A Joint Analysis of BICEP2/Keck Array and Planck Data


1School of Physics and Astronomy, Cardiff University, Queens Buildings, The Parade, Cardiff, CF24 3AA, U.K.
2Institut d’Astrophysique Spatiale, CNRS (UMR8617) Université Paris 11, Bâtiment 121, Orsay, France
3Department of Physics, Stanford University, Stanford, California 94305, U.S.A.
4California Institute of Technology, Pasadena, California, U.S.A.
5Harvard-Smithsonian Center for Astrophysics, 60 Garden Street MS 42, Cambridge, Massachusetts 02138, U.S.A.
6Laboratoire AIM, IRFU/Service d’Astrophysique - CEA/DSM - CNRS - Université Paris Diderot, Bât. 709, CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France
7SISSA, Astrophysics Sector, via Bonomea 265, 34136, Trieste, Italy
8Université de Toulouse, UPS-OMP, IRAP, F-31028 Toulouse cedex 4, France
I. INTRODUCTION

The cosmic microwave background (CMB) [1], is an essential source of information about all epochs of the Universe. In the past several decades, characterization of the temperature and polarization anisotropies of the CMB has helped to establish the standard cosmological model (ΛCDM) and to measure its parameters to high precision (see for example Refs. [2, 3]).

An extension to the standard big bang model, inflation, postulates a short period of exponential expansion in the very early Universe, naturally setting the initial conditions required by ΛCDM, as well as solving a number of additional problems in standard cosmology. Inflation’s basic predictions regarding the Universe’s large-scale geometry and structure have been borne out by cosmological measurements to date (see Ref. [4] for a review). Inflation makes an additional prediction, the existence of a background of gravitational waves, or tensor mode perturbations [5–8]. At the recombination epoch, the inflationary gravitational waves (IGW) contribute to the anisotropy of the CMB in both total intensity and linear polarization. The amplitude of tensors is conventionally parameterized by $r$, the tensor-to-scalar ratio at a fiducial scale. Theoretical predictions of the value of $r$ cover a very wide range. Conversely, a measurement of $r$ can discriminate between models of inflation.

Tensor modes produce a small increment in the temperature anisotropy power spectrum over the standard ΛCDM scalar perturbations at multipoles $\ell \lesssim 60$; measuring this increment requires the large sky coverage traditionally achieved by space-based experiments, and an understanding of the other cosmological parameters. The effects of tensor perturbations on $B$-mode polarization is less ambiguous than on temperature or $E$-mode polarization over the range $\ell \lesssim 150$. The $B$-mode polarization signal produced by scalar perturbations is very small and is dominated by the weak lensing of $E$-mode polarization on small angular scales, making the detection of an IGW contribution possible [9–12].

Planck [13] was the third generation CMB space mission, which mapped the full sky in polarization in seven bands centered at frequencies from 30 GHz to 353 GHz to a resolution of 33 to 5 arcminutes [14, 15]. The Planck collaboration has published the best limit to date on tensor modes using CMB data alone [3]: $r_{0.05} < 0.12$ at 95% confidence using a combination of Planck, SPT and ACT temperature data, plus WMAP polarization, although the Planck $r$ limit is model-dependent, with running of the scalar spectral index or additional relativistic degrees of freedom being well-known degeneracies which allow larger values of $r$.

Interstellar dust grains produce thermal emission, the brightness of which increases rapidly from the 100–150 GHz frequencies favored for CMB observations, becoming dominant at $\geq 350$ GHz even at high galactic latitude. The dust grains align with the Galactic magnetic field to produce emission with a degree of linear polarization [16]. The observed degree of polarization depends on the structure of the Galactic magnetic field along the line
of sight, as well as the properties of the dust grains (see for example Refs. [17, 18]). This polarized dust emission results in both E-mode and B-mode, and acts as a potential contaminant to a measurement of \( r \). Galactic dust polarization was detected by Archeops [19] at 353 GHz and by WMAP [2, 20] at 90 GHz.

BICEP2 was a specialized, low angular resolution experiment, which operated from the South Pole from 2010 to 2012, concentrating 150 GHz sensitivity comparable to Planck on a roughly 1% patch of sky at high Galactic latitude [21]. The BICEP2 Collaboration published a highly significant detection of B-mode polarization in excess of the \( r=0 \) lensed-\( \Lambda \)CDM expectation over the range \( 30 < \ell < 150 \) in Ref. [22, hereafter BK-I]. Modest evidence against a thermal Galactic dust component dominating the observed signal was presented based on the cross-spectrum against 100 GHz maps from the previous BICEP1 experiment. The detected B-mode level was higher than that projected by several existing dust models [23, 24] although these did not claim any high degree of reliability.

The Planck survey released information on the structure of the dust polarization sky at intermediate latitudes [25], and the frequency dependence of the polarized dust emission at frequencies relevant to CMB studies [26]. Other papers argued that the BICEP2 region is significantly contaminated by dust [27, 28]. Finally Planck released information on dust polarization at high latitude [29, hereafter PIP-XXX], and in particular examined a field centered on the BICEP2 region (but somewhat larger than it) finding a level of polarized dust emission at 353 GHz sufficient to explain the 150 GHz excess observed by BICEP2, although with relatively low signal-to-noise.

Keck Array is a system of BICEP2-like receivers also located at the South Pole. During the 2012 and 2013 seasons Keck Array observed the same field as BICEP2 in the same 150 GHz frequency band [30, hereafter BK-V]. Combining the BICEP2 and Keck Array maps yields \( Q \) and \( U \) maps with rms noise of 57 kN in nominal 1 deg\(^2\) pixels—by far the deepest made to date.

In this paper, we take cross-spectra between the joint BICEP2/Keck maps and all the polarized bands of Planck. The structure is as follows. In Sec. II we describe the preparation of the input maps, the expectations for dust, and the power spectrum results. In Sec. III the main multi-frequency cross-spectrum likelihood method is introduced and applied to the data, and a number of variations from the selected fiducial analysis are explored. Sec. IV describes validation tests using simulations as well as an alternate likelihood. In Sec. V we investigate whether there could be decorrelation between the Planck and BICEP2/Keck maps due to the astrophysics of dust and/or instrumental effects. Finally we conclude in Sec. VI.

