

## BICEP/Keck XII: Constraints on axionlike polarization oscillations in the cosmic microwave background

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

(BICEP/Keck Collaboration)

<sup>1</sup>*School of Physics and Astronomy, Cardiff University, Cardiff CF24 3AA, United Kingdom*

<sup>2</sup>*Kavli Institute for Particle Astrophysics and Cosmology, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, California 94025, USA*

<sup>3</sup>*Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada*

<sup>4</sup>*Harvard-Smithsonian Center for Astrophysics,*

*60 Garden Street, MS42, Cambridge, Massachusetts 02138, USA*

<sup>5</sup>*Department of Physics, California Institute of Technology, Pasadena, California 91125, USA*

<sup>6</sup>*Department of Physics, University of Cincinnati, Cincinnati, Ohio 45221, USA*

<sup>7</sup>*Jet Propulsion Laboratory, Pasadena, California 91109, USA*

<sup>8</sup>*Minnesota Institute for Astrophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA*

<sup>9</sup>*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA*

<sup>10</sup>*National Institute of Standards and Technology, Boulder, Colorado 80305, USA*

<sup>11</sup>*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*

<sup>12</sup>*Department of Physics, Stanford University, Stanford, California 94305, USA*

<sup>13</sup>*Service des Basses Températures, Commissariat à l'Energie Atomique, 38054 Grenoble, France*

<sup>14</sup>*Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

<sup>15</sup>*Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA*

<sup>16</sup>*Department of Physics, University of California at San Diego, La Jolla, California 92093, USA*

<sup>17</sup>*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

<sup>18</sup>*Department of Physics, University of Toronto, Toronto, Ontario M5S 1A7, Canada*

<sup>19</sup>*Aix-Marseille Université, CNRS/IN2P3, CPPM, 13288 Marseille, France*

<sup>20</sup>*Department of Physics, Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA*



(Received 17 November 2020; accepted 7 January 2021; published 12 February 2021)

We present a search for axionlike polarization oscillations in the cosmic microwave background (CMB) with observations from the *Keck Array*. A local axion field induces an all-sky, temporally sinusoidal rotation of CMB polarization. A CMB polarimeter can thus function as a direct-detection experiment for axionlike dark matter. We develop techniques to extract an oscillation signal. Many elements of the method are generic to CMB polarimetry experiments and can be adapted for other datasets. As a first demonstration, we process data from the 2012 observing season to set upper limits on the axion-photon coupling constant in the mass range  $10^{-21}$ – $10^{-18}$  eV, which corresponds to oscillation periods on the order of hours to months. We find no statistically significant deviations from the background model. For periods larger than 24 hr (mass  $m < 4.8 \times 10^{-20}$  eV), the median 95% confidence upper limit is equivalent to a

\*Corresponding author.  
ajcukier@stanford.edu

rotation amplitude of  $0.68^\circ$ , which constrains the axion-photon coupling constant to  $g_{\phi\gamma} < (1.1 \times 10^{-11} \text{ GeV}^{-1})m/(10^{-21} \text{ eV})$ , if axionlike particles constitute all of the dark matter. The constraints can be improved substantially with data already collected by the BICEP series of experiments. Current and future CMB polarimetry experiments are expected to achieve sufficient sensitivity to rule out unexplored regions of the axion parameter space.

DOI: [10.1103/PhysRevD.103.042002](https://doi.org/10.1103/PhysRevD.103.042002)

## I. INTRODUCTION

With many astrophysical and cosmological measurements establishing the existence of dark matter, an understanding of its particle properties is one of the main aspirations of modern physics [1,2]. A promising dark matter candidate is the QCD (quantum chromodynamics) axion, which we define here to be the pseudo-Nambu-Goldstone degree of freedom associated with the Peccei-Quinn mechanism that has been proposed to solve the strong  $CP$  problem [3–9]. In this work, we consider the much larger class of *axionlike particles* (sometimes abbreviated as ALPs), which are light, bosonic particles with couplings to the Standard Model (SM) that are similar to those of the QCD axion, but with important differences. Whereas the QCD axion requires a specific coupling to the QCD field strength that generically gives rise to a relationship between the mass of the QCD axion and its coupling to the SM, albeit with some model dependence, the axionlike particles considered in this work lack this coupling to QCD. They are, therefore, not related to solutions of the strong  $CP$  problem and generally have no fixed relationship between their mass and coupling to the SM. Because of this, they occupy a much larger area in the mass-coupling parameter space. For simplicity, we will hereafter refer to axionlike particles as *axions*.

Very light axions can have astrophysically large de Broglie wavelengths, which have macroscopic consequences for the formation of structure. Such dark matter candidates are sometimes referred to as *fuzzy dark matter* (FDM) [10].

An important property of an axion field is that it creates an effective birefringence for opposite-helicity photons. Linear polarizations are, therefore, rotated, and the amount of rotation is proportional to the change in the axion field between the point of emission and the point of absorption [11–14]. In particular, we emphasize that the rotation of polarization responds to the harmonic oscillations of the axion field that occur with a frequency  $m/(2\pi)$ , where  $m$  is the axion mass. In this paper, we consider axion masses in the range of  $10^{-21}$ – $10^{-18}$  eV, which roughly corresponds to oscillation periods of hours to months.

Recently, Fedderke *et al.* proposed two axion observables accessible by current and future cosmic microwave background (CMB) polarimetry experiments [15]. The first is an overall suppression of CMB polarization, which is

referred to as the “washout” effect and can be constrained by measurements of the  $TT$ ,  $TE$ , and  $EE$  power spectra. The second is a time-varying global rotation of CMB polarization with angular frequency  $m$ . The latter observable is called the “AC oscillation” and is the main focus of this work.

The washout effect is due to the axion field evolution during the epoch of recombination. With this observable, Fedderke *et al.* used publicly available *Planck* data to rule out regions of the axion parameter space, which we show below in Fig. 6. The washout is, ultimately, cosmic variance limited, because it relies on the statistics of power spectra. The limits already set with the washout effect are within an order of magnitude of the cosmic variance limit.

The AC oscillation is a sinusoidal global rotation of CMB polarization with an angular frequency  $m$ . Whereas the washout effect is sensitive to axion dark matter present during the epoch of recombination, the AC oscillation is sensitive to axion dark matter at the location of the experiment. The temporal change in CMB polarization is a direct probe of the oscillation of the *local* axion field. The measurement of axionlike polarization oscillations in the CMB is a form of direct dark matter detection. The expected coherence time is  $\sim 2\pi/(mv^2)$ , where  $v \sim 10^{-3}$  is the Galactic virial velocity. The associated coherence length is  $\sim 2\pi/(mv)$ . For oscillation periods shorter than  $\sim 1$  day, existing axion limits are stronger than what can be achieved with the current generation of CMB instruments, so we set a minimum oscillation period of 2 hr for our search. For  $2\pi/m = 2$  hr, the coherence time is  $\sim 200$  yr, and the coherence length is  $\sim 0.07$  pc. We can, therefore, take the oscillation to be in phase for all CMB experiments. Furthermore, the signal should be in phase across all photon frequencies. The data from experiments operating at different times, locations, and wavelengths can be combined to search for a coherent polarization oscillation. As the signal persists in time, there is no cosmic variance limit. In the long term, therefore, the oscillation effect will likely be a more sensitive observable than the washout effect.

We denote by  $Q(\hat{\mathbf{n}}, t)$  and  $U(\hat{\mathbf{n}}, t)$  the Stokes parameters that are measured at sky coordinate  $\hat{\mathbf{n}}$  at time  $t$ . We denote by  $Q_0(\hat{\mathbf{n}})$  and  $U_0(\hat{\mathbf{n}})$  the Stokes parameters that would be measured if the axion were completely decoupled from photons, which is also the limit in which the CMB polarization field is static. The axion-photon coupling

constant is  $g_{\phi\gamma}$ . The amplitude of the axion field averaged over the CMB visibility function is denoted  $\langle\phi_*\rangle$ , which we take to be isotropic. The amplitude of the local axion field today is denoted  $\phi_0$ . We allow for an arbitrary phase  $\alpha$  in the oscillation. Then the observed Stokes parameters are related to the decoupled limit by [15]

$$\begin{aligned} Q(\hat{\mathbf{n}}, t) \pm iU(\hat{\mathbf{n}}, t) &= J_0(g_{\phi\gamma}\langle\phi_*\rangle) \exp[\pm ig_{\phi\gamma}\phi_0 \cos(mt + \alpha)] \\ &\times (Q_0(\hat{\mathbf{n}}) \pm iU_0(\hat{\mathbf{n}})). \end{aligned} \quad (1)$$

All of the combinations of the form  $g_{\phi\gamma}\phi_x$  for  $x \in \{*, 0\}$  are small and dimensionless. The Bessel function can be expanded as  $J_0(x) \approx 1 - x^2/4$  and represents an overall suppression of CMB polarization—i.e., the washout effect described above. Expanding the complex exponential to first order and defining the time-averaged polarization fields  $\langle Q(\hat{\mathbf{n}}) \rangle$  and  $\langle U(\hat{\mathbf{n}}) \rangle$ , we have

$$\begin{pmatrix} \langle Q(\hat{\mathbf{n}}, t) \rangle \\ \langle U(\hat{\mathbf{n}}, t) \rangle \end{pmatrix} = \begin{pmatrix} 1 & -f(t) \\ f(t) & 1 \end{pmatrix} \begin{pmatrix} \langle Q(\hat{\mathbf{n}}) \rangle \\ \langle U(\hat{\mathbf{n}}) \rangle \end{pmatrix}, \quad (2)$$

where

$$f(t) \equiv g_{\phi\gamma}\phi_0 \cos(mt + \alpha). \quad (3)$$

To linear order, the polarization oscillation causes a mixing of Stokes parameters. The parameter  $f(t)$  is small compared with unity, so the mixing matrix in Eq. (2) can be viewed as a *rotation* matrix expanded to leading order. The time-averaged fields  $\langle Q(\hat{\mathbf{n}}) \rangle$  and  $\langle U(\hat{\mathbf{n}}) \rangle$  are rotated into each other by an angle  $f(t)$ . This is equivalent to an on-sky rotation of the polarization pseudovectors by an angle  $f(t)/2$ . In Eq. (2), the washout effect has been absorbed into the definition of the time-averaged polarization fields, i.e.,  $\langle Q(\hat{\mathbf{n}}) \rangle = J_0(g_{\phi\gamma}\langle\phi_*\rangle)Q_0(\hat{\mathbf{n}})$ , and similarly for the Stokes  $U$ . Since the time-averaged polarization fields are direct observables of CMB polarimetry experiments, we can use Eq. (2) to search for the Stokes mixing  $f(t)$  without referring to the decoupled limit—i.e., to  $Q_0(\hat{\mathbf{n}})$  and  $U_0(\hat{\mathbf{n}})$ .

Although relatively faint compared with many polarized astrophysical sources, the CMB provides a number of serendipitous advantages in the search for axionlike polarization oscillations. Current-generation CMB experiments have deployed thousands of photon-noise-limited detectors that scan CMB-dominated patches of sky repetitively for years. The steady increase in detector count in CMB experiments translates directly to an increase in the statistical weight of each instantaneous measurement. Since the signal is a coherent all-sky rotation of polarization angles, every optically active detector can contribute to the measurement. The detectors are observing the CMB for a substantial fraction of each year, which provides temporal sensitivity on timescales of hours to years. By repetitively

scanning the same patch of sky, the time-averaged maps  $\langle Q(\hat{\mathbf{n}}) \rangle$  and  $\langle U(\hat{\mathbf{n}}) \rangle$  can be well estimated and used as templates to search for time variability as in Eq. (2). The oscillation signal is coherent both over the sky and over wavelength, so all CMB instruments can contribute independent of angular resolution and observing frequency.

The CMB has a theoretical advantage, in that the axion field at the point of emission is effectively zero [15]. The surface of last scattering represents an era much longer [ $\mathcal{O}(10^4)$  yr] than the axion oscillation periods under consideration in this work [ $\mathcal{O}(1)$  yr]. Emissions from different redshifts occur with different axion field values that, taken together, average to approximately zero along all lines of sight. An oscillation observed in the CMB today is, therefore, a direct measure of the *local* axion field only—i.e., the field at the point of *absorption*.

Polarization oscillations may also be observed in astrophysical sources such as pulsars [16,17], the jets of active galaxies [18], protoplanetary disks [19], and strong gravitational lens systems [20]. A complication in setting constraints with polarized astrophysical sources is the uncertainty in both the amplitude and the phase of the axion field at the point of *emission*, and these amplitudes and phases are, in general, different for different sources, which may introduce a large number of free parameters. A CMB-based search is constrained to have the same amplitude and phase across the entire sky, at every wavelength and at every observing site.

One of the main challenges in many CMB polarimetry experiments is contamination from Galactic foregrounds. These foregrounds are less problematic in an axion oscillation search for two reasons. The first is that the oscillations affect *all* CMB polarization—i.e., both  $E$  and  $B$  modes. While the Galactic foregrounds dominate at, e.g., 150 GHz in  $B$  modes even in the cleanest patches of sky, they are subdominant in  $E$  modes. In the BICEP observation patch, the polarization power in foregrounds is  $\sim 10\%$  as strong as the CMB power. By considering the much brighter  $E$  modes, we make the foreground contamination relatively weaker. The second reason is that the foregrounds do not present an all-sky coherent polarization oscillation. Some polarized signals from the Galaxy may be emitted from regions with a substantially different axion field value, and these signals would oscillate. The amplitude and phase of these oscillations, however, would depend on the axion field value at the point of emission, and these axion field values would not be coherent across the entire observing region. By constraining our search to polarization oscillations that are both global and coherent, we suppress contamination from Galactic foregrounds. For these reasons, we consider foreground contamination to be a minor concern.

Polarization oscillations are sensitive to the product  $g_{\phi\gamma}\phi_0$  [Eq. (3)] and, therefore, depend on both the axion-photon coupling constant  $g_{\phi\gamma}$  and the axion mass  $m$ . The latter dependence comes from the axion field strength

$$\phi_0 = (2.1 \times 10^9 \text{ GeV}) \left( \frac{m}{10^{-21} \text{ eV}} \right)^{-1} \times \left( \frac{\kappa \rho_0}{0.3 \text{ GeV/cm}^3} \right)^{1/2}, \quad (4)$$

where  $\rho_0$  is the local density of dark matter and  $\kappa$  is the fraction of dark matter composed of axionlike particles [15]. The  $m$  dependence of  $\phi_0$  implies that oscillation-derived limits on the coupling constant will roughly follow  $g_{\phi\gamma} \propto m$ . The coupling constant can be probed approximately independently of mass in a number of ways—e.g., by conversion of solar axions to x rays in strong laboratory magnetic fields [21], by conversion of supernova-produced axions to gamma rays in Galactic magnetic fields [22], from the x-ray transparency of the intracluster medium [23] (though we note that this bound has been challenged [24]), and by conversion of axions produced in Wolf-Rayet stars to x rays in Galactic magnetic fields [25]. At the same time, the axion mass can be constrained approximately independently of the coupling constant through considerations of small-scale structure—e.g., in the Lyman- $\alpha$  forest [26], in the population of Milky Way satellite galaxies [27], and with the subhalo mass function [28]. These investigations have set similar bounds in the range of  $m \gtrsim 2 \times 10^{-21}$  eV, which suggests a maximum oscillation period of  $\sim 20$  days. Ultimately, the axion parameter space will be constrained by a variety

of probes, each subject to a different set of systematic uncertainties.

