

This is the accepted manuscript made available via CHORUS. The article has been published as:

## Covariance of CMB anomalies

Jessica Muir, Saroj Adhikari, and Dragan Huterer

Phys. Rev. D **98**, 023521 — Published 17 July 2018

DOI: [10.1103/PhysRevD.98.023521](https://doi.org/10.1103/PhysRevD.98.023521)

# The Covariance of CMB anomalies

Jessica Muir, Saroj Adhikari, and Dragan Huterer

*Department of Physics, University of Michigan, 450 Church St, Ann Arbor, MI 48109-1040,  
Leinweber Center for Theoretical Physics, University of Michigan, 450 Church St, Ann Arbor, MI 48109-1040*

Several unexpected features are observed at large angular scales in the CMB temperature anisotropy measurements by both WMAP and Planck. These include the lack of both variance and correlation, alignment of the lowest multipole moments with one another, hemispherical power asymmetry, and an odd-to-even parity excess. In this work, we study the statistics of eight representative large-angle CMB features in order to evaluate their covariance in the standard  $\Lambda$ CDM model. We do so using two sets of simulated CMB temperature maps; an ensemble of 100,000 simple Gaussian simulations, and 1000 Full Focal Plane (FFP) simulations provided by the Planck collaboration. In measuring feature probabilities, we pay particular attention to analysis choices, making sure that we can reproduce previous results in the literature, and explain differences where appropriate. The covariance structure we find is consistent with expectations given that many of the features studied are functions of the angular power spectrum. Notably, we find significant differences in the covariance entries associated with the quadrupole-octopole alignments derived from the Gaussian and FFP simulations. We additionally perform a principal component analysis to quantitatively gauge what combinations of features capture the most information about how simulation measurements vary, and to provide an alternative assessment of the ways in which the real sky is anomalous. The first four principal components explain about 90% of the simulations' variance, with the first two roughly quantifying the lack of large-angle correlations, and the next two quantifying the phase-dependent anomalies (multipole alignments and power asymmetry). Though the results of this analysis are fairly unsurprising, its comprehensive approach serves to tie together a number of previous results, and will therefore provide context for future studies of large-angle anomalies.

## I. INTRODUCTION

The spectacular maps of the cosmic microwave background (CMB) anisotropy that have been made over the past few decades have revolutionized our understanding of the universe, and rejuvenated efforts to test fundamental processes in the early and late universe. The CMB maps are overall in a very good agreement with the six-parameter spatially flat  $\Lambda$ CDM model specified by the energy densities of dark matter and baryons, the amplitude and spectral index of primordial scalar fluctuations, the reionization optical depth, and the expansion rate (Hubble constant) [1]. Shortly after the WMAP experiment's data were released, however, several surprising coincidences were noticed at large angular scales. In particular, the WMAP maps of temperature anisotropies exhibit low variance, a lack of correlation on the largest angular scales, alignment between various low multipole moments [2], alignment between those low multipole moments and the motion and geometry of the Solar System [3], a hemispherical power asymmetry [4], a preference for odd parity modes [5], and an unexpectedly large cold spot in the Southern hemisphere [6]. Planck data [7] largely confirmed the presence of these features. For a review of the CMB anomalies, see Ref. [8].

While these anomalies have remained an active area of study over the years, it is difficult to draw firm conclusions from the study of features of the CMB at very large angles, mainly due to the significant cosmic variance at those scales. Moreover, the *a posteriori* nature of their observation, as well as the generally good fit of data to the standard cosmological model, means that large-angle

anomalies do not in themselves provide compelling evidence for beyond- $\Lambda$ CDM physics [9].

Rather, in a time where nearly all cosmological observations have been in remarkable agreement with the predictions of  $\Lambda$ CDM, the statistically unlikely large-angle features have attracted attention because of the tantalizing possibility that one or some of them might have cosmological origins [10–13]. If that were the case, due to e.g. an isotropy-breaking mechanism in the early universe, the feature in question could provide insight into the physics of inflation. It is important, however, to additionally consider other explanations for anomalous large-scale CMB features: they could be artifacts of instrumental or astrophysical systematics or they could simply be unlikely fluctuations in the standard isotropic model. Much of the study of large-angle anomalies has thus been focused on disentangling these three logical possibilities: whether large-angle CMB anomalies are cosmological, are due to systematics, or are statistical flukes. Better understanding of the anomalies will be driven in the future by observations of new quantities on very large spatial scales, such as CMB polarization [14–18] and lensing [19], as well as large-scale structure [20]. Whether or not new insights about the early universe become readily apparent, studying large-angle anomalies has and will thus continue to provide an opportunity to build a deeper understanding of our measurements of the large-angle CMB.

One largely unexplored question is how the observed anomalies are related to one another: If we observe one unlikely feature, does that make us less surprised to find another? Roughly speaking, (positively) correlated

anomalies imply a smaller overall joint significance than if they are uncorrelated. Full understanding of the anomalies thus enables an accurate accounting of the likelihood for the joint observation of unexpected features. Since the *a posteriori* nature of the choice of anomaly statistics remains, we would certainly not advocate for using a measure of the combined tension of all observed anomalies as a test of  $\Lambda$ CDM. However, for studies focusing on pairs or groups of large-angle features as tests for phenomenological models of inflation, knowing their joint probability distribution in  $\Lambda$ CDM will prevent double-counting (or, if the anomalies are anti-correlated, under-counting) of the anomalies' significance. Furthermore, the knowing the covariance of large-angle features associated with anomalies allows us to quantitatively separate anomaly 'atoms', i.e. a set of independent features out of which drive the anomalous observations [8].

Previous work on the covariance of CMB temperature anisotropy anomalies has mostly been limited to studying pairs of anomalies. For example in Ref. [21], the authors show that missing power at large scales quantified by  $S_{1/2}$  and the quadrupole-octopole alignment are not correlated in  $\Lambda$ CDM (such a conclusion was also reached, albeit for the full-sky-only analysis, by Ref. [22]). The lack of correlation between hemispherical power asymmetry and the quadrupole-octopole alignment in  $\Lambda$ CDM are demonstrated in Ref. [23]. In Ref. [24], the authors indicate a possible connection between the lack of power and the odd-multipole preference anomaly. In particular, they claim — based on an analytical argument — that the odd-multipole preference can be a phenomenological cause of the lack of large-angle correlation. Additionally, in Ref. [25] the authors find a correlation between the low large-angle power and the low value of the CMB quadrupole. Ref. [26] explores the relationship between low power at large angles and the amplitude of the CMB quadrupole and octopole, while Ref. [27] studies its relationship with the quadrupole and octopole phases. Here we aim to take a more global view by studying the relationship between all of these features simultaneously.

