consistent with ΛCDM plus a remarkably simple dust only foreground model. Nevertheless as we move forward to even higher levels of sensitivity dust decorrelation, and foreground complexity more generally, will remain a serious concern. In addition, we are already in the regime where the sample variance of the lensing component dominates $\sigma(r)$. However, the lensing $B$-modes can be spatially separated from a primordial component and in this regard we have recently demonstrated a path forward by adding a "lensing template" derived from SPTpol and *Planck* data to the BK14 analysis, resulting in an improved constraint on $r$ [49].

The *Keck Array* mount has now been replaced by a larger, more capable machine and we are in the process of upgrading to a new system we call BICEP Array [50, 51]. A BICEP3 class receiver is now operating in the 30/40 GHz band and in the coming years additional receivers will be installed at 95, 150 and 220/270 GHz. The system is projected to reach $\sigma(r) \sim 0.003$ within five years with delensing in conjunction with SPT3G.

## ACKNOWLEDGMENTS

The BICEP/*Keck* projects have been made possible through a series of grants from the National Science Foundation including 0742818, 0742592, 1044978, 1110087, 1145172, 1145143, 1145248, 1639040, 1638957, 1638978, & 1638970, and by the Keck Foundation. The development of antenna-coupled detector technology was supported by the JPL Research and Technology Development Fund, and by NASA Grants 06-ARPA206-0040, 10-SAT10-0017, 12-SAT12-0031, 14-SAT14-0009 & 16-SAT-16-0002. The development and testing of focal planes were supported by the Gordon and Betty Moore Foundation at Caltech. Readout electronics were supported by a Canada Foundation for Innovation grant to UBC. Support for quasi-optical filtering was provided by UK STFC grant ST/N000706/1. The computations in this paper were run on the Odyssey/Cannon cluster supported by the FAS Science Division Research Computing Group at Harvard University. The analysis effort at Stanford and SLAC is partially supported by the U.S. DOE Office of Science. We thank the staff of the U.S. Antarctic Program and in particular the South Pole Station without whose help this research would not have been possible. Most special thanks go to our heroic winter-overs Robert Schwarz, Steffen Richter, Sam Harrison, Grantland Hall and Hans Boenish. We thank all those who have contributed past efforts to the BICEP/*Keck* series of experiments, including the BICEP1 team. We also thank the *Planck* and WMAP teams for the use of their data, and are grateful to the *Planck* team for helpful discussions.

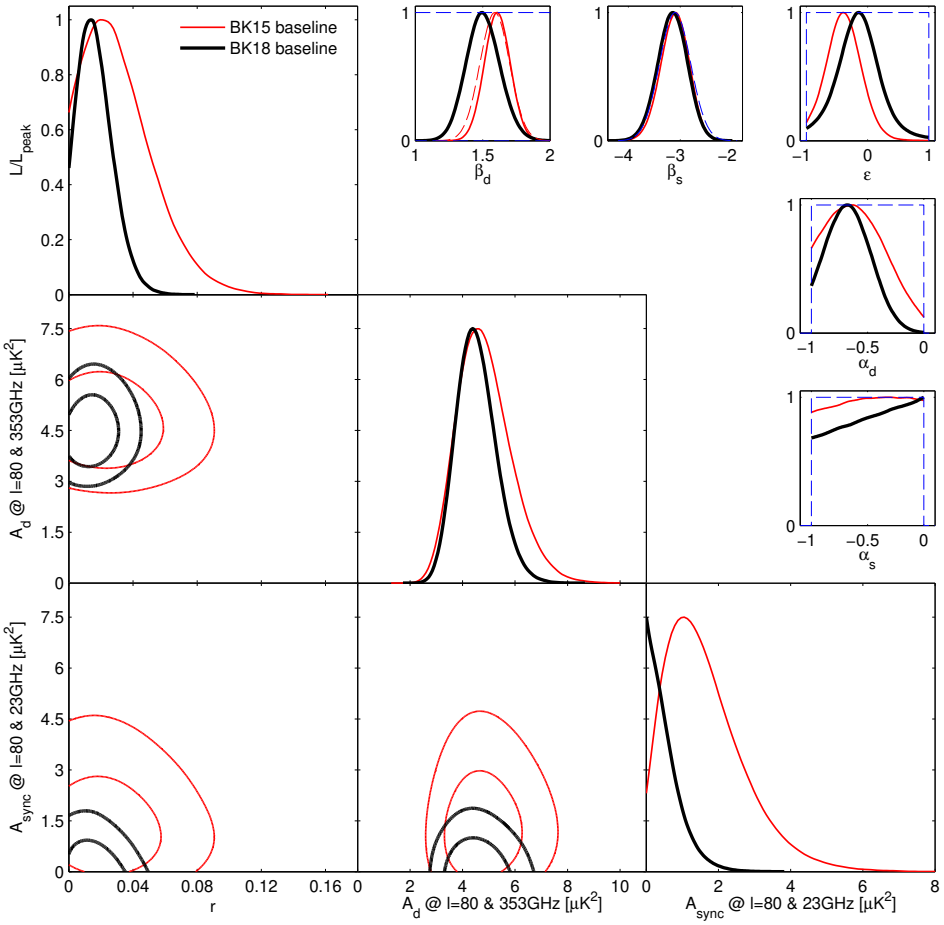


FIG. 4. Results of a multicomponent multi-spectral likelihood analysis of BICEP/*Keck*+WMAP/*Planck* data. The red faint curves are the baseline result from the previous BK15 paper (the black curves from Fig. 4 of that paper). The bold black curves are the new baseline BK18 results, adding a large amount of additional data at 95 and 220 GHz taken by BICEP3 and *Keck Array* during the 2016–2018 observing seasons. The upper limit on the tensor-to-scalar ratio tightens to $r_{0.05} < 0.036$ at 95% confidence. The parameters $A_{\rm d}$ and $A_{\rm sync}$ are the amplitudes of the dust and synchrotron $B$-mode power spectra, where $\beta$ and $\alpha$ are the respective frequency and spatial spectral indices. The correlation coefficient between the dust and synchrotron patterns is $\epsilon$. In the $\beta$, $\alpha$ and $\epsilon$ panels the dashed lines show the priors placed on these parameters (either Gaussian or uniform). Note that the Gaussian prior on $\beta_{\rm d}$ has been removed going from BK15 to BK18.

* pryke@physics.umn.edu


[1] A. A. Penzias and R. W. Wilson, Astrophys. J. **142**, 419 (1965).
[2] Planck Collaboration 2018 VI, Astron. Astrophys. **641**, A6 (2020), arXiv:1807.06209.
[3] M. Kamionkowski and E. D. Kovetz, Annual Review of Astronomy and Astrophysics **54**, 227 (2016), arXiv:1510.06042.
[4] CMB-S4 Collaboration, ArXiv e-prints (2016), arXiv:1610.02743.
[5] M. Zaldarriaga and U. Seljak, Phys. Rev. D **58**, 023003 (1998), astro-ph/9803150.
[6] U. Seljak, Astrophys. J. **482**, 6 (1997), astro-ph/9608131.
[7] M. Kamionkowski, A. Kosowsky, and A. Stebbins, Phys. Rev. Lett. **78**, 2058 (1997), astro-ph/9609132.
[8] U. Seljak and M. Zaldarriaga, Phys. Rev. Lett. **78**, 2054 (1997), astro-ph/9609169.
[9] BICEP2 Collaboration I, Physical Review Letters **112**, 241101 (2014), arXiv:1403.3985.
[10] BICEP2/*Keck* and *Planck* Collaborations, Physical Review Letters **114**, 101301 (2015), arXiv:1502.00612.
[11] *Keck Array* and BICEP2 Collaborations VI, Physical Review Letters **116**, 031302 (2016), arXiv:1510.09217.
[12] *Keck Array* and BICEP2 Collaborations X, Phys. Rev. Lett. **121**, 221301 (2018), arXiv:1810.05216.
[13] BICEP2 Collaboration II, Astrophys. J. **792**, 62 (2014), arXiv:1403.4302.
[14] *Keck Array* and BICEP2 Collaborations V, Astrophys. J. **811**, 126 (2015), arXiv:1502.00643.

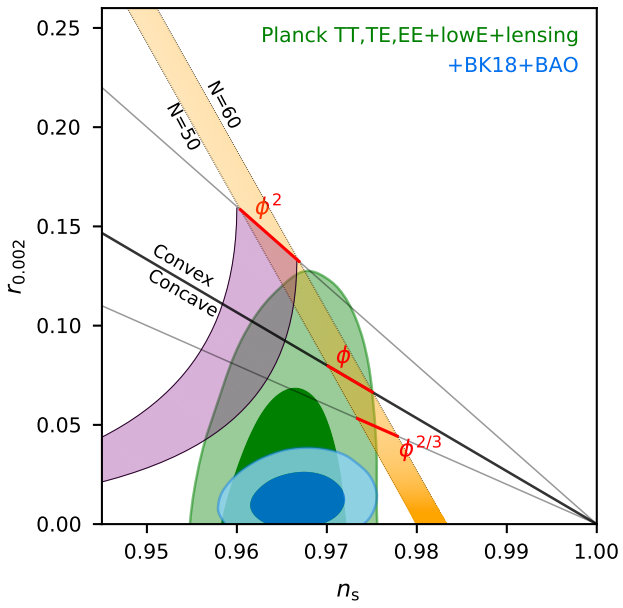


FIG. 5. Constraints in the $r$ vs. $n_s$ plane for the *Planck* 2018 baseline analysis, and when also adding BICEP/*Keck* data through the end of the 2018 season plus BAO data to improve the constraint on $n_s$. The constraint on $r$ tightens from $r_{0.05} < 0.11$ to $r_{0.05} < 0.035$. This figure is adapted from Fig. 28 of Ref. [2] with the green contours being identical. Some additional inflationary models are added from Fig. 8 of Ref. [35] with the purple region being natural inflation.

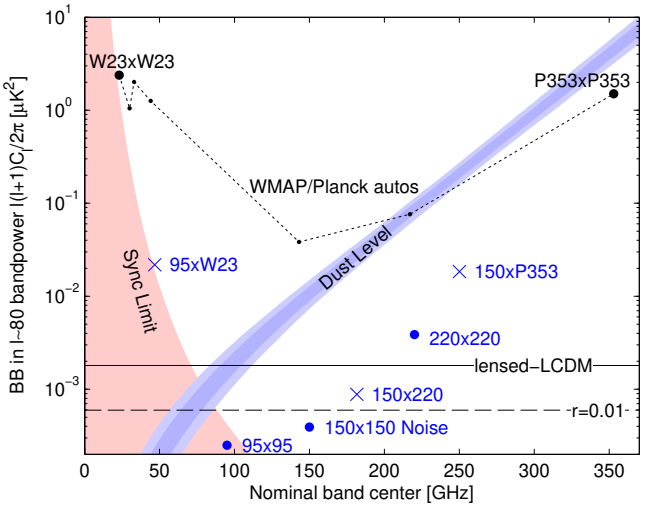


FIG. 6. Expectation values and noise uncertainties of the $\ell \sim 80$ $BB$ bandpower in the BICEP/*Keck* field. The solid and dashed black lines show the expected signal power of lensed-ΛCDM and $r_{0.05} = 0.01$. Since CMB units are used, the levels corresponding to these are flat with frequency. The blue bands show the 1 and $2\sigma$ ranges of dust, and the red shaded region shows the 95% upper limit on synchrotron in the baseline analysis including the uncertainties in the amplitude and frequency spectral index parameters ($A_{\rm sync,23}, \beta_{\rm s}$ and $A_{\rm d,353}, \beta_{\rm d}$). The BICEP/*Keck* auto-spectrum noise uncertainties are shown as large blue circles, and the noise uncertainties of the used WMAP/*Planck* single-frequency spectra evaluated in the BICEP/*Keck* field are shown in black. The blue crosses show the noise uncertainty of selected cross-spectra, and are plotted at horizontal positions such that they can be compared vertically with the dust and sync curves.