The paper is organized as follows: In Sec. II, we provide an overview of the BICEP program, including details of the instruments and the integrated dataset. The next several sections outline our analysis method to search for axionlike polarization oscillations. Some of the analysis choices are specific to the *Keck Array*, and some expectations are stated on the basis of experience with the BICEP dataset. Many elements of the method, however, can be adapted for other CMB polarimetry experiments. An overview of the analysis structure is shown in Fig. 1. There are three main components of the analysis. The first step is to make maps, both real and simulated, which we do in the same way as for a standard CMB analysis [29]. These maps are our best estimates for the time-averaged Stokes parameters  $\langle Q(\hat{n}) \rangle$  and  $\langle U(\hat{n}) \rangle$ . The main departure from standard CMB mapmaking is the reobservation of a simulation component that represents a global polarization rotation. We call this component the “rotated CMB map,” which we will often abbreviate as rCMB, and its computational utility is described in Sec. III G. In Sec. III, we describe our method for extracting estimates of the global polarization rotation angle  $f(t)/2$  as a function of time. For this purpose, we introduce the correlation method of Sec. III E. From our ensemble of simulations, we construct a model distribution and an associated likelihood function. The statistical analysis is presented in Sec. IV. We impose conservative

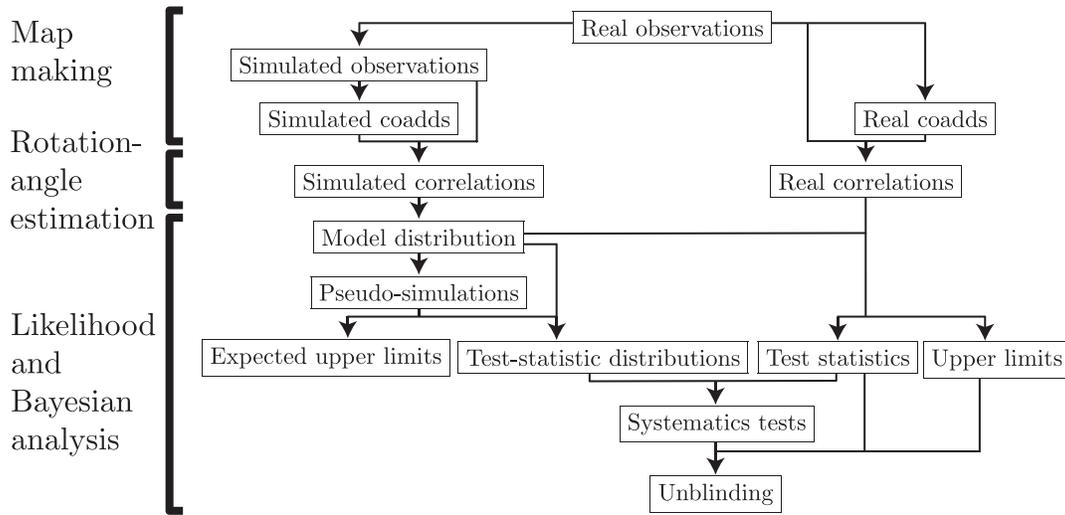


FIG. 1. Flow diagram of the axion oscillation analysis pipeline. The three main components are indicated on the left, and more detail is presented in the rest of the figure. The mapmaking step is similar to that of a standard CMB analysis, though the reobservation must now include a component that represents polarization oscillations (Secs. III B, III D, and III G). The coadds are estimates of the time-averaged Stokes parameters  $\langle Q(\hat{n}) \rangle$  and  $\langle U(\hat{n}) \rangle$ . The global rotation angle  $f(t)/2$  is obtained through the *correlation method* of Sec. III E with some of the computational speedups of Sec. III G. The likelihood and Bayesian analysis are described in Sec. IV. The ensemble of simulations is used to construct a model distribution, which we can resample to form large numbers of Gaussian pseudosimulations (Sec. IV A 1). The model distribution also implies a likelihood function (Sec. IV B) that is used to set Bayesian upper limits (Sec. IV C). At the same time, we estimate test statistics to check consistency with the background model (Sec. IV D) and to test for spurious systematic signals (Sec. V). Based on the results of the systematics checks, which we call “jackknife tests,” we unblind the real nonjackknife data (Sec. VI B).

prior distributions on the oscillation parameters and extract Bayesian upper limits on the oscillation amplitude. Additionally, we search for systematics in our data with several “jackknife tests” described in Sec. V. Having presented our analysis methods, we then use the 2012 observing season of the *Keck Array* for a first demonstration of the techniques, and the results are given in Sec. VI. We close in Sec. VII with some expectations for future results from BICEP and from other current and planned CMB experiments.

## II. INSTRUMENT OVERVIEW

The *Keck Array* made observations from the South Pole from 2012 to 2019 and consisted of a single mount with five microwave receivers, each similar to the precursor BICEP2 [30]. Each receiver is an independent refracting telescope with cryogenic lenses and an aperture diameter of 25 cm. The entire mount can be rotated to change the boresight orientation or *deck angle* of all five receivers. Each focal plane consists of 512 dual-polarized slot-dipole antenna arrays coupled to transition edge sensor (TES) bolometers [31]. A SQUID-based time-division multiplexing system is used to read out the TESs [32]. Each pixel is coupled to two TESs, one for each linear polarization, and the difference in signals is a measure of the on-sky polarization.

The main observing region occupies  $\sim 1\%$  of the sky centered on RA 0h, Dec.  $-57.5^\circ$ . For the 2012 and 2013 observing seasons, all five receivers observed at 150 GHz. In subsequent seasons, some receivers were switched to other observing frequencies. In the results below, however, we consider data from only the 2012 season.

The main science goal of the BICEP program is the search for *B*-mode polarization from primordial gravitational waves [33,34]. In combination with *Planck* [35] and WMAP [36], the BICEP2/*Keck* experiments have detected *B* modes from gravitational lensing [37] at  $8.8\sigma$  significance and constrained the tensor-to-scalar ratio to  $r_{0.05} < 0.06$  using data through the 2015 season [38]. The polarization map depth achieved by this dataset at 150 GHz is  $2.9 \mu\text{K}_{\text{CMB}}$  arcmin. The results presented in Sec. VI use only data from a single season, but we intend to extend this axion oscillation analysis to the full BICEP dataset in the future and take advantage of the full sensitivity of these polarization maps. The complete dataset also includes maps at 95 and 220 GHz, which have achieved polarization map depths of 5.2 and  $26 \mu\text{K}_{\text{CMB}}$  arcmin, respectively. All frequencies with CMB sensitivity can be used for the axion oscillation search.

Although designed for other purposes, the BICEP instruments, scan strategy, and data processing are compatible with an axion oscillation search. The methods and results presented below rely on data taken from standard observations targeted at CMB *B*-mode polarization.

No change to the scan strategy or low-level data processing is necessary for an axion oscillation analysis.

## III. GLOBAL POLARIZATION ROTATION ESTIMATION

The BICEP experiment does not measure Stokes parameters instantaneously but instead measures *pair differences*—i.e., the difference in power between two orthogonally polarized detectors. The average Stokes parameters are constructed only after repeated observations have been made with multiple detector orientations. From a single pair-difference measurement, it is not possible to construct Stokes parameters, but we will argue that it is not necessary for estimating  $f(t)$ .

### A. Pair difference

We parametrize the polarization orientation of detector pair  $i$  by the angle  $\psi_i(\hat{\mathbf{n}}, t)$ .<sup>1</sup> The pair difference is then related to the Stokes parameters by

$$D_i(\hat{\mathbf{n}}, t) = Q(\hat{\mathbf{n}}, t)c_i(\hat{\mathbf{n}}, t) + U(\hat{\mathbf{n}}, t)s_i(\hat{\mathbf{n}}, t), \quad (5)$$

where we introduce the shorthand

$$c_i(\hat{\mathbf{n}}, t) \equiv \cos[2\psi_i(\hat{\mathbf{n}}, t)], \quad s_i(\hat{\mathbf{n}}, t) \equiv \sin[2\psi_i(\hat{\mathbf{n}}, t)]. \quad (6)$$

In the limit that the polarization field is dominated by the CMB, we can decompose the pair difference into two components: a static component  $D_i^{(\text{static})}(\hat{\mathbf{n}}, t)$  that depends on the average Stokes parameters and an oscillating component  $D_i^{(\text{osc})}(\hat{\mathbf{n}}, t)$  that is induced by the local axion field. Combining Eqs. (2) and (5), we write

$$D_i(\hat{\mathbf{n}}, t) = D_i^{(\text{static})}(\hat{\mathbf{n}}, t) + D_i^{(\text{osc})}(\hat{\mathbf{n}}, t), \quad (7)$$

where

$$D_i^{(\text{static})}(\hat{\mathbf{n}}, t) = \langle Q(\hat{\mathbf{n}}) \rangle c_i(\hat{\mathbf{n}}, t) + \langle U(\hat{\mathbf{n}}) \rangle s_i(\hat{\mathbf{n}}, t) \quad (8)$$

and

$$D_i^{(\text{osc})}(\hat{\mathbf{n}}, t) = f(t)(\langle Q(\hat{\mathbf{n}}) \rangle s_i(\hat{\mathbf{n}}, t) - \langle U(\hat{\mathbf{n}}) \rangle c_i(\hat{\mathbf{n}}, t)). \quad (9)$$

The “static” component depends on time only through the time dependence of the polarization orientation  $\psi_i(\hat{\mathbf{n}}, t)$ . We call it “static,” because the underlying Stokes

<sup>1</sup>There is some redundancy in the expression for the polarization orientation  $\psi_i(\hat{\mathbf{n}}, t)$ , since a knowledge of the detector pointing as a function of time immediately implies a value for  $\hat{\mathbf{n}}$  given  $t$  and  $i$ . We will, however, bin observations in time. In each time bin, labeled by some mean time  $\tau$ , a detector pair produces a map that covers many sky coordinates  $\hat{\mathbf{n}}$ , so it is useful to keep track of both the mean time  $\tau$  and the sky coordinate  $\hat{\mathbf{n}}$ .

parameters are static. The average Stokes parameters  $\langle Q(\hat{\mathbf{n}}) \rangle$  and  $\langle U(\hat{\mathbf{n}}) \rangle$  are standard data products of the BICEP experiment and can be considered, at least approximately, known quantities. The trigonometric factors  $c_i(\hat{\mathbf{n}}, t)$  and  $s_i(\hat{\mathbf{n}}, t)$  depend on detector pointing and orientation and are also known. The full pair difference  $D_i(\hat{\mathbf{n}}, t)$  is measured, so the only unknown quantity is  $f(t)$ .

We define the *rotated map*

$$r_i(\hat{\mathbf{n}}, t) \equiv \langle Q(\hat{\mathbf{n}}) \rangle s_i(\hat{\mathbf{n}}, t) - \langle U(\hat{\mathbf{n}}) \rangle c_i(\hat{\mathbf{n}}, t) \quad (10)$$

and note that

$$D_i^{(\text{osc})}(\hat{\mathbf{n}}, t) = f(t)r_i(\hat{\mathbf{n}}, t). \quad (11)$$

The rotated map  $r_i(\hat{\mathbf{n}}, t)$  is a template that can be used to search for an oscillating component of the measured pair difference. We correlate the rotated map  $r_i(\hat{\mathbf{n}}, t)$  with the pair difference  $D_i(\hat{\mathbf{n}}, t)$  to estimate  $f(t)$ .

Comparing Eqs (5) and (10), we see that the rotated map  $r_i(\hat{\mathbf{n}}, t)$  is the pair difference that would be measured from Stokes parameters that are orthogonal (in  $Q/U$ -space) to the time-averaged values  $\langle Q(\hat{\mathbf{n}}) \rangle$  and  $\langle U(\hat{\mathbf{n}}) \rangle$ . This orthogonal component contributes  $D_i^{(\text{osc})}(\hat{\mathbf{n}}, t)$  [Eq. (11)] to the instantaneous pair difference  $D_i(\hat{\mathbf{n}}, t)$  when there is a global rotation of the time-averaged polarization field by an angle  $f(t)/2$ .

## B. Multiple components

The above discussion assumes that the polarization field is dominated by the CMB. We now take a more realistic approach and include noise and foregrounds.<sup>2</sup> Then the time-dependent Stokes fields can be expressed as

$$\begin{aligned} Q(\hat{\mathbf{n}}, t) &= \langle Q^{(\text{CMB})}(\hat{\mathbf{n}}) \rangle + Q^{(\text{fg})}(\hat{\mathbf{n}}) + Q^{(\text{N})}(\hat{\mathbf{n}}, t) \\ &+ Q^{(\text{osc})}(\hat{\mathbf{n}}, t), \end{aligned} \quad (12)$$

and similarly for  $U(\hat{\mathbf{n}}, t)$ , where the four terms on the right-hand side represent time-averaged CMB, static foregrounds, time-varying noise, and time-varying polarization oscillations, respectively. The pair difference is linearly related to these components and can also be written as a sum of the four contributions:

$$\begin{aligned} D_i(\hat{\mathbf{n}}, t) &= D_i^{(\text{CMB})}(\hat{\mathbf{n}}, t) + D_i^{(\text{fg})}(\hat{\mathbf{n}}, t) + D_i^{(\text{N})}(\hat{\mathbf{n}}, t) \\ &+ D_i^{(\text{osc})}(\hat{\mathbf{n}}, t). \end{aligned} \quad (13)$$

<sup>2</sup>In Sec. I, we argued that foreground contamination is expected to be negligible. We therefore make the simplifying approximation that the polarized foregrounds are not themselves subject to axion-induced polarization oscillations.

The instantaneous pair difference  $D_i(\hat{\mathbf{n}}, t)$  is dominated by the noise term  $D_i^{(\text{N})}(\hat{\mathbf{n}}, t)$ , so we approximate the variance in our measurement of  $D_i(\hat{\mathbf{n}}, t)$  to be entirely due to noise. The next-largest contribution is from the time-averaged CMB and is mainly in the form of  $E$  modes. Next, we have foregrounds. For 100–300 GHz, the dominant foreground is Galactic dust, which is suppressed relative to the CMB  $E$  modes by approximately an order of magnitude in the BICEP observation region. Finally, we have the oscillation signal  $D_i^{(\text{osc})}(\hat{\mathbf{n}}, t)$ . It will be useful to define the quantities

$$Q^{(\text{rCMB})}(\hat{\mathbf{n}}) \equiv -\langle U^{(\text{CMB})}(\hat{\mathbf{n}}) \rangle, \quad (14)$$

$$U^{(\text{rCMB})}(\hat{\mathbf{n}}) \equiv \langle Q^{(\text{CMB})}(\hat{\mathbf{n}}) \rangle, \quad (15)$$

as well as the pair difference formed from these “rotated” Stokes parameters, which is given by the usual formula

$$\begin{aligned} D_i^{(\text{rCMB})}(\hat{\mathbf{n}}, t) &\equiv Q^{(\text{rCMB})}(\hat{\mathbf{n}})c_i(\hat{\mathbf{n}}, t) \\ &+ U^{(\text{rCMB})}(\hat{\mathbf{n}})s_i(\hat{\mathbf{n}}, t). \end{aligned} \quad (16)$$

Then the oscillating component of the pair difference is

$$D_i^{(\text{osc})}(\hat{\mathbf{n}}, t) = f(t)D_i^{(\text{rCMB})}(\hat{\mathbf{n}}, t). \quad (17)$$

This is a useful recasting, since the quantity  $D_i^{(\text{rCMB})}(\hat{\mathbf{n}}, t)$  depends only on the time-averaged CMB polarization field and the detector pointing. This will become especially convenient in our discussion of efficient simulation schemes. The quantity  $D_i^{(\text{rCMB})}(\hat{\mathbf{n}}, t)$  is on the order of the static CMB, and the smallness of the oscillating component  $D_i^{(\text{osc})}(\hat{\mathbf{n}}, t)$  is therefore made manifest by the factor  $f(t)$ . Since an oscillation amplitude of  $0.1^\circ \approx 2 \times 10^{-3}$  has already been ruled out with *Planck* data [15], the oscillating component of the pair difference is suppressed relative to the CMB  $E$  modes by at least  $\sim 10^2$ .

Our goal is to extract the component  $D_i^{(\text{osc})}(\hat{\mathbf{n}}, t)$  from the total pair difference  $D_i(\hat{\mathbf{n}}, t)$ . We argued above that the oscillating component of the pair difference is proportional to  $r_i(\hat{\mathbf{n}}, t)$  in the limit that the time-averaged maps are CMB dominated, which is a good approximation at the map depths achieved by the BICEP/*Keck* experiments. The rotated map is constructed from known and measured quantities. To extract the oscillating component, then, we correlate the instantaneous pair difference with the rotated map—i.e., we correlate  $D_i(\hat{\mathbf{n}}, t)$  with  $r_i(\hat{\mathbf{n}}, t)$ .

## C. Time binning

The fundamental unit of observation in the BICEP/*Keck* experiments is the *scanset*, which consists of approximately 45 minutes of constant-elevation scanning. It is, therefore, convenient to bin observations by scanset. This makes the analysis insensitive to oscillation periods shorter than

roughly 1 hr, but the constraints on the axion-photon coupling from measurements of SN1987A already rule out an observable polarization oscillation at these time-scales given current-generation sensitivities [22].<sup>3</sup> Consequently, there is little motivation to extend the analysis to these short periods.