II. MAPS TO POWER SPECTRA

A. Maps and preparation

We primarily use the BICEP2/Keck combined maps, as described in BK-V. We also use the BICEP2-only and Keck-only maps as a cross check. The Planck maps used for cross-correlation with BICEP2/Keck are the full-mission polarized maps from the PR2 Planck science release [31, 32], a subset of which was presented in PIP-XXX. We compute Planck single-frequency spectra by taking crosses between data-split maps; we consider detector-set maps, yearly maps, and half-ring maps [33]. To evaluate uncertainties due to Planck instrumental noise, we use 500 noise simulations of each map; these are the standard set of time-ordered data noise simulations projected into sky maps (the FFP8 simulations defined in Ref. [34]).

While the Planck maps are filtered only by the instrument beam (the effective beam defined in Refs. [35] and [36]), the BICEP2/Keck maps are in addition filtered due to the observation strategy and analysis process. In particular, large angular scales are suppressed anisotropically in the BICEP2/Keck mapmaking process to avoid atmospheric and ground-fixed contamination; this suppression is corrected in the power spectrum estimate. In order to facilitate comparison, we therefore prepare “Planck as seen by BICEP2/Keck” maps. In the first step we use the ANAFAST, ALTERALM and SYNFAST routines from the HEALPIX [37] package [38] to resmooth the Planck maps with the BICEP2/Keck beam profile, assuming azimuthal symmetry of the beam. The coordinate rotation from Galactic to celestial coordinates of the T, Q, and U maps is performed using the ALTERALM routine in the HEALPIX package. The sign of the Stokes \( U \) map is flipped to convert from the HEALPIX to the IAU polarization convention. Next we pass these through the “observing” matrix \( R \) described in Section VI.B of BK-I, to produce maps that include the filtering of modes occurring in the data processing pipeline (including polynomial filtering and scan-synchronous template removal, plus deprojection of beam systematics).

Figure 1 shows the resmoothed Planck 353 GHz T, Q, and U maps before and after filtering. In both cases the BICEP2/Keck inverse variance apodization mask has been applied. This figure emphasizes the need to account for the filtering before any comparison of maps is attempted, either qualitative or quantitative.

B. Expected spatial and frequency spectra of dust

Before examining the power spectra it is useful to review expectations for the spatial and frequency spectra of dust. Figure 2 of PIP-XXX shows that the dust BB (and EE) angular power spectra are well fit by a simple power law \( D_\ell \propto \ell^{-0.42} \), where \( D_\ell = C_\ell \ell (\ell + 1) / 2\pi \), when averaging over large regions of sky. Section 5.2 of
the same paper states that there is no evidence for departure from this behavior for 1% sky patches, although the signal-to-noise ratio is low for some regions. Presumably we expect greater fluctuation from the mean behavior than would be expected for a Gaussian random field.

The spectral energy distribution (SED) of dust polarization was measured in Ref. [26] for 400 patches with 10° radius at intermediate Galactic latitudes. The SED is well fit by a modified blackbody spectrum with $T_d = 19.6$ K and $\beta_d = 1.59 \pm 0.17$, where the dust temperature is obtained from a fit to the SED of total intensity, and the uncertainty on the spectral index represents the 1σ dispersion of the individual patch measurements. The uncertainty is an upper limit, since some fluctuation is due to noise rather than real variation on the sky. This SED is confirmed to be a good match to data when averaging over 24% of the cleanest high latitude sky in Fig. 6 of PIP-XXX.

C. Power spectrum estimation and results

The power spectrum estimation proceeds exactly as in BK-I, including the matrix based purification operation to prevent $E$ to $B$ mixing. Figure 2 shows the results for BICEP2/Keck and Planck 353 GHz for $TT$, $TE$, $EE$, and $BB$. In all cases the error bars are the standard deviations of lensed-$\Lambda$CDM+noise simulations [39] and hence contain no sample variance on any other component. The results in the left column are auto-spectra, identical to those given in BK-I and BK-V—these spectra are consistent with lensed-$\Lambda$CDM+noise except for the excess in $BB$ for $\ell < 200$.

The right column of Fig. 2 shows cross-spectra between two halves of the Planck 353 GHz data set, with three different splits shown. The Planck collaboration prefers the use of cross-spectra even at a single frequency to gain additional immunity to systematics and to avoid the need to noise debias auto-spectra. The $TT$ spectrum is higher than $\Lambda$CDM around $\ell = 200$—presumably due to a dust
FIG. 2. Single- and cross-frequency spectra between BICEP2/Keck maps at 150 GHz and Planck maps at 353 GHz. The left column shows single-frequency spectra of the BICEP2, Keck Array and combined BICEP2/Keck maps. The BICEP2 spectra are identical to those in BK-I, while the Keck Array and combined are as given in BK-V. The center column shows cross-frequency spectra between BICEP2/Keck maps and Planck 353 GHz maps. The right column shows Planck 353 GHz data-split cross-spectra. In all cases the error bars are the standard deviations of lensed-ΛCDM+noise simulations and hence contain no sample variance on any other component. For EE and BB the χ² and χ (sum of deviations) versus lensed-ΛCDM for the nine bandpowers shown is marked at upper/lower left (for the combined BICEP2/Keck points and DS1×DS2). In the bottom row (for BB) the center and right panels have a scaling applied such that signal from dust with the fiducial frequency spectrum would produce signal with the same apparent amplitude as in the 150 GHz panel on the left (as indicated by the right-side y-axes). We see from the significant excess apparent in the bottom center panel that a substantial amount of the signal detected at 150 GHz by BICEP2 and Keck Array indeed appears to be due to dust.

contribution. The EE and BB spectra are noisy, but both appear to show an excess over ΛCDM for ℓ < 150—again presumably due to dust. We note that these spectra do not appear to follow the power-law expectation mentioned in Sec. II B, but we emphasize that the error bars contain no sample variance on any dust component (Gaussian or otherwise).