In this work, we will use ensembles of simulated CMB temperature maps to empirically characterize, in the context of  $\Lambda$ CDM, the covariance between a collection of features associated with commonly-studied large-angle anomalies. Our analysis proceeds in three general steps. First, we will measure the quantities associated with those features and confirm that the comparison between our measurements of the real CMB sky and simulations reproduce previous findings. Next, we will study the distribution of the simulation ensembles in the space defined by the “anomaly feature” quantities to find their covariances. In doing so, we will investigate the impact of foregrounds and survey properties by comparing the results obtained from simple Gaussian simulations of the CMB temperature map to the more realistic Planck Full Focal Plane simulations [28]. Finally, we will use the measured feature covariances perform a principal component analysis in order to further characterize the ways in which

large-angle CMB map properties are expected to vary in  $\Lambda$ CDM, and in which the observed CMB sky is unusual. We emphasize that the goal of this analysis is to gain a deeper understanding of the predictions of  $\Lambda$ CDM, rather than to do any explicit model-comparison.

The rest of the paper is organized as follows. In Sec. II we introduce our methods in detail, including the description of maps and masks adopted, of our simulation ensembles and of our power spectrum measurements. In Sec. III, we outline the statistical description of the eight large-scale features that we study in this work and report the statistical significance of the features with respect to our simulation ensembles. The main results of our work — the measurement of covariances between the anomalies — is presented in Sec. IV, along with discussion of a PCA analysis. We summarize and conclude in Sec. V.

## II. METHODS

We begin by introducing basic terminology and notation for describing CMB anisotropies. Temperature fluctuations can be expanded in a spherical harmonic series as follows

$$T(\hat{\mathbf{n}}) = \sum_{\ell} \sum_{m=-\ell}^{\ell} a_{\ell m} Y_{\ell m}(\hat{\mathbf{n}}), \quad (1)$$

where  $\hat{\mathbf{n}}$  is the direction on the sky and the complex coefficients  $a_{\ell m}$  contain all information about the temperature field. For statistically isotropic fluctuations, the expectation of the two-point correlation between coefficients  $a_{\ell m}$  drastically simplifies and only depends on  $\ell$ :

$$\langle a_{\ell m} a_{\ell' m'}^* \rangle = \delta_{\ell \ell'} \delta_{m m'} C_{\ell}. \quad (2)$$

If the fluctuations are statistically isotropic and Gaussian, the angular power spectrum,  $C_{\ell}$ , contains all statistical information about the temperature field.

It is useful to additionally define the real-space angular correlation function as

$$C(\theta) = \langle T(\hat{\mathbf{n}}_1) T(\hat{\mathbf{n}}_2) \rangle \quad (3)$$

$$= \frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_{\ell} P_{\ell}(\cos \theta) \quad (4)$$

where  $\hat{\mathbf{n}}_1 \cdot \hat{\mathbf{n}}_2 = \cos \theta$  and  $P_{\ell}(x)$  is a Legendre polynomial of order  $\ell$ .

### A. Maps and masks

In the course of this analysis we use several data products from the Planck 2015 data release, provided on the Planck Legacy Archive<sup>1</sup>. For transparency and reproducibility, when relevant we identify the names of specific files used in footnotes.

<sup>1</sup> [pla.esac.esa.int](http://pla.esac.esa.int)

Though the primary product of this paper will be a study of the covariance between anomalies as measured from simulation ensembles, we also use Planck data to quantify the values of the anomaly statistics described above for the real, observed CMB sky. Our purpose in doing this will be twofold. First, it will allow us to compare our assessment of the anomalousness of features of the observed CMB sky against the probabilities reported in the literature. Additionally, measuring the same statistics from the real Planck data as from our simulations will allow us to place the real CMB sky in the multi-dimensional feature space examined in Sec. III and IV.

For map-based statistics we use the SMICA [29] map<sup>2</sup> from the 2015 Planck data release. Though the Commander map should more properly be used for the analysis of very large scale features, past studies [7, 30, 31] have found that the significance of the various anomalies does not depend strongly on which component separation method is used. Therefore, using the SMICA map should be sufficient for our purposes.

Because we care only about large-angle features, we work with maps that are at a resolution of  $N_{\text{SIDE}} = 64$ , smoothed with a Gaussian beam of 160 arcmin. We downgrade the SMICA CMB temperature map, which is provided at  $N_{\text{SIDE}} = 1024$ , following the prescription described in Ref. [30]. We do this by first extracting its spherical components  $a_{\ell m}$  using the `healpy`<sup>3</sup> [32] function `map2alm`. Then, again using `healpy`, we get the harmonic space representation of the Gaussian beam  $b_\ell$  and pixel window functions  $p_\ell$  corresponding to the full width half maximum (FWHM) and pixel resolution, respectively, of both the input and output maps. By combining these together, we obtain the downgraded harmonic coefficients,

$$a_{\ell m}^{\text{out}} = \frac{b_\ell^{(\text{out})} p_\ell^{(\text{out})}}{b_\ell^{(\text{in})} p_\ell^{(\text{in})}} a_{\ell m}^{\text{in}}. \quad (5)$$

We then use the `healpy` function `alm2map` to convert back to pixel space, obtaining the downgraded map. We refer to Table 1 in Ref. [30] for the appropriate beam FWHM values: 160 arcmin for  $N_{\text{SIDE}} = 64$ , and 10 arcmin for  $N_{\text{SIDE}} = 1024$ . (For other parts of this study we will also use the conversions 5 arcmin for  $N_{\text{SIDE}} = 2048$  and 640 arcmin for  $N_{\text{SIDE}} = 16$ .)

When we use a mask, we adopt the UT78 common mask<sup>4</sup>, which is identified in Ref. [7] as the mask that should be used for the analysis of Planck temperature maps. UT78 is the union of the masks for Planck’s four component-separation methods (SMICA,

NILC, SEVEM, and Commander). This mask is provided as a map of zeros and ones at resolution  $N_{\text{SIDE}} = 2048$ , where zeros represent masked pixels and ones signify unmasked pixels. To downgrade the mask to  $N_{\text{SIDE}} = 64$ , we follow the procedure described in Eq. (5), then threshold the resulting map so that all pixels with a value  $\leq 0.9$  are marked as masked. This reduces the mask from its original  $f_{\text{sky}} = 0.78$  to 0.67. (When we use  $N_{\text{SIDE}} = 16$  maps for one of the anomalies studied below, the UT78 mask’s sky coverage reduces further to  $f_{\text{sky}} = 0.58$ .)