In each time bin, we construct a pair-difference map for each detector, which we call a *pairmap*. We take the variable  $\tau$  to label the mean time of a scanset, and we take the sky coordinate  $\hat{\mathbf{n}}$  to be discretized according to the BICEP map pixelization [29]. Then, we denote by  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  the pairmap constructed for detector  $i$  during the scanset that occurred at mean time  $\tau$ . This map will cover only a fraction of the full observation patch. The map pixels that are covered may have been visited multiple times over the course of the observation, and  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  represents a weighted average of these repeated measurements. We also include in the definition of  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  any timestream filtering, so  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  is the quantity that is coadded over all detectors and observations in a standard CMB analysis to form the final map.<sup>4</sup>

If the scan strategy involves multiple visits to the same map pixels, then the polarization angle  $\psi_i(\hat{\mathbf{n}}, t)$  may be different each time due to sky or, in principle, instrument rotation. If the pair differences from all of the visits are then averaged, the trigonometric factors  $c_i(\hat{\mathbf{n}}, t)$  and  $s_i(\hat{\mathbf{n}}, t)$  will beat against each other and wash out the polarization signal. An advantage of observing from the South Pole is that a constant-elevation scan is also approximately a constant-declination scan, so repeated visits to the same map pixel have nearly the same  $\psi_i(\hat{\mathbf{n}}, t)$ , as long as no boresight rotation has been performed. We denote the weighted average of the trigonometric factors by  $\bar{c}_i(\hat{\mathbf{n}}, \tau)$  and  $\bar{s}_i(\hat{\mathbf{n}}, \tau)$ . The weights are the same as those used to form the scanset map  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$ , though there is no filtering.

For oscillation periods of the same order as the time binning, there will be a suppression of the axion signal. For a time bin of length  $\Delta t$ , the average Stokes mixing angle is

$$\bar{f}(\tau) = f(\tau) \sin c\left(\frac{m\Delta t}{2}\right), \quad (18)$$

where the notation  $\bar{f}(\tau)$  indicates the average value of  $f(t)$  in the scanset of mean time  $\tau$ . The shortest oscillation period we will consider is 2 hr, for which a 45-minute

<sup>3</sup>As discussed explicitly in Sec. VI E, the oscillation analysis sets limits that scale as  $g_{\phi\gamma} \propto m$ , so sensitivity to the coupling constant  $g_{\phi\gamma}$  is worse at higher masses (shorter oscillation periods).

<sup>4</sup>The pair difference is not itself coadded. When a map pixel  $\hat{\mathbf{n}}$  has been visited more than once with different polarization orientations  $\psi_i(\hat{\mathbf{n}}, t)$ , it is possible to extract Stokes  $Q$  and  $U$  from the pair-difference measurements. The quantities that are actually used for the coaddition are  $c_i(\hat{\mathbf{n}}, t)D_i(\hat{\mathbf{n}}, t)$  and  $s_i(\hat{\mathbf{n}}, t)D_i(\hat{\mathbf{n}}, t)$ . With multiple visits at different orientations, these quantities can be inverted to recover  $Q(\hat{\mathbf{n}})$  and  $U(\hat{\mathbf{n}})$ .

observation produces a suppression of 22%. This is arguably small but should be accounted for in referring our measurements to an oscillation amplitude, since we would otherwise claim greater sensitivity at these short periods than is justified.

## D. Observables

In Sec. III B, we outlined a method of correlating the time-averaged CMB maps  $\langle Q^{(\text{CMB})}(\hat{\mathbf{n}}) \rangle$  and  $\langle U^{(\text{CMB})}(\hat{\mathbf{n}}) \rangle$  with the instantaneous pair difference  $D_i(\hat{\mathbf{n}}, t)$  as a means of extracting the polarization rotation angle. We do not have direct access to the true time-averaged CMB maps. Instead, we construct coadded maps, which are nontrivially impacted by filtering operations. As these coadded maps are dominated by CMB polarization, we will use them as templates. We use the notation  $\bar{Q}(\hat{\mathbf{n}})$  to denote the measured coaddition for Stokes  $Q$  and similarly for  $U$ . These ‘‘coadds’’ include all of the biases introduced by weighting and filtering observations but are the best available approximations to the time-averaged sky. While the true time-averaged values of noise and oscillations are zero, our coadds will, in general, contain nonzero contributions from these components. Our coadd for Stokes  $Q$  is, then,

$$\bar{Q}(\hat{\mathbf{n}}) = \bar{Q}^{(\text{CMB})}(\hat{\mathbf{n}}) + \bar{Q}^{(\text{fg})}(\hat{\mathbf{n}}) + \bar{Q}^{(\text{N})}(\hat{\mathbf{n}}) + \bar{Q}^{(\text{osc})}(\hat{\mathbf{n}}), \quad (19)$$

and similarly for Stokes  $U$ . Our approximation for the rotated map  $r_i(\hat{\mathbf{n}}, t)$  is, then,

$$\bar{r}_i(\hat{\mathbf{n}}, \tau) \equiv \bar{Q}(\hat{\mathbf{n}})\bar{s}_i(\hat{\mathbf{n}}, \tau) - \bar{U}(\hat{\mathbf{n}})\bar{c}_i(\hat{\mathbf{n}}, \tau), \quad (20)$$

where  $\tau$  is the nearest mean subobservation time to  $t$ .

## E. Correlation method

We define the correlation

$$\rho(\tau) \equiv \frac{1}{W(\tau)} \sum_{i, \hat{\mathbf{n}}} \bar{r}_i(\hat{\mathbf{n}}, \tau) \bar{D}_i(\hat{\mathbf{n}}, \tau) w_i(\hat{\mathbf{n}}, \tau) v_i(\hat{\mathbf{n}}, \tau), \quad (21)$$

where

$$\begin{aligned} w_i(\hat{\mathbf{n}}, \tau) &\equiv \frac{1}{\text{Var}[\bar{D}_i(\hat{\mathbf{n}}, \tau)]}, \\ v_i(\hat{\mathbf{n}}, \tau) &\equiv \frac{1}{\text{Var}[\bar{r}_i(\hat{\mathbf{n}}, \tau)]}, \\ W(\tau) &\equiv \sum_{i, \hat{\mathbf{n}}} w_i(\hat{\mathbf{n}}, \tau) v_i(\hat{\mathbf{n}}, \tau). \end{aligned} \quad (22)$$

The quantity  $\rho(\tau)$  is a weighted correlation of the measured pair difference  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  with our approximation to the rotated map  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$ . The weights are the inverse variances of the two maps. In general, the variance in the per-scanset pair difference will be relatively large, being dominated by atmospheric fluctuations at the time of observation.

This variance  $w_i(\hat{\mathbf{n}}, \tau)$  is a direct observable, which we use to downweight noisy pair-difference measurements in our standard mapmaking pipeline. At the same time, we use  $v_i(\hat{\mathbf{n}}, \tau)$  to downweight regions of the map for which we have relatively poor estimates of the CMB polarization field. The efficacy of the analysis depends on having time-averaged maps that are dominated by CMB polarization. In general, it will be the edges of the coadded maps that show significant residual noise, since these map pixels are visited much less frequently than those at the center. The particular linear combination of time-averaged Stokes parameters with which we are correlating is  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$ , so we downweight by its variance

$$\text{Var}[\bar{r}_i(\hat{\mathbf{n}}, \tau)] = \text{Var}[\bar{Q}(\hat{\mathbf{n}})]\bar{s}_i^2(\hat{\mathbf{n}}, \tau) + \text{Var}[\bar{U}(\hat{\mathbf{n}})]\bar{c}_i^2(\hat{\mathbf{n}}, \tau) - 2 \text{Cov}[\bar{Q}(\hat{\mathbf{n}}), \bar{U}(\hat{\mathbf{n}})]\bar{c}_i(\hat{\mathbf{n}}, \tau)\bar{s}_i(\hat{\mathbf{n}}, \tau). \quad (23)$$

All of the variances and covariances on the right-hand side are standard data products that are produced alongside the coadded maps. The quantity  $W(\tau)$  is nothing more than a normalization. These weights can be directly estimated from the data and can be considered known quantities.

An inverse-variance weighting may be suboptimal for the rotated map  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$ , because the  $E$  modes in our coadded maps  $\bar{Q}(\hat{\mathbf{n}})$  and  $\bar{U}(\hat{\mathbf{n}})$  tend to be stronger than the noise. In the signal-dominated limit, it is better to use all of the available modes. A possible improvement to the analysis is to weight by  $1/(1 + N/S)$ , where  $S/N$  is a figure of merit for the  $E$ -mode signal-to-noise ratio. For noisy pixels, this is essentially an inverse-variance weighting. For large signal-to-noise ratios, however, we achieve nearly equal weighting. The prescription amounts to an approximate Wiener filter.

The weighting in Eq. (21) does not include covariances between detectors  $i$  or between map pixels  $\hat{\mathbf{n}}$ . One consequence is that the weight  $W(\tau)$  does not exactly predict the true inverse variance of  $\rho(\tau)$ . It is computationally simpler, however, to calibrate the variances through Monte Carlo simulations as described below in Sec. IV A. Omitting covariances in the definition of  $\rho(\tau)$  [Eq. (21)] may degrade the sensitivity of the analysis, but it does not bias the results. With some simplifying approximations, it may be computationally practical to estimate such covariances, and this is a possible avenue for improvement in a future iteration of an axion oscillation search.

We can model  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  as a linear combination of constituent components, i.e.,

$$\bar{D}_i(\hat{\mathbf{n}}, \tau) = \bar{D}_i^{(\text{CMB})}(\hat{\mathbf{n}}, \tau) + \bar{D}_i^{(\text{fg})}(\hat{\mathbf{n}}, \tau) + \bar{D}_i^{(\text{N})}(\hat{\mathbf{n}}, \tau) + \bar{D}_i^{(\text{osc})}(\hat{\mathbf{n}}, \tau), \quad (24)$$

where  $\bar{D}_i^{(s)}(\hat{\mathbf{n}}, \tau)$  is the weighted, filtered, and binned scanset map formed from  $D_i^{(s)}(\hat{\mathbf{n}}, t)$  for any component  $s$ .

We have direct access to these constituent components only in simulation. In the limit that the coadded maps are good representations of the true time-averaged polarization field and the pair-difference measurements are good probes of the instantaneous polarization field, we have, from Eq. (11),

$$\bar{D}_i^{(\text{osc})}(\hat{\mathbf{n}}, \tau) \approx f(\tau)\bar{r}_i(\hat{\mathbf{n}}, \tau), \quad (25)$$

when the oscillation period is sufficiently long to treat  $f(t)$  as constant over the scanset of mean time  $\tau$ . We will discuss the case of faster oscillations in Sec. III F. The correlation  $\rho(\tau)$  defined in Eq. (21) picks out this component by correlating  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  with  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$ . On average, the correlation with the nonoscillating components of  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  vanishes approximately. Exact orthogonality is not necessary, though it improves the efficacy of the correlation method. An illustrative example of the correlation method is given in Fig. 2, where we show how the template  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$  picks out the oscillating component  $\bar{D}_i^{(\text{osc})}(\hat{\mathbf{n}}, \tau)$  from the full pair difference  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$ . The example shows only the results for a single detector  $i$  from a single scanset  $\tau$ , and we see that the noise variance dominates. These noise fluctuations, however, average down when we combine many detectors over many scansets, while the correlation with the oscillating component, because it is biased positive, does not. In this way, we increase the signal-to-noise ratio as the dataset grows.

We define the normalizing constant

$$R(\tau) \equiv \frac{1}{W(\tau)} \sum_{i, \hat{\mathbf{n}}} \bar{r}_i^2(\hat{\mathbf{n}}, \tau) w_i(\hat{\mathbf{n}}, \tau) v_i(\hat{\mathbf{n}}, \tau), \quad (26)$$

in terms of which we form the estimator

$$\hat{f}(\tau) \equiv \frac{\rho(\tau)}{R(\tau)}. \quad (27)$$

Due to residual noise in the coadded maps, filtering, weighting, binning, and imperfect orthogonality of the pair-difference components of Eq. (24), the estimator must be calibrated through simulation. When these effects are negligible, however,  $\hat{f}(\tau)$  is an unbiased estimator of  $f(\tau)$ .

### 1. Optimality

In the limit of Gaussian noise, negligible covariances, and CMB-dominated template maps  $\bar{Q}(\hat{\mathbf{n}})$  and  $\bar{U}(\hat{\mathbf{n}})$ , it can be shown that  $\hat{f}(\tau)$  [Eq. (27)] is the maximum-likelihood estimator. We expect the assumption of Gaussian noise to be a good approximation. We expect detector-detector and pixel-pixel covariances to be small, but potentially worth including in future iterations of the analysis. For one season of data from the *Keck Array*, however, the noise power in the coadded polarization maps is about half as strong as the

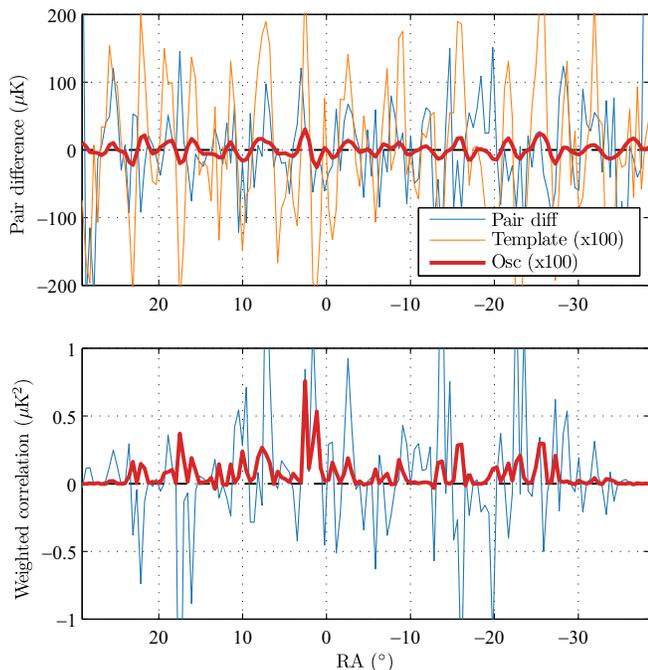


FIG. 2. An illustration of the correlation method introduced in Sec. III E for a simulation of a single detector and a single observation. Top: the pair difference  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  (blue) from a constant-elevation scan is plotted against right ascension (RA). The pair difference is dominated by atmospheric loading and fluctuates strongly. In this simulation, a global rotation of  $3^\circ$  has been imposed. The oscillating component  $\bar{D}_i^{(\text{osc})}(\hat{\mathbf{n}}, \tau)$  (red) is hidden underneath the atmospheric fluctuations. The rotated map  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$  (orange) is used as a template to pick out global polarization rotations. Bottom: the correlation  $\rho(\tau)$  [Eq. (21) summed over detector  $i$  only] between the template and the full pair difference (blue) has a large variance but a mean of zero. The correlation with the oscillating component (red) is biased positive.

CMB power. When we extend the analysis to include more seasons of BICEP data, the assumption of CMB domination will become significantly better. For the preliminary results presented in this work, we accept the sensitivity hit from having an estimator that maximizes the likelihood only approximately. There may be some gains from accounting for covariances between detectors and between map pixels. As mentioned above, we also expect an improvement in sensitivity by Wiener filtering instead of inverse-variance weighting the rotated map  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$ .

By comparing with results from the  $EB$  nulling procedure of our standard CMB analysis pipeline [39], we can roughly check the optimality of the correlation method as defined above. To absolutely calibrate the polarization angle of our receivers, we perform a global rotation to minimize the  $EB$  and  $TB$  cross spectra, which are expected to vanish in the CMB in the absence of cosmic birefringence. Finding the rotation angle  $\alpha_{EB}$  that minimizes  $EB$  only is similar to a search for axionlike polarization

oscillations with  $m = 0$ —i.e., a temporally constant offset in  $\hat{f}(\tau)$ .<sup>5</sup> The uncertainty in  $\alpha_{EB}$  is a measure of the sensitivity of the  $EB$  nulling procedure to an  $m = 0$  oscillation. For the 2012 observing season of the *Keck Array*, we find  $\Delta\alpha_{EB} = 0.21^\circ$ . In Sec. IV D below, we outline a procedure for finding the best-fit oscillation amplitude and phase for each angular frequency  $m$ . We can estimate an uncertainty by calculating the standard deviation of the constant offset for an ensemble of simulations. The result for  $m = 0$  is 0.27. A comparison with the  $EB$  result suggests that the correlation method may be  $\sim 30\%$  suboptimal, if none of the above improvements are implemented. The constant offset can be considered one oscillation mode, and each nonzero frequency  $m \neq 0$  represents a single oscillation mode as well. We expect roughly similar sensitivity to each mode. We compare to the rms rotation angle  $\theta_m$  for modes that pass through many oscillation periods in our dataset. If the rotation amplitude is  $\hat{A}_m/2$ , then the rms rotation angle (averaged over time) is  $\theta_m = \hat{A}_m/2^{3/2}$ . For oscillation periods between 1 and 30 days, the real data show an rms (averaged over  $m$ ) of  $\Delta\theta = 0.28^\circ$ , which again suggests a  $\sim 30\%$  degradation relative to the  $EB$  result. These comparisons are meant to give some indication of the possible margin for improvement. A full simulation-based study will be necessary to evaluate the true gains in sensitivity from, e.g., the changes suggested above, and we intend to report on this exploration in future publications.