[15] BICEP/*Keck* Collaboration XV, Submitted to Astrophys. J. (2021).
[16] BICEP2/*Keck* and Spider Collaborations, Astrophys. J. **812**, 176 (2015), arXiv:1502.00619.
[17] C.-L. Kuo, Astrophys. J. **848**, 64 (2017), arXiv:1707.08400.
[18] *Keck Array* and BICEP2 Collaborations VII, Astrophys. J. **825**, 66 (2016), arXiv:1603.05976.
[19] See `http://lambda.gsfc.nasa.gov/product/map/dr5/m_products.cfm`.
[20] C. L. Bennett, D. Larson, J. L. Weiland, N. Jarosik, G. Hinshaw, N. Odegard, K. M. Smith, R. S. Hill, B. Gold, M. Halpern, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, J. Dunkley, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright, Astrophys. J. Suppl. Ser. **208**, 20 (2013), arXiv:1212.5225.
[21] See `https://irsa.ipac.caltech.edu/data/Planck/release_3/ancillary-data/HFI_Products.html`.
[22] Planck Collaboration Int. LVII, Astron. Astrophys. **643**, A42 (2020), arXiv:2007.04997.
[23] Planck Collaboration Int. XXX, Astron. Astrophys. **586**, A133 (2016), arXiv:1409.5738.
[24] Planck Collaboration 2018 XI, Astron. Astrophys. **641**, A11 (2020), arXiv:1801.04945.
[25] POLARBEAR Collaboration, Astrophys. J. **897**, 55 (2020), arXiv:1910.02608.
[26] S. Hamimeche and A. Lewis, Phys. Rev. D **77**, 103013 (2008), arXiv:0801.0554.
[27] A. Lewis and S. Bridle, Phys. Rev. D **66**, 103511 (2002), astro-ph/0205436.
[28] CMB-S4 Collaboration, submitted to Astrophys. J. (2020), arXiv:2008.12619.
[29] U. Fuskeland, I. K. Wehus, H. K. Eriksen, and S. K. Næss, Astrophys. J. **790**, 104 (2014), arXiv:1404.5323.
[30] N. Krachmalnicoff, E. Carretti, C. Baccigalupi, G. Bernardi, S. Brown, B. M. Gaensler, M. Haverkorn, M. Kesteven, F. Perrotta, S. Poppi, and L. Staveley-Smith, Astron. Astrophys. **618**, A166 (2018), arXiv:1802.01145.
[31] U. Fuskeland, K. J. Andersen, R. Aurlien, R. Banerji, M. Brilenkov, H. K. Eriksen, M. Galloway, E. Gjerløw, S. K. Næss, T. L. Svalheim, and I. K. Wehus, Astron. Astrophys. **646**, A69 (2021), arXiv:1909.05923.
[32] J. Dunkley, A. Amblard, C. Baccigalupi, M. Betoule, D. Chuss, A. Cooray, J. Delabrouille, C. Dickinson, G. Dobler, J. Dotson, H. K. Eriksen, D. Finkbeiner, D. Fixsen, P. Fosalba, A. Fraisse, C. Hirata, A. Kogut, J. Kristiansen, C. Lawrence, A. M. Magalhães, M. A. Miville-Deschenes, *et al.*, AIP Conf. Proc. **1141**, 222 (2009), arXiv:0811.3915.
[33] S. K. Choi and L. A. Page, J. Cosmol. Astropart. Phys. **12**, 020 (2015), arXiv:1509.05934.
[34] BICEP2 Collaboration III, Astrophys. J. **814**, 110 (2015), arXiv:1502.00608.

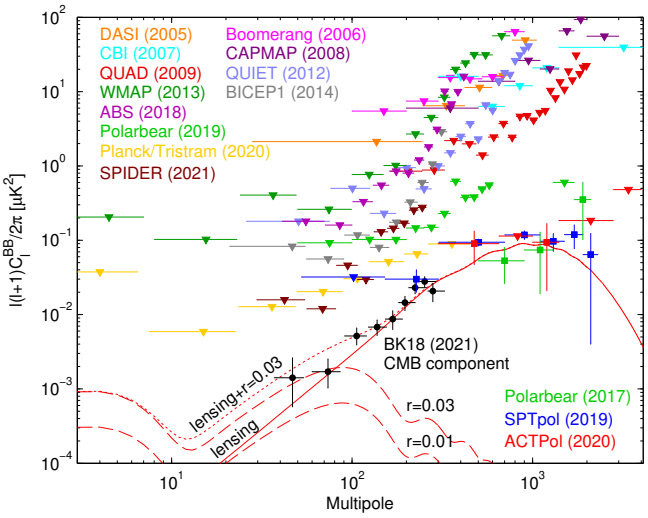


FIG. 7. Summary of CMB $B$-mode polarization upper limits [20, 25, 36–45] and detections [46–48]. Theoretical predictions are shown for the lensing $B$-modes (solid red) which peak at arcminute scales (multipole $\ell \sim 1000$), and for gravitational wave $B$-modes (dashed red) for two values of $r$ peaking at degree scales ($\ell \sim 80$). The BK18 data are shown after removing Galactic foregrounds.


[35] Planck Collaboration 2018 X, Astron. Astrophys. **641**, A10 (2020), arXiv:1807.06211.

[36] E. M. Leitch, J. M. Kovac, N. W. Halverson, J. E. Carlstrom, C. Pryke, and M. W. E. Smith, Astrophys. J. **624**, 10 (2005), astro-ph/0409357.

[37] T. E. Montroy, P. A. R. Ade, J. J. Bock, J. R. Bond, J. Borrill, A. Boscaleri, P. Cabella, C. R. Contaldi, B. P. Crill, P. de Bernardis, G. De Gasperis, A. de Oliveira-Costa, G. De Troia, G. di Stefano, E. Hivon, A. H. Jaffe, T. S. Kisner, W. C. Jones, A. E. Lange, S. Masi, *et al.*, Astrophys. J. **647**, 813 (2006), astro-ph/0507514.

[38] J. L. Sievers, C. Achermann, J. R. Bond, L. Bronfman, R. Bustos, C. R. Contaldi, C. Dickinson, P. G. Ferreira, M. E. Jones, A. M. Lewis, B. S. Mason, J. May, S. T. Myers, N. Oyarce, S. Padin, T. J. Pearson, M. Pospieszalski, A. C. S. Readhead, R. Reeves, A. C. Taylor, and S. Torres, Astrophys. J. **660**, 976 (2007), astro-ph/0509203.

[39] CAPMAP Collaboration, Astrophys. J. **684**, 771 (2008), arXiv:0802.0888.

[40] QUaD Collaboration, Astrophys. J. **705**, 978 (2009), arXiv:0906.1003.

[41] QUIET Collaboration, Astrophys. J. **760**, 145 (2012), arXiv:1207.5034.

[42] BICEP1 Collaboration, Astrophys. J. **783**, 67 (2014), arXiv:1310.1422.

[43] A. Kusaka, J. Appel, T. Essinger-Hileman, J. A. Beall, L. E. Campusano, H.-M. Cho, S. K. Choi, K. Crowley, J. W. Fowler, P. Gallardo, M. Hasselfield, G. Hilton, S.-P. P. Ho, K. Irwin, N. Jarosik, M. D. Niemack, G. W. Nixon, M. Nolta, L. A. Page, Jr., G. A. Palma, L. Parker, S. Raghunathan, C. D. Reintsema, J. Sievers, S. M. Simon, S. T. Staggs, K. Visnjic, and K.-W. Yoon, J. Cosmol. Astropart. Phys. **9**, 005 (2018), arXiv:1801.01218.

[44] M. Tristram, A. J. Banday, K. M. Górski, R. Keskitalo, C. R. Lawrence, K. J. Andersen, R. B. Barreiro, J. Borrill, H. K. Eriksen, R. Fernandez-Cobos, T. S. Kisner, E. Martínez-González, B. Partridge, D. Scott, T. L. Svalheim, H. Thommesen, and I. K. Wehus, Astron. Astrophys. **647**, A128 (2021), arXiv:2010.01139.

[45] SPIDER Collaboration, submitted to Astrophys. J. (2021), arXiv:2103.13334.

[46] J. T. Sayre, C. L. Reichardt, J. W. Henning, P. A. R. Ade, *et al.*, Phys. Rev. D **101**, 122003 (2020), arXiv:1910.05748.

[47] S. K. Choi, M. Hasselfield, S.-P. P. Ho, B. Koopman, M. Lungu, M. H. Abitbol, G. E. Addison, P. A. R. Ade, *et al.*, J. Cosmol. Astropart. Phys. **2020**, 045 (2020), arXiv:2007.07289.

[48] POLARBEAR Collaboration, Astrophys. J. **848**, 121 (2017), arXiv:1705.02907.

[49] BICEP2/*Keck* and SPTPOL Collaborations, Phys. Rev. D **103**, 022004 (2021), arXiv:2011.08163.

[50] H. Hui and the BICEP/*Keck* Collaboration, in *SPIE Conference Series*, Vol. 10708 (2018) p. 1070807, arXiv:1808.00568.

[51] L. Moncelsi and the BICEP/*Keck* Collaboration, in *SPIE Conference Series*, Vol. 11453 (2020) p. 1145314, arXiv:2012.04047.

[52] *Keck Array* and BICEP2 Collaborations XI, Astrophys. J. **884**, 114 (2019), arXiv:1904.01640.

[53] J. Willmert, *Constraining Inflationary B-modes with the BICEP/Keck Array Telescopes*, Ph.D. thesis, University of Minnesota (2019).

[54] Planck Collaboration Int. XXII, Astron. Astrophys. **576**, A107 (2015), arXiv:1405.0874.

[55] V. Pelgrims, S. E. Clark, B. S. Hensley, G. V. Panopoulou, V. Pavlidou, K. Tassis, H. K. Eriksen, and I. K. Wehus, Astron. Astrophys. **647**, A16 (2021), arXiv:2101.09291.

[56] G. V. Panopoulou and D. Lenz, Astrophys. J. **902**, 120 (2020), arXiv:2004.00647.

[57] Planck Collaboration Int. L, Astron. Astrophys. **599**, A51 (2017), arXiv:1606.07335.

[58] B. Thorne, J. Dunkley, D. Alonso, and S. Næss, Mon. Not. R. Astron. Soc. **469**, 2821 (2017), arXiv:1608.02841.

[59] A. G. Kritsuk, S. D. Ustyugov, and M. L. Norman, New Journal of Physics **19**, 065003 (2017), arXiv:1705.01912.

[60] A. G. Kritsuk, R. Flauger, and S. D. Ustyugov, Phys. Rev. Lett. **121**, 021104 (2018), arXiv:1711.11108.

[61] G. Martínez-Solaeche, A. Karakci, and J. Delabrouille, Mon. Not. R. Astron. Soc. **476**, 1310 (2018), arXiv:1706.04162.

[62] F. Vansyngel, F. Boulanger, T. Ghosh, B. Wandelt, J. Aumont, A. Bracco, F. Levrier, P. G. Martin, and L. Montier, Astron. Astrophys. **603**, A62 (2017), arXiv:1611.02577.

[63] *Keck Array* and BICEP2 Collaborations IV, Astrophys. J. **806**, 206 (2015), arXiv:1502.00596.

[64] R. A. Battye, I. W. A. Browne, M. W. Peel, N. J. Jackson, and C. Dickinson, Mon. Not. R. Astron. Soc. **413**, 132 (2011), arXiv:1003.5846.

### Appendix A: Maps

Figures 8, 9 & 10 show $T/Q/U$ maps at 95 GHz from BICEP3, at 150 GHz from BICEP2/*Keck* and at 220 GHz from *Keck*. The right side of each figure shows realizations of noise created by randomly flipping the sign of data subsets while coadding the map—see Sec. V.B of Ref. [9] for further details. Due to the larger instantaneous field of view BICEP3 naturally maps a larger area of sky. We have perturbed the boresight "scan-box" on the sky to higher declination such that most of the additional area lies on that side of the BICEP2/*Keck* observed region. Comparing the three frequencies we see almost identical patterns in the $T$ maps since ΛCDM signal is dominant. For $Q$ and $U$ the patterns are again extremely similar between 95 and 150 GHz, and dominated by the ΛCDM $E$-mode signal. However, at 220 GHz we start to see dust contamination become comparable in amplitude to the ΛCDM $E$-modes at the edges of the field.