## B. Simulated ensembles

Our primary simulation ensemble will be a set of 100,000 noiseless Gaussian CMB temperature maps generated using the `synfast` function in `healpy`. Gaussian temperature map realizations are drawn using the Planck best-fit theory prediction for the power spectrum.<sup>5</sup> The maps are produced at  $N_{\text{SIDE}} = 64$  with FWHM=160 arcmin Gaussian smoothing, and with the `pixwin` argument set to `True`. These settings were chosen to make the simulated maps have properties consistent with the downgraded SMICA temperature maps described above. These straightforward-to-implement simulations, which we will henceforth refer to as the “synfast simulations,” will allow us to obtain the statistics of fluctuations associated with the CMB signal only.

In order to explore whether foreground and survey-related effects that are present in the Planck SMICA map influence the relationship between anomalies, we repeat our analysis on the publicly available ensemble of Planck Full Focal Plane (FFP) simulations [28]. Specifically, we use the FFP8.1 CMB sky and noise maps that have been processed using the SMICA component separation pipeline, which we add together before downgrading to  $N_{\text{SIDE}} = 64$ . The FFP simulations include the physical effects of astrophysical foregrounds, gravitational lensing, Doppler modulation, and frequency-dependent Rayleigh scattering effects. They also model the Planck mission’s scanning strategy, detector response, beam shape, and data reduction pipeline. Additionally, a small, frequency-dependent intensity quadrupole has been added to the FFP simulations to account for an uncorrected residual in the data from the dipole-induced Doppler quadrupole identified in Ref. [33]. Note that because we use the FFP8.1 rather than FFP8 simulations, we do not need to rescale the CMB components of the simulations by the factor of 1.0134 that was applied in Ref. [7].

<sup>2</sup> COM\_CMB\_IQU-smica.1024.R2.02.full.fits

<sup>3</sup> `healpy` is the Python implementation of HEALPix, described at <http://healpix.sourceforge.net>.

<sup>4</sup> COM\_Mask\_CMB-IQU-common-field-MaskInt.2048.R2.01.fits, field 0.

<sup>5</sup> COM\_PowerSpect\_CMB-base-plikHM-TT-lowTEB-minimum...-theory.R2.02.txt

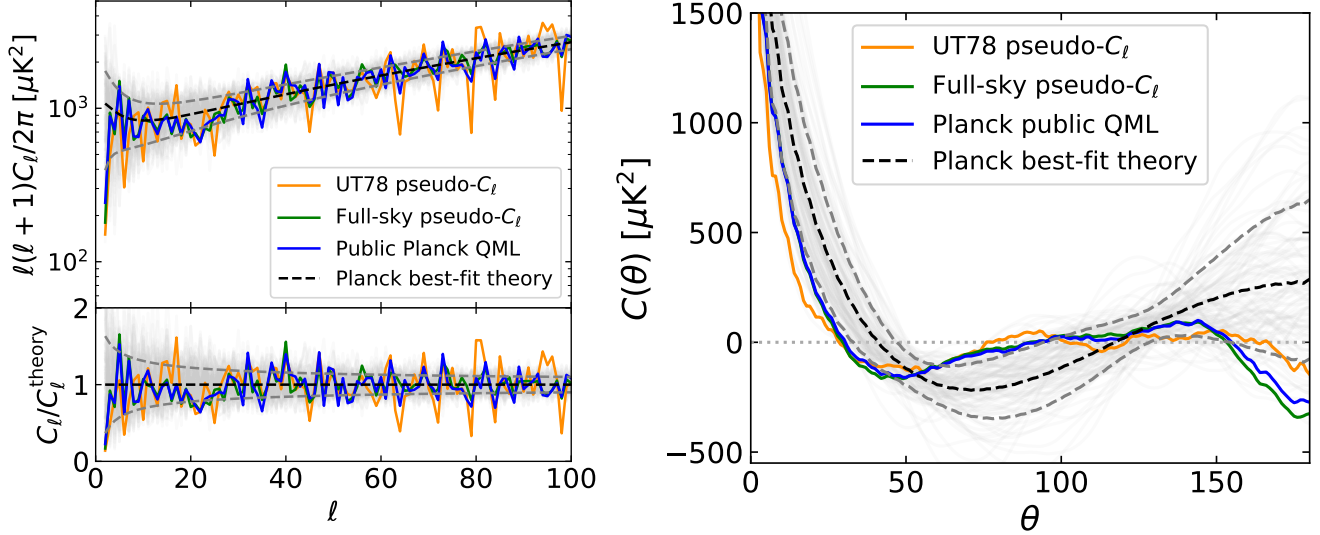


FIG. 1. Comparison of the different angular power spectrum measurements described in Sec. II C (left) and their corresponding angular correlation functions (right). The collection of grey lines behind them are from the full-sky  $C_\ell$  measurements of the first 100 synfast simulations. The black dotted line shows theoretical expectation, and the gray dotted lines show the 68% confidence level cosmic-variance errors.

### C. Power spectrum measurements

There are several methods that one can use to measure the angular power spectrum of a map of CMB temperature fluctuations. Using different methods generally will cause variations in the estimate for  $C_\ell$ , and thus the choice of how to measure  $C_\ell$  can impact anomaly statistics.

Our analysis mainly relies on pseudo- $C_\ell$  estimates for the power spectrum. For full-sky measurements these will give unbiased estimates of the true power spectrum simply by averaging the observed spherical harmonics,

$$\hat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=-\ell}^{m=+\ell} |a_{\ell m}|^2. \quad (6)$$

If a mask is used to remove contaminated parts of the map, additional care must be taken, as is described in e.g. Ref. [31]. In practice we measure cut-sky pseudo- $C_\ell$ 's using the `polspice` algorithm<sup>6</sup> [34], which removes the monopole and dipole of the masked map, measures the angular correlation function  $C(\theta)$  of the unmasked part of the sky, and then integrates to obtain  $C_\ell$ ,

$$C_\ell = 2\pi \int_{-1}^1 C(\theta) P_\ell(\cos \theta) d \cos \theta. \quad (7)$$

In contrast to these cut-sky pseudo- $C_\ell$ 's, which only estimate the statistical properties of the unmasked parts of the sky, quadratic maximum likelihood (QML) methods can be used to estimate the statistical properties of the entire sky. QML power spectrum estimators are unbiased, and have a smaller variance than pseudo- $C_\ell$  estimators [35]. It will make the most sense for us to study the statistical properties of certain large-angle features in terms of the observed SMICA map's QML power spectrum. For this, we do not implement our own QML power spectrum estimator, but instead use the public QML spectrum provided by the Planck team [36]. The Planck QML power spectrum was obtained using a Blackwell-Rao estimator [37] applied to the Commander component-separated map (and mask) at NSIDE = 16 for multipoles  $\ell = 2 - 29$ , and the `Plik` likelihood applied to measured pseudo- $C_\ell$ 's for  $\ell \geq 30$ . The low- $\ell$  power spectrum estimation uses the Commander mask, which has  $f_{\text{sky}} = 0.94$  and therefore leaves available much more of the sky available for analysis than the UT78 common mask. The high- $\ell$  power spectrum likelihood uses galactic masks, described in Appendix A of Ref. [36], which leave less available sky than those used with component-separated CMB maps.