## 2. Correlation matrices

In simulation, it is useful to keep track of the independent components contributing to both the per-scanset pair difference  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  and the rotated map  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$ . We decomposed  $\bar{D}_i(\hat{\mathbf{n}}, \tau)$  in Eq. (24) into four components: time-averaged CMB, foregrounds, time-varying noise, and time-varying polarization oscillations. As the mapmaking process is linear in the map components, we can decompose the rotated map, which is built from coadded maps, in the same way. We make the approximation that the contribution of polarization oscillations to the coadd is negligible, since we expect it to be small and to average down over all the observations included in the coadd. Then the rotated map can be decomposed as

$$\bar{r}_i(\hat{\mathbf{n}}, \tau) = \bar{r}_i^{(\text{CMB})}(\hat{\mathbf{n}}, \tau) + \bar{r}_i^{(\text{fg})}(\hat{\mathbf{n}}, \tau) + \bar{r}_i^{(\text{N})}(\hat{\mathbf{n}}, \tau), \quad (28)$$

<sup>5</sup>A possible scheme for an axion oscillation search is to estimate  $EB$  rotation angles from small subsets of observations and look for time variability in  $\alpha_{EB}$ . We considered but avoided this approach, because the incomplete map coverage of each scanset makes it awkward to form nonlocal quantities like the  $E$  and  $B$  modes. Additionally, we considered the  $E/B$  decomposition and the associated cross-spectra to be unnecessary computational expenses for the purposes of detecting a time-varying global polarization rotation.

where

$$\bar{r}_i^{(s)}(\hat{\mathbf{n}}, \tau) = \bar{Q}^{(s)}(\hat{\mathbf{n}})\bar{s}_i(\hat{\mathbf{n}}, \tau) - \bar{U}^{(s)}(\hat{\mathbf{n}})\bar{c}_i(\hat{\mathbf{n}}, \tau). \quad (29)$$

We approximate the weights to be dominated by the noise component—i.e.,

$$w_i(\hat{\mathbf{n}}, \tau) \approx \frac{1}{\text{Var}[\bar{D}_i^{(N)}(\hat{\mathbf{n}}, \tau)]}, \quad (30)$$

$$v_i(\hat{\mathbf{n}}, \tau) \approx \frac{1}{\text{Var}[\bar{r}_i^{(N)}(\hat{\mathbf{n}}, \tau)]}. \quad (31)$$

In particular, we note that these weights are independent of the non-noise components.

With all of these constituent components, the correlation  $\rho(\tau)$  can be written as a sum of cross terms, i.e.,

$$\rho(\tau) = \sum_{s_r, s_D} \rho^{(s_r, s_D)}(\tau), \quad (32)$$

where

$$\begin{aligned} \rho^{(s_r, s_D)}(\tau) \\ \equiv \frac{\sum_{i, \hat{\mathbf{n}}} \bar{r}_i^{(s_r)}(\hat{\mathbf{n}}, \tau) \bar{D}_i^{(s_D)}(\hat{\mathbf{n}}, \tau) w_i(\hat{\mathbf{n}}, \tau) v_i(\hat{\mathbf{n}}, \tau)}{W(\tau)}. \end{aligned} \quad (33)$$

On a per-observation basis, we expect  $\rho(\tau)$  to be dominated by  $\rho^{(\text{CMB}, \text{N})}(\tau)$ , since the rotated map is dominated by  $\bar{r}_i^{(\text{CMB})}(\hat{\mathbf{n}}, \tau)$  and the pair difference by  $\bar{D}_i^{(N)}(\hat{\mathbf{n}}, \tau)$ . On average, we expect  $\rho^{(\text{CMB}, \text{N})}(\tau)$  to vanish, but it has the largest variance of all the elements in the  $\rho$  matrix. The oscillation signal is in  $\rho^{(\text{CMB}, \text{osc})}(\tau)$ . The other elements make relatively minor contributions.

We also decompose the normalization constant as

$$R(\tau) = \sum_{s_{r_1}, s_{r_2}} R^{(s_{r_1}, s_{r_2})}(\tau), \quad (34)$$

where

$$\begin{aligned} R^{(s_{r_1}, s_{r_2})}(\tau) \\ \equiv \frac{\sum_{i, \hat{\mathbf{n}}} \bar{r}_i^{(s_{r_1})}(\hat{\mathbf{n}}, \tau) \bar{r}_i^{(s_{r_2})}(\hat{\mathbf{n}}, \tau) w_i(\hat{\mathbf{n}}, \tau) v_i(\hat{\mathbf{n}}, \tau)}{W(\tau)}. \end{aligned} \quad (35)$$

We can consider  $R^{(s_{r_1}, s_{r_2})}(\tau)$  to be a symmetric matrix at each observation time  $\tau$ . The matrix is dominated by the diagonal terms, which are essentially autocorrelations of the constituent components of the rotated map  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$ . As the CMB  $E$  modes dominate the coadded maps, we expect the largest contribution to be from  $R^{(\text{CMB}, \text{CMB})}(\tau)$ .

The next-largest contributions will be from the residual map noise  $R^{(\text{N}, \text{N})}(\tau)$  and the foregrounds  $R^{(\text{fg}, \text{fg})}(\tau)$ .

The correlation matrices are dominated by a relatively small minority of elements. Roughly, we expect

$$\rho(\tau) \approx \rho^{(\text{CMB}, \text{osc})}(\tau) + \rho^{(\text{CMB}, \text{N})}(\tau) + \rho^{(\text{N}, \text{N})}(\tau) \quad (36)$$

and

$$R(\tau) \approx R^{(\text{CMB}, \text{CMB})}(\tau) + R^{(\text{N}, \text{N})}(\tau). \quad (37)$$

Since  $D_i^{(\text{osc})}(\hat{\mathbf{n}}, t) = f(t)r_i^{(\text{CMB})}(\hat{\mathbf{n}}, t)$ , we expect  $\bar{D}_i^{(\text{osc})}(\hat{\mathbf{n}}, \tau) \approx \bar{f}(\tau)\bar{r}_i^{(\text{CMB})}(\hat{\mathbf{n}}, \tau)$ , and therefore  $\rho^{(\text{CMB}, \text{osc})}(\tau) \approx \bar{f}(\tau)R^{(\text{CMB}, \text{CMB})}(\tau)$ . Then, our estimator returns

$$\hat{f}(\tau) \approx \frac{\bar{f}(\tau)R^{(\text{CMB}, \text{CMB})}(\tau) + \rho^{(\text{CMB}, \text{N})}(\tau) + \rho^{(\text{N}, \text{N})}(\tau)}{R^{(\text{CMB}, \text{CMB})}(\tau) + R^{(\text{N}, \text{N})}(\tau)}. \quad (38)$$

The main nonidealities are due to noise. In the numerator, the high-variance but mean-zero correlations  $\rho^{(\text{CMB}, \text{N})}(\tau)$  and  $\rho^{(\text{N}, \text{N})}(\tau)$  cause a large scatter in measurements of  $\hat{f}(\tau)$ . In the denominator, the noise autocorrelation  $R^{(\text{N}, \text{N})}(\tau)$ , which is due to residuals in the coadded maps, causes an overall suppression. In the limit of negligible noise, we find  $\hat{f}(\tau) \approx \bar{f}(\tau)$ .

## F. Signal transfer function

For a single season of data from the *Keck Array*, the most significant bias in  $\hat{f}(\tau)$  comes from residual noise in the coadded maps. As shown in Eq. (38), the bias tends to suppress the signal strength. For one season, the suppression is at the level of  $\sim 30\%$  with percent-level variation across observations. By constructing coadded maps from multiple seasons of *Keck Array* observations, this effect can be reduced substantially in future analyses.

For oscillation periods comparable to the length of a scanset, we expect a signal suppression similar to an averaging of the oscillation as in Eq. (18). Instantaneously, we have  $D_i^{(\text{osc})}(\hat{\mathbf{n}}, t) = f(t)D_i^{(\text{rCMB})}(\hat{\mathbf{n}}, t)$ . If  $f(t)$  is changing over the course of the scanset, then the binning, weighting, and filtering required to produce  $\bar{D}_i^{(\text{osc})}(\hat{\mathbf{n}}, \tau)$  will act on both  $f(t)$  and  $D_i^{(\text{rCMB})}(\hat{\mathbf{n}}, t)$ —i.e., we cannot treat  $f(t)$  merely as an overall constant scaling. The analysis, however, depends only on the correlation  $\rho^{(s_r, \text{osc})}(\tau)$ , which involves averaging over thousands of detectors and map pixels. When the oscillation period is much longer than the scanset, we can treat  $f(t)$  as approximately constant and write

$$\rho^{(s_r, \text{osc})}(\tau) \approx f(\tau)\rho^{(s_r, \text{rCMB})}(\tau), \quad (39)$$

where  $\rho^{(s_r, \text{rCMB})}(\tau)$ , as defined in Eq. (33), is the correlation between  $\bar{r}_i^{(s_r)}(\hat{\mathbf{n}}, \tau)$  and the unphysical but well-defined

$\bar{D}_i^{(\text{rCMB})}(\hat{\mathbf{n}}, \tau)$ . For shorter oscillation periods, we can subdivide the scanset into shorter time bins until the approximation is valid. Then the scanset-level correlation  $\rho^{(s_r, \text{osc})}(\tau)$  is a weighted average of the correlations from the subdivisions. As the time-weighting is approximately uniform in BICEP observations that pass standard selection criteria, we approximate the weighted average with the uniform-weight average  $\bar{f}(\tau)$  from Eq. (18) and write

$$\rho^{(s_r, \text{osc})}(\tau) \approx \bar{f}(\tau) \rho^{(s_r, \text{rCMB})}(\tau). \quad (40)$$

This averaging suppresses the signal for oscillation periods on the order of a scanset. The suppression could be reduced by binning more finely in time. We argued in Sec. III C that finer time binning is unmotivated given the constraints set by SN1987A.

### G. Simulations

The analysis has been cast in terms of the correlation quantities  $\rho(\tau)$  and  $R(\tau)$ . We showed in Sec. III E 2 how each correlation can be decomposed as a matrix, each element of which gives the correlation between two contributing components of the polarization field. The components we consider are static CMB, static foregrounds, noise, and polarization oscillations. The first three are standard simulation products and are independent. The oscillations are derived from the static CMB according to Eq. (2) and depend on three parameters: amplitude  $A$ , phase  $\alpha$ , and angular frequency  $m$ —i.e., the input oscillation is parametrized as

$$f(t) = A \cos(mt + \alpha), \quad (41)$$

where, by assumption,  $A \ll 1$ .

Simulations are constructed from *reobservations* of input maps using the real detector pointings, data cuts, and weighting. For the static CMB maps, we use simulations of lensed  $\Lambda$ CDM cosmology, since that is the closest representation to the foreground-cleaned sky signal. Consistent with the standard BICEP simulation pipeline, we use Gaussian dust to simulate Galactic foregrounds.

For noise, we use the real data with random sign flips assigned at each scanset and with the average value subtracted out. This ensures that the coadded noise-simulation map has the same noise properties as the real data, including detector covariances, with negligible residuals from CMB and foregrounds. By flipping signs randomly, any true oscillation signal is scrambled. Furthermore, the template in simulation is constructed from a different CMB realization than the real data, so the correlation method would not pick up an oscillation in the sign-flip noise, even if one were present. A sign-flip noise scheme is computationally efficient, since the noise realizations are simply drawn from the real data rather than from an additional set of reobservations.

For computational efficiency, a major goal of the analysis is to minimize the number of independent reobservations. In the *Keck Array*, each scanset corresponds to roughly 50 minutes of observation time. In a full year of data, there are  $\sim 4 \times 10^3$  scansets. Many of the scansets are approximately identical in terms of detector pointings, so signal-only simulations need only be run on a minimal set of independent observations, which typically consists of only  $\sim 200$  scansets.

Sampling the oscillation parameters  $A$ ,  $\alpha$ , and  $m$  would be computationally infeasible through reobservations. Instead, we note from Eq. (40) that  $\rho^{(s_r, \text{osc})}(\tau)$  is always proportional to  $\rho^{(s_r, \text{rCMB})}(\tau)$ , which depends only on the input static CMB field and the detector pointing. We can then reobserve the unphysical rotated CMB map to create  $\bar{D}_i^{(\text{rCMB})}(\hat{\mathbf{n}}, \tau)$  for each scanset and then correlate with  $\bar{r}_i^{(s_r)}(\hat{\mathbf{n}}, \tau)$  to construct  $\rho^{(s_r, \text{rCMB})}(\tau)$ . This need only be done once per realization. For the physical oscillation signal, we need only scale the result by  $\bar{f}(\tau)$ , which is related to the oscillation parameters through Eqs. (18) and (41).

For the results presented below, we reobserved 110 realizations of lensed  $\Lambda$ CDM CMB, Gaussian dust, and the rotated version of lensed  $\Lambda$ CDM CMB. For the noise simulations, we use 110 independent sign-flip sequences on the real data. For each realization, we compute the elements of the correlation matrices  $\rho^{(s_r, s_D)}(\tau)$  and  $R^{(s_{r_1}, s_{r_2})}(\tau)$ , where  $s_r \in \{\text{CMB}, \text{fg}, \text{N}\}$  and  $s_D \in \{\text{CMB}, \text{fg}, \text{N}, \text{rCMB}\}$ . The efficiency is increased further by noting that  $R^{(s_{r_1}, s_{r_2})}(\tau)$  is symmetric.

The correlations are saved to disk for each detector pair independently, so we can apply detector selections without redundant computation. We define the per-detector correlation matrix elements

$$\rho_i^{(s_r, s_D)}(\tau) \equiv \frac{\sum_{\hat{\mathbf{n}}} \bar{r}_i^{(s_r)}(\hat{\mathbf{n}}, \tau) \bar{D}_i^{(s_D)}(\hat{\mathbf{n}}, \tau) w_i(\hat{\mathbf{n}}, \tau) v_i(\hat{\mathbf{n}}, \tau)}{W_i(\tau)} \quad (42)$$

and

$$R_i^{(s_{r_1}, s_{r_2})}(\tau) \equiv \frac{\sum_{\hat{\mathbf{n}}} \bar{r}_i^{(s_{r_1})}(\hat{\mathbf{n}}, \tau) \bar{r}_i^{(s_{r_2})}(\hat{\mathbf{n}}, \tau) w_i(\hat{\mathbf{n}}, \tau) v_i(\hat{\mathbf{n}}, \tau)}{W_i(\tau)}, \quad (43)$$

where

$$W_i(\tau) \equiv \sum_{\hat{\mathbf{n}}} w_i(\hat{\mathbf{n}}, \tau) v_i(\hat{\mathbf{n}}, \tau). \quad (44)$$

With these definitions, we can construct the all-detector matrix elements with

$$\rho^{(s_{r_1}, s_{r_2})}(\tau) = \frac{1}{W(\tau)} \sum_i \rho_i^{(s_{r_1}, s_{r_2})}(\tau) W_i(\tau), \quad (45)$$

$$R^{(s_{r_1}, s_{r_2})}(\tau) = \frac{1}{W(\tau)} \sum_i R_i^{(s_{r_1}, s_{r_2})}(\tau) W_i(\tau) \quad (46)$$

and

$$W(\tau) = \sum_i W_i(\tau). \quad (47)$$

To accommodate the scan-direction jackknife test described below in Sec. V, we separate the results for left- and right-going scans. We save the per-detector correlation to disk in this separated form and combine scan directions in a weighted average only once a jackknife test has been chosen that does not depend on scan direction.