### Appendix B: Internal Consistency Tests of New Data

A powerful internal consistency test are map level data split difference tests which we refer to as "jackknives". As well as the full coadd signal maps we also form many pairs of split maps where the splits are chosen such that one might expect different systematic contamination in the two halves of the split. The split halves are differenced and the power spectra taken. We then take the normalized deviations of the real bandpowers from the mean of signal+noise simulations and form $\chi^2$ and $\chi$ (sum of deviations) statistics. In this section we perform tests of the new 2016, 2017 and 2018 data which are the same as those performed on the previous BICEP/*Keck* data in Sec. VII.C of Ref. [9], Sec. 6.3 of Ref. [14], Appendix B of BK14, and Appendix B of BK15. In this paper we are adding three new years of data and we choose to test each band/year separately. The fourteen data splits themselves remain as described previously and we again do each $\chi^2$ and $\chi$ test twice using the lowest five and lowest nine bandpowers ($\ell < 200$ and $\ell < 300$), and for each of the $EE$, $BB$ and $EB$ spectra.

The two nominally 220 GHz receivers added before the 2015 season actually have passbands centered at $\approx 230$ GHz, whereas the two added before the 2016 season have passbands centered at $\approx 220$ GHz. We made separate maps using these receivers and perform separate jackknife tests on these. (We have chosen to coadd these sub-bands in the main analysis to reduce the number of cross-spectra.) For 95 GHz we have 3 years $\times 14$ data-splits $\times 2$ statistics ($\chi^2$ or $\chi$) $\times 2$ bandpower-ranges $\times 3$ spectra ($EE/BB/EB$) $= 504$ tests, and for 220/230 GHz twice as many again for 1008 tests. The $\chi^2$ and $\chi$ jackknife statistics do not actually have the nominal theory distributions since they are summations of bandpower values which are themselves $\chi^2$ rather than Gaussian distributed. We deal with this by computing PTE values versus the simulations which should have the correct distributions. Since there are now more tests than simulation realizations (499) we expect a few zero and unity PTE values—there are 2 for BICEP3 and 3 for *Keck* 220/230 GHz. However, we note that in these cases the real data value is only just beyond the end of the simulated distribution, and when compared to the distribution of simulated values aggregated over all jackknifes the values are within the expected range. Figures 11 & 12 show histograms of the PTE values breaking down into the $\chi^2$ and $\chi$ tests over the two bandpower ranges. Problems with the noise debias and/or estimated fluctuation would be expected to result in an excess number of values in the edge bins.

In the initial pass of the jackknife analysis there were a few $\chi^2$ and $\chi$ values sitting sufficiently beyond the tails of the simulation distributions that it was unlikely to happen by chance. Through a long process we determined that removing two small sub-sets of data moved these values in such that there were no longer clear problems. Mostly data is not taken during the "station open" period from early November through mid-February, and when it has been it has typically not been included in science analysis. In 2016 data was taken with *Keck Array* for the full month of November and initially this was included in the analysis. However, we found evidence for failure of the season split jackknife when including this data which lead us to exclude it. In addition we found it necessary to remove BICEP3 data from focal plane tile 1 in order to pass the tile jackknife. Due to readout limitations the partner tile across the boresight was not populated and was instead blanked off with a reflective plate. The hypothesis is that this led to reflections and ghost beams for these detectors—see Ref. [15] for further details. We note that we did not look at the non-jackknife spectra until the above data removals had been done and the null tests determined to be satisfactory.

### Appendix C: 95 & 220 GHz Spectral Stability

We next test the mutual compatibility of the previously published BK15 spectra and the new BK18 spectra. In 2014 and 2015 *Keck Array* operated with two receivers at 95 GHz producing a relatively low sensitivity map. In 2016, 2017 and 2018 BICEP3 observed in the same frequency band over an overlapping but somewhat larger sky area. The upper part of Fig. 13 compares the two $BB$ spectra—the large improvement in sensitivity is clear. Since the sky regions overlap the bandpowers are correlated. To gauge consistency we compare the differences of the real bandpowers to the differences of simulations which share common input skies. The lower part of the figure shows that this (not very stringent) test is passed.

The BK15 220 GHz map came from two *Keck* receivers

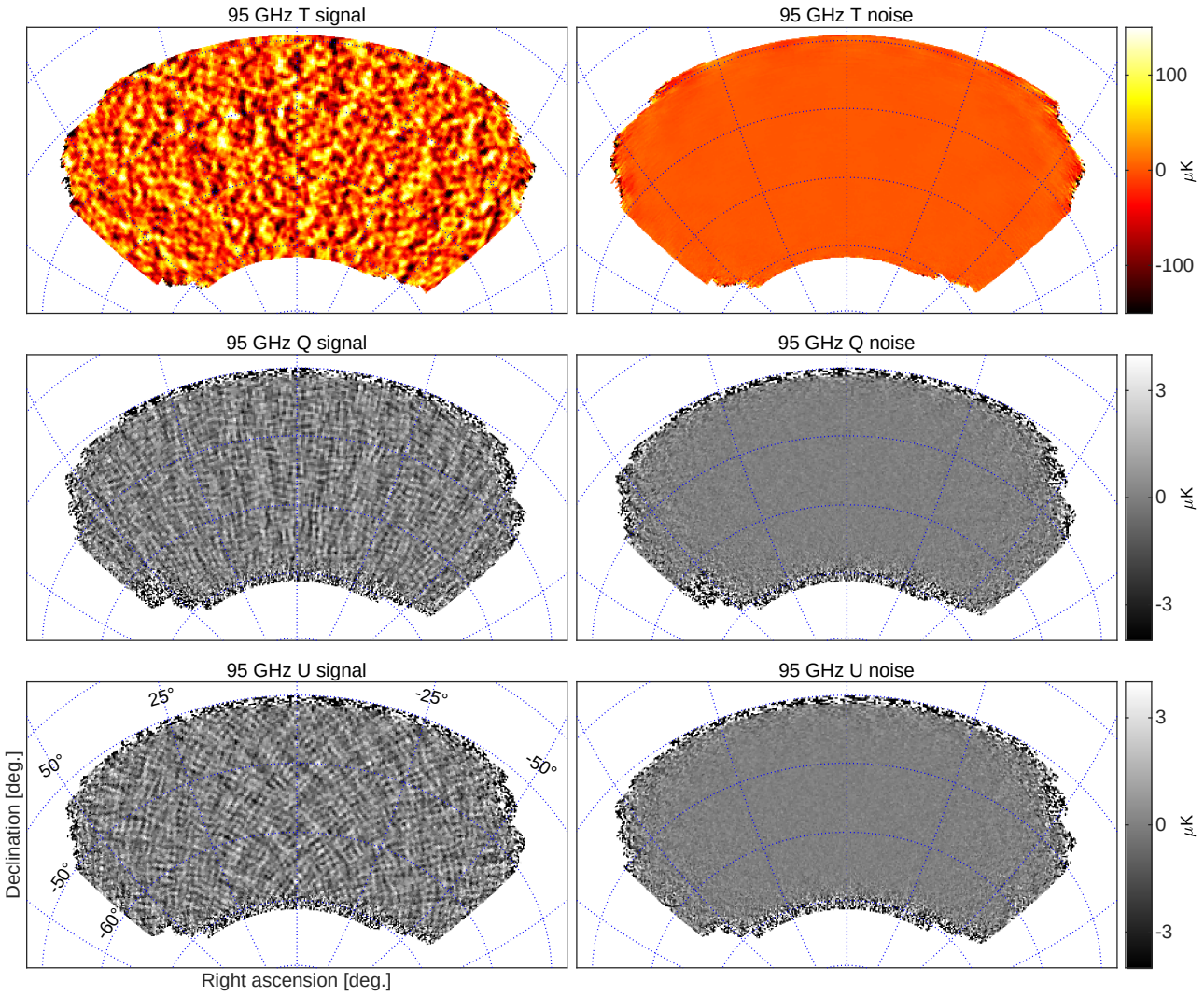


FIG. 8. $T$, $Q$, $U$ maps at 95 GHz using data taken by BICEP3 during the 2016–2018 seasons—we refer to these maps as BK18$_{B95}$. The left column shows the real data maps with 0.25° pixelization as output by the reduction pipeline. The right column shows a noise realization made by randomly assigning positive and negative signs while coadding the data. These maps are filtered by the instrument beam (24 arcmin FWHM [15]), timestream processing, and (for $Q$ & $U$) deprojection of beam systematics. Note that the horizontal/vertical and 45° structures seen in the $Q$ and $U$ signal maps are expected for an $E$-mode dominated sky.

running in 2015. Before the 2016 season we added two more receivers and ran all four through 2016, 2017 and 2018. The upper part of Fig. 14 compares the two $BB$ spectra—a large improvement in sensitivity is again clear, and the BK18 220 GHz spectrum is sample variance limited out to $\ell \sim 200$. The lower panel again shows no evidence for inconsistency between the two spectra.

### Appendix D: Multi Frequency Spectra

Since the BKP paper our analyses have relied on comparing a multi-component parametric model of lensed-ΛCDM+foregrounds+noise to the ensemble of auto- and cross-spectra between the BICEP/*Keck* and WMAP/*Planck* bands. Through BK15 the coverage pattern of the BICEP/*Keck* bands on the sky was very similar. The apodization mask chosen was the geometric mean of the inverse noise variance maps, and this was also applied to the WMAP/*Planck* bands. BICEP3 maps a larger area of sky as can be seen comparing Figures 8 and 9. In this paper we define the baseline BK18 analysis as being one where the BICEP3 maps are apodized by their "natural" larger inverse noise variance pattern, and the WMAP/*Planck* maps are also apodized with this pattern. Using different apodizations suppresses the ampli-

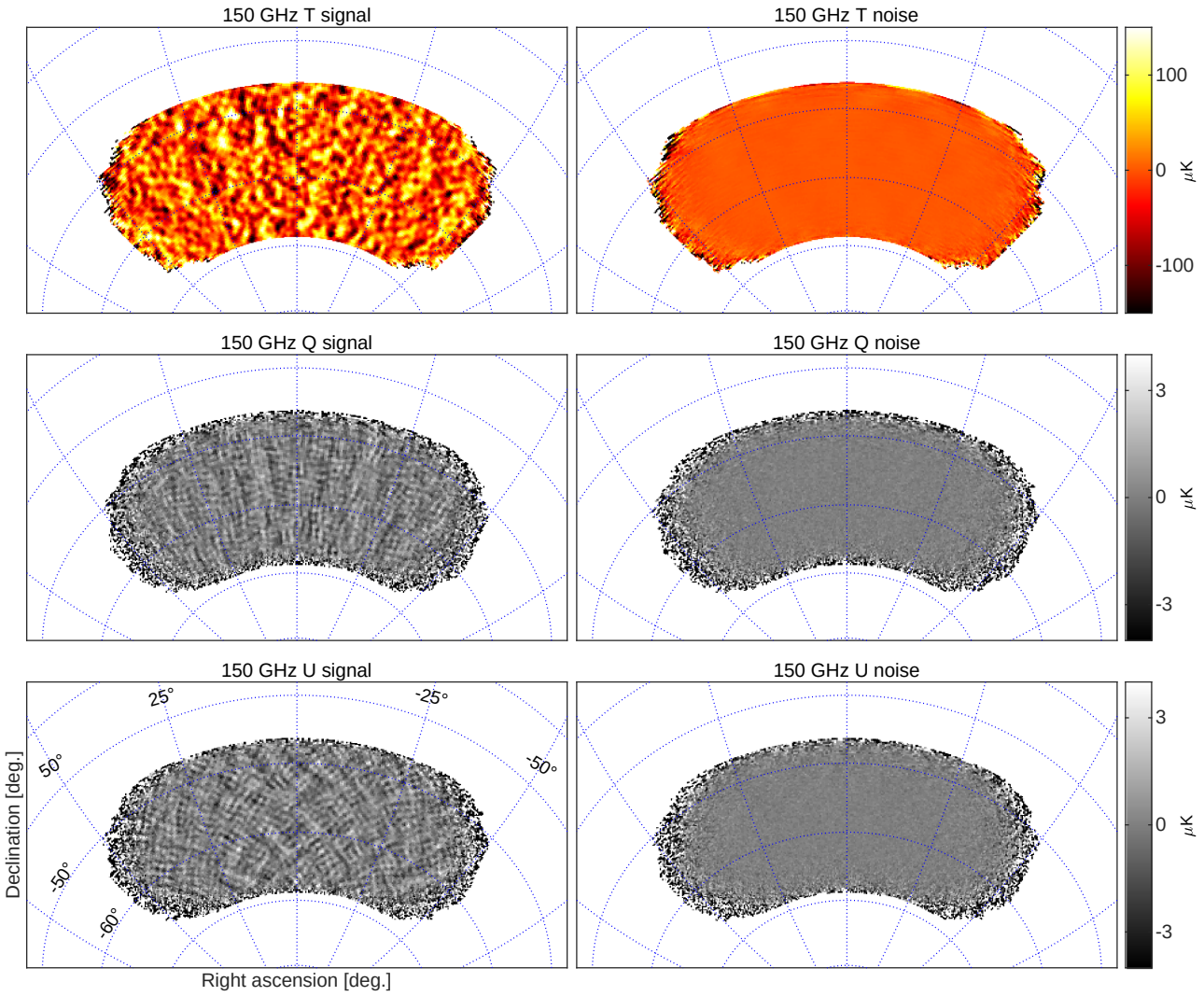


FIG. 9. $T$, $Q$, $U$ maps at 150 GHz using data taken by BICEP2 and *Keck Array* during the 2010–2016 seasons—we refer to these maps as $\mathrm{BK18}_{150}$. These maps are directly analogous to the 95 GHz maps shown in Fig. 8 except that the instrument beam filtering is in this case 30 arcmin FWHM [52]. The smaller coverage region is due to the smaller instantaneous field of view of the BICEP2/*Keck* receivers.