To summarize, the three  $C_\ell$  measurement strategies we will examine are:

- **Full-sky  $C_\ell$ 's:** We computed them using `polspice` based on a map with NSIDE = 64, with the monopole and dipole subtracted.
- **UT78 pseudo- $C_\ell$ 's:** We computed them using `polspice` with the same settings as the full-sky

<sup>6</sup> The `polspice` software can be found at <http://www2.iap.fr/users/hivon/software/PolSpice/>. we run it using the settings `subav=YES`, `subdipole=YES`, `apodizesigma=NO`, and `pixelfile=NO`.

case, using the NSIDE = 64 version of the Planck UT78 common mask.

- **Planck public QML  $C_\ell$ 's:** These estimates are provided for low<sup>7</sup> and high<sup>8</sup>- $\ell$  on the Planck Legacy archive.

In Fig. 1 we compare the angular power spectra derived from the SMICA CMB temperature map using these three methods, as well as the corresponding angular correlation functions derived using Eq. (3). We also show theoretical predictions using the Planck best-fit model, along with the 68% confidence region for cosmic variance.

Our NSIDE = 64 resolution implies that we can study multipoles up to  $\ell_{\max} = 3\text{NSIDE} - 1 = 191$  (though for practical purposes, pixelization effects become apparent for pseudo- $C_\ell$  measurements at  $\ell \sim 150$  for full-sky measurements and  $\ell \sim 100$  for cut-sky). Since we will be focusing on scales  $\ell < 100$ , this choice of NSIDE is sufficiently high.

### III. FEATURES STUDIED

Here we study eight characteristics of the large-angle CMB temperature maps. These features, which are summarized in Table I, include some of the most prominently discussed CMB anomalies. This set of features is not intended to be comprehensive<sup>9</sup> Rather, our intention is to focus on a representative sample that will allow us to develop an understanding of the large-angle CMB's statistical properties in  $\Lambda$ CDM.

Broadly, we classify features based on whether they depend entirely on information in the (isotropic) two-point statistics, or whether they require map or  $a_{\ell m}$ -based information. We adopt this classification to aid in our interpretation of their covariances: because the two-point function anomalies are all functions of the same angular power spectrum, their respective definitions directly imply some *a priori* expectations for their covariances. The same is not necessarily true for the isotropy-breaking anomalies, due to the stochastic nature of the  $a_{\ell m}$  coefficients.

As we introduce each feature, we will define the quantity that we use to measure it, briefly introduce relevant findings from previous studies, and discuss how those findings compare to our measurements. We will mainly perform these comparisons against Ref. [7], the Planck 2015 paper on the isotropy and statistics of the CMB, which we will henceforth refer to as Planck XVI (I&S).

Unless otherwise noted, the anomaly measurements in that work were done using a QML  $C_\ell$  estimator on UT78 cut-sky maps evaluated at NSIDE = 64. Their measurements of the real sky were done on the SMICA temperature map, and they evaluated statistics based on the FFP8 simulations. Note that because we use a different ensemble of simulations (FFP8.1 rather than FFP8), as well as a different power spectrum measurement technique, we expect our findings for anomaly statistics to be similar to the Planck XVI (I&S) results, but not necessarily to exactly match them.

Following Planck XVI (I&S), we will quantify how unusual (or not) the SMICA temperature map appears compared to simulations using  $p$ -values, defined to be equal to the fraction of simulations that return more extreme values than the real sky. As part of each feature description, we will note whether and how measurement choices (between Planck QML vs. pseudo- $C_\ell$ , cut-sky vs. full-sky) affect those statistics, and will take care to identify which of those choices are used in our anomaly covariance studies presented in the next Section (IV). These single-feature results are summarized in Fig. 3, which is described in more detail in Sec. III C.

#### A. Features depending on two-point functions only

We first study the six features that are fundamentally a function of the angular clustering power.

##### 1. $S_{1/2}$ : Large-angle power

First, we measure power in large angular scales of the temperature map using the  $S_{1/2}$  statistic, defined as the integral of the square of the angular correlation function  $C(\theta)$  over angles between  $60^\circ$  and  $180^\circ$  [40]

$$S_{1/2} = \int_{-1}^{1/2} [C(\theta)]^2 d(\cos \theta). \quad (8)$$

It measures the deviation of  $C(\theta)$  from zero at angles greater than  $60^\circ$ . The inclusion of this statistic is motivated by the lack of power at large angular scales  $\theta \gtrsim 60^\circ$  first observed by COBE [41], and later confirmed by WMAP [40] and Planck [7, 31]. Though there are several ways of quantifying this lack of large-angle correlation, we adopt  $S_{1/2}$  because it is the most commonly studied in the literature.

In practice, to measure  $S_{1/2}$  for a temperature map, we first measure the angular power spectrum and then calculate it in harmonic space via [26]

$$S_{1/2} = \frac{1}{(4\pi)^2} \sum_{\ell, \ell'} (2\ell + 1)(2\ell' + 1) C_\ell I_{\ell, \ell'} \left(\frac{1}{2}\right) C_{\ell'}. \quad (9)$$

Here the matrix  $I_{\ell, \ell'}$  is defined as

$$I_{\ell, \ell'}(x) = \int_{-1}^x P_\ell(x') P_{\ell'}(x') dx', \quad (10)$$

<sup>7</sup> COM\_PowerSpect\_CMB-TT-loL-full\_R2.02.txt

<sup>8</sup> COM\_PowerSpect\_CMB-TT-hiL-full\_R2.02.txt

<sup>9</sup> For example, we do not include the statistics from Ref. [38] which quantify alignments between different combinations of even and odd parity multipoles. We also do not include the ‘‘cold spot’’ [39] which, being a localized feature at smaller scales, does not naturally belong to the set of large-angle features studied here.

Depends on	Quantity	Description	Multipoles	Section
Two-point functions only	$S_{1/2}$	Amount of angular power at $\theta > 60^\circ$	2 – 100	III A 1
	$C_2$	Quadrupole amplitude	2	III A 2
	$C_3$	Octopole amplitude	3	III A 3
	$\sigma_{16}^2$	Variance of temperature fluctuations at $N_{\text{side}} = 16$	2 – 47	III A 4
	$R_{27}$	Ratio of power between even and odd multipoles	2 – 27	III A 5
	$C(\pi)$	Angular correlation at $\theta = 180^\circ$	2 – 191	III A 6
Phases of $a_{\ell m}$	$S_{\text{QO}}$	Quadrupole-octopole alignment	2, 3	III B 1
	$A_{\text{LV}}$	Hemispherical power asymmetry	2 – 191	III B 2

TABLE I. Summary of quantities studied in this work.

but is in practice computed using the recursion relation in Appendix A of Ref. [26]. We sum over values  $\ell = 2 - 100$ .