An example of a simulated time series is shown in Fig. 3, where we also isolate the contributions from background and from axion-like polarization oscillations. The data are dominated by background fluctuations, and the oscillation is a small perturbation. As discussed in Sec. III F, the estimator  $\hat{f}(\tau)$  returns a slightly suppressed version of the true signal  $f(t)$ .

The template maps  $\bar{Q}(\hat{\mathbf{n}})$  and  $\bar{U}(\hat{\mathbf{n}})$  are constructed from all detectors observing at the same photon frequency.

For this reason, the detector-related systematics tests described in Sec. V are *partial* jackknives, since the template is constructed from *all* detectors, but the correlation sums in Eqs. (21) and (26) cover only *half* of the detectors.

#### IV. LIKELIHOOD AND BAYESIAN UPPER LIMITS

We compare the data  $\hat{f}(\tau)$  to a model consisting of static CMB, static foregrounds, noise, and a single oscillating component; i.e., we assume there is only one axion mass  $m$ . From simulation, we construct the model distribution and estimate a likelihood for each candidate value of  $m$  independently. By imposing prior distributions on the amplitude  $A$  and phase  $\alpha$ , we can set Bayesian upper limits on the axion-photon coupling constant, which is directly related to the oscillation amplitude  $A$ . At the same time, we fit for amplitude and phase and form a test statistic to check for consistency with the background-only model, where we take the background to consist of static CMB, static foregrounds, and noise.

##### A. Model distribution from simulation

In simulation, we can decompose the estimator  $\hat{f}(\tau)$  [Eq. (27)] as a matrix by

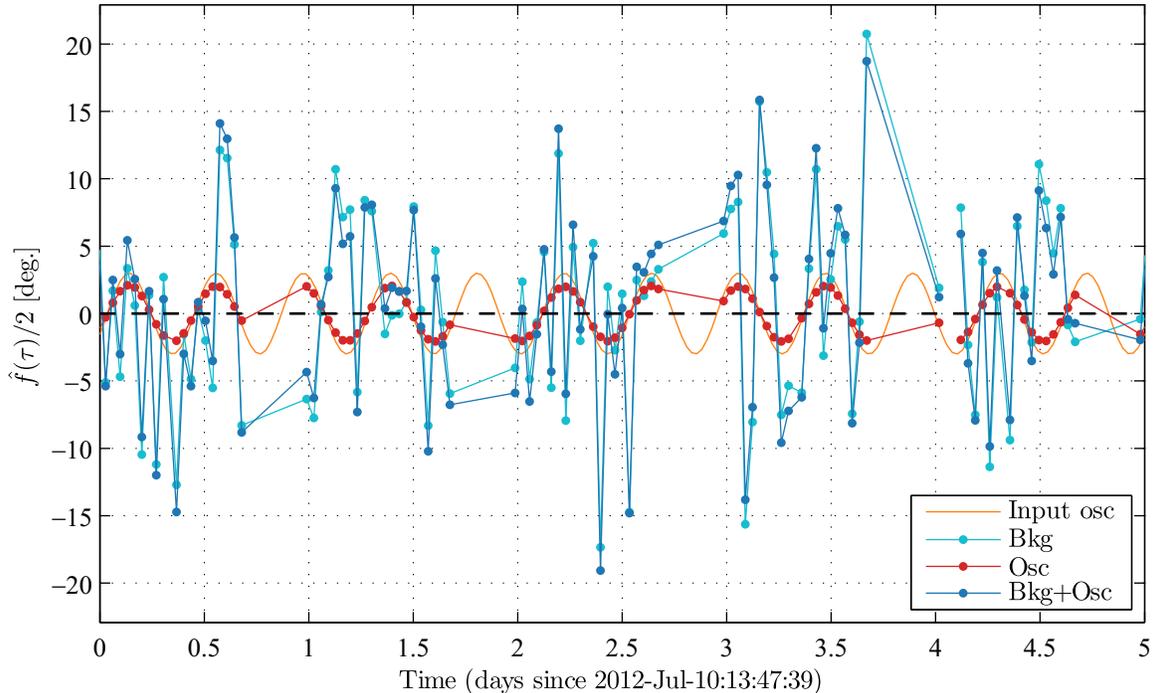


FIG. 3. A simulated time series  $\hat{f}(\tau)$  [blue] with an input rotation amplitude  $A/2 = 3^\circ$ , chosen to be relatively large in order to illustrate the effect more clearly. The background  $\hat{f}^{(\text{bkg})}(\tau)$  [cyan, defined explicitly in Eq. (52)] dominates over the oscillating component  $\hat{f}^{(\text{osc})}(\tau)$  [red, Eq. (51)]. The template maps  $\bar{Q}(\hat{\mathbf{n}})$  and  $\bar{U}(\hat{\mathbf{n}})$  are constructed from standard BICEP simulations of only the 2012 observing season of the *Keck Array*, so the underlying true oscillation  $f(t)$  (orange) is not recovered at full strength but is instead suppressed by  $\sim 30\%$  as described in Sec. III F.

$$\hat{f}^{(s_r, s_D)}(\tau) \equiv \frac{\rho^{(s_r, s_D)}(\tau)}{R(\tau)}, \quad (48)$$

where  $\rho^{(s_r, s_D)}(\tau)$  is defined in Eq. (33) and  $R(\tau)$  in Eq. (26). The denominator, then, contains contributions from all of the nonoscillating map components. From Eq. (40), we can pull out the dependence on  $\bar{f}(\tau)$  and rewrite the oscillating elements in terms of the rotated map, i.e.,

$$\hat{f}^{(s_r, \text{osc})}(\tau) = \bar{f}(\tau) \hat{f}^{(s_r, \text{rCMB})}(\tau). \quad (49)$$

This is a convenient factorization, since we save  $\rho^{(s_r, \text{rCMB})}(\tau)$  to disk, but not  $\rho^{(s_r, \text{osc})}(\tau)$ , as described in Sec. III G. We define the *dynamic* mixing angle

$$\hat{f}^{(\text{dyn})}(\tau) \equiv \sum_{s_r} \hat{f}^{(s_r, \text{rCMB})}(\tau), \quad (50)$$

where  $s_r \in \{\text{CMB}, \text{fg}, \text{N}\}$ . The oscillating component of the mixing angle is, then,

$$\hat{f}^{(\text{osc})}(\tau) \equiv \bar{f}(\tau) \hat{f}^{(\text{dyn})}(\tau). \quad (51)$$

Independent of any particular choice for oscillation parameters  $A$ ,  $\alpha$ , and  $m$ , we can use the dynamic mixing angle  $\hat{f}^{(\text{dyn})}(\tau)$  to investigate and precompute statistics of the signal transfer function. We define the *background* mixing angle by

$$\hat{f}^{(\text{bkg})}(\tau) \equiv \sum_{s_r, s_D} \hat{f}^{(s_r, s_D)}(\tau), \quad (52)$$

where  $s_r, s_D \in \{\text{CMB}, \text{fg}, \text{N}\}$ . Then, the full mixing angle is

$$\begin{aligned} \hat{f}(\tau) &= \hat{f}^{(\text{bkg})}(\tau) + \hat{f}^{(\text{osc})}(\tau) \\ &= \hat{f}^{(\text{bkg})}(\tau) + \bar{f}(\tau) \hat{f}^{(\text{dyn})}(\tau). \end{aligned} \quad (53)$$

We simulate many realizations of  $\hat{f}(\tau)$ . The mean over realizations is

$$\langle \hat{f}(\tau) \rangle = \langle \hat{f}^{(\text{bkg})}(\tau) \rangle + \bar{f}(\tau) \langle \hat{f}^{(\text{dyn})}(\tau) \rangle, \quad (54)$$

so we need save to disk only  $\langle \hat{f}^{(\text{bkg})}(\tau) \rangle$  and  $\langle \hat{f}^{(\text{dyn})}(\tau) \rangle$ , neither of which depends on oscillation parameters. To explore the oscillation parameter space, we scale the dynamic component by  $\bar{f}(\tau)$  and take the linear combination with the background component. In the limit of noiseless coadded maps, the dynamic component is close to unity. From Eq. (38), however, we see that residual noise in the coadded maps will suppress the signal. The quantity  $\langle \hat{f}^{(\text{dyn})}(\tau) \rangle$  is a measure of this suppression. We find  $\langle \hat{f}^{(\text{dyn})}(\tau) \rangle \approx 70\%$  with one season of 150 GHz data from the *Keck Array*, and the variance over  $\tau$  is at the percent level. The variance in  $\hat{f}^{(\text{dyn})}(\tau)$  over realizations for a given

scanset  $\tau$  is also at the percent level. In general, the consequence of residual noise is a consistent suppression of the signal by roughly 30%. The suppression can be lessened by using more seasons of data to form the coadded maps.

By assumption, the true background mean vanishes. Even though the sample mean over realizations may be nonzero, we set the background model mean to zero. Then the full model mean—i.e., including oscillations—is

$$\mu(\tau) \equiv \bar{f}(\tau) \langle \hat{f}^{(\text{dyn})}(\tau) \rangle. \quad (55)$$

The variance over realizations is dominated by the background, so we take the model variance to be

$$\sigma^2(\tau) \equiv \langle (\hat{f}^{(\text{bkg})}(\tau))^2 \rangle. \quad (56)$$

With the ensemble of simulations, one can check that the standardized variable

$$s_\tau \equiv \frac{\hat{f}(\tau) - \mu(\tau)}{\sigma(\tau)} \quad (57)$$

is Gaussian distributed with mean zero and unit variance when  $\hat{f}(\tau)$  is created with the same oscillation parameters that define  $\mu(\tau)$ . Additionally, no evidence is found for covariances between scansets. We take it as a model assumption, then, that  $s_\tau$  is drawn independently for each scanset time  $\tau$  from a standard Gaussian distribution.

### 1. Pseudosimulations

We can quickly form pseudosimulations of the time series  $\hat{f}(\tau)$  by resampling from a Gaussian distribution with mean  $\mu(\tau)$  [Eq. (55)] and standard deviation  $\sigma(\tau)$  [Eq. (56)]. In this way, we can avoid the computational expense of a large number of reobservations. In the results presented below, we use 110 reobservations to estimate  $\mu(\tau)$  and  $\sigma(\tau)$  for each scanset time  $\tau$ . To estimate  $p$  values, however, we fill out the distribution with a larger number of pseudo-simulations:  $2 \times 10^4$  realizations to test for consistency with the background model, and  $5 \times 10^3$  realizations to test for spurious systematic excesses.

### B. Likelihood function

We form a Gaussian likelihood for a three-parameter oscillation model. Let  $\mu(\tau; m, A, \alpha)$  be the model mean [Eq. (55)] formed when  $f(t) = A \cos(mt + \alpha)$ . Then we form the test statistic

$$q_m(A, \alpha) \equiv \sum_{\tau} \left[ \frac{\hat{f}(\tau) - \mu(\tau; m, A, \alpha)}{\sigma(\tau)} \right]^2, \quad (58)$$

where the estimator  $\hat{f}(\tau)$  is formed from the input data. We remove from the analysis any scansets whose model

variance  $\sigma^2(\tau)$  is more than two standard deviations from the mean model variance. The associated observation times  $\tau$  simply do not contribute to the sum in Eq. (58). As the model is assumed to be Gaussian, the quantity  $q_m(A, \alpha)$  is a  $\chi^2$  test statistic with  $n$  degrees of freedom, where  $n$  is the number of scansets contributing to the sum.

In general, we will consider each value of  $m$  separately. For each  $m$ , we will allow amplitude  $A$  and phase  $\alpha$  to vary, so we form an ensemble of likelihoods, one for each value of  $m$ :

$$L_m(A, \alpha) = N \exp\left[-\frac{q_m(A, \alpha)}{2}\right], \quad (59)$$

where

$$N \equiv \frac{1}{\sqrt{(2\pi)^n \prod_{\tau} \sigma^2(\tau)}} \quad (60)$$

is a normalization coefficient. Crucially, there is no dependence on oscillation parameters in  $N$ , so it will be convenient to consider likelihood *ratios*, for which the  $N$  dependence drops out.

For computational efficiency, it is convenient to express the test statistic  $q_m(A, \alpha)$  from Eq. (58) as a linear combination of  $\tau$  sums, each of which depends only on  $m$  and *not* on  $A$  or  $\alpha$ . This can be accomplished by rewriting Eq. (41) as

$$f(t) = A[\cos(mt) \cos \alpha - \sin(mt) \sin \alpha]. \quad (61)$$

The decomposition carries through Eqs. (18) and (55), which enters in the test statistic  $q_m(A, \alpha)$ . By expanding the square in the summand of Eq. (58), the  $A$ - and  $\alpha$ -dependent factors can be pulled out of the  $\tau$  sums. These  $\tau$  sums can be evaluated for a chosen set of  $m$  values and saved to disk. The two-dimensional parameter space of amplitude  $A$  and phase  $\alpha$  can then be explored quickly by forming linear combinations of the  $m$ -dependent terms.

### C. Bayesian upper limits

Having computed the likelihood  $L_m(A, \alpha)$ , we marginalize over the phase  $\alpha$ , which carries no information about axion properties and is expected to be random. We set a uniform prior on the phase and define the marginal likelihood

$$L_m(A) \equiv \frac{1}{2\pi} \int_0^{2\pi} d\alpha L_m(A, \alpha). \quad (62)$$

For the amplitude  $A$ , we impose a prior distribution  $P(A)$ . The prior could, in principle, vary with  $m$ . For example, the prior could incorporate axion constraints from other datasets. For this analysis, however, we set a uniform,  $m$ -independent prior.

The posterior distribution for each  $m$  value is

$$P_m(A|\{\hat{f}(\tau)\}) = \frac{P(A)L_m(A)}{\int dA P(A)L_m(A)}, \quad (63)$$

i.e., the probability density for amplitude  $A$  given the angular frequency  $m$  and the data  $\{\hat{f}(\tau)\}$ . We set a linearly uniform prior

$$P(A) = \frac{1}{A_{\max}} [0 \leq A \leq A_{\max}], \quad (64)$$

where  $A_{\max}$  is set conservatively above the current constraints in the axion mass range under consideration. In the results below, we have used  $A_{\max}/2 = 4^\circ$  in the results below.<sup>6</sup> We integrate the posterior to estimate a cumulative distribution function (CDF), which we again compute for each  $m$  value independently:

$$F_m(A) = \int_0^A dA' P_m(A'|\{\hat{f}(\tau)\}). \quad (65)$$

We set a Bayesian 95% credible interval by finding the amplitude  $A$  that satisfies the condition  $F_m(A) = 95\%$ . This is our upper limit on the oscillation amplitude at each  $m$  value.

### D. Background consistency

To check for consistency with the background, we form a  $\Delta\chi^2$  test statistic. First, we evaluate the test statistic [Eq. (58)] for the background model—i.e., with  $A = 0$ :

$$q_0 \equiv q_m(0, \alpha). \quad (66)$$

Then, we find the amplitude  $\hat{A}_m$  and phase  $\hat{\alpha}_m$  that minimize  $q_m(A, \alpha)$ . The test statistic for background consistency is

$$\Delta q_m \equiv q_0 - q_m(\hat{A}_m, \hat{\alpha}_m), \quad (67)$$

which is expected to be  $\chi^2$ -distributed with 2 degrees of freedom. This expectation is confirmed in simulations, but we do not rely on it in the results presented below. Instead, all  $p$  values are calibrated with an ensemble of simulations.

We evaluate the test statistic  $\Delta q_m$  from Eq. (67) for each  $m$  value. In the results presented below, we consider  $\sim 10^4$  values, so it is necessary to account for a trials factor. We use

$$\Delta \hat{q} \equiv \max_{m>0}(\Delta q_m) \quad (68)$$

<sup>6</sup>We will often be interested in the quantity  $A/2$ , since it is the amplitude of the on-sky oscillation of polarization angles. The quantity  $A$  is the amplitude of the mixing of Stokes  $Q$  and  $U$ .

as a global test statistic. We exclude  $m = 0$  from the maximization, since this term represents a constant offset and not an oscillation. Let  $\hat{p}$  be the associated probability to exceed (PTE) or  $p$  value for  $\Delta\hat{q}$ . In the results presented below, we have calibrated  $\hat{p}$  with  $2 \times 10^4$  realizations and can therefore estimate statistical tension up to the level of  $\sim 3\sigma$ .

## V. SYSTEMATICS

We perform a set of data splits to test for systematic effects that could create spurious oscillation signals. The set of splits is identical to those performed in previous BICEP/Keck analyses (cf. Ref. [40]), but the implementation and interpretation are different.