tude of the cross-spectra as compared to the auto-spectra but this is automatically taken into account in the bandpower window function calculations [18, 53]. Using different apodizations also in principle reduces the correlations of the auto- and cross-spectra which are key to the separation of foreground and CMB signals. However, in the current case the penalty turns out to be negligible. We consider some alternate apodizations in Appendix E 2 below.

Fig. 2 shows only a small subset of the spectra which are used in the likelihood analysis and included in the provided `COSMOMC` file. In this paper we use four BICEP/*Keck* maps, two WMAP maps (23 and 33 GHz), and five *Planck* maps (30, 44, 143, 217 and 353 GHz) resulting in 11 auto- and 55 cross-spectra. (To keep down the total number of spectra we drop the other *Planck* bands since these provide a negligible amount of additional information.) In Fig. 15 we show all of the included spectra together with the maximum likelihood model from the baseline analysis whose parameters were quoted above. Most of the spectra not already shown in Fig. 2 have low signal-to-noise, although a few of them carry interesting additional information on the possible level of synchrotron.

To quantify the absolute goodness-of-fit of the data to the maximum likelihood model we form statistics from the real data bandpowers and compare these to corresponding values from the simulation realizations. For the $9 \times 66 = 594$ bandpowers shown in Fig. 15, $\chi^2 = (d-m)^T C^{-1} (d-m) = 541.6$, where $d$ are the band-

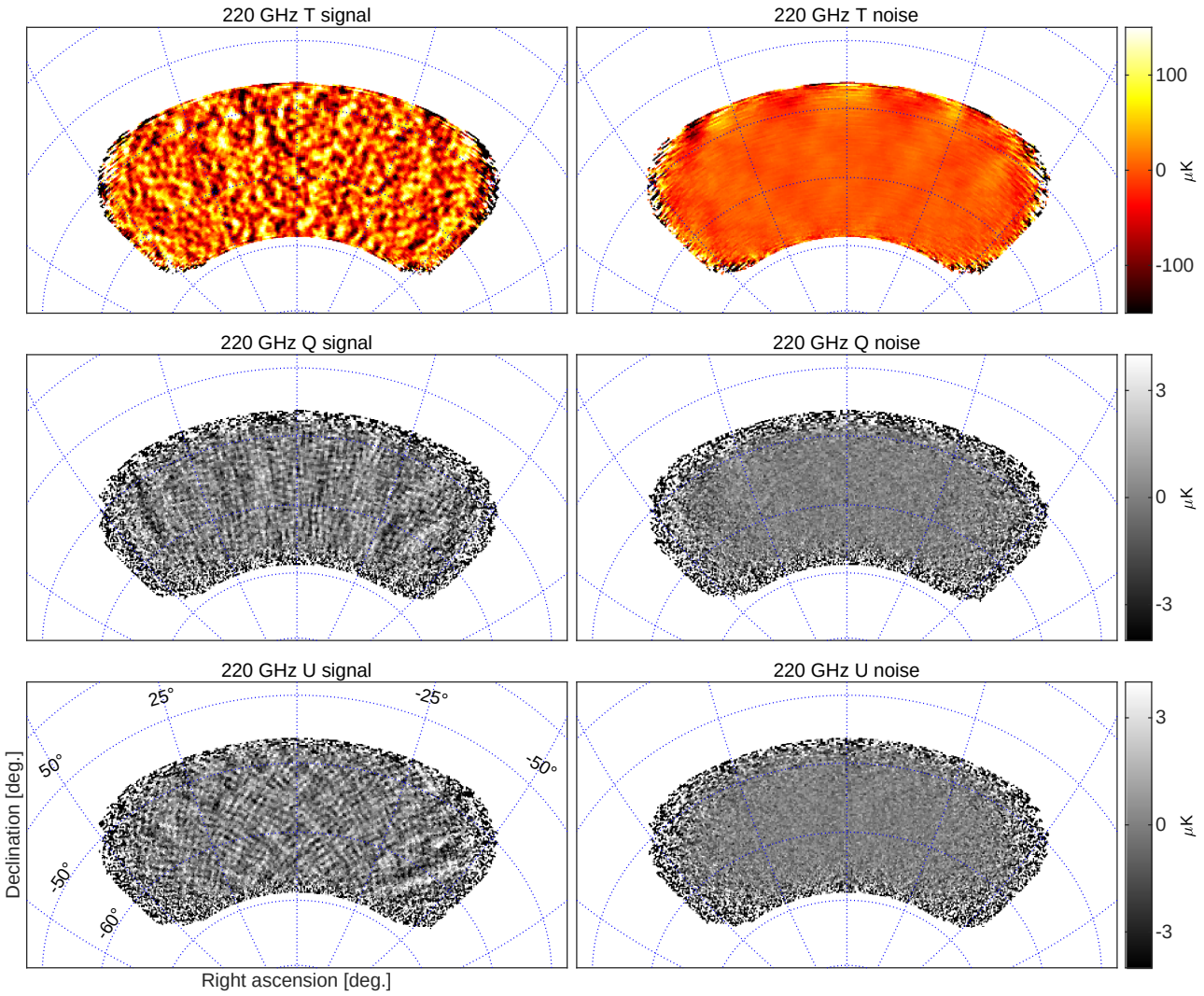


FIG. 10. $T$, $Q$, $U$ maps at $\approx 220\,\mathrm{GHz}$ using 14 receiver-years of data taken by *Keck Array* during the 2015–2018 seasons—we refer to these maps as $\mathrm{BK18}_{220}$. These maps are directly analogous to the 95 GHz maps shown in Fig. 8 except that the instrument beam filtering is in this case 20 arcmin FWHM [52]. Note the polarized dust emission which is visible in the right part of the $Q/U$ signal maps.

power values, $m$ are the model expectation values, and $C$ is the bandpower covariance matrix for the maximum likelihood model. This has a PTE versus the simulations of 0.94. If instead we take the sum of the normalized deviations $\chi = \sum((d-m)/e)$ where $e$ is the square-root of the diagonal of $C$, we find that the PTE versus the simulations is 0.22. The parametric model which we are using, including the approximation of Gaussian fluctuation of the dust (and synchrotron) sky patterns, remains an adequate description of the presently available data.

As in BK15 we also run a likelihood analysis to find the CMB and foreground contributions on a bandpower-by-bandpower basis. The baseline analysis is a single fit to all 9 bandpowers across 66 spectra with 8 parameters. Instead we now perform 9 separate fits—one for each bandpower—across the 66 spectra, with 6 parameters in each fit. These 6 parameters are the amplitudes of CMB, dust and synchrotron plus $\beta_\mathrm{d}$, $\beta_\mathrm{s}$ and $\epsilon$. While the Planck-derived $\beta_\mathrm{d}$ prior ($\beta_\mathrm{d} = 1.59 \pm 0.11$) is no longer necessary for the baseline analysis, it is needed for this bin-by-bin analysis which otherwise does not have enough information to constrain the dust spectrum in the higher $\ell$ bins. We repeat the analysis with foreground amplitudes defined at 150 GHz, where most of the BICEP2 and *Keck Array* sensitivity is concentrated, and at 95 GHz, the BICEP3 observing frequency. The results are shown in Fig. 16—the resulting CMB values are consistent with lensed-ΛCDM while the dust values are consistent with the level of dust found in the baseline analysis. Synchrotron is tightly limited in all the

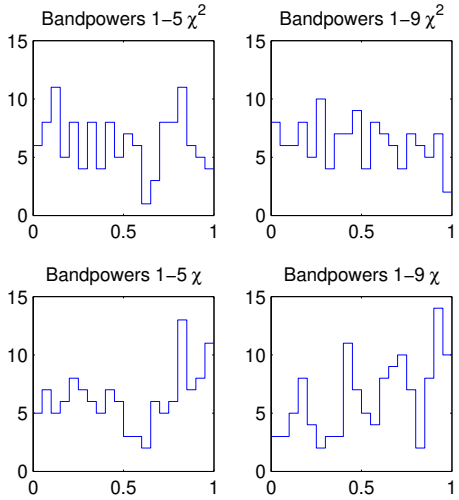


FIG. 11. Distributions of the jackknife $\chi^2$ and $\chi$ PTE values for the BICEP3 2016, 2017, & 2018 95 GHz data. This figure is analogous to Fig. 10 of BK15.

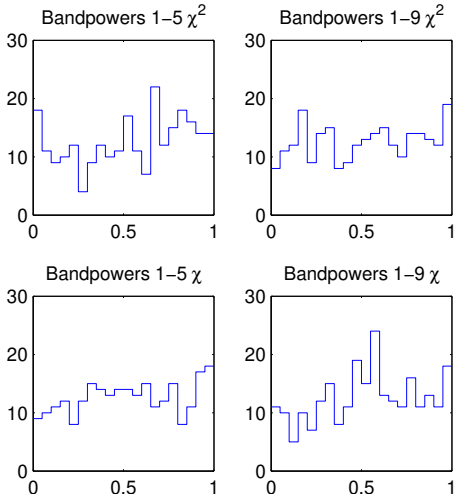


FIG. 12. Distributions of the jackknife $\chi^2$ and $\chi$ PTE values for the *Keck Array* 2016, 2017, & 2018 220 GHz data. This figure is analogous to Fig. 11 of BK15.

multipole ranges, and not detected in any of them. We note that this figure offers empirical backing for an assumption which we make in our mainline analysis: that the power spectrum of the polarized dust emission in our field follows a power law.