When analyzing the SMICA map at  $N_{\text{SIDE}} = 64$  with the UT78 mask, Planck XVI (I&S) reports a low value for  $S_{1/2}$  with a lower tail probability<sup>10</sup> of  $p = 0.4\%$ . That means they find that only 0.4% of simulations have a lower value of  $S_{1/2}$  than the SMICA map. Our cut-sky  $S_{1/2}$  measurements give similar probabilities:  $p = 0.7\%$  for the fiducial synfast simulations and  $p = 0.5\%$  for the FFP simulations.<sup>11</sup>

The lower-tail probability of the observed sky's  $S_{1/2}$  value depends dramatically on the method used to measure the angular power spectrum, increasing to 8% for full-sky  $C_\ell$ 's and to 6% for Planck public QML  $C_\ell$ 's (which, recall, effectively reconstruct the full-sky anisotropy field). This is consistent with results from previous studies [26, 31, 42] which have shown that the relatively small amount of (non-zero) correlations on the full sky are dominated by contributions from pixels close to the galactic mask.

## 2. $C_2$ : Quadrupole amplitude

We additionally study  $C_2$ , the quadrupole of temperature fluctuations, which was first found to be low in COBE [43] data, and later in WMAP [2, 44] and Planck [36]. Analyses have shown that the lowness of the quadrupole is not particularly significant [37, 45, 46], so its value or lower-tail probabilities are not generally reported explicitly in the literature. Given this, we do not directly compare our measurement of  $C_2$  to previous results, but do include it as one of our statistics in order

to study its covariance with the low angular power at large angles and other features.

Our one-dimensional study of  $C_2$ 's statistics reflect the findings in the literature. Our fiducial choice for the quadrupole is to adopt the Planck QML  $C_2$  to represent the observed value, while for the simulations we use the quadrupole from the full-sky  $C_\ell$ 's. In order to make the simulation measurements more directly comparable to the QML power spectrum, we apply a correction for the  $N_{\text{SIDE}} = 64$  maps' beam and pixel window functions via

$$C_\ell = (b_\ell^{(64)} p_\ell^{(64)})^{-2} C_\ell^{\text{polspice}}. \quad (11)$$

Here,  $b_\ell^{(64)}$  and  $p_\ell^{(64)}$  are the harmonic components of the beam and pixel window functions for the  $N_{\text{SIDE}} = 64$  input map.

We find that  $C_2$  has a lower-tail probability of 5% using the synfast simulations, and 6% compared to the FFP simulations. Pseudo- $C_\ell$  measurements of the SMICA map give slightly lower probabilities, with  $p \sim 2\%$  for either full- or cut-sky measurements.

## 3. $C_3$ : Octopole amplitude

Though the value of the observed CMB temperature octopole amplitude is not anomalous (see e.g. [46]), we also include it in our study because its behavior in relation to other features has the potential to be interesting. For example, Ref. [27] points out that contributions from the quadrupole and octopole seem to be canceling the power from the rest of the sky, and that a measure of large-angle power becomes less anomalous when their contributions to the correlation function are removed. Additionally, Ref. [26] finds that the relationship between several of the lowest multipoles, certainly more than the just the quadrupole, is responsible for the low observed  $S_{1/2}$ . Given this, we include  $C_3$  in our analysis because the relationship between  $C_3$ ,  $C_2$  and  $S_{1/2}$  may reveal some interesting structure.

We perform our fiducial measurement of  $C_3$  in the same way as for  $C_2$ : we use the Planck QML  $C_\ell$ 's for the observed temperature map, and the beam-and-pixelization-

<sup>10</sup> Value from Table 13 of Ref. [7]

<sup>11</sup> These  $p$ -values are weakly sensitive to whether the  $C_\ell$ 's are corrected for resolution according to Eq. (11): with that correction, the  $p$ -values for  $S_{1/2}$  go down to 0.6% for the synfast simulations and to 0.4% for the FFP simulations. We opt not to make that correction when computing  $S_{1/2}$  and  $C(\pi)$  because doing so introduces significant noise contributions at high multipoles and makes the sums involved overly sensitive to our choice of  $\ell_{\text{max}}$ .

corrected (according to Eq. (11)) full-sky  $C_\ell$  measurements from simulations. Compared to the synfast and FFP simulations, the  $p$ -values for both the QML and full-sky SMICA measurements are 47 – 49%, while the cut-sky octopole is lower, with  $p \sim 15\%$ .

#### 4. $\sigma_{16}^2$ : Variance at NSIDE = 16

We study another indicator of large-angle power via  $\sigma_{16}^2$ , the variance of unmasked pixels of a low resolution, NSIDE = 16 temperature map. The variance of CMB temperature maps, especially at low spatial resolution, has been observed to be anomalously low in analyses of both WMAP [47, 48] and Planck [7] data. Planck measured the variance of unmasked pixel values with various resolutions, and found that the lowest investigated value of  $N_{\text{side}} = 16$ , with a  $p$ -value<sup>12</sup> of 0.5%, produced the most anomalously low variance. They found that the variance tends to become lower as the mask is extended to cover more of the sky, and that the statistical significance of its lowness persists when different foreground subtraction methods are applied.

To measure  $\sigma_{16}^2$  for a given CMB temperature map, we first downgrade the map from NSIDE = 64 to NSIDE = 16. We also downgrade the UT78 mask, but go directly from the original NSIDE = 2048 resolution to NSIDE = 16 in order to make the resulting sky fraction consistent with that used in the Planck study. We then simply compute the variance of all unmasked pixels.

Though we measure  $\sigma_{16}^2$  through a pixel-based method, given an angular power spectrum  $C_\ell$  we can predict its expectation value for full-sky measurements via

$$\langle \sigma_{16}^2 \rangle (C_\ell) = \frac{1}{4\pi} \sum_{\ell} (2\ell + 1) C_\ell (b_\ell^{(16)} p_\ell^{(16)})^2, \quad (12)$$

where  $b_\ell^{(16)}$  and  $p_\ell^{(16)}$  are the beam and pixel window functions corresponding to NSIDE = 16. This expression will allow us to compare our map-based measurements of  $\sigma_{16}^2$  to the predictions from the Planck best fit theory  $C_\ell$ 's as well as the Planck public QML  $C_\ell$ 's.