The template maps  $\bar{Q}(\hat{\mathbf{n}})$  and  $\bar{U}(\hat{\mathbf{n}})$  are constructed from the full dataset regardless of which jackknife test is under consideration. The rotated map  $\bar{r}_i(\hat{\mathbf{n}}, \tau)$  [defined in Eq. (20)] depends on the orientation of detector  $i$  at scanset time  $\tau$  but is otherwise simply a linear combination of the template maps. For each scanset  $\tau$ , we form the per-detector correlation matrix elements  $\rho_i^{(s_{r_1}, s_{r_2})}(\tau)$  and  $R_i^{(s_{r_1}, s_{r_2})}(\tau)$  defined in Eqs. (42) and (43), respectively. The correlations are constructed separately for left- and right-going scans, of which there are approximately 50 each per scanset. Having saved to disk the per-detector correlations split by scan direction for each scanset, we can perform jackknife tests by selecting subsets of these quantities to form a time series  $\hat{f}(\tau)$  and a likelihood  $L_m(A)$ . For the nontemporal jackknife tests, we form the mixing-angle estimator  $\hat{f}(\tau)$  for each scanset  $\tau$  from half of the data, either from only one of the scan directions or from only half of the detectors. For the temporal jackknife tests, we form the mixing-angle estimator  $\hat{f}(\tau)$  from all of the data available at each scanset, but we form the likelihood  $L_m(A)$  from only half of the scansets.

### A. Nontemporal jackknife tests

The nontemporal jackknives test whether the estimator  $\hat{f}(\tau)$  is a good statistical representation of the data collected during scanset  $\tau$ . We split the data either by the direction the telescope is slewing or by the contributing detectors, and we search for a systematic difference in the results. We perform the following nine nontemporal data splits defined in Sec. 8 of Ref. [40]: *Scan direction*, *Tile*, *Tile/deck*, *Focal plane inner/outer*, *Tile top/bottom*, *Tile inner/outer*, *Mux column*, *Mux row*, and *Differential pointing best/worst*. The scan-direction jackknife is considered nontemporal for the axion oscillation analysis, since the left- and right-going scans are interleaved on timescales much smaller than the oscillation periods of interest. The other jackknives split the data to expose potential nonidealities in optical, detector, or readout properties.

For these jackknife tests, it is possible to cancel the time-domain signal at the scanset level. We form a time series for the jackknife difference

$$\hat{f}^{(\text{jk})}(\tau) \equiv \frac{\hat{f}^{(1)}(\tau) - \hat{f}^{(2)}(\tau)}{2}, \quad (69)$$

where  $\hat{f}^{(i)}(\tau)$  is the mixing-angle estimator formed from the  $i$ th half of the data split. We can treat  $\hat{f}^{(\text{jk})}(\tau)$  as an ordinary time series. In simulation, one can check the efficacy of the signal cancellation by looking at  $\langle \hat{f}^{(\text{dyn})}(\tau) \rangle$  (introduced in Sec. IV A) constructed from many realizations of  $\hat{f}^{(\text{jk})}(\tau)$ . Recall that  $\langle \hat{f}^{(\text{dyn})}(\tau) \rangle \sim 70\%$  for undifferenced data. With differencing, there is a substantial reduction in  $\langle \hat{f}^{(\text{dyn})}(\tau) \rangle$ , though there is variation among the jackknife tests. As the noise level is expected to be comparable to current limits on the axion-photon coupling constant, we need only require a relatively modest signal cancellation. All of the tests reduce  $\langle \hat{f}^{(\text{dyn})}(\tau) \rangle$  by more than a factor of 20 with percent-level scatter over  $\tau$ . The scan-direction jackknife cancels the signal significantly better than all other tests, while the *Focal plane inner/outer* test cancels worst, with a few-percent bias away from zero.

Treating  $\hat{f}^{(\text{jk})}(\tau)$  as a measured rotation-angle time series, we can evaluate the test statistic [Eq. (67)] for background consistency:

$$\Delta q_m^{(\text{jk})} \equiv \Delta q_m(\{\hat{f}^{(\text{jk})}(\tau)\}). \quad (70)$$

Even though we expect any time-variable signal to be cancelled, we must perform this consistency test as a function of  $m$ . If we only used, e.g., the test statistic  $q_0$  [Eq. (66)] to check for background consistency, we would not pick up small residual oscillations. Sinusoidal fitting has much greater sensitivity to oscillatory signals, so we use the test statistic  $\Delta q_m$  [Eq. (67)] for these jackknife tests as well. The  $\Delta q_m$  test statistic is formed by comparing to the *undifferenced* model distribution. We do not compare to a model distribution based on the jackknife difference, because the signal transfer function for oscillations is close to zero. If  $\hat{f}^{(1)}(\tau)$  and  $\hat{f}^{(2)}(\tau)$  have equal variances, then  $\hat{f}^{(\text{jk})}(\tau)$  has the same variance as the background. In that case,  $\Delta q_m^{(\text{jk})}$  follows a  $\chi^2$  distribution with 2 degrees of freedom just like  $\Delta q_m$  from Eq. (67). In general, the variances are *not* equal, but the test statistic  $\Delta q_m^{(\text{jk})}$  can be scaled by an  $\mathcal{O}(1)$  factor, which can be fit for, to map it onto a  $\chi^2$  distribution with 2 degrees of freedom. We do not, however, rely on the  $\chi^2$  expectation for any results and instead calibrate  $p$  values for  $\Delta q_m^{(\text{jk})}$  through simulation.

Just as discussed for the background consistency test in Sec. IV D, we must account for the large number of  $m$  values being tested. We define

$$\Delta\hat{q}^{(\text{jk})} \equiv \max_{m>0}(\Delta q_m^{(\text{jk})}), \quad (71)$$

the most extreme signal-like excess, as a global test statistic for consistency with simulations. We exclude  $m = 0$  from the maximization, since this value represents a constant rotational offset, which may indicate an inefficiency in the experiment but does not produce spurious signals. We estimate a  $p$  value for this test statistic by comparing to a distribution of background-only pseudosimulations differenced in the same manner. Since the signal mostly cancels, the background-only simulations give approximately the same results as simulations with signal included.

### B. Temporal jackknife tests

Whereas the nontemporal jackknives test for statistical consistency *within* a scanset, the temporal jackknives test for consistency *among* scansets. We split the scansets into two groups and search for a systematic difference in oscillation signals. We perform the following five temporal data splits defined in Sec. 8 of Ref. [40]: *Deck angle*, *Alternative deck*, *Temporal split*, *Azimuth*, and *Moon up/down*. These tests are designed to expose pickup from far sidelobes and nonidealities of the optical performance. Because we are searching for a time-varying signal, it is not possible for temporal jackknives to cancel the signal in the time domain. Instead, we cancel the signal in the frequency domain by considering the best-fit oscillation amplitude constructed from each half of the temporal split. We form the test statistic

$$\hat{A}_m^{(\text{jk})} \equiv \left| \frac{\hat{A}_m^{(1)} - \hat{A}_m^{(2)}}{2} \right|, \quad (72)$$

where  $\hat{A}_m^{(i)}$  is the best-fit amplitude formed from the  $i$ th half of the data split. We take the absolute value of the difference, so signal-like systematics appear on only one side of the test-statistic distribution. Since the temporal data splits impose different window functions on the time series  $\hat{f}(\tau)$ , the best-fit amplitudes may vary even when the true frequency content is the same. This is similar to apodization effects in Fourier transforms, though we are not computing a Fourier transform here. This test statistic cancels the signal by more than an order of magnitude, but the residual has a larger variance than the background. We accept this increased variance from potential signals but still require the real jackknife results to match a background-only model. The test statistic  $\hat{A}_m^{(\text{jk})}$  is distributed approximately as a one-sided Gaussian, but we estimate all  $p$  values by comparing with simulations.

Due to the signal transfer function from time binning, the variance in  $\hat{A}_m^{(\text{jk})}$  increases with  $m$ . To keep the same normalization across the entire  $m$  range, we divide by the standard deviation as measured from background-only

simulations. The test statistic we use for estimating statistical deviations from the model distribution is, then,

$$a_m^{(\text{jk})} \equiv \frac{\hat{A}_m^{(\text{jk})}}{\text{std}[\hat{A}_m^{(\text{jk})}]}. \quad (73)$$

As in Secs. IV D and V A, we must account for the large number of  $m$  values under consideration. We take the largest value over the  $m$  range, which is the most extreme signal-like excess:

$$\hat{a}^{(\text{jk})} \equiv \max_{m>0}(a_m^{(\text{jk})}). \quad (74)$$

We exclude  $m = 0$  from the maximization, since this term indicates a constant rotational offset, which produces spurious signals in a way that is dealt with in Sec. V B 1. We estimate a  $p$  value for this test statistic by comparing to a distribution of background-only pseudosimulations that have been subjected to the same temporal jackknife. Although more signal can leak through the temporal jackknives than the nontemporal jackknives, we require the real data to be statistically consistent with the background-only simulations. This is a stricter requirement than is necessary to test for spurious systematic signals.

#### 1. Constant offset

The temporal jackknives test for two different types of systematics. First, we want to check that any oscillations are appearing at the same level in both halves of the temporal split. Second, we want to know if there is a systematic bias in the rotation angle that depends on a time-variable scan parameter and can, therefore, produce spurious oscillation signals at frequencies related to the observing schedule. The test statistic  $\hat{a}^{(\text{jk})}$  addresses the first type of concern. The second concern is addressed by privileging the  $m = 0$  jackknife difference—i.e., the difference in constant offset between the two halves of the data split. If there is a systematic bias that depends on, e.g., deck angle, then a naive analysis may detect an oscillation signal at frequencies related to the deck rotation schedule. A jackknife split, however, will discover the bias as a statistically anomalous value for  $\hat{A}_0^{(\text{jk})}$ . We therefore include  $\hat{A}_0^{(\text{jk})}$  as a test statistic for the *Deck angle*, *Alternative deck*, *Azimuth*, and *Moon up/down* jackknife tests. The reason we omit the *Temporal split* is that a nonzero value for  $\hat{A}_0^{(\text{jk})}$  could also be produced by a long-period oscillation. In fact, this jackknife test represents a minor unblinding, since a true oscillation, if it happens to be roughly synchronized with one of the temporal jackknife timescales, could appear in  $\hat{A}_0^{(\text{jk})}$ . Since this is unlikely, we proceed with the jackknife test and treat any deviation from the background-only simulations as evidence for a systematic bias.

### C. Global systematics assessment

For the nine nontemporal jackknives listed in Sec. VA, we calculate the test statistic  $\Delta\hat{q}^{(\text{jk})}$  [Eq. (74)]. For the five temporal jackknives listed in Sec. VB, we calculate the test statistic  $\hat{a}^{(\text{jk})}$  [Eq. (72)]. For the four temporal jackknives selected in Sec. VB 1, we calculate the test statistic  $\hat{A}_0^{(\text{jk})}$  [Eq. (72)]. In total, then, we are performing 18 tests for consistency with simulations. For each test, we form a  $p$  value, which we will denote  $p_i$ , where  $i$  is an index labeling each of the 18 tests. All of these  $p$  values are calibrated by comparing with an ensemble of  $5 \times 10^3$  background-only pseudosimulations (Sec. IVA 1). A simulation ensemble of this size allows us to estimate  $p$  values down to the level of  $\sim 10^{-3}$ . It is not important to precisely estimate smaller values, since we consider values below this level to indicate unacceptable inconsistency with the model. If we obtained such extreme values, we would consider our measurements to be systematically biased and would investigate the source before unblinding the undifferenced data.

Because we perform 18 systematics tests, we must account for a trials factor in determining the statistical significance of the most extreme result. We take the minimum jackknife  $p$  value

$$\hat{p}^{(\text{sys})} \equiv \min_i p_i \quad (75)$$

as a global test statistic for consistency with the simulation ensemble. While  $\hat{p}^{(\text{sys})}$  tests for signal-like systematics, we also check the sensitivity of the jackknife tests with the test statistic

$$\hat{c}^{(\text{sys})} \equiv 1 - \max_i p_i, \quad (76)$$

which is a measure of how well the model variances have been estimated. A small value for  $\hat{c}^{(\text{sys})}$  indicates that the variances have been overestimated, which would degrade the sensitivity of the jackknife tests.

The quantities  $\hat{p}^{(\text{sys})}$  and  $\hat{c}^{(\text{sys})}$  are drawn from an ensemble of  $p$  values but are to be regarded as test statistics. We use the ensemble of 110 reobservations to estimate  $p$  values for these test statistics. The reason we do not use the ensemble of  $5 \times 10^3$  pseudosimulations is that these do not include covariances between jackknife tests. The reobservations show negligible covariances among the 18 test statistics, but we do not wish to depend on that statistical independence for any of our results. Let  $p_p$  be the  $p$  value of the test statistic  $\hat{p}^{(\text{sys})}$  and  $p_c$  the  $p$  value for  $\hat{c}^{(\text{sys})}$ . Because these  $p$  values are estimated from an ensemble of only 110 realizations, we provide only two significant figures instead of the three significant figures provided for the individual jackknife tests.

We require both  $\hat{p}^{(\text{sys})}$  and  $\hat{c}^{(\text{sys})}$  to lie within the central  $2\sigma$  region of the background distribution. Although there

are two test statistics, we do not account for a trials factor. Our requirement, therefore, is more stringent than simply requiring overall  $2\sigma$  consistency.

## VI. RESULTS

For a first demonstration of the methods described in Secs. III–V, we selected the 2012 observing season of the *Keck Array*. During this season, all five receivers observed at 150 GHz, and the dataset has been thoroughly vetted by the standard BICEP CMB analyses [38]. The data volume is small enough for relatively quick iteration but large enough to understand computational scalings. The 2012 season represents only a small fraction of the total BICEP dataset, and we intend to extend this analysis to include more data in future publications.

An important element in the analysis is the rotated map  $\tilde{r}_i(\hat{\mathbf{n}}, \tau)$  [Eqs. (10) and (20)], which is constructed from the coadded maps  $\hat{Q}(\hat{\mathbf{n}})$  and  $\hat{U}(\hat{\mathbf{n}})$ . In principle, these coadded maps could be constructed from the full BICEP dataset, while the time series  $\hat{f}(\tau)$  could be restricted to the 2012 season. For computational speed, however, we used only 2012 data in all components of the analysis, which produces a moderate but non-negligible signal suppression (Sec. III F).

### A. Mass coverage

Our upper limits (Sec. IV C) are estimated for each value of  $m$  independently. The set of  $m$  values represents a discrete sampling in mass space rather than a binning. These  $m$  values can, in principle, be chosen arbitrarily. We wish, however, to achieve approximately continuous coverage over as broad a mass range as possible. Unlike a discrete Fourier transform (DFT), we aim to have some redundancy between neighboring mass values in order to justify interpolation. The methods of Secs. IV and V do not require the results from different mass values to be independent.

Since we bin our results by  $\sim 45$ -minute scansets (Sec. III C), we take the minimum oscillation period considered in the analysis to be 2 hr. As described in Sec. III C, the constraints on  $g_{\phi\gamma}$  from SN1987A are sufficiently strong that there is little motivation to push to oscillation periods smaller than 2 hr. This sets a maximum frequency for the analysis  $\nu_{\text{max}} = 0.5 \text{ hr}^{-1}$  and an associated maximum mass  $m_{\text{max}} = 2\pi\nu_{\text{max}}$ . Let  $T$  be the total time range covered by the time series  $\hat{f}(\tau)$ . Each season, science observations for the *Keck Array* typically lasted from early March until late October, so  $T \approx 8$  mo for a single season. We set a frequency resolution  $\Delta\nu \equiv 1/(\beta T)$ , where  $\beta$  is a factor that determines the amount of oversampling relative to a conventional DFT. In the results shown below, we use  $\beta = 3$ . With this frequency resolution, we consider the mass range  $0 \leq m \leq m_{\text{max}}$ . For the 2012 *Keck Array* season, this amounts to 8638  $m$

values. The  $m = 0$  results are used only for the temporal jackknife test statistic  $\hat{A}_0^{(\text{jk})}$  and are explicitly excluded from all other data products. We also ignore the results for oscillation periods longer than 30 days in order to satisfy the approximation that the coadded maps contain only a negligible contribution from polarization oscillations [Eq. (28)]. This last condition removes only 23  $m$  values, but it limits the low-frequency extent of our results. In a future iteration of the analysis, it may be computationally feasible to account for the oscillation residual in the coadded maps and set limits at arbitrarily low frequencies, though we expect degraded sensitivity when the oscillation period is on the order of or larger than the total observing time.