The center column of Fig. 2 shows cross-spectra between BICEP2/Keck and Planck maps. For TE one can use the T-modes from BICEP2 and the E-modes from Planck or vice versa and both options are shown. Since the T-modes are very similar between the two experiments, these TE spectra look similar to the single-experiment TE spectrum which shares the E-modes. The EE and BB cross-spectra are the most interesting—there appears to be a highly significant detection of correlated B-mode power between 150 and 353 GHz, with the pattern being much brighter at 353, consistent with the expectation from dust. We also see hints of detection in the EE spectrum—while dust E-modes are subdominant to the cosmological signal at 150 GHz, the weak dust contribution enhances the BK150×P353 cross-spectrum.
at $\ell \approx 100$.

The polarized dust SED model mentioned in Sec. II B implies that dust emission is approximately 25 times brighter in the Planck 353 GHz band than it is in the BICEP2/Keck 150 GHz band (integrating appropriately over the instrumental bandpasses). The expectation for a dust-dominated spectrum is thus that the BK150×P353 cross-spectrum should have an amplitude 25 times that of BK150×BK150, and P353×P353 should be 25 times higher again. The $y$-axis scaling in the bottom row of Fig. 2 has been adjusted so that a dust signal obeying this rule will have equal apparent amplitude in each panel. We see that a substantial amount of the BK150 signal indeed appears to be due to dust.

To make a rough estimate of the significance of deviation from lensed-ΛCDM, we calculate $\chi^2$ and $\chi$ (sum of deviations) for each of the $EE$ and $BB$ spectra and show these in Fig. 2. For the nine bandpowers used the expectation value/standard-deviation for $\chi^2$ and $\chi$ are 9/4.2 and 0/3 respectively. We see that BK150 and BK150×P353 are highly significant in $BB$, while P353×P353 has modest significance in both $EE$ and $BB$.

Figure 3 shows $EE$ and $BB$ cross-spectra between BICEP2/Keck and all of the polarized frequencies of Planck (also including the BICEP2/Keck auto-spectra). For the five bandpowers shown the expectation value/standard-deviation for $\chi^2$ and $\chi$ are 5/3.1 and 0/2.2 respectively. As already noted, the BK150×BK150 and BK150×P353 $BB$ spectra show highly significant excesses. Additionally, there is evidence for excess $BB$ in BK150×P217 spectrum, and for excess $EE$ in BK150×P353. The other spectra in Fig. 3 show no strong evidence for excess, although we note that only one of the $\chi$ values is negative. There is weak evidence for excess in the BK150×P70 $BB$ spectrum but none in BK150×P30 so this is presumably just a noise fluctuation.

There are a large number of additional Planck-only spectra, which are not plotted here. The noise on these is large and all are consistent with ΛCDM, with the possible exception of P217×P353, where modest evidence for an excess is seen in both $EE$ and $BB$ (see e.g., Figure 10 of PIP-XXX).

D. Consistency of BICEP2 and Keck Array spectra

The $BB$ auto-spectra for BICEP2 and Keck Array in the lower left panel of Fig. 2 appear to differ by more than might be expected, given that the BICEP2 and Keck maps cover almost exactly the same region of sky. However, the error bars in this figure are the standard deviations of lensed-ΛCDM+noise simulations; while the signal is largely common between the two experiments the noise is not, and the signal-noise cross terms produce substantial additional fluctuation of the difference. The correct way to quantify this is to compare the difference of the real data to the pair-wise differences of simulations, using common input skies that have power similar to that of the real data to the pair-wise differences of simulations, using common input skies that have power similar to that of deviations) for each of the $EE$ and $BB$ spectra and show these in Fig. 2. For the nine bandpowers used the expectation value/standard-deviation for $\chi^2$ and $\chi$ are 9/4.2 and 0/3 respectively. We see that BK150 and BK150×P353 are highly significant in $BB$, while P353×P353 has modest significance in both $EE$ and $BB$.

Figure 3 shows $EE$ and $BB$ cross-spectra between BICEP2/Keck and all of the polarized frequencies of Planck (also including the BICEP2/Keck auto-spectra). For the five bandpowers shown the expectation value/standard-deviation for $\chi^2$ and $\chi$ are 5/3.1 and 0/2.2 respectively. As already noted, the BK150×BK150 and BK150×P353 $BB$ spectra show highly significant excesses. Additionally, there is evidence for excess $BB$ in BK150×P217 spectrum, and for excess $EE$ in BK150×P353. The other spectra in Fig. 3 show no strong evidence for excess, although we note that only one of the $\chi$ values is negative. There is weak evidence for excess in the BK150×P70 $BB$ spectrum but none in BK150×P30 so this is presumably just a noise fluctuation.

There are a large number of additional Planck-only spectra, which are not plotted here. The noise on these is large and all are consistent with ΛCDM, with the possible exception of P217×P353, where modest evidence for an excess is seen in both $EE$ and $BB$ (see e.g., Figure 10 of PIP-XXX).

D. Consistency of BICEP2 and Keck Array spectra

The $BB$ auto-spectra for BICEP2 and Keck Array in the lower left panel of Fig. 2 appear to differ by more than might be expected, given that the BICEP2 and Keck maps cover almost exactly the same region of sky. However, the error bars in this figure are the standard deviations of lensed-ΛCDM+noise simulations; while the signal is largely common between the two experiments the noise is not, and the signal-noise cross terms produce substantial additional fluctuation of the difference. The correct way to quantify this is to compare the difference of the real data to the pair-wise differences of simulations, using common input skies that have power similar to that of deviations) for each of the $EE$ and $BB$ spectra and show these in Fig. 2. For the nine bandpowers used the expectation value/standard-deviation for $\chi^2$ and $\chi$ are 9/4.2 and 0/3 respectively. We see that BK150 and BK150×P353 are highly significant in $BB$, while P353×P353 has modest significance in both $EE$ and $BB$.