It is worth noting that our method of measuring  $\sigma_{16}^2$  is different from that used to quantify map variance in the WMAP and Planck analyses. Those analyses use an estimator [47] to isolate the cosmological contribution to the variance of a normalized version of the temperature map, in which each pixel value has been divided by its expected dispersion from both cosmological temperature fluctuations and noise. Because of this, our reported numbers for  $\sigma_{16}^2$  will be much larger than the normalized variances reported in Planck XVI (I&S). Nonetheless, the statistical distribution of variances should be similar, to

the extent that noise contributions to variance can be approximated as direction-independent.<sup>13</sup>

For measurements of  $\sigma_{16}^2$  we would like to exclude pixels that may contain residual foregrounds, so we focus on its cut-sky value for both from the SMICA map and simulations. We find the SMICA  $\sigma_{16}^2$  to be low compared to simulations, with single-tail probability of  $p = 0.8\%$  and  $0.5\%$  for the **synfast** and FFP simulations, respectively. Thus, our cut-sky FFP  $p$ -value exactly matches that in Planck XVI (I&S). The  $\sigma_{16}^2$  expectation value from the Planck public QML  $C_\ell$ 's and our full-sky  $C_\ell$  measurements are very similar, with  $p = 20\%$  when compared to either simulation ensemble.

#### 5. $R_{27}$ : Parity asymmetry at low $\ell$

We use the statistic  $R_{27}$  to quantify large-angle parity asymmetry of the CMB temperature map. It has been noted that, at low  $\ell$ , the CMB maps have more power in odd multipoles than even. This was observed in the WMAP 3, 5, and 7 year data [49–51] as well as in Planck [7]. We quantify this asymmetry using the same estimator as Planck XVI (I&S),

$$R_{\ell_{\text{max}}} = \frac{D_+(\ell_{\text{max}})}{D_-(\ell_{\text{max}})} \quad (13)$$

where

$$D_{+,-} = \frac{1}{\ell_{\text{tot}}^{+,-}} \sum_{\ell=2, \ell_{\text{max}}}^{+,-} \frac{\ell(\ell+1)}{2\pi} C_\ell, \quad (14)$$

and the plus and minus indicate sums over even (parity-symmetric) and odd (parity-antisymmetric) multipoles, respectively. The  $R_{\ell_{\text{max}}}$  statistic is therefore a ratio of the parity-even over parity-odd multipole band-powers. The factor of  $\ell(\ell+1)/(2\pi)$  is used because the theoretical prediction for  $\ell(\ell+1)(2\pi)^{-1}C_\ell$  is approximately scale-independent out to multipoles of  $\ell \lesssim 50$ , and thus gives  $R_{\ell_{\text{max}}} \sim 1$  over that range.

Because  $R_{\ell_{\text{max}}}$  is directly based on the power spectrum, we will focus on its measurements from the Planck QML power spectrum, and compare them to full-sky  $C_\ell$  measurements in simulations. As in the case of  $C_2$  and  $C_3$ , we correct for the impact of the simulations' resolution on the power spectrum using Eq. (11).

For our covariance studies, we will focus on the behavior for  $\ell_{\text{max}} = 27$ , as that multipole range gives the most anomalously low value of  $R_{\ell_{\text{max}}}$  in the Planck XVI (I&S)

<sup>12</sup> This value is taken from Table 12 of Ref. [7].

<sup>13</sup> In principle the noise dispersion can vary with position on the sky due to beam effects and weights used to construct component separated maps, but those effects are expected to be small [48].



analysis<sup>14</sup>, with a single-tail probability  $p = 0.2\%$  for the SMICA map. We find the SMICA map's  $R_{27}$  to be notably less anomalous: measurements of the Planck QML power spectrum give single tail probabilities of  $p = 3\%$  and  $2\%$  when compared to the **synfast** and FFP simulations, respectively.

Given this  $p$ -value discrepancy, we investigated how  $R_{\ell_{\max}}$  depends on  $\ell_{\max}$  and the power spectrum measurement technique. Results of this investigation, shown in Fig. 2, reveal that the significance of parity asymmetry heavily depends on the choice of mask and power spectrum measurement method. The discrepancy between our results and those of Planck XVI (I&S) can be explained by two main differences [52]. The first and primary cause is that the Planck XVI (I&S) measurements of  $R_{\ell_{\max}}$  for the real sky are based on QML  $C_\ell$  measurements of the UT78 cut-sky map, degraded to  $N_{\text{SIDE}} = 32$ , which leaves  $f_{\text{sky}} = 0.64$  unmasked. This has a smaller sky fraction than the maps used to produce the Planck public QML power spectrum, which was obtained using the Commander  $\chi^2$ -based LM93 mask with  $f_{\text{sky}} = 0.93$  [53]. The other difference is that our simulation measurements are based on pseudo- $C_\ell$  measurements of the full sky, while those in Planck XVI (I&S) use QML simulation measurements of the cut sky, but we find that this has little impact on the probability distribution for  $R_{\ell_{\max}}$ . We confirmed that we were able to replicate the Planck results when using their UT78 QML  $C_\ell$  values, which are an intermediate data product of their analysis, but restrict all results presented in this paper to only publicly available data.

We obtain a  $p$ -value closest to that reported for  $R_{27}$  in Planck XVI (I&S) using full-sky  $C_\ell$  measurement of the SMICA map compared to FFP simulations ( $0.6\%$ ). However, in order to be consistent with our treatment of the other purely power-spectrum based features, we will use the public Planck QML  $C_\ell$ 's compared to full-sky  $C_\ell$  simulation measurements as our fiducial choices for measuring  $R_{27}$ .

#### 6. $C(\pi)$ : Two-point correlation at $\theta = 180^\circ$

We next consider the angular correlation function of CMB temperature evaluated at  $180^\circ$ , which we will refer to as  $C(\pi)$ . We include it in the hope that it will help clarify the relationship between other features. Our motivation comes from the fact that  $C(\theta)$ , which is otherwise fairly flat at large angles, drops to negative values at  $\theta \simeq \pi$ . This dip has been observed in both WMAP and Planck data, and can be seen in the colored curves on the