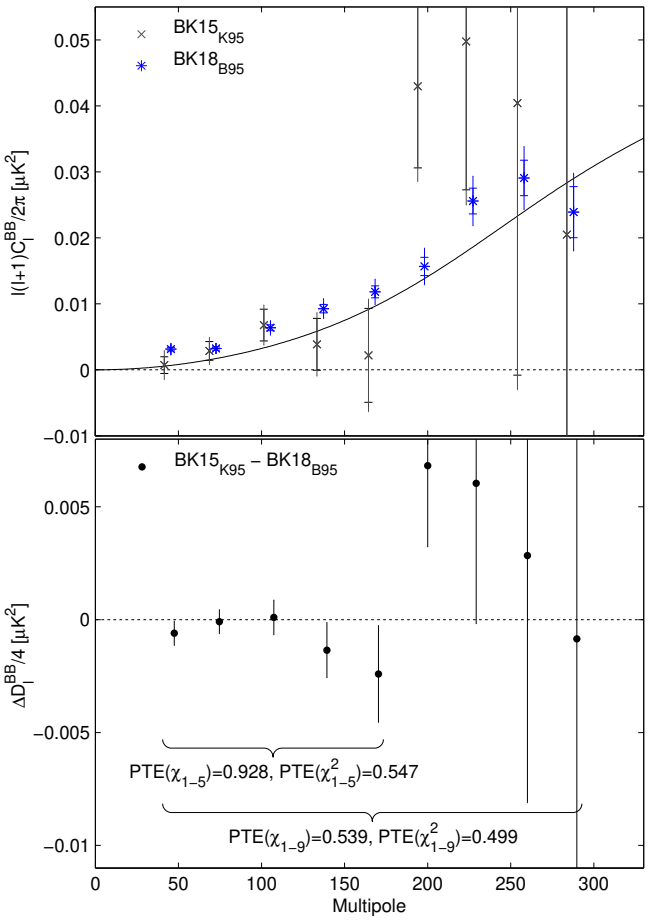


FIG. 13. *Upper:* Comparison of the previously published 95 GHz $BB$ auto-spectrum from the *Keck Array* receivers (BK15$_{K95}$) to the new BICEP3 spectrum (BK18$_{B95}$). The BICEP3 sky coverage region is a super-set of the *Keck* region. The inner error bars are the standard deviation of the noise realizations, while the outer error bars also include signal sample variance (lensed-ΛCDM+dust). Neither of these uncertainties are appropriate for comparison of the band power values—for this see the lower panel. The solid line shows the lensing $B$-mode expectation. (For clarity the sets of points are offset horizontally.) *Lower:* The difference of the spectra shown in the upper panel divided by a factor of four. The error bars are the standard deviation of the pairwise differences of signal+noise simulations which share common input skies (the simulations used to derive the outer error bars in the upper panel). Comparison of these points with null is an appropriate test of the compatibility of the spectra, and the PTE of $\chi$ and $\chi^2$ are shown. This figure is similar to Fig. 12 of BK15.

## Appendix E: Likelihood Variation and Validation

### 1. Likelihood Evolution

Fig. 17 shows the sequence of steps from the BK15 baseline analysis to the new baseline. Starting with the

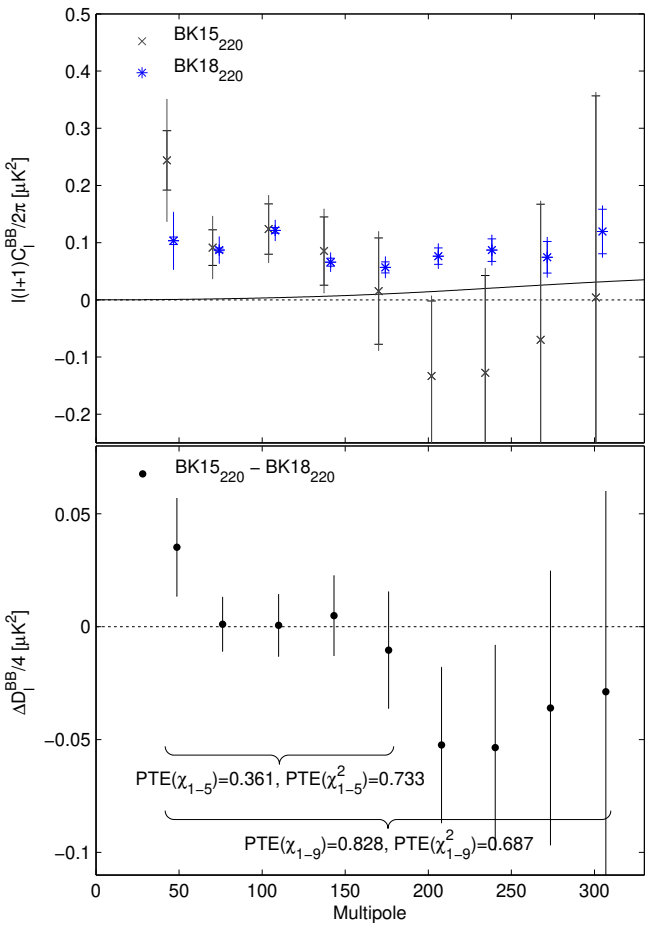


FIG. 14. Comparison of the previously published 220 GHz $BB$ auto-spectrum ($\mathrm{BK15}_{220}$) to the new BK18 version ($\mathrm{BK18}_{220}$). This figure is analogous to Fig. 13. In this case the sky coverage regions fully overlap but there is a small amount of common data between the two maps (2 receiver-years of the 14 receiver-years in BK18).

BK15 data set we first update the version of COSMOMC from Nov2016 to July2019 (green to magenta) which makes very little difference. We next switch from the BK15 data set to the BK18 one—i.e. including the improved sensitivity at 220 GHz and the improved sensitivity and sky area at 95 GHz (magenta to yellow). This makes a substantial difference: the $r$ constraint narrows and its peak position shifts down. In addition the synchrotron amplitude constraint $A_{\rm sync}$ shifts down to peak at zero.

The BKP, BK14 and BK15 analyses all placed a prior on the frequency spectral index of dust of $\beta_{\rm d} = 1.59 \pm 0.11$, using a Gaussian prior with the given $1\sigma$ width, this being an upper limit on the patch-to-patch variation [10, 54]. In BK15 the need for this prior was becoming weaker with the data able to constrain $\beta_{\rm d}$ to $1.65 \pm 0.20$. With the improved sensitivity of BK18 the constraint becomes $\beta_{\rm d} = 1.49^{+0.13}_{-0.12}$. In Fig. 17 dropping the $\beta_{\rm d}$ prior corresponds to the transition from yellow to red curves—$r$ and $A_{\rm sync}$ show very little change while $A_{\rm d}$ gets just slightly wider. Since there is little downside we choose to drop the $\beta_{\rm d}$ prior in the baseline analysis for BK18, gaining a little confidence that we are not inducing any bias due to the value of $\beta_{\rm d}$ in our particular sky patch differing from the central value of the prior. Interestingly the peak value of $\beta_{\rm d} = 1.50$ selected by our data is close to the mean value from *Planck* 2018 analysis of larger regions of sky $\beta_{\rm d}^P = 1.53 \pm 0.02$ [24].

The BKP, BK14 and BK15 likelihood analyses all assumed a lensing $B$-mode spectrum based on Planck 2013 cosmological parameters. The predicted lensing $BB$ spectrum for the Planck 2018 baseline cosmological parameters [2, Table 2] has shifted down by 5% (for $\ell < 500$). In Fig. 17 switching to the updated Planck 2018 spectrum corresponds to the transition from red to black—the $r$ constraint shifts slightly up indicating that this is a marginally relevant change at BK18 sensitivity levels. We define this analysis as the "BK18 baseline" and it is this that is shown in Fig. 4.

Since there is residual uncertainty in the *Planck* 2018 cosmological parameters there is also residual uncertainty in the prediction of the ΛCDM lensing $BB$ spectrum. In a simple study we find this to be 2.2% within the multipole range $20 < \ell < 500$ (with a distribution close to Gaussian). To marginalize over this uncertainty we run a joint analysis of the *Planck* 2018 TT,TE,EE+lowE+lensing likelihood and the BK18 likelihood producing the shift in Fig. 17 from heavy black to blue. By doing this we also include the small additional constraining power on $r$ which comes from the *Planck* TT,TE,EE+lowE+lensing likelihood. It is this joint likelihood which is shown in Fig. 5.

### 2. Likelihood Variation

*Dust Decorrelation.*—Spatial or line-of-sight variations of $\beta_{\rm d}$ will produce variations in the dust polarization pattern with frequency. This will result in suppression of the cross-spectra with respect to the geometric mean of the auto-spectra—a phenomenon referred to as decorrelation. The *Planck* 2018 dust analysis [24] did not show any evidence for dust decorrelation. Recently Ref. [55] reported evidence for line-of-sight decorrelation in *Planck* data by using HI data to select sight lines which integrate over multiple dust clouds. This study uses maps of the effective number of clouds taken from Ref. [56] which in Sec. 4.2 specifically examines the BICEP/*Keck* patch and states that "one cloud dominates the column density in the majority of pixels in the BICEP/*Keck* region". Nonetheless, decorrelation surely exists in our sky patch at some level—the question is whether it is relevant as compared to the BK18 experimental noise. In Appendix F of BK15 we added a decorrelation parameter $\Delta_{\rm d}$ to our parametric model, defined as the ratio of the cross-spectrum $217 \times 353$ to the geometric mean of the

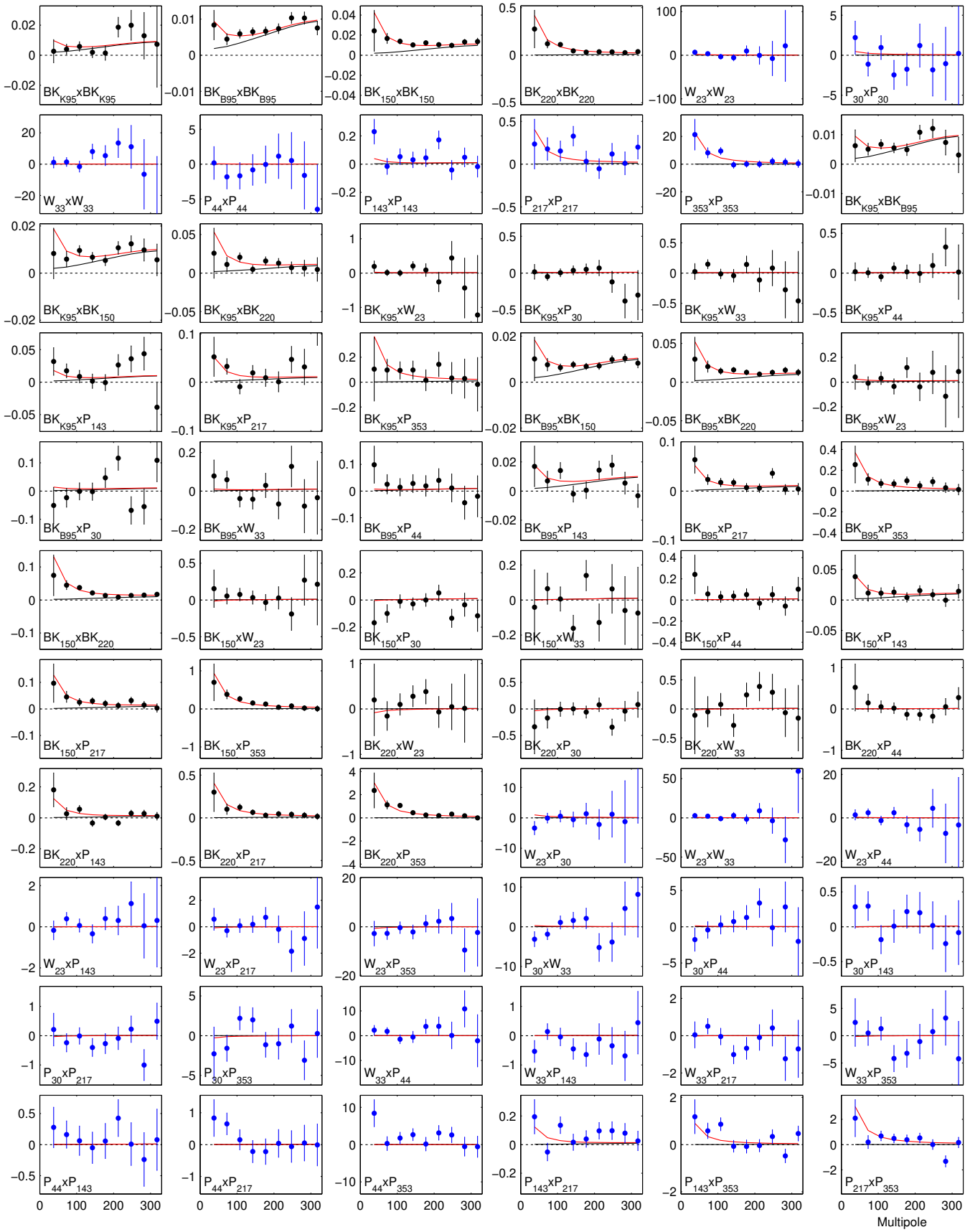


FIG. 15. The full set of $BB$ auto- and cross-spectra from which the joint model likelihood is derived. In all cases the quantity plotted is $100\ell C_\ell/2\pi$ ($\mu$K$^2$). Spectra involving BICEP/*Keck* data are shown as black points while those using only WMAP/*Planck* data are shown as blue points. The black lines show the expectation values for lensed-$\Lambda$CDM, while the red lines show the expectation values of the maximum likelihood lensed-$\Lambda$CDM+$r$+dust+synchrotron model ($r = 0.011$, $A_{\rm d,353} = 4.4\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.5$, $\alpha_{\rm d} = -0.66$, $A_{\rm sync,23} = 0.6\,\mu{\rm K}^2$, $\beta_{\rm s} = -3.0$, $\alpha_{\rm s} = 0.00$, $\epsilon = -0.11$), and the error bars are scaled to that model.