Figure 3 shows $EE$ and $BB$ cross-spectra between BICEP2/Keck and all of the polarized frequencies of Planck (also including the BICEP2/Keck auto-spectra). For the five bandpowers shown the expectation value/standard-deviation for $\chi^2$ and $\chi$ are 5/3.1 and 0/2.2 respectively. As already noted, the BK150×BK150 and BK150×P353 $BB$ spectra show highly significant excesses. Additionally, there is evidence for excess $BB$ in BK150×P217 spectrum, and for excess $EE$ in BK150×P353. The other spectra in Fig. 3 show no strong evidence for excess, although we note that only one of the $\chi$ values is negative. There is weak evidence for excess in the BK150×P70 $BB$ spectrum but none in BK150×P30 so this is presumably just a noise fluctuation.

There are a large number of additional Planck-only spectra, which are not plotted here. The noise on these is large and all are consistent with ΛCDM, with the possible exception of P217×P353, where modest evidence for an excess is seen in both $EE$ and $BB$ (see e.g., Figure 10 of PIP-XXX).
observed in the real data. This was done in Section 8 of BK-V and the BICEP2 and Keck maps were shown to be statistically compatible. In an analogous manner we can also ask if the B150×P353 and K150×P353 BB cross-spectra shown in the bottom middle panel of Fig. 2 are compatible. Figure 4 shows the results. We calculate the $\chi^2$ and $\chi$ statistics on these difference spectra and compare to the simulated distributions exactly as in BK-V. The probability to exceed (PTE) the observed values is given in the figure for bandpowers 1–5 ($20 < \ell < 200$) and 1–9 ($20 < \ell < 330$). There is no evidence that these spectra are statistically incompatible.

E. Alternative power spectrum estimation

We check the reliability of the power spectrum estimation with an alternative pipeline. The filtered and purified Planck and BICEP2 maps used to make the spectra shown in Fig. 2 are transformed back into the HEALPix pixelization using cubic spline interpolation. The $B$-mode cross-power is then computed with the XL pipeline using cubic spline interpolation. The $B$-mode power spectrum is then computed with the XL pipeline. We check the reliability of the power spectrum estimation with an alternative pipeline. The filtered and purified Planck and BICEP2 maps used to make the spectra shown in Fig. 2 are transformed back into the HEALPix pixelization using cubic spline interpolation. The $B$-mode cross-power is then computed with the XL pipeline using cubic spline interpolation. The $B$-mode power spectrum is then computed with the XL pipeline.

III. LIKELIHOOD ANALYSIS

A. Algorithm

While it is conventional in plots like Fig. 2 to present bandpowers with symmetric error bars, it is important to appreciate that this is an approximation. The likelihood of an observed bandpower for a given model expectation value is generally an asymmetric function, which can be computed given knowledge of the noise level(s). To compute the joint likelihood of an ensemble of measured bandpower values it is of course necessary to consider their full covariance—this is especially important when using spectra taken at different frequencies on the same field, where the signal covariance can be very strong.

We compute the bandpower covariance using full simulations of signal-cross-signal, noise-cross-noise, and signal-cross-noise. From these, we can construct the covariance matrix for a general model containing multiple signal components with any desired set of SEDs. When we do this we deliberately exclude terms whose expectation value is zero, in order to reduce noise in the resulting matrix due to the limited number of simulated realizations.

To compute the joint likelihood of the data for any given proposed model we use the Hamimeche-Lewis [42] approximation (HL; see Section 9.1 of Ref. [43] for implementation details). Here we extend the method to deal with single- and cross-frequency spectra, and the covariances thereof, in an analogous manner to the treatment of, for example, $TT$, $TE$, and $EE$ in the standard HL method. The HL formulation requires that the bandpower covariance matrix be determined for only a single “fiducial model.” We compute multi-dimensional grids of models explicitly and/or use COSMOMC [44] to sample...
the parameter space.

B. Fiducial analysis

As an extension of the simplest lensed-$\Lambda$CDM paradigm, we initially consider a two component model of IGW with amplitude $r$, plus dust with amplitude $A_d$ (specified at 353 GHz and $\ell = 80$). (Here we assume that the spectral index of the tensor modes ($n_t$) is zero, and a scalar pivot scale of 0.05 Mpc$^{-1}$; all values of $r$ quoted in this paper are $r_{0.05}$ unless noted otherwise.) Figure 6 shows the results of fitting such a model to $BB$ bandpowers taken between BICEP2/Keck and the 217 and 353 GHz bands of Planck, using bandpowers 1–5 ($20 < \ell < 200$). For the Planck single-frequency case, the cross-spectrum of detector-sets (DS1×DS2) is used, following PIP-XXX. The dust is modeled as a power law $D_\ell \propto \ell^{-0.42}$, with free amplitude $A_d$ and scaling with frequency according to the modified blackbody model.

A tight Gaussian prior $\beta_d = 1.59 \pm 0.11$ is imposed, since this parameter is not well constrained from these data alone, but does appear to be stable across the sky (see Secs. II B and V A). For the frequency range of interest here variations in the two parameters of the modified blackbody are highly degenerate and the choice is made to hold $b_d$ fixed while allowing $\beta_d$ to be free. The prior assumes that the SED of dust polarization at intermediate latitudes [26] applies to the BICEP2/Keck field, where the signal-to-noise ratio of the Planck data is too low to determine it directly. From dust astrophysics, we expect variations of the dust SED in intensity and polarization to be correlated [18]. We thus tested our assumption by measuring the spectral index of the dust total intensity in the BICEP2/Keck field using the template fitting analysis described in Ref. [45], and find the same value. The uncertainty on $\beta_d$ is scaled from the dispersion of spectral indices at intermediate Galactic latitudes in Ref. [26], as explained in PIP-XXX.

In Fig. 6 we see that the BICEP2 data produce a $r$ likelihood that peaks higher than that for the Keck Array data. This is because for $\ell < 120$ the auto-spectrum B150×B150 is higher than for K150×K150, while the cross-spectrum B150×P353 is lower than K150×P353 (see Fig. 2). However, recall that both pairs of spectra B150×B150/K150×K150 and B150×P353/K150×P353 have been shown to be consistent within noise fluctuation (see Sec. II D). In Sec. IV A these likelihood results are also found to be compatible. Given the consistency between the two experiments, the combined result gives the best available measurement of the sky.