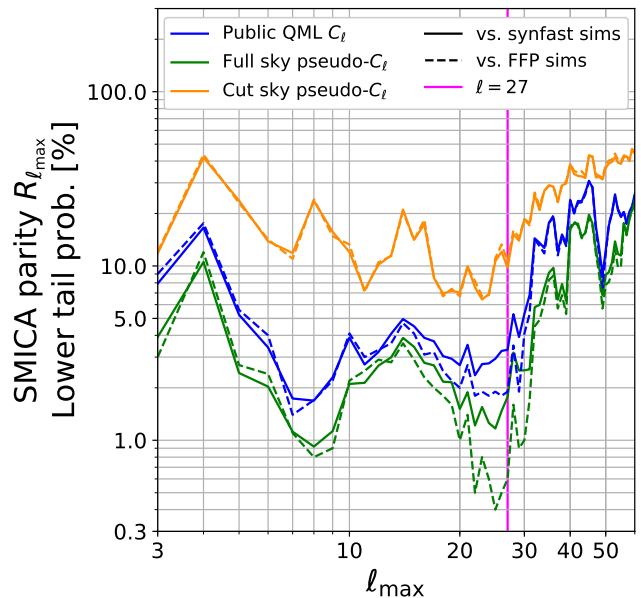


FIG. 2. Plot of the lower-tail probability  $p$  for the parity asymmetry as a function of the largest multipole  $\ell_{\max}$  considered, for various SMICA power spectrum measurement and simulation ensemble combinations. Solid lines show the probabilities for the SMICA map assessed relative to the synfast simulations, and dashed lines show them relative to the FFP simulations. For the QML and full-sky  $C_\ell$  SMICA measurements, the simulations are measured using full-sky  $C_\ell$ 's. For the cut-sky pseudo- $C_\ell$  SMICA measurements, cut-sky pseudo- $C_\ell$  simulation measurements are used. The vertical line denotes the  $\ell_{\max}$  value at which where Planck XVI (I&S) found the most anomalous parity statistic; see text for details.

right-hand side of Fig. 1. By its definition we expect the value  $C(\pi)$  to be related to the missing large-angle correlations statistic  $S_{1/2}$ , as well as to the  $R_{27}$  measurement of parity asymmetry. This can be seen by comparing the definition for  $R_{\ell_{\max}}$  in Eq. (13) to

$$C(\pi) = \sum_{\ell=2}^{\infty} (-1)^\ell \frac{2\ell+1}{4\pi} C_\ell. \quad (15)$$

We will of course investigate these correlations quantitatively further below.

We measure  $C(\pi)$  by using cut-sky pseudo- $C_\ell$  measurements of the SMICA map and simulations to obtain measured power spectra (as we do for  $S_{1/2}$ ), transforming to real space angular correlations via Eq. (15). In practice, we compute the sum over multipoles  $\ell = 2 - 100$ . For the cut-sky measurements, the SMICA map's  $C(\pi)$  is lower than for the majority of simulations, but is not particularly anomalous, with  $p \sim 11\%$  for both the synfast and FFP simulation ensembles. The  $p$ -value goes down to  $5 - 6\%$  if we instead compare SMICA map measurements using the public Planck QML  $C_\ell$ 's to the same set of simulation measurements, and  $3 - 4\%$  for full-sky

<sup>14</sup> Value taken from text associated with Fig. 20 of Ref [7]. Though that text actually reports  $\ell = 28$  to give the lowest  $R_{\ell_{\max}}$   $p$ -value, this is due to a typographical error, and we confirmed with that section's author that the minimum  $p$ -value is actually at  $\ell = 27$ . [52]

$C_\ell$ 's. As this measure is not commonly studied in the literature, we do not compare this to any reported values.

## B. Features depending on $a_{\ell m}$ phases

We also consider two features that depend on the phases of CMB temperature harmonic coefficients  $a_{\ell m}$  — i.e. which cannot be measured solely from the two-point correlations of the maps. These features are associated with reported anomalies that could indicate a possible departure from the assumption of statistical isotropy of CMB temperature fluctuations: the quadrupole-octopole alignment and the hemispherical power asymmetry.

### 1. $S_{\text{QO}}$ : Quadrupole-octopole alignment

The CMB temperature's quadrupole and octopole were first observed to be planar and aligned in Ref. [2]. There are a number of possible ways to denote the directionality, and thus alignment, of multipoles. To do so, we will follow the approach presented in Ref. [54] and make use of Maxwell multipole vectors. Multipole vectors are a representation of a function on a sphere; while they are at some level equivalent to spherical harmonics in that role, their relationship to the  $a_{\ell m}$  is highly nonlinear in a way that makes them particularly well suited to studying the directionality of patterns on the sky.

For each multipole  $\ell$ , there are  $\ell$  corresponding multipole vectors  $\mathbf{v}^{(\ell, i)}$ , where  $i \in \{1, 2, \dots, \ell\}$ . Roughly speaking, the more planar the temperature fluctuations associated with a given multipole are, the more its associated multipole vectors will be confined to a plane, and the more the oriented-area vectors defined by their cross products,

$$\mathbf{w}^{(\ell, i, j)} \equiv \pm(\mathbf{v}^{(\ell, i)} \times \mathbf{v}^{(\ell, j)}), \quad (16)$$

will line up in a direction normal to that plane. Moreover, planarity (as opposed to simply orientation along a direction) of the temperature multipole will cause the multipole vectors  $\mathbf{v}^{(\ell, i)}$  and  $\mathbf{v}^{(\ell, j)}$  to be at large angles relative to each other, enhancing the magnitude of  $\mathbf{w}^{(\ell, i, j)}$ . Thus, we can use the extent to which the object oriented vectors for two multipoles point in similar directions to measure how much the power from those  $\ell$ -modes are aligned.

The statistic  $S_{\text{QO}}$  takes advantage of this property to quantify the quadrupole-octopole alignment. It is the normalized sum of the dot products of the quadrupole oriented-area vector  $\mathbf{w}^{(2, 1, 2)}$  with the three octopole oriented-area vectors  $\mathbf{w}^{(3, i, j)}$ : [55, 56]

$$S_{\text{QO}} = \frac{1}{3} \sum_{\{i, j\}} |\mathbf{w}^{(2, 1, 2)} \cdot \mathbf{w}^{(3, i, j)}| \quad (17)$$

where  $\{i, j\}$  can be  $\{1, 2\}$ ,  $\{2, 3\}$ , or  $\{3, 1\}$ . Given this, larger values of  $S_{\text{QO}}$  indicate more alignment and pla-

nararity in the  $\ell = 2$  and  $\ell = 3$  modes of the temperature maps.

Because multipole vectors are defined in terms of  $a_{\ell m}$ , this measurement can only be done on full-sky maps. To measure  $S_{\text{QO}}$  for a temperature map, we first use the HEALPix function `map2alm` to measure its  $a_{\ell m}$ , and then use the procedure<sup>15</sup> described in Appendix A of Ref. [54] to extract the multipole vectors for  $\ell = 2 - 3$ . We then combine them via Eqs. (16) and (17) to get  $S_{\text{QO}}$ .