### B. Unblinding procedure

All real data products were kept blinded until the jackknife tests had been designed and shown in simulation to substantially suppress oscillation signals. From that point on, the results of real jackknife tests were unblinded. When it was concluded that there was no evidence for systematic

effects in the jackknife tests, we agreed on a decision tree for unblinding the undifferenced data. Since the *Keck Array* has collected data for eight seasons, the results from the 2012 season shown below represent a *partial* unblinding of the full dataset. To prepare for the possibility of a signal-like excess, we decided *before* unblinding that a measurement of  $\hat{p} \leq 6.2 \times 10^{-3}$  (see Sec. IV D), which would indicate tension with the background model in excess of  $2.5\sigma$ , would trigger an analysis of an additional season of data and allow for unblinded investigation of systematic effects in 2012 data. If the excess persisted, it would trigger an analysis of all four of the seasons of *Keck Array* observations contributing to Ref. [38]. This strategy would allow us to distinguish between a real signal and a statistical fluctuation while also providing more opportunities and data to search for systematic effects. We measured  $\hat{p} = 0.14$ , which indicates  $1.1\sigma$  signal-like tension with the background model. As this tension is significantly below the  $2.5\sigma$  threshold, we present results from only the 2012 season below. The results from additional seasons are being processed, and we intend to present them in subsequent publications.

TABLE I. Nontemporal jackknife  $p$  values for the test statistic  $\Delta\hat{q}^{(\text{jk})}$  [Eq. (71)]. Several  $p$  values are close to 1, and this may be due to overestimates of the model variances, which would weaken but not invalidate the jackknife tests. We assess the statistical significance of the most extreme value with the test statistic  $\hat{c}^{(\text{sys})}$  [Eq. (76)], which gives a  $p$  value  $p_c = 0.044$  and is, therefore, in tension with the background model at the level of  $1.7\sigma$ .

Jackknife test	$p(\Delta\hat{q}^{(\text{jk})})$
Scan direction	0.954
Tile	0.709
Tile/deck	0.333
Focal plane inner/outer	0.434
Tile top/bottom	0.965
Tile inner/outer	0.534
Mux column	0.970
Mux row	0.994
Differential pointing best/worst	0.999

TABLE II. Temporal jackknife  $p$  values for the test statistics  $\hat{a}^{(\text{jk})}$  [Eq. (74)] and  $\hat{A}_0^{(\text{jk})}$  [Eq. (72) and Sec. V B 1]. As discussed in Sec. V B 1, we do not consider  $\hat{A}_0^{(\text{jk})}$  for the temporal split.

Jackknife test	$p(\hat{a}^{(\text{jk})})$	$p(\hat{A}_0^{(\text{jk})})$
Deck angle	0.481	0.220
Alternative deck	0.330	0.621
Temporal split	0.127	...
Azimuth	0.760	0.927
Moon up/down	0.621	0.191

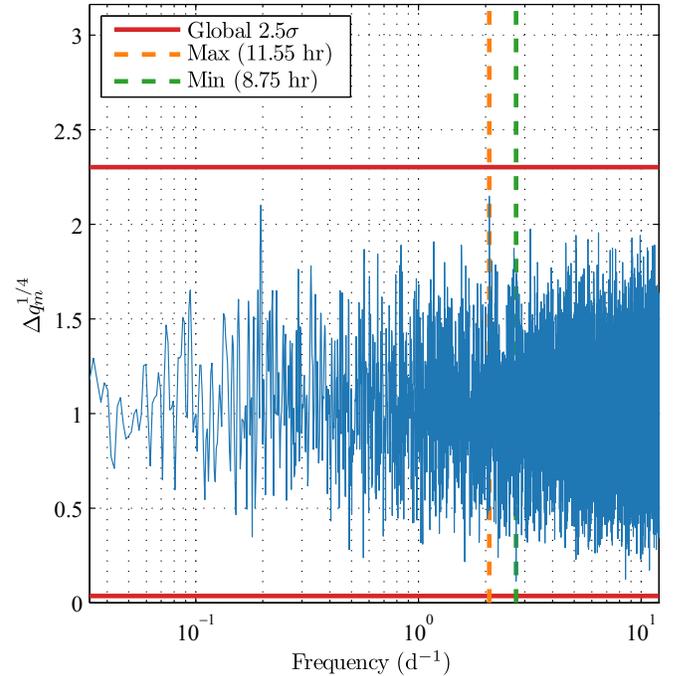


FIG. 4. The test statistic  $\Delta q_m$  for consistency with the background model [Eq. (67)] for real data from the 2012 observing season of the *Keck Array*. We plot  $\Delta q_m^{1/4}$  on the vertical axis in order to compress the distribution for visual purposes, and we plot frequency  $m/(2\pi)$  in units of inverse days ( $\text{d}^{-1}$ ) on the horizontal axis. The maximum and minimum values are indicated in the legend with their corresponding oscillation periods. The levels for global  $2.5\sigma$  fluctuations in both directions are indicated by horizontal red lines—i.e., there is a 1.2% probability in the background model that at least one value of  $\Delta q_m$  will lie outside the region bounded by the red lines.

### C. Systematics

The results of the 18 jackknife tests described in Sec. V are shown in Tables I and II. We find  $p_i \geq 12.7\%$  for all tests  $i$ , and this indicates that no individual test has revealed a spurious signal. Two of the nontemporal jackknife tests (Table I) show  $p_i > 99\%$ , which may be interpreted as statistically anomalous agreement with the background model. These large  $p$  values suggest possible overestimates of the model variances, which lower the sensitivity of the jackknife tests to systematic effects. We provide a global assessment of the jackknife results by considering the test statistics  $\hat{p}^{(\text{sys})}$  [Eq. (75)] and  $\hat{c}^{(\text{sys})}$  [Eq. (76)]. The associated  $p$  values come to  $p_p = 0.45$  and  $p_c = 0.044$ , respectively. The latter value is relatively low but lies within the central  $2\sigma$  region ( $0.0228 \leq p \leq 0.9772$ ), and we therefore conclude that there is no statistically significant tension with the background model.

A possible improvement for a future iteration of the systematics analysis is to consider, in addition to the most extreme  $p$  values  $\hat{p}^{(\text{sys})}$  and  $\hat{c}^{(\text{sys})}$ , the  $p$ -value distribution, though it should be noted that the introduction of additional test statistics dilutes the sensitivity of each.

### D. Background consistency

To check for consistency with the background model, we consider the test statistic  $\Delta q_m$  (Sec. IV D), which is plotted in Fig. 4 for real data from the 2012 observing season of the *Keck Array*. Since there are roughly  $10^4$  frequencies included in our analysis, we use the test statistic  $\Delta \hat{q}$  [Eq. (68)], which is simply the maximum value of  $\Delta q_m$ , to estimate a global  $p$  value  $\hat{p}$ . In Fig. 4, we show the  $2.5\sigma$  levels for  $\Delta \hat{q}$ , and we see that the entire spectrum lies within this region. We find  $\hat{p} = 0.14$ , which indicates a  $1.1\sigma$  signal-like fluctuation in  $\Delta \hat{q}$ . As the statistical significance of this fluctuation is far below the threshold set in Sec. VI B, we claim no evidence for tension with the background model.

### E. Upper limits

The direct observable in this analysis is the Stokes mixing amplitude  $A$ . For  $A \ll 1$ , which is a good approximation in this case, the amplitude of polarization rotations on the sky is  $A/2$ . Following the convention of Ref. [15], we express our upper limits in terms of the rotation amplitude  $A/2$ . We follow the prescription of Sec. IV C

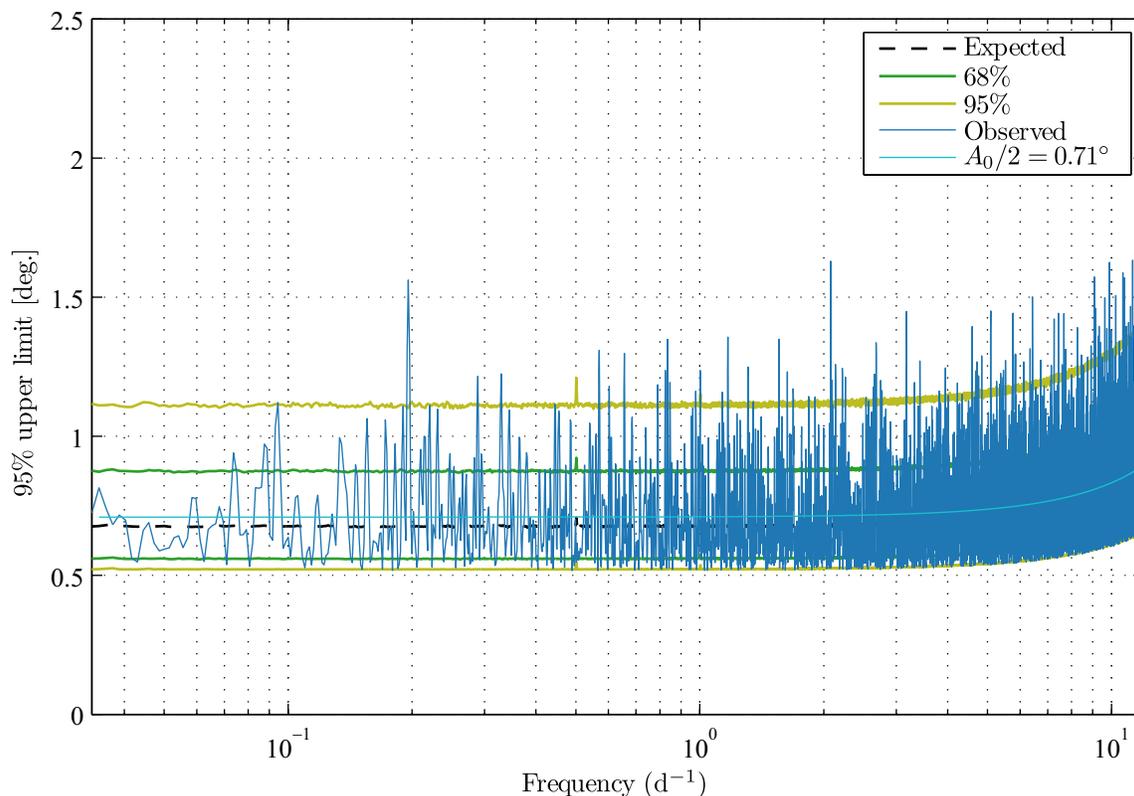


FIG. 5. Bayesian 95% confidence upper limits on rotation amplitude  $A/2$  (Sec. IV C). We also provide the median expectation (black dashes) from background-only simulations as well as  $1\sigma$  (green) and  $2\sigma$  (yellow) regions. These expectations represent local rather than global percentiles. With nearly  $10^4$  frequencies under consideration, we expect several values outside of the  $2\sigma$  region. The question of background consistency is addressed by Fig. 4 and the test statistic  $\Delta \hat{q}$  [Eq. (68)]. The median limit for oscillation periods larger than 24 hr (frequency less than  $1 \text{ d}^{-1}$ ) is  $0.68^\circ$ . For shorter periods (larger frequencies), the limits are degraded due to binning observations in  $\sim 1$  hr scansets (Sec. III C). Additionally, we plot a smoothed approximation to our upper limits [Eq. (78)] in cyan.

to compute 95% confidence upper limits, and we present the results in Fig. 5. For oscillation periods longer than one day ( $m/(2\pi) < 1 \text{ d}^{-1}$ ), the median limit is

$$A/2 < 0.68^\circ. \quad (77)$$

For visual comparison, we also show the expected distribution of upper limits as implied by background-only simulations.

For periods shorter than one day [ $m/(2\pi) > 1 \text{ d}^{-1}$ ], the limits degrade by  $\sim 20\%$  as we approach the binning timescale (Sec. III C). Over the entire frequency range,

we can obtain a smoothed approximation to our upper limits by performing a least-squares fit to

$$\frac{A}{2} < \frac{A_0}{2 \text{sinc}(m\Delta t/2)} \quad (78)$$

with  $\Delta t = 44.2 \text{ min}$ , which is the median scanset duration, and  $A_0$  as a free parameter. The sum over  $m \neq 0$  of squared residuals is minimized with  $A_0/2 = 0.71^\circ$ .

To convert our limits on rotation amplitude to the axion parameter space, we identify

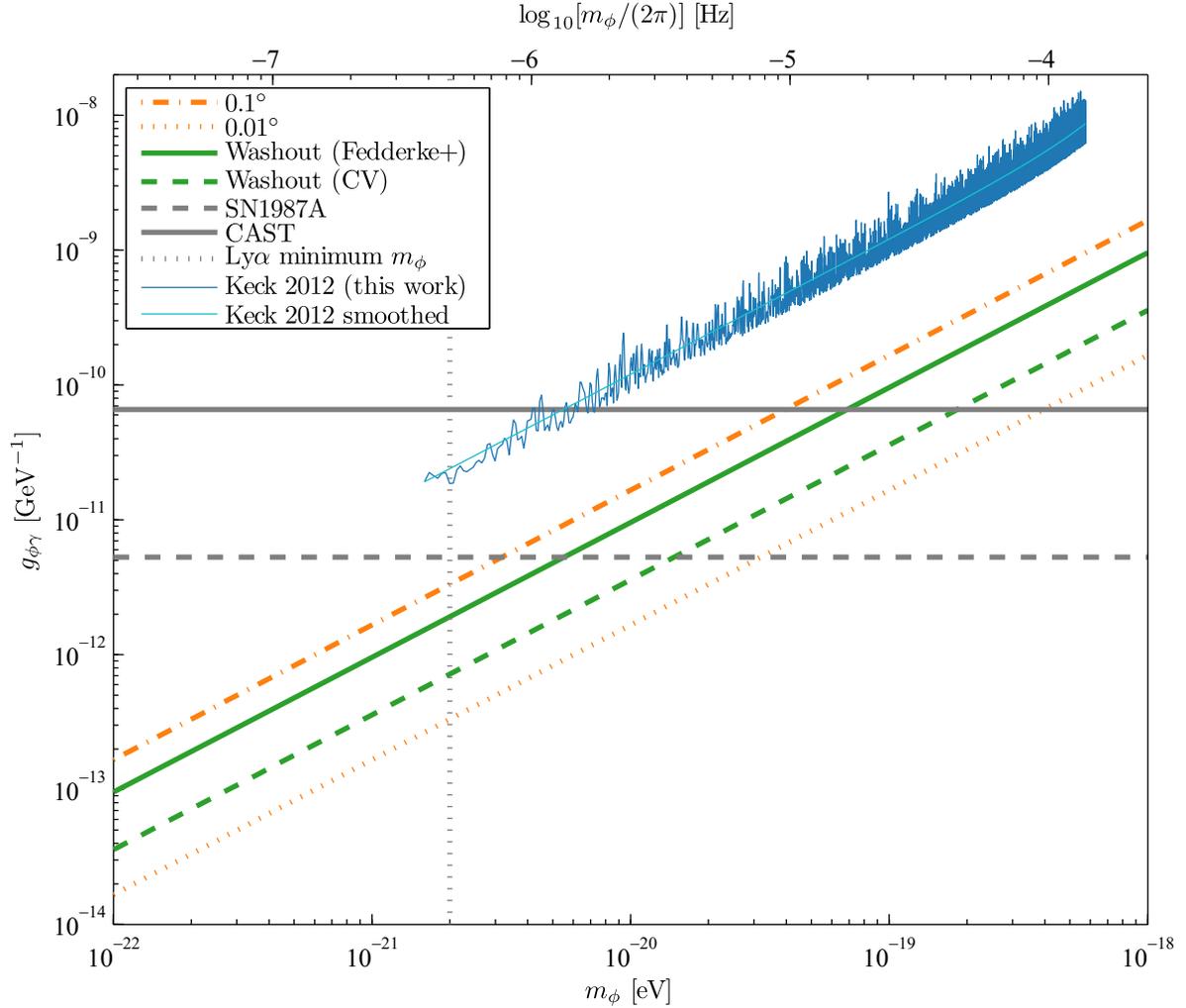


FIG. 6. Excluded regions in the mass-coupling parameter space for axionlike dark matter (cf. Fig. 3 in Ref. [15]). All constraints push the allowed regions to larger masses and smaller coupling constants—i.e., toward the bottom right of the figure. If the dark matter is assumed to consist entirely of axionlike particles—i.e., if  $\kappa = 1$ , then our constraints (blue) are immediately implied by Eq. (4) and the results of Fig. 5. A smoothed approximation is shown in cyan [Eq. (80)]. The orange dot-dashed and dotted lines show the constraints that would be achieved if the rotation amplitude were constrained to  $0.1^\circ$  and  $0.01^\circ$ , respectively. The green solid line shows the constraint set by Fedderke *et al.* [15] by searching for the washout effect (Sec. I) in publicly available *Planck* power spectra. The dashed green line shows the cosmic variance limit for the washout effect. The dashed gray horizontal line shows the limit from searching for a gamma-ray excess from SN1987A [22]. The solid gray horizontal line is the limit set by the CAST experiment [21]. The dotted gray vertical line is a constraint on the minimum axion mass from observations of small-scale structure in the Lyman- $\alpha$  forest [26], though we note that several similar bounds have also been set by other considerations of small-scale structure [27,28].