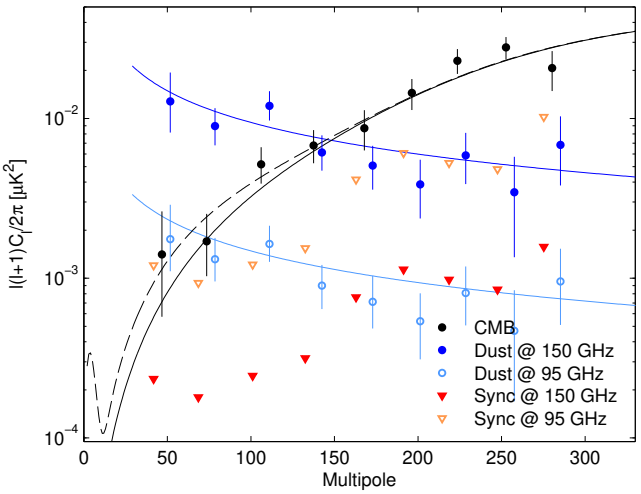


FIG. 16. Spectral decomposition of the $BB$ data into synchrotron (red), CMB (black) and dust (blue) components at 150 GHz (filled points) or 95 GHz (open points). The decomposition is calculated independently in each bandpower, marginalizing over $\beta_{\rm d}$, $\beta_{\rm s}$ and $\epsilon$, with priors. Error bars denote 68% credible intervals, with the point marking the most probable value. For synchrotron, which is not detected in this analysis, the downward triangles denote 95% upper limits. (For clarity the sets of points are offset horizontally.) The solid black line shows lensed-ΛCDM with the dashed line adding on top an $r_{0.05} = 0.011$ tensor contribution. The blue curves show the maximum-likelihood dust model from the baseline analysis ($A_{\rm d,353} = 4.4\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.5$, and $\alpha_{\rm d} = -0.66$) as evaluated at 150 GHz (dark blue) or 95 GHz (light blue).

corresponding auto-spectra, and varying with frequency in a manner suggested by Ref. [57]. Fig. 18 shows the result of allowing this parameter to vary freely (for the case where $\Delta_{\rm d}$ is assumed to scale as $\ell/80$). We see that the $r$ constraint shifts down slightly and the $A_{\rm sync}$ constraint shifts up slightly. In contrast to Fig. 21 of BK15 the $\Delta_{\rm d}$ curve peaks at unity (no decorrelation). Note that $\Delta_{\rm d}$ is a biasing parameter with respect to $r$—in BK15 it was found that allowing $\Delta_{\rm d}$ to vary freely resulted in 72% of the $r$ curves peaking at zero in simulations where there was no decorrelation. We therefore do not allow $\Delta_{\rm d}$ to vary freely in our baseline analysis at this time. (To avoid having to define dust decorrelation and dust/sync correlation simultaneously when we allow $\Delta_{\rm d}$ to vary freely we fix $\epsilon = 0$.)

*Free Lensing Amplitude.*—The *Planck* 2018 cosmological parameters predict a specific lensing $BB$ spectrum as used in the "BK18 baseline" analysis described above. If we introduce an artificial scaling parameter on this spectrum $A_{\rm L}^{\rm BB}$ and repeat the analysis we get the results shown in Fig. 19. With a uniform prior on $A_{\rm L}^{\rm BB}$ (and marginalizing over all other parameters) we obtain $A_{\rm L}^{\rm BB} = 1.03^{+0.08}_{-0.09}$ showing no tension with ΛCDM. The $r$ constraint hardly changes due to the fact that the lensing amplitude prefered by the BK18 data is very close to that predicted by the *Planck* 2018 cosmological parameters. This is an interesting result given the *Planck* 2018 TT,TE,EE+lowE result $A_{\rm L} = 1.18 \pm 0.065$ [2, Eqn. 36b]. In addition in the right panel of Fig. 19 we see that there is very little degeneracy between $r$ and $A_{\rm L}^{\rm BB}$ implying that all one now needs from ΛCDM is the template shape of the lensing $B$-mode spectrum, and not the amplitude.

*Sky Coverage.*—As seen in Figures 8 and 9 the BICEP3 sky coverage is larger than the BICEP2/*Keck* region. As already mentioned in Appendix D for the baseline BK18 analysis we apodize each map with its "natural" inverse variance coverage pattern and then take auto- and cross-spectra. The WMAP and *Planck* maps are apodized with the BICEP3 coverage pattern. Running maximum likelihood searches on simulations indicates that this results in $\sigma(r)$ as low as any other option with no detectable bias on $r$. Fig. 20 shows two alternatives: "small field only" apodizes all maps with the BICEP2/*Keck* coverage pattern and "BICEP3 field only" drops the BICEP2/*Keck* 95, 150 and 220 GHz maps and just uses BICEP3 at 95 GHz plus the WMAP/*Planck* bands. Interestingly the peak position of $A_{\rm d}$ only shifts a very small amount under these variations showing no evidence for variation of the polarized dust amplitude as averaged across the BICEP3 coverage pattern versus the subset BICEP2/*Keck* field. It is very interesting that dropping the 150 and 220 GHz bands does not substantially increase the width of the $r$ constraint, although it is unsurprising that it does increase the widths of the $A_{\rm d}$ and $\beta_{\rm d}$ constraints. Note that for the small field only version the $A_{\rm sync}$ peak shifts above zero as it was in BK15.

*Including $EE$.*—Fig. 20 also shows a case when including the $EE$ spectra (and hence also the $EB$ spectra) under the artificial assumption that the $EE/BB$ ratios for dust and synchrotron are both exactly 2, as is shown to be close to the case in Refs [24] and [30]. The effects are similar to those seen in BK15—the $r$ peak position shifts up, the $A_{\rm d}$ curve narrows, and the $A_{\rm sync}$ curve peaks strongly away from zero. It is unclear how much patch-to-patch variation we should in fact allow in the $EE/BB$ ratio so these results should not be overinterpreted at this time.

*External data variations.*—Fig. 21 shows variations in the WMAP and *Planck* data being included. With $\beta_{\rm d}$ now free, the *Planck* 353 GHz band provides a significant part of the $\beta_{\rm d}$ constraint (red curve wider than black in the rightmost panel). Dropping the other HFI bands makes smaller additional changes (red to green). It appears that there is some tension between the WMAP bands (23 and 33 GHz) and the LFI bands (30 and 44 GHz) in terms of their cross correlation with the BICEP/*Keck* maps. Dropping WMAP results in $A_{\rm sync}$ peaking strongly at zero whereas dropping LFI results in $A_{\rm sync}$ peaking well above zero. Unsurprisingly dropping both WMAP and LFI results in a highly degraded constraint on $A_{\rm sync}$ (cyan). BICEP/*Keck* alone (blue) constrains $A_{\rm sync}$ and $\beta_{\rm d}$ poorly with the $r$ constraint peaking hard at zero.

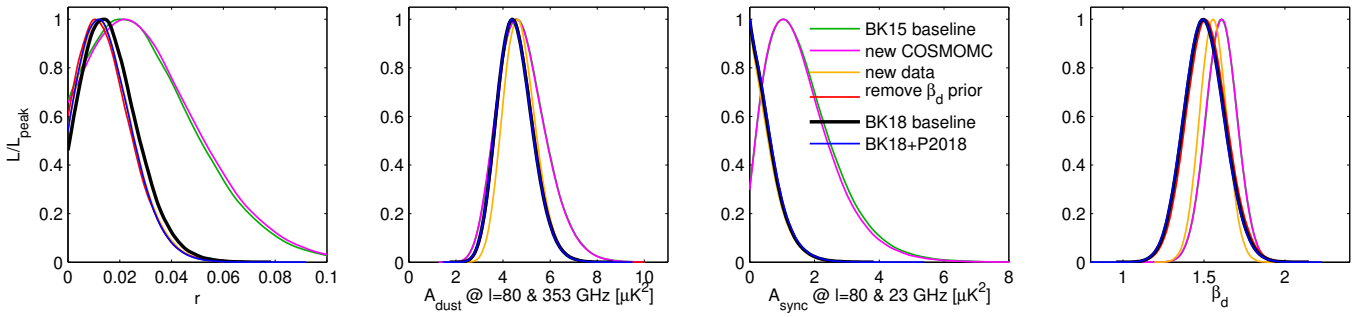


FIG. 17. Evolution of the BK15 analysis to the new BK18 "baseline" analysis as defined in this paper. See Appendix E 1 for further details.

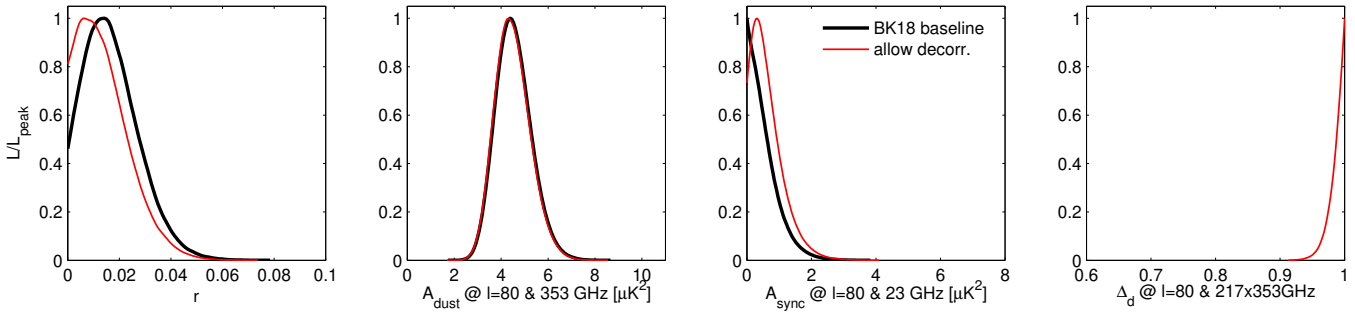


FIG. 18. Likelihood results when freeing dust decorrelation parameter $\Delta_{\rm d}$. See Appendix E 2 for details.

*Adding and removing fit parameters.*—In Appendix D above we saw that the absolute goodness-of-fit of the baseline model to the data is adequate. In addition we can ask whether the data shows evidence of the need for the baseline parameters, or preference for additional ones, by considering $\Delta(\chi^2)$. Performing maximum likelihood searches on the real data and taking $-2\log(L)$ as a proxy for $\chi^2$, Table I shows the shifts which occur. Allowing decorrelation parameter $\Delta_{\rm d}$ to vary freely while fixing $\epsilon = 0$ leaves the nominal number of parameters unchanged and as expected hardly changes the mean value of $-2\log(L)$ for the simulations. The reduction of 0.3 for the real data offers no significant evidence for detection of decorrelation.

Removing synchrotron from the fit (i.e. fixing $As = 0$, $\beta_{\rm s} = 0$, $\alpha_{\rm s} = 0$ and $\epsilon = 0$) corresponds to a nominal reduction in the number of fit parameters of 4, but the increase in the mean $-2\log(L)$ for the simulations (which contain no synchrotron) suggests only 2.5 effective free parameters. The increase of 1.5 for the real data is smaller than the average increase in the simulations indicating no evidence that synchrotron is present.

Fixing $r = 0$ results in an increase of 1.0 for the mean of simulations (which have $r = 0$). The increase for the real data is 0.6 indicating no significant evidence for $r > 0$. Removing synchrotron and $r$ we are down to the 3 parameter model ($A_{\rm d}$, $\beta_{\rm d}$, $\alpha_{\rm d}$) from which the simulations were generated. The mean of simulations responds exactly as the sum of the "−sync" and "$-r$" cases. The real data shows an increase almost exactly equal to that in the mean of simulations—i.e. the data is perfectly compatible with the lensed-ΛCDM expectation plus the simple 3 parameter dust model.