The combined curves (BK+P) in the left and center panels of Fig. 6 yield the following results: $r = 0.048^{+0.035}_{-0.032}$, $r < 0.12$ at 95% confidence, and $A_d = 3.3^{+0.9}_{-0.8}$, $A_d < 30$ for $r$ zero-to-peak likelihood ratio is 0.38. For the $r$ zero-to-peak likelihood ratio is 0.38. Taking $\frac{1}{2}(1 - f(2\log L_0/L_{peak}))$, where $f$ is the $\chi^2$ cdf (for one degree of freedom), we estimate that the probability to get a number smaller than this is 8% if in fact $r = 0$. For $A_d$ the zero-to-peak ratio is $1.8 \times 10^{-6}$ corresponding to a smaller-than probability of $1.4 \times 10^{-7}$, and a 5.1$\sigma$ detection of dust power.

The maximum likelihood model on the grid has parameters $r = 0.05$, $A_d = 3.30\mu$K$^2$ (and $\beta_d = 1.6$). Computing the bandpower covariance matrix for this model, we obtain a $\chi^2$ of 40.9. Using 28 degrees of freedom—5 bandpowers times 6 spectra, minus 2 fit parameters (since $\beta_d$ is not really free)—gives a PTE of 0.06. The largest contributions to $\chi^2$ come from the P353×P353 spectrum shown in the lower right panel of Fig. 2.

C. Variations from the fiducial data set and model

We now investigate a number of variations from the fiducial analysis to see what difference these make to the constraint on $r$.

- **Choice of Planck single-frequency spectra:** switching the Planck single-frequency spectra to use one of the alternative data splits (yearly or half-ring instead of detector-set) makes little difference (see Fig. 7).

- **Using only 150 and 353 GHz:** dropping the spectra involving 217 GHz from consideration also has little effect (see Fig. 7).

- **Using only BK150×BK150 and BK150×P353:** also excluding the 353 GHz single-frequency spectrum from consideration makes little difference. The statistical weight of the BK150×BK150 and BK150×P353 spectra dominate (see Fig. 7).

- **Extending the bandpower range:** going back to the base data set and extending the range of bandpowers considered to 1–9 (corresponding to $20 < \ell < 330$) makes very little difference—the dominant statistical weight is with the lower bandpowers (see Fig. 7).

- **Including EE spectra:** we can also include in the fits the EE spectra shown in Fig. 3. PIP-XXX (figures 5 and A.3) shows that the level of EE from Galactic dust is on average around twice the level of BB. However, there are substantial variations in this ratio from sky-patch to sky-patch. Setting $EE/BB = 2$ we find that the constraint on $A_d$ narrows, while the $r$ constraint changes little; this latter result is also shown in Fig. 7.

- **Relaxing the $\beta_d$ prior:** relaxing the prior on the dust spectral index to $\beta_d = 1.59 \pm 0.33$ pushes the peak of the $r$ constraint up (see Fig. 7). However, it is not clear if this looser prior is self consistent; if the frequency spectral index varied significantly across the sky it would invalidate cross-spectral analysis, but there is strong evidence against such
variation at high latitude, as explained in Sec. V A. Nevertheless, it is important to appreciate that the $r$ constraint curves shown in Fig. 6 shift left (right) when assuming a lower (higher) value of $\beta_d$. For $\beta_d = 1.3 \pm 0.11$ the peak is at $r = 0.021$ and for $\beta_d = 1.9 \pm 0.11$ the peak is at $r = 0.073$.

- **Varying the dust power spectrum shape:** in the fiducial analysis the dust spectral power spectrum is assumed to be a power law with $D_\ell \propto \ell^{-0.42}$. Marginalizing over spectral indices in the range $-0.8$ to 0 we find little change in the $r$ constraint (see also Sec. IV B for an alternate relaxation of the assumptions regarding the spatial properties of the dust pattern).

- **Using Gaussian determinant likelihood:** the fiducial analysis uses the HL likelihood approximation, as described in Sec. III A. An alternative is to recompute the covariance matrix $C$ at each point in parameter space and take $L = \det(C)^{-1/2} \exp\left(-\langle d^TC^{-1}d\rangle/2,\right)$ where $d$ is the deviation of the observed bandpowers from the model expectation values. This results in an $r$ constraint which peaks slightly lower, as shown in Fig. 7. Running both methods on the simulated realizations described in Sec. IV A, indicates that such a difference is not unexpected and that there may be a small systematic downward bias in the Gaussian determinant method.

- **Varying the HL fiducial model:** as mentioned in Sec. III A the HL likelihood formulation requires that the expectation values and bandpower covariance matrix be provided for a single “fiducial model” (not to be confused with the “fiducial analysis” of Sec. III B). Normally we use the lensed-$\Lambda$CDM+dust simulations described in Sec. IV A below. Switching this to lensed-$\Lambda$CDM+$r=0.2$ produces no change on average in the simulations, although it does cause any given realization to shift slightly—the change for the real data case is shown in Fig. 7.

- **Adding synchrotron:** BK-I took the WMAP K-band (23 GHz) map, extrapolated it to 150 GHz according to $\nu^{-3.3}$ (mean value within the BICEP2 field of the MCMC “Model f” spectral index map provided by WMAP [2]), and found a negligible predicted contribution ($r_{\text{sync,150}} = 0.0008 \pm 0.0041$). Figure 3 does not offer strong motivation to reexamine this finding—the only significant detections of correlated BB power are in the BK150×P353 and, to a lesser extent, BK150×P217 spectra. However, here we proceed to a fit including all the polarized bands of Planck (as shown in Fig. 3) and adding a synchrotron component to the base lensed-$\Lambda$CDM+noise+$r$+dust model. We take synchrotron to have a power law spectrum $D_\ell \propto \ell^{-0.6}$ [23], with free amplitude $A_{\text{sync}}$, where $A_{\text{sync}}$ is the amplitude at $\ell = 80$ and at 150 GHz, and scaling with frequency according to $\nu^{-3.3}$. In such a scenario we can vary the degree of correlation that is assumed between the dust and synchrotron sky patterns. Figure 8 shows results for the uncorrelated and fully correlated cases. Marginalizing over $r$ and $A_d$ we find $A_{\text{sync}} < 0.0003 \mu K^2$ at 95% confidence for the uncorrelated case, and many times smaller for the correlated. This last is because once one has a detection of dust it effectively becomes a template for the synchrotron. This synchrotron limit is driven by the Planck 30 GHz band—we obtain almost identical results when adding only this band, and a much softer limit when not including it. If we instead assume synchrotron scaling of $\nu^{-3.0}$ the limit on $A_{\text{sync}}$ is approximately doubled for the uncorrelated case and reduced for the correlated. (Because the DS1×DS2 data-split is not available for the Planck LFI bands we switch to Y1×Y2 for