$$A = g_{\phi\gamma}\phi_0 \quad (79)$$

from Eqs. (3) and (41). The  $m$  dependence of the axion field strength  $\phi_0$  [Eq. (4)] implies that our limits on the coupling constant will roughly follow  $g_{\phi\gamma} \propto m$ . In Fig. 6, we present our constraints on the parameter space of axionlike particles from the 2012 observation season of the *Keck Array*. Combining Eqs. (4), (78), and (79), we can approximate our limits on the coupling constant by

$$g_{\phi\gamma} < (1.2 \times 10^{-11} \text{ GeV}^{-1}) \text{sinc}^{-1} \left( \frac{m}{5.0 \times 10^{-19} \text{ eV}} \right) \times \left( \frac{m}{10^{-21} \text{ eV}} \right) \left( \frac{\kappa\rho_0}{0.3 \text{ GeV/cm}^3} \right)^{-1/2}. \quad (80)$$

For periods greater than 24 hr, which corresponds to  $m < 4.8 \times 10^{-20} \text{ eV}$ , we can convert the median limit from Eq. (77) to

$$g_{\phi\gamma} < (1.1 \times 10^{-11} \text{ GeV}^{-1}) \left( \frac{m}{10^{-21} \text{ eV}} \right) \times \left( \frac{\kappa\rho_0}{0.3 \text{ GeV/cm}^3} \right)^{-1/2}. \quad (81)$$

For comparison, we include in Fig. 6 the constraints from other probes. Our constraints from only 2012 data do not exclude new regions of parameter space, but we note that the time-domain polarization-oscillation observable is distinct from all others, and consequently subject to a different set of possible systematic biases. Furthermore, we emphasize that the 2012 observing season of the *Keck Array* represents only a small subset of the total CMB data collected to date, and that more sensitive observations will be conducted in the future.

## VII. CONCLUSIONS AND OUTLOOK

We have presented a method to search for axionlike polarization oscillations in the CMB, and we have demonstrated the use of this method with data from the 2012 observing season of the *Keck Array*. The search is compatible with the design and operation of experiments targeting primordial  $B$  modes and can be continued by current and future projects with no change to scan strategy or to low-level data processing.

With only 2012 data from the *Keck Array*, we do not exclude any new regions of the parameter space. We note, however, that we have analyzed only a relatively small fraction of the total BICEP dataset. The *Keck Array* observed for eight seasons, and we have in this work analyzed only one season. Additionally, BICEP3 has been observing at 95 GHz since 2015 with more than twice the mapping speed of the entire *Keck Array* [41]. The full BICEP dataset has a survey weight more than an order of magnitude greater than that of the 2012 season.

When more of the BICEP dataset is included in an axion oscillation analysis, we expect improvements in sensitivity for two reasons. The first is a decrease in residual map noise in the template maps  $\bar{Q}(\hat{n})$  and  $\bar{U}(\hat{n})$ . With a better template, we more efficiently extract an oscillationlike signal from the pairmaps. Preliminary investigations indicate that the elimination of residual map noise can improve the per-scanset signal-to-noise ratio by  $\sim 15\%$  for the 150 GHz observations analyzed in this work. For frequencies above  $\sim 200$  GHz, which tend to be significantly noisier due to stronger atmospheric fluctuations, the signal-to-noise ratio can be improved by more than a factor of 2 by using lower-frequency maps as the CMB templates. The second improvement in sensitivity will come from the increased sample size. We have verified through simulations that, when the template maps are held constant, our expected upper limits scale approximately as  $1/\sqrt{n}$ , where  $n$  is the number of scansets included in the analysis. With existing BICEP data, we conservatively anticipate an improvement in upper limits by at least a factor of 3.

Current and future BICEP observations will allow for even more sensitive measurements. We are continuing observations with BICEP3 at 95 GHz. The *BICEP Array* has begun a staged deployment of four new receivers of similar size to BICEP3 [42]. The first receiver, which observes at 30 and 40 GHz, achieved first light in February 2020. The second and third receivers will observe at 150 and 95 GHz, respectively, and the fourth receiver will observe at 220 and 270 GHz.

Additional improvements in sensitivity can be achieved by correlating in the time domain with other CMB experiments. The South Pole Observatory is a formal partnership between the BICEP collaboration and the South Pole Telescope (SPT) Collaboration. The current generation of SPT, which is called SPT-3G [43], has been observing from the South Pole since 2017. While the BICEP dataset has greater integrated polarization sensitivity, SPT has greater angular resolution, and is therefore sensitive to more polarization modes. All else being equal, a higher-resolution CMB experiment is more sensitive to polarization oscillations due to the increased number of modes, though this advantage is less significant for multipoles larger than  $\ell \sim 2000$ , where the CMB anisotropies are suppressed.

The CMB Stage 4 (CMB-S4) project will contain more than an order of magnitude more detectors than any current-generation experiment, and this will provide yet another boost in sensitivity [44,45]. An axion oscillation search imposes few requirements on the design and scan strategy of CMB-S4, since the main elements are nothing more than sensitive, repetitive measurements of CMB polarization. The search is more sensitive at CMB-dominated frequency bands like 95 and 150 GHz, since the global oscillation affects only the CMB component of the polarization field. To take full advantage of the polarization information in the CMB and thereby increase the signal-to-noise ratio,

higher-resolution instruments are preferred, e.g., with aperture diameters of 5–10 m, which allow for sensitivity to polarization modes into the CMB damping tail.

The methods presented in this work can be adapted with relatively minor alterations to analyze data from other CMB polarimetry experiments. Some of our analysis choices take advantage of unique characteristics of the *Keck Array*, and we have attempted to draw attention to those experiment-specific assumptions. As the signal is coherent over large time and length scales (Sec. I), observations from several CMB experiments can be combined to protect against systematics and improve sensitivity.

### ACKNOWLEDGMENTS

We thank Adam J. Anderson, Aviv R. Cukierman, Michael A. Fedderke, and Ethan O. Nadler for useful conversations. The BICEP/*Keck* projects have been made possible through a series of grants from the National Science Foundation including Grants No. 0742818, No. 0742592, No. 1044978, No. 1110087, No. 1145172, No. 1145143, No. 1145248, No. 1639040, No. 1638957, No. 1638978, No. 1638970, and No. 1836010, and by the

Keck Foundation. The development of antenna-coupled detector technology was supported by the Jet Propulsion Laboratory (JPL) Research and Technology Development Fund and Grants No. 06-ARPA206-0040 and No. 10-SAT10-0017 from the NASA Astrophysics Research and Analysis program (APRA) and Strategic Astrophysics Technology (SAT) programs. 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. The computations in this paper were run on the Odyssey cluster supported by the FAS Science Division Research Computing Group at Harvard University. The analysis effort at Stanford and S. L. A. C. was partially supported by the Department of Energy, Contract No. DE-AC02-76SF00515. 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 and Steffen Richter. We thank all those who have contributed past efforts to the BICEP/*Keck* series of experiments, including the BICEP1 team.

- 
- [1] L. Bergström, Non-baryonic dark matter: Observational evidence and detection methods, *Rep. Prog. Phys.* **63**, 793 (2000).
  - [2] G. Bertone, D. Hooper, and J. Silk, Particle dark matter: evidence, candidates and constraints, *Phys. Rep.* **405**, 279 (2005).
  - [3] R. D. Peccei and H. R. Quinn, CP Conservation in the Presence of Pseudoparticles, *Phys. Rev. Lett.* **38**, 1440 (1977).
  - [4] R. D. Peccei and H. R. Quinn, Constraints imposed by CP conservation in the presence of pseudoparticles, *Phys. Rev. D* **16**, 1791 (1977).
  - [5] S. Weinberg, A New Light Boson? *Phys. Rev. Lett.* **40**, 223 (1978).
  - [6] F. Wilczek, Problem of Strong  $p$  and  $t$  Invariance in the Presence of Instantons, *Phys. Rev. Lett.* **40**, 279 (1978).
  - [7] J. Preskill, M. B. Wise, and F. Wilczek, Cosmology of the invisible axion, *Phys. Lett.* **120B**, 127 (1983).
  - [8] L. Abbott and P. Sikivie, A cosmological bound on the invisible axion, *Phys. Lett.* **120B**, 133 (1983).
  - [9] M. Dine and W. Fischler, The not so harmless axion, *Phys. Lett.* **120B**, 137 (1983).
  - [10] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, Ultralight scalars as cosmological dark matter, *Phys. Rev. D* **95**, 043541 (2017).
  - [11] S. M. Carroll, G. B. Field, and R. Jackiw, Limits on a Lorentz- and parity-violating modification of electrodynamics, *Phys. Rev. D* **41**, 1231 (1990).
  - [12] S. M. Carroll and G. B. Field, Einstein equivalence principle and the polarization of radio galaxies, *Phys. Rev. D* **43**, 3789 (1991).
  - [13] D. Harari and P. Sikivie, Effects of a Nambu-Goldstone boson on the polarization of radio galaxies and the cosmic microwave background, *Phys. Lett. B* **289**, 67 (1992).
  - [14] S. M. Carroll, Quintessence and the Rest of the World: Suppressing Long-Range Interactions, *Phys. Rev. Lett.* **81**, 3067 (1998).
  - [15] M. A. Fedderke, P. W. Graham, and S. Rajendran, Axion dark matter detection with CMB polarization, *Phys. Rev. D* **100**, 015040 (2019).
  - [16] A. Caputo, L. Sberna, M. Frías, D. Blas, P. Pani, L. Shao, and W. Yan, Constraints on millicharged dark matter and axionlike particles from timing of radio waves, *Phys. Rev. D* **100**, 063515 (2019).
  - [17] T. Liu, G. Smoot, and Y. Zhao, Detecting axionlike dark matter with linearly polarized pulsar light, *Phys. Rev. D* **101**, 063012 (2020).
  - [18] M. Ivanov, Y. Kovalev, M. Lister, A. Panin, A. Pushkarev, T. Savolainen, and S. Troitsky, Constraining the photon coupling of ultra-light dark-matter axion-like particles by polarization variations of parsec-scale jets in active galaxies, *J. Cosmol. Astropart. Phys.* **02** (2019) 059.
  - [19] T. Fujita, R. Tazaki, and K. Toma, Hunting Axion Dark Matter with Protoplanetary Disk Polarimetry, *Phys. Rev. Lett.* **122**, 191101 (2019).

- [20] A. Basu, J. Goswami, D. J. Schwarz, and Y. Urakawa, Searching for axion-like particles under strong gravitational lenses, [arXiv:2007.01440](#).
- [21] V. Anastassopoulos *et al.*, New CAST limit on the axion-photon interaction, *Nat. Phys.* **13**, 584 (2017).
- [22] A. Payez, C. Evoli, T. Fischer, M. Giannotti, A. Mirizzi, and A. Ringwald, Revisiting the SN1987a gamma-ray limit on ultralight axion-like particles, *J. Cosmol. Astropart. Phys.* **02** (2015) 006.
- [23] C. S. Reynolds, M. D. Marsh, H. R. Russell, A. C. Fabian, R. Smith, F. Tombesi, and S. Veilleux, Astrophysical limits on very light axion-like particles from Chandra grating spectroscopy of NGC 1275, *Astrophys. J.* **890**, 59 (2020).
- [24] M. Libanov and S. Troitsky, On the impact of magnetic-field models in galaxy clusters on constraints on axion-like particles from the lack of irregularities in high-energy spectra of astrophysical sources, *Phys. Lett. B* **802**, 135252 (2020).
- [25] C. Dessert, J. W. Foster, and B. R. Safdi, X-Ray Searches for Axions from Super Star Clusters, *Phys. Rev. Lett.* **125**, 261102 (2020).
- [26] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton, and G. D. Becker, First Constraints on Fuzzy Dark Matter from Lyman- $\alpha$  Forest Data and Hydrodynamical Simulations, *Phys. Rev. Lett.* **119**, 031302 (2017).
- [27] E. O. Nadler, V. Gluscevic, K. K. Boddy, and R. H. Wechsler, Constraints on dark matter microphysics from the Milky Way satellite population, *Astrophys. J. Lett.* **878**, L32 (2019).
- [28] K. Schutz, Subhalo mass function and ultralight bosonic dark matter, *Phys. Rev. D* **101**, 123026 (2020).
- [29] P. A. R. Ade *et al.* (BICEP2 Collaboration), Detection of *B*-Mode Polarization at Degree Angular Scales by BICEP2, *Phys. Rev. Lett.* **112**, 241101 (2014).
- [30] P. A. R. Ade *et al.* (BICEP2 Collaboration), BICEP2. II. Experiment and three-year data set, *Astrophys. J.* **792**, 62 (2014).
- [31] P. A. R. Ade *et al.* (BICEP2, Keck Array, and SPIDER Collaborations), Antenna-coupled TES bolometers used in BICEP2, Keck Array, and spider, *Astrophys. J.* **812**, 176 (2015).
- [32] P. A. J. de Korte, J. Beyer, S. Deiker, G. C. Hilton, K. D. Irwin, M. MacIntosh, S. W. Nam, C. D. Reintsema, L. R. Vale, and M. E. Huber, Time-division superconducting quantum interference device multiplexer for transition-edge sensors, *Rev. Sci. Instrum.* **74**, 3807 (2003).
- [33] M. Kamionkowski, A. Kosowsky, and A. Stebbins, A Probe of Primordial Gravity Waves and Vorticity, *Phys. Rev. Lett.* **78**, 2058 (1997).
- [34] U. Seljak and M. Zaldarriaga, Signature of Gravity Waves in the Polarization of the Microwave Background, *Phys. Rev. Lett.* **78**, 2054 (1997).
- [35] R. Adam *et al.* (Planck Collaboration), Planck 2015 results. I. Overview of products and scientific results, *Astron. Astrophys.* **594**, A1 (2016).
- [36] C. L. Bennett *et al.*, Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Final maps and results, *Astrophys. J. Suppl. Ser.* **208**, 20 (2013).
- [37] M. Zaldarriaga and U. Seljak, Gravitational lensing effect on cosmic microwave background polarization, *Phys. Rev. D* **58**, 023003 (1998).
- [38] P. A. R. Ade *et al.* (BICEP2 and Keck Array Collaboration), Constraints on Primordial Gravitational Waves Using Planck, WMAP, and New BICEP2/Keck Observations through the 2015 Season, *Phys. Rev. Lett.* **121**, 221301 (2018).
- [39] J. P. Kaufman *et al.*, Self-calibration of BICEP1 three-year data and constraints on astrophysical polarization rotation, *Phys. Rev. D* **89**, 062006 (2014).
- [40] P. A. R. Ade *et al.* (Bicep2 Collaboration), Bicep2 III: Instrumental systematics, *Astrophys. J.* **814**, 110 (2015).
- [41] J. H. Kang *et al.*, 2017 upgrade and performance of BICEP3: A 95 GHz refracting telescope for degree-scale CMB polarization, *Proc. SPIE Int. Soc. Opt. Eng.* **10708**, 107082N (2018) [[arXiv:1808.00567](#)].
- [42] A. Schillaci *et al.*, Design and performance of the first BICEP array receiver, *J. Low Temp. Phys.* **199**, 976 (2020).
- [43] A. N. Bender *et al.*, Year two instrument status of the SPT-3G cosmic microwave background receiver, *Proc. SPIE Int. Soc. Opt. Eng.* **10708**, 1070803 (2018) [[arXiv:1809.00036](#)].
- [44] K. Abazajian *et al.*, CMB-S4 Decadal Survey APC White Paper, [arXiv:1908.01062](#).
- [45] K. Abazajian *et al.*, CMB-S4 science case, reference design, and project plan, [arXiv:1907.04473](#).