TABLE I. Values of $-2\log(L)$ (as a proxy for $\chi^2$) for the BK18 dataset as the model is varied. The last two rows are toy cases—see text for details.

| Model | mean sims | real data | Δ sims | Δ data | Δ nom. |
|---|---|---|---|---|---|
| baseline | 589.65 | 535.99 | | | |
| $+\Delta_{\rm d} - \epsilon$ | 589.73 | 535.69 | 0.08 | -0.29 | 0 |
| −sync | 592.19 | 537.52 | 2.54 | 1.52 | 4 |
| $-r$ | 590.71 | 536.58 | 1.06 | 0.59 | 1 |
| −sync$-r$ | 593.24 | 539.24 | 3.59 | 3.24 | 5 |
| $+A_L - \epsilon$ | 589.70 | 536.12 | 0.05 | 0.13 | 0 |
| $+A_L = 0.5$ | 626.44 | 566.70 | 36.79 | 30.71 | 1 |
| $+A_{\rm d} = 2\,\mu{\rm K}^2$ | 602.55 | 553.80 | 12.90 | 17.81 | 1 |

The $+A_L - \epsilon$ case shows no preference for varying the lensing scale factor away from unity for either the real data (as expected given the results in Fig. 19 above) or the simulations. The last two lines of the table are intended simply as test cases. If we fix the lensing spectrum rescale factor $A_{\rm L}^{\rm BB} = 0.5$ we see very large in-

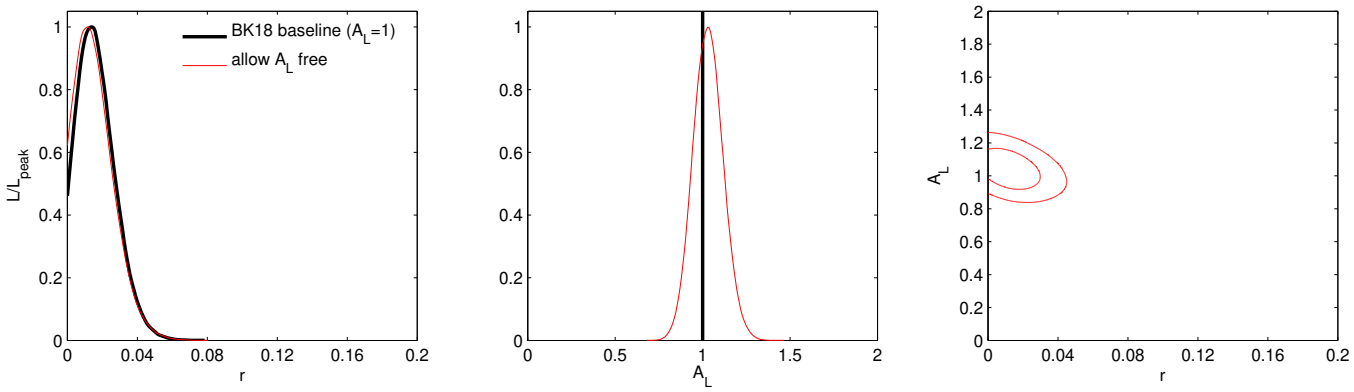


FIG. 19. Likelihood results when allowing the lensing amplitude to be a free parameter—see Appendix E 2 for details.

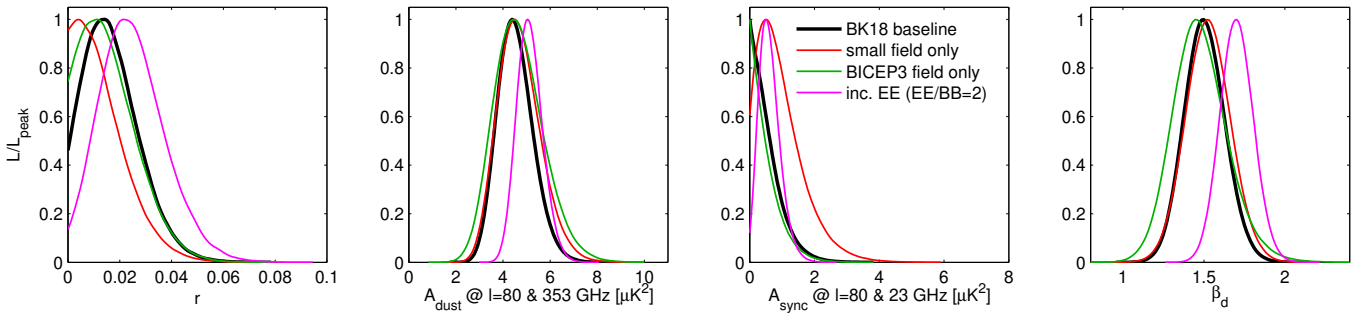


FIG. 20. Likelihood results when varying the internal data set selection. The result including $EE$ should not be overinterpreted. See Appendix E 2 for details.

creases in $-2\log(L)$ for simulations and data indicating strong incompatibility with such a model. Likewise setting $A_{\rm d,353} = 2\,\mu{\rm K}^2$ is strongly incompatible with the simulations and data.

### 3. Likelihood Validation

To validate the likelihood analysis on the real data we run full `COSMOMC` runs on the ensemble of lensed-ΛCDM+dust+noise simulations. Fig. 22 is an update of Fig. 19 of BK15. The left panel shows the $r$ constraint curves for the simulations, while the right panel compares the CDF of the zero-to-peak likelihood ratios to the simple analytic ansatz $\frac{1}{2}\left(1 - f\left(-2\log L_0/L_{\rm peak}\right)\right)$ where $f$ is the $\chi^2$ CDF (for one degree of freedom). We find that 60% of the simulations peak at zero, and 11% have a lower zero-to-peak ratio than the real data—i.e. show more evidence for $r$ when the true value is in fact zero.

An alternate (and much faster) likelihood validation exercise is to run maximum likelihood searches, with non-physical parameter values allowed (such as negative $r$). When running on simulations generated according to the model being re-fit, we then have an a priori expectation that the input parameter values should be recovered in the mean. Fig. 23 shows the results when running on the standard lensed-ΛCDM+dust+noise simulations. Only the four well constrained parameters are shown, the others being prior dominated. The input values are recovered in the mean although formally there is a small bias detectable on $A_{\rm d}$. We prefer this $\sigma(r)$ measure of the intrinsic constraining power of the experiment since it is independent of the particular noise fluctuation that is present in the real data.

### 4. Exploration of Alternate Foreground Models

We now extend the maximum likelihood validation study to simulations using alternate foreground models, updating Appendix E.4 of BK15, and giving results in Table II. Since the BK18 baseline analysis now has $\beta_{\rm d}$ free we provide results for this case only. Our basic set of simulations includes Gaussian realizations of dust generated according to the parametric model which we use when re-fitting, and these are thus expected and to give unbiased results—as we see they do in the first row of the table. We also make decorrelated simulations according to

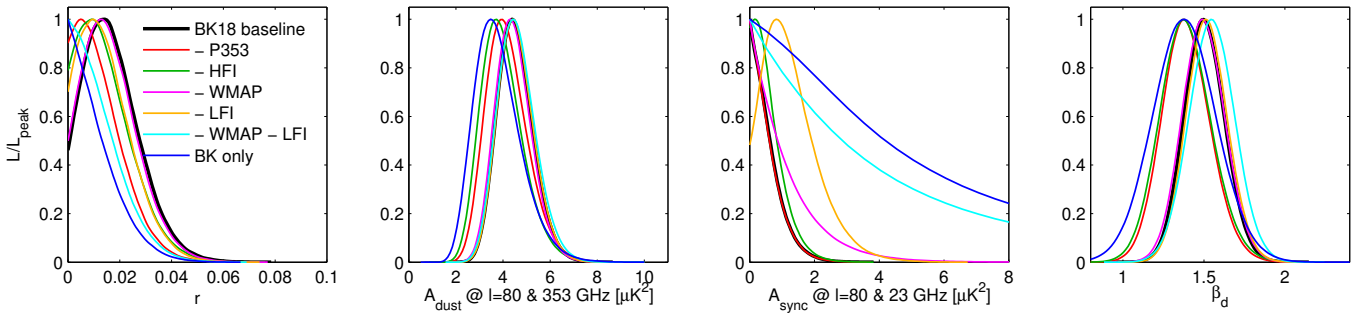


FIG. 21. Likelihood results when removing some or all of the external WMAP and *Planck* bands—see Appendix E 2 for details.

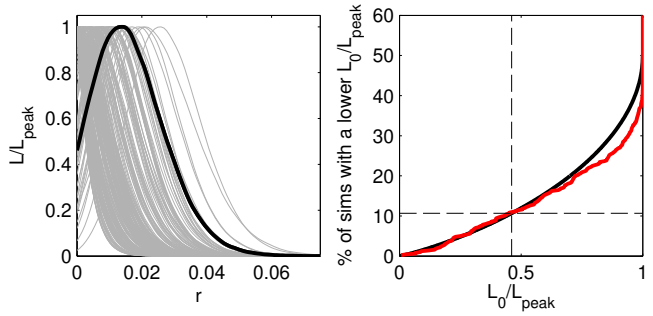


FIG. 22. *Left:* Likelihood curves for $r$ when running the baseline analysis on 200 of the lensed-ΛCDM+dust+noise simulations. The real data curve is shown overplotted in heavy black. *Right:* The CDF of the zero-to-peak ratio (red) of the curves shown at right as compared to the simple analytic ansatz (solid black) $\frac{1}{2}\left(1 - f\left(-2\log L_0/L_{\rm peak}\right)\right)$ where $f$ is the $\chi^2$ CDF (for one degree of freedom). About one tenth of the simulations offer more evidence for non-zero $r$ than the real data when the true value is actually zero (dashed black).

our parameterization of decorrelation, and these are thus also expected to give unbiased results when re-fit with that extension to the parameterization—as we see they do in the second line of the table. In addition we make so called "amplitude modulated Gaussian" dust simulations where Gaussian full sky realizations are multiplied by the square root of maps of degree scale $BB$ power measured from small patches of the *Planck* 353 GHz map in a similar manner to Fig. 8 of Ref. [23]. Such variation of the dust amplitude across the field can potentially produce bias on $r$ given that our re-fit model assumes a single $A_{\rm d}$ value applies to both the smaller BICEP2/*Keck* field and the larger BICEP3 field. However, in practice the third line of the table shows no detectable bias.

We also consider a suite of third party foreground models which are constructed in a variety of ways and which do not necessarily conform to any specific parametric model. Hence they may potentially produce bias in $r$ at levels relevant compared to the noise. The fourth and subsequent rows of Table II summarize the results.

TABLE II. Uncertainty and bias on $r$ in simulations using a variety of foreground models. For the strongly decorrelated model bias is expected when refit without a decorrelation parameter so this case is in parentheses.

| | $\overline{A_d}$ | $\overline{A_s}$ | $\sigma(r)$, $\overline{r}/\sigma(r)$ | |
|---|---|---|---|---|
| Model | ($\mu$K$^2$) | ($\mu$K$^2$) | no decorr. | with decorr. |
| Gaussian | 3.9 | 0.1 | 0.009, $0.0\sigma$ | 0.010, $0.0\sigma$ |
| G. Decorr. | 5.1 | 0.1 | (0.012, $+2.1\sigma$) | 0.014, $-0.1\sigma$ |
| G. amp. mod. | 4.4 | 0.0 | 0.009, $0.0\sigma$ | 0.010, $0.0\sigma$ |
| PySM 1 | 11.3 | 0.9 | 0.010, $+0.1\sigma$ | 0.012, $+0.2\sigma$ |
| PySM 2 | 25.6 | 0.8 | 0.011, $0.0\sigma$ | 0.012, $0.0\sigma$ |
| PySM 3 | 11.6 | 0.9 | 0.011, $0.0\sigma$ | 0.013, $-0.1\sigma$ |
| MHDv3 | 3.2 | 7.1 | 0.012, $-0.1\sigma$ | 0.013, $-0.4\sigma$ |
| MKD | 3.9 | 0.1 | 0.009, $0.1\sigma$ | 0.010, $0.0\sigma$ |
| Vansyngel | 5.5 | 0.1 | 0.009, $-0.1\sigma$ | 0.010, $0.0\sigma$ |

These third-party models provide only a single realization of the foreground sky, and we add it on top of each of the lensed-ΛCDM+noise realizations that are used in the standard simulations. The PySM models 1, 2 and 3 [58] are unchanged from BK15. They have $A_{\rm d}$ in our sky region which is much greater that the actual level and this modestly increases $\sigma(r)$, although interestingly by an amount that is fractionally less than the increase seen in BK15. The MHD model [59, 60] is also unchanged from BK15 and has a level of synchrotron which is now in strong conflict with the data.