FIG. 6. Likelihood results from a basic lensed-$\Lambda$CDM+$r$+dust model, fitting BB auto- and cross-spectra taken between maps at 150 GHz, 217, and 353 GHz. The 217 and 353 GHz maps come from Planck. The primary results (heavy black) use the 150 GHz combined maps from BICEP2/Keck. Alternate curves (light blue and red) show how the results vary when the BICEP2 and Keck Array only maps are used. In all cases a Gaussian prior is placed on the dust frequency spectrum parameter $\beta_d = 1.59 \pm 0.11$. In the right panel the two dimensional contours enclose 68% and 95% of the total likelihood.
As expected, approximately 50% of the realizations have a ratio between zero and peak of \(3 \times 10^{-12}\), equivalent to a 7.0\(\sigma\) detection of lensing in the BB spectrum. We note this is the most significant to-date direct measurement of lensing in B-mode polarization.

**IV. LIKELIHOOD VALIDATION**

**A. Validation with simulations**

We run the algorithm used in Sec. III B on ensembles of simulated realizations to check its performance. We first consider a model where \(r = 0\) and \(A_d = 3.6 \mu K^2\), this latter being close to the value favored by the data in a dust-only scenario [46]. We generate Gaussian random realizations using the fiducial spatial power law \(D_\ell \propto \ell^{-0.42}\), scale these to the various frequency bands using the modified blackbody law with \(T_d = 19.6\) K and \(\beta_d = 1.59\), and add to the usual realizations of lensed-\(\Lambda\)CDM+noise. Figure 10 shows the resulting \(r\) and \(A_d\) constraint curves, with the result for the real data from Fig. 6 overplotted. As expected, approximately 50% of the \(r\) likelihoods peak above zero. The median 95\% upper limit is \(r < 0.075\). We find that 8\% of the realizations have a ratio \(L_0/L_{\text{peak}}\) less than the 0.38 observed in the real data, in agreement with the estimate in Sec. III B. We can reject \(r = 0\) with 7.0\(\sigma\) confidence.

The above simulations assume that the dust component follows on average the fiducial \(D_\ell \propto \ell^{-0.42}\) spatial power law, and fluctuates around it in a Gaussian manner. To obtain sample dust sky patterns that may deviate from this behavior in a way which better reflects reality, we take the pre-launch version of the Planck Sky Model (PSM; version 1.7.8 run in “simulation” mode) [24] evaluated in the Planck 353 GHz band and pull out the same 352 [b] > 35° partially overlapping regions used in PIP-XXX. We then scale these to the other bands and proceed as before. Figure 11 presents the results. Some of the regions have dust power orders of magnitude higher than the real data and we cut them out (selecting 139 regions with peak \(A_d < 20 \mu K^2\)). The \(r\) likelihoods will broaden.

Varying lensing amplitude: in the fiducial analysis the amplitude of the lensing effect is held fixed at the \(\Lambda\)CDM expectation (\(A_L = 1\)). Using their own and other data, the Planck Collaboration quote a limit on the amplitude of the lensing effect versus the \(\Lambda\)CDM expectation of \(A_L = 0.99 \pm 0.05\) [3]. Allowing \(A_L\) to float freely, and using all nine bandpowers, we obtain the results shown in Fig. 9—there is only weak degeneracy between \(A_L\) and both \(r\) and \(A_d\). Marginalizing over \(r\) and \(A_d\) we find \(A_L = 1.13 \pm 0.18\) with a likelihood ratio between zero and peak of \(3 \times 10^{-11}\). Using the expression given in Sec. III B this corresponds to a smaller-than probability of \(2 \times 10^{-12}\), equivalent to a 7.0\(\sigma\) detection of lensing in the BB spectrum. We note this is the most significant to-date direct measurement of lensing in B-mode polarization.
as the level of $A_d$ increases, and we should therefore not be surprised if the fraction of realizations peaking at a value higher than the real data is increased compared to the simulations with mean $A_d = 3.6 \mu K^2$. However we still expect that on average 50% will peak above zero and approximately 8% will have an $L_0/L_{\text{peak}}$ ratio less than the 0.38 observed in the real data. In fact we find 57% and 7%, respectively, consistent with the expected values. There is one realization which has a nominal (false) detection of non-zero $r$ of $3.3\sigma$, although this turns out to also have one of the lowest $L_0/L_{\text{peak}}$ ratios in the Gaussian simulations shown in Fig. 10 (with which it shares the CMB and noise components), so this is apparently just a relatively unlikely fluctuation.

FIG. 10. Likelihoods for $r$ and $A_d$, using BICEP2/Keck and Planck, as plotted in Fig. 6, overplotted on constraints obtained from realizations of a lensed-ΛCDM+noise+dust model with dust power similar to that favored by the real data ($A_d = 3.6 \mu K^2$). Half of the $r$ curves peak at zero as expected.

FIG. 11. Constraints obtained when adding dust realizations from the Planck Sky Model version 1.7.8 to the base lensed-ΛCDM+noise simulations. (Curves for 139 regions with peak $A_d < 20 \mu K^2$ are plotted). We see that the results for $r$ are unbiased in the presence of dust realizations which do not necessarily follow the $\ell^{-0.42}$ power law or have Gaussian fluctuations about it.

As previously mentioned, the modified blackbody model predicts that dust emission is 4% as bright in the BICEP2 band as it is in the Planck 353 GHz band. Therefore, taking the auto- and cross-spectra of the combined BICEP2/Keck maps and the Planck 353 GHz maps, as shown in the bottom row of Fig. 2, and evaluating $(BK \times BK - \alpha BK \times P)/(1 - \alpha)$, at $\alpha = \alpha_{\text{fid}}$ cleans out the dust contribution (where $\alpha_{\text{fid}} = 0.04$). The upper panel of Fig. 12 shows the result.

As an alternative to the full likelihood analysis presented in Sec. III B, we can instead work with the differenced spectra from above, a method we denote the “cleaning” approach. If $\alpha_{\text{fid}}$ were the true value, the ex-
pected value of this combination over CMB and noise would have no dust contribution. However, dust would still contribute to its variance, but only through its 2-point function. In practice, we do not know $\alpha$ perfectly, and this uncertainty needs to be accounted for in a likelihood constructed from the differenced spectra. Our approach is to treat the differenced spectra as a form of data compression, and to compute the expectation value as a function of $r$, $A_d$, and $\beta_d$ at each point in parameter space (the dust dependence enters for $\alpha(\beta_d) \neq \alpha_{fid}$). We use the method of Ref. [42], with a fiducial covariance matrix, to build a likelihood for the difference spectra, and marginalize over $A_d$ and $\beta_d$, and hence $\alpha$, adopting the prior $\beta_d = 1.59 \pm 0.11$. This alternative likelihood has the advantage of being less sensitive to non-Gaussianity of the dust, since only the 2-point function of the dust affects the covariance of the differenced spectra close to $\alpha_{fid}$, while the full analysis may, in principle, be affected by the non-Gaussianity of the dust through 4-point contributions to power spectra covariances. This cleaning approach does, however, ignore the (small amount of) additional information available at other frequencies. The lower panel of Fig. 12 compares the result to the fiducial analysis with the full multi-spectra likelihood. It is clear from the widths of the likelihood curves that compressing the spectra to form the cleaned difference results in very little loss of information on $r$. The difference in peak values arises from the different data treatments and is consistent with the scatter seen across simulations. Finally, we note that one could also form a combination $(BK \times BK - 2\alpha BK \times P + \alpha^2 P \times P)/(1-\alpha)^2$ in which dust does not enter at all for $\alpha = \alpha_{fid}$. However, the variance of this combination of spectra is large due to the $Planck$ noise levels, and likelihoods built from this combination are considerably less constraining.

V. POSSIBLE CAUSES OF DECORRELATION

Any systematic error that suppresses the $BK_{150} \times P_{353}$ cross-frequency spectrum with respect to the $BK_{150} \times BK_{150}$ and $P_{353} \times P_{353}$ single-frequency spectra would cause a systematic upward bias on the $r$ constraint. Here we investigate a couple of possibilities.

A. Spatially varying dust frequency spectrum

If the frequency dependence of polarized dust emission varied from place to place on the sky, it would cause the 150 GHz and 353 GHz dust sky patterns to decorrelate and suppress the $BK_{150} \times P_{353}$ cross-spectrum. The assumption made so far in this paper is that such decorrelation is negligible. In fact PIP-XXX implicitly tests for such variation in their Figure 6, where the $Planck$ single- and cross-frequency spectra are compared to the modified blackbody model (with the cross-frequency spectra plotted at the geometric mean of their respective frequencies). This plot is for an average over a large region of low foreground sky (24%); however, note that if there were spatial variation of the spectral behavior anywhere in this region it would cause suppression of the single-frequency spectra with respect to the cross-frequency spectra.

PIP-XXX also tests explicitly for evidence of decorrelation of the dust pattern across frequencies. Their figure E.1 shows the results for large and small sky patches. The signal-to-noise ratio is low in clean regions, but no evidence of decorrelation is found.

As a further check, we artificially suppress the amplitude of the $BK_{150} \times P_{353}$ spectra in the Gaussian dust-only simulations (see Sec. IV A) by a conservative 10% (PIP-XXX sets a 7% upper limit). We find that the maximum likelihood value for $r$ shifts up by an average of 0.018, while $A_d$ shifts down by an average of 0.43 $\mu K^2$, with the size of the shift proportional to the magnitude of the dust power in each given realization. This behavior is readily understandable—since the $BK_{150} \times BK_{150}$ and $BK_{150} \times P_{353}$ spectra dominate the statistical weight, a decrease of the latter is interpreted as a reduction in dust power, which is compensated by an increase in $r$. The bias on $r$ will be linearly related to the assumed decorrelation factor.

B. Calibration, analysis etc.

Figure 3 shows that the $EE$ spectrum $BK_{150} \times BK_{150}$ is extremely similar to that for $BK_{150} \times P_{143}$. We can compare such spectra to set limits on possible decorrelation between the BICEP2/$Keck$ and $Planck$ maps arising from any instrumental or analysis related effect, including differential pointing, polarization angle mis-characterization, etc. Taking the ratio of $BK_{150} \times P_{143}$ to the geometric mean of $BK_{150} \times BK_{150}$ and $P_{143H1} \times P_{143H2}$, we find that for $TT$ the decorrelation is approximately 0.1%. For $EE$ the signal-to-noise ratio is lower, but decorrelation is limited to below 2%, and consistent with zero when compared to the fluctuation of signal+noise simulations.

VI. CONCLUSIONS

BK-I reported a highly significant detection of $B$-mode polarization, at 150 GHz, in excess of the lensed-$LCDM$ expectation over the range $30 < \ell < 150$. This excess has been confirmed by additional data on the same field from the successor experiment $Keck$ Array. PIP-XXX found that the level of dust power in a field centered on the BICEP2/$Keck$ region (but somewhat larger than it) is of the same magnitude as the reported excess, but noted that, “the present uncertainties are large,” and that a joint analysis was required.

In this paper we have performed this joint analysis, using the combined BICEP2/$Keck$ maps. Cross-correlating these maps against all of the polarized frequency bands
of Planck we find a highly significant B-mode detection only in the cross spectrum with 353 GHz. We emphasize that this 150×353 GHz cross-spectrum has a much higher signal-to-noise ratio than the 353 GHz single-frequency spectrum that PIP-XXX analyzed.