The last two models in Table II are added for this paper. The MKD model [61] is a three-dimensional model of polarized galactic dust emission that takes into account the variation of the dust density, spectral index and temperature along the line of sight, and contains randomly generated small scale polarization fluctuations. This model is constrained to match observed dust emission on large scales, and match on smaller scales extrapolations of observed intensity and polarization power spectra. The Vansyngel model [62] is also multi-layer. Each layer has the same intensity (constrained by the Planck intensity map), but different magnetic field realizations. It produces $Q/U$ by integrating along the line of sight

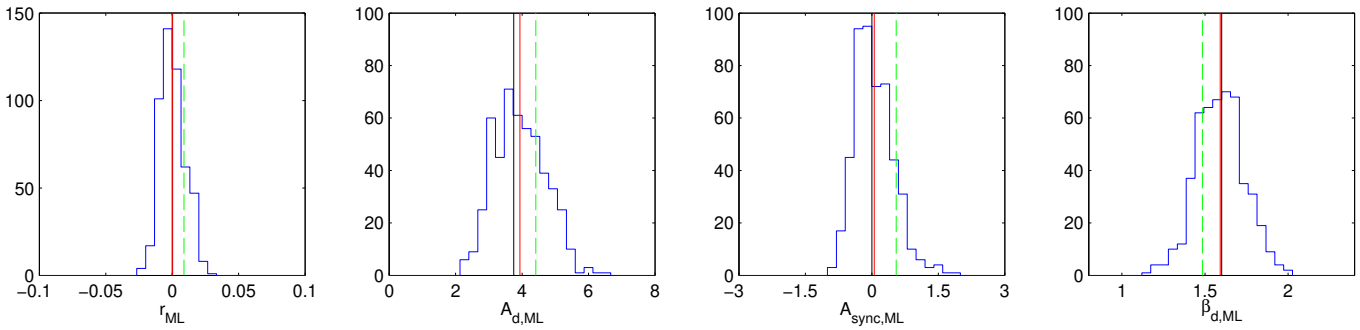


FIG. 23. Results of a validation test running maximum likelihood search on the standard lensed-ΛCDM+dust+noise simulations ($r = 0$, $A_{\rm d,353} = 3.75\,\mu{\rm K}^2$, $\beta_{\rm d} = 1.6$, $\alpha_{\rm d} = -0.4$, $A_{\rm sync} = 0$). The blue histograms are the recovered maximum likelihood values with the red lines marking their means and the black lines showing the input values. The green dashed lines are the real data values for the BK18 baseline data set. In the left panel $\sigma(r) = 0.009$. See Appendix E 3 for details.

over these multiple layers of magnetic fields. Neither of these models produces detectable bias on $r$.

## Appendix F: Systematics

Systematics effects in BICEP/*Keck* datasets have been studied in detail in Refs. [13, 34, 52, 63]. In this section, we assess the impact of the most important of these effects that could bias our current constraints on $r$. We estimate these biases by looking at the shift in maximum-likelihood parameters derived from simulations and/or the real data. For the case of bandpass uncertainty, we also calculate the full multi-dimensional likelihood of the real data for a model with additional nuisance parameters, and compare that to the BK18 baseline result. For all effects considered, the magnitude of the estimated bias on $r$ is significantly smaller than our statistical uncertainty, and not all of the biases have the same sign, which should lead to partial cancelation.

### 1. Temperature-to-Polarization Leakage from Residual Beam Mismatch

Following Ref. [28] systematic effects may be classified as additive, characterized by contamination in polarization maps not correctly estimated by the noise model, or non-additive, which includes most signal calibration errors. The most important additive effect for our datasets is temperature-to-polarization leakage resulting from residual beam mismatch, crosstalk, or similar effects. Following the methodology developed in Ref. [52], we evaluate the impact of $T \to P$ leakage from undeprojected main beam mismatch on $r$ recovery. We analyze the shift in maximum likelihood $r$ estimation for a set of 499 simulations which have an added bias corresponding to an estimate of $T \to P$ leakage. This leakage estimate comes from specialized "beam map simulations", as described in Sec. 5 of Ref. [52]. We report the median and 1-$\sigma$ standard deviation of the realization-to-realization recovered values of $r$.

Fig. 24 shows the beam map simulation $BB$ auto spectrum and cross with real for BICEP3. The corresponding spectra for the *Keck* 95 GHz, 150 GHz and 220 GHz bands were found to be very similar to the BK15 results shown in Ref. [52].

We report results corresponding to the "CMB data-driven" scenario presented in Ref. [52], for which leakage templates are added only on auto-frequency spectra, as the leakage templates show no evidence of a common-mode component that would bias the cross-frequency spectra. Specifically, we add a leakage contribution whose mean amplitude is set by the cross spectra of beam map simulations with real BK18 maps (blue circles in Fig. 24), plus a random component with the standard deviation of the cross spectra between the beam map simulations the standard lensed-ΛCDM+dust+noise simulations (error bars in Fig. 24).

The recovered bias is $\Delta(r) = 1.5 \pm 1.1 \times 10^{-3}$, which is subdominant compared to our statistical uncertainty $\sigma(r) = 0.009$. Compared to BK15 results, the reduction in $T \to P$ bias ($\Delta(r) = 2.7 \pm 1.9 \times 10^{-3} \to 1.5 \pm 1.1 \times 10^{-3}$) nearly matches the reduction in statistical uncertainty ($\sigma(r) = 0.020 \to 0.009$). The driving factor in this improvement in the recovered bias is the inclusion of BICEP3, which demonstrates lower differential beam power than *Keck* and BICEP2, both before and after deprojection. Note that this bias estimate naturally includes the effect of $T \to P$ leakage due to crosstalk arising from our time-domain multiplexed readout. Beam simulations with crosstalk deprojected (as described in Ref. [34]) show that the contribution to $\Delta(r)$ from crosstalk is small.

This analysis highlights the importance of taking high-fidelity far-field beam map (FFBM) measurements *in situ*, which are essential in validating the removal of leading order difference modes via deprojection, and in ensuring that the bias on $r$ due to undeprojected residual leakage is properly controlled. In this leakage analysis

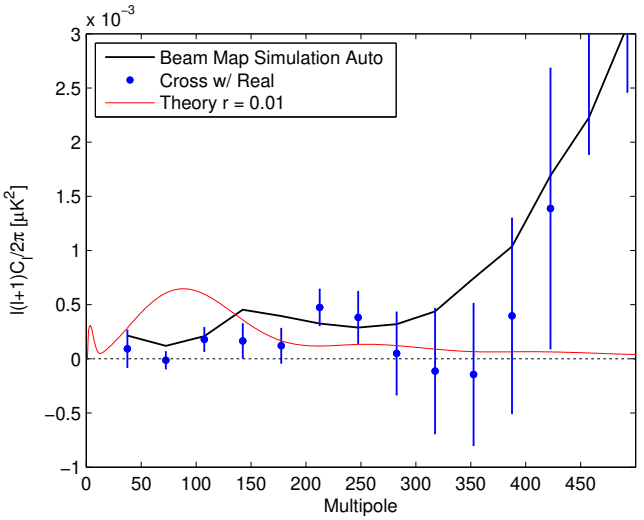


FIG. 24. $BB$ power spectra corresponding to $T \rightarrow P$ leakage in BICEP3, as predicted by beam map simulations. The black line shows the auto-spectrum, which has been noise debiased (from noise in the beam map measurement). The blue circles show the cross of the beam map simulation with real BK18 data, with error bars derived from the cross between the fixed beam map simulation with 499 lensed-ΛCDM+dust+noise simulations.

for BK18, care was taken to minimize systematic contamination in the composite beam maps from the FFBM analysis pipeline itself, e.g. by minimizing non-Gaussian noise and excess smoothing. Further improvements can be made in the future, by developing improved noise estimates that accurately capture per-map pixel uncertainty in the beam maps, especially where the coverage from the redirecting mirror varies. Higher S/N beam maps out to larger radii can also be constructed by combining FFBM measurements made using the thermal chopper with those made using a high-powered noise source [63]. This allows a quantification of leakage due to mismatch not only of the main beam, but of any region of the beam within $< 20°$ of the main beam. Improvements such as these will ensure that $T \rightarrow P$ leakage does not become a limiting effect as $\sigma(r)$ further declines with the deployment of the new BICEP Array receivers.

### 2. Other Additive Systematics Effects

Other additive systematic effects including EMI, magnetic, thermal, and other scan-synchronous sources of residual contamination, and ghost beams, were studied in Ref. [34] and upper limits placed on their contributions equivalent to $r < 10^{-4}$, except in the case of EMI contamination from the satellite uplink transmitter, which was bounded at $r < 1.7\times10^{-3}$ with 95% confidence. Levels of these potential sources of contamination have been monitored throughout collection of the current dataset (including EMI reduction from additional shielding at the satellite uplink), and contributions from these effects remain negligible. More significant, but still subdominant for the current analysis, are uncertainties from the contributions of polarized flux from point sources in our field, which estimates suggest may yield an upward bias on $r$ of 1–3$\times10^{-3}$ [9, 64, and recent internal work], and imperfections in our noise bias estimates as they interact with leakage deprojection filtering, which recent preliminary work suggests may yield a downward bias on $r$ of $2 \times 10^{-3}$. While small compared to our present uncertainties, we expect in future rounds of analysis to be able to mitigate both these effects.

### 3. Bandpass Uncertainty

Given the importance of multi-frequency component separation to the current result, the most important non-additive systematic effect is spectral bandpass uncertainty. Bandpass parameters (center and width) for the BICEP3 (95 GHz) and *Keck* (95, 150 and 220 GHz) receivers are determined using a Fourier-Transform Spectrometer (FTS). We consider various sources of uncertainty in these measurements and estimate them to be small, at the 1% level or less. As a conservative upper limit, we estimate the impact of errors up to 2% for the band centers of the BICEP3 and *Keck* data sets.

First, we simulate all possible combinations of ± 2% bandcenter shifts on the BICEP3 and *Keck* bands. We use maximum likelihood searches on 499 lensed-ΛCDM+dust+noise simulations to derive upper bounds for biases on recovered cosmological parameters. We show that, for all 8 parameters that we fit for in our maximum likelihood framework, the expected biases are low and within statistical uncertainties. In particular, we show that the worst case scenario would result in a bias on $r$ of $|\Delta(r)| = 8.4 \pm 5 \times 10^{-4}$, well below the statistical uncertainty $\sigma(r) = 0.009$

Secondly, we include nuisance parameters on BICEP3 and *Keck* bands in the CosmoMC likelihood to marginalize over errors in the bandpass measurement. Similar to BK15, we consider one nuisance parameter per frequency band that represents a fractional shift in the band center, and use a Gaussian prior with mean/standard deviation of (0/0.02) on each of these parameters. Compared to the BK18 baseline, the shift in the 95% upper limit on $r$ due to the addition of these nuisance parameters is $3 \times 10^{-5}$, negligible compared to the statistical uncertainty. The shifts on dust parameters ($A_d$ and $\beta_d$) are a few percent, and still within statistical uncertainty and upper bounds predicted using maximum likelihood searches on simulations. Additionally, we compare the shifts seen from the real data to those obtained from 21 simulations on which we run the CosmoMC likelihood analysis, with and without bandpass nuisance parameters. We compute the realization-to-realization shift in the likelihood peak and found none of the shifts observed in real data for $r$ and dust parameters to be statistically significant compared to random fluctuations in simulations.