We have analyzed the data using a multi-frequency, multi-component fit. In this fit it is necessary to impose a strong prior on the variation of the brightness of the dust emission with observing frequency, since the available data are unable to constrain this alone, due to the relatively low signal-to-noise ratio in B-mode polarization at 353 GHz. However PIP-XXX clearly shows that the frequency spectrum of dust emission is quite uniform across the entire high latitude sky—a necessary condition for the type of analysis employed here.

We have shown that the final constraint on the tensor-to-scalar ratio r is very stable when varying the frequency bands used, as well as the model priors. The result does differ when using the BICEP2 and Keck Array data alone rather than in combination, but the difference is compatible with noise fluctuation. Expanding the model to include synchrotron emission, while also including lower Planck frequencies, does not change the result.

Allowing the amplitude of lensing to be free, we obtain $A_L = 1.13 \pm 0.18$, with a significance of detection of 7.0σ. This is the most significant direct detection to-date of lensing in B-mode polarization, even compared to experiments with higher angular resolution. The POLAR-BEAR experiment has reported a detection of B-mode lensing on smaller angular scales ($500 < \ell < 2100$), rejecting the $A_L = 0$ hypothesis at 97.2% confidence [47]. Additionally, ACT [48] and SPT [49] have reported lensing detections in polarization in cross-correlation with some other tracer of the dark matter distribution on the sky.

We have validated the main likelihood analysis on simulations of a dust-only model and performed a simple subtraction of scaled spectra, which approximates a map-based dust cleaning (obtaining an r constraint curve that peaks somewhat lower). Finally we investigated the possibility of astrophysical or instrumental decorrelation of the sky patterns between experiments or frequencies and find no evidence for relevant bias.

The final result is expressed as a likelihood curve for r, and yields an upper limit $r < 0.12$ at 95% confidence. The median limit in the lensed-ΛCDM+noise+dust simulations is $r < 0.075$. It is interesting to compare this latter to dust-free simulations using only BICEP2/Keck where the median limit is $r < 0.03$—the difference represents the limitation due to noise in the Planck maps, when marginalizing over dust. The r constraint curve peaks at $r = 0.05$ but disfavors zero only by a factor of 2.5. This is expected by chance 8% of the time, as confirmed in simulations of a dust-only model. We emphasize that this significance is too low to be interpreted as a detection of primordial B-modes. Transforming the Planck temperature-only 95% confidence limit of $r_{0.002} < 0.11$ [3] to the pivot scale used in this paper yields $r_{0.05} < 0.12$ compatible with the present result.

A COSMOMC module containing the EE and BB band-powers for all cross-spectra between the combined BICEP2/Keck maps and all of the frequencies of Planck is available for download at http://wheretogobig.

In order to further constrain or detect IGW, additional data are required. The Planck Collaboration may be able to make progress alone using the large angular scale "reionization bump," if systematics can be appropriately controlled [50]. To take small patch recombination bump" studies of the type pursued here to the next level, data with signal-to-noise comparable to that achieved by BICEP2/Keck at 150 GHz are required at more than one frequency. Figure 13 summarizes the situation. The BICEP2/Keck noise is much lower in the BICEP2/Keck field than the Planck noise. However, since dust emission is dramatically brighter at 353 GHz, it is detected in the cross-spectrum between BICEP2/Keck and Planck 353 GHz. Synrotron is not detected and the crossover frequency with dust is $\lesssim 100$ GHz. Planck’s PR2 data release [51] shows that for the cleanest 73% of the sky, at 40’ scales, the polarized foreground minimum is at 80-90 GHz. During the 2014 season, two of the Keck Array receivers observed in the 95 GHz band and these data are under active analysis. BICEP3 will add substantial additional sensitivity at 95 GHz in the 2015, and especially 2016, seasons. Meanwhile many other ground-based and sub-orbital experiments are making measurements at a variety of frequencies and sky coverage fractions.

ACKNOWLEDGMENTS

BICEP2 was supported by the U.S. National Science Foundation under Grants No. ANT-0742818 and ANT-1044978 (Caltech and Harvard) and ANT-0742592 and ANT-1110087 (Chicago and Minnesota). The development of antenna-coupled detector technology was supported by the JPL Research and Technology Development Fund and Grants No. 06-ARPA206-0040 and 10-SAT10-0017 from the NASA APRA and SAT programs. The Keck Array project was supported by the National Science Foundation under Grants ANT-1145172 (Harvard), ANT-1145143 (Minnesota) and ANT-1145248 (Stanford), and from the Keck Foundation (Caltech). 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. The development of Planck has been supported by: ESA; CNES and CNRS/INSU-INS23-INP (France); ASI, CNR, and INAF (Italy); NASA and DoE (USA); STFC and UKSA (UK); CSIC, MICINN, JA and RES (Spain); Tekes, AoF and CSC (Finland); DLR and MPG (Germany); CSA (Canada); DTU Space (Denmark); SER/SSO (Switzerland); RCN (Norway); SFI (Ireland); FCT/MCTES (Portugal); and ERC and PRACE (EU). A description of the Planck Collaboration and a list of its
FIG. 13. Expectation values, and uncertainties thereon, for the $\ell \sim 80$ BB bandpower in the BICEP2/Keck field. The green and magenta lines correspond to the expected signal power of lensed-ΛCDM and $r = 0.05$. Since CMB units are used, the levels corresponding to these are flat with frequency. The grey band shows the best fit dust model (see Section III B) and the blue shaded region shows the allowed region for synchrotron (see Sec. III C). The BICEP2/Keck noise uncertainty is shown as a single starred point, and the noise uncertainties of the Planck single-frequency spectra evaluated in the BICEP2/Keck field are shown in red. The blue points show the noise uncertainty of the cross-spectra taken between BICEP2/Keck and, from left to right, Planck 30, 44, 70, 100, 143, 217 & 353 GHz, and plotted at horizontal positions such that they can be compared vertically with the dust and sync curves.

[13] Planck (http://www.esa.int/planck) is a project of the European Space Agency (ESA) with instruments provided by two scientific consortia funded by ESA member states (in particular the lead countries, France and Italy) with contributions from NASA (USA), and telescope reflectors provided in a collaboration between ESA and a scientific consortium led and funded by Denmark.