We find that the  $S_{\text{QO}}$  value measured from the SMICA map is larger than that from most simulations, with a  $p$ -value (here, upper-tail probability) of 0.4% when compared with either our synfast or FFP simulation ensembles. This is consistent with the results in Ref. [57]. They found the SMICA map from the Planck 2013 data release [58] had a larger  $S_{\text{QO}}$ , with a  $p$ -value<sup>16</sup> of 0.54% compared to an ensemble of  $10^6$  simulations analogous to our synfast simulations, but which use constrained realizations to in-paint masked regions. The fact that our  $p$ -value is so similar to theirs indicates that simply measuring  $S_{\text{QO}}$  on full-sky maps (as we do) rather than doing in-painting does not significantly affect the large-scale alignment behavior.

### 2. $A_{\text{LV}}$ : Hemispherical power asymmetry via local-variance dipole

Finally, we include a measure of the level of asymmetry in temperature power between two hemispheres of the sky. This is studied because one hemisphere of the observed CMB sky has been noted to have more power than the other [4, 59], which can be modeled by a dipole modulation of temperature fluctuations at large angular scales [60, 61]. Following Refs. [62, 63], we quantify hemispherical power asymmetry using a local-variance map, which measures the size of temperature fluctuations within disks of radius  $\theta$  centered on each of its pixels. By measuring the dipole of a local-variance map, we can quantify the direction and magnitude of any hemispherical power asymmetry in a computationally inexpensive way. Additionally, we can probe the scale dependence of the effect by varying the angular size  $\theta$  of the disks used to create the local-variance map.

More formally, if  $\hat{\mathbf{n}}_i$  is the location of the  $i$ th pixel of the input temperature map  $T(\hat{\mathbf{n}})$  from which the monopole and dipole of unmasked pixels have been removed, we can write the local-variance map  $\sigma_\theta^2(\hat{\mathbf{n}})$  as

$$\sigma_\theta^2(\hat{\mathbf{n}}) = \frac{1}{N[\mathcal{D}_\theta(\hat{\mathbf{n}})]} \sum_{i \in \mathcal{D}_\theta(\hat{\mathbf{n}})} [T(\hat{\mathbf{n}}_i) - \bar{T}_\theta(\hat{\mathbf{n}})]^2, \quad (18)$$

<sup>15</sup> Calculations were performed using code `mpd.decomp.py` provided at <http://www.phys.cwru.edu/projects/mpvectors/>.

<sup>16</sup> Value from Table 7 of Ref. [57]

where  $\mathcal{D}_\theta(\hat{\mathbf{n}})$  is the set of unmasked pixels within angle  $\theta$  of direction  $\hat{\mathbf{n}}$ ,  $N[\mathcal{D}_\theta(\hat{\mathbf{n}})]$  is the number of pixels in that set, and  $\bar{T}_\theta(\hat{\mathbf{n}})$  is their average temperature.

In practice, we measure the dipole of a dimensionless, weighted version of the local-variance map,

$$\tilde{\sigma}_\theta^2(\hat{\mathbf{n}}) = \frac{w(\hat{\mathbf{n}})}{\bar{w}} \times \frac{\sigma_\theta^2(\hat{\mathbf{n}}) - \mu_\theta(\hat{\mathbf{n}})}{\mu_\theta(\hat{\mathbf{n}})}. \quad (19)$$

In this expression,  $w(\hat{\mathbf{n}})$  is a dimensionless weight map (defined below),  $\bar{w}$  is its average over unmasked pixels, and  $\mu_\theta(\hat{\mathbf{n}})$  is the mean computed by averaging the local-variance maps of an ensemble of simulated CMB temperature maps.

We choose the weight function  $w(\hat{\mathbf{n}})$  to be

$$w(\hat{\mathbf{n}}) = \frac{1}{\text{Var}[\sigma_\theta^2(\hat{\mathbf{n}})]} \times \left\{ \frac{1}{N_{\text{pix}}} \sum_{i=1}^{N_{\text{pix}}} \text{Var}[\sigma_\theta^2(\hat{\mathbf{n}}_i)] \right\} \quad (20)$$

where we define, if  $\alpha$  labels the simulation realization and  $N_{\text{sim}}$  is the number of simulations,

$$\text{Var}[\sigma_\theta^2(\hat{\mathbf{n}})] = \frac{1}{N_{\text{sim}}} \sum_{\alpha=1}^{N_{\text{sim}}} \left[ \frac{\sigma_\theta^{2(\alpha)}(\hat{\mathbf{n}}) - \mu_\theta(\hat{\mathbf{n}})}{\mu_\theta(\hat{\mathbf{n}})} \right]^2. \quad (21)$$

This inverse variance weighting suppresses the impact of noisy regions of the input temperature map, as long as those noise contributions are modeled in the simulations. In the limit that noise properties are direction-independent, the weight factor  $w(\hat{\mathbf{n}})/\bar{w}$  will approach 1.

In our work, we measure  $A_{\text{LV}}$  from the UT78 cut-sky for both the SMICA and simulation maps. We fix the disk radius to be  $\theta = 8^\circ$ , which is the scale previously found to produce the most anomalous local-variance dipole, and compute the local-variance maps at a resolution of  $N_{\text{SIDE}} = 16$ . The amplitude of the dipole of a normalized local-variance map,  $A_{\text{LV}}$ , is then obtained by using the Healpix function `remove_dipole`. Following Planck XVI (I&S), we include only disks for which at least 90% of the input pixels are unmasked.

Our measurements return a local-variance dipole amplitude with  $A_{\text{LV}} = 0.22$  when  $\sigma_\theta^2(\hat{\mathbf{n}})$  is normalized using the synfast simulations, and  $A_{\text{LV}} = 0.21$  using the FFP simulations. These values give an upper tail probability of  $p = 1\%$  when compared to either set of simulations. Both of these values are notably larger than the Planck XVI (I&S) findings of  $A_{\text{LV}} \sim 0.044$  for the SMICA map, with a  $p$ -value of  $0.1\%$ .<sup>17</sup> In investigating this discrepancy, we found that the value of  $A_{\text{LV}}$  is highly sensitive to the resolution of the input temperature maps, with lower resolution input maps tending to give larger dipole amplitudes. Because we compute local variances using

our fiducial set of  $N_{\text{SIDE}} = 64$  maps, while Planck XVI (I&S) uses  $N_{\text{SIDE}} = 2048$  input maps, we believe this resolution dependence explains the difference.

We also note that Doppler dipole modulation included in the FFP simulations [64] (but not the synfast simulations) will generate a small power asymmetry which contributes to the local-variance power asymmetry. However, that contribution is expected to be negligible for the scales  $\ell \lesssim 191$  that we are investigating [65]. The fact that the  $p$ -values from comparisons to the synfast and FFP simulations are nearly identical is in line with that expectation.

### C. Summary: Individual anomaly measurements

To summarize, we have defined quantities associated with eight properties of the CMB temperature map which are either found to be statistically unlikely in the observed sky or which are expected to shed light on the relationship between statistically unlikely features. For each feature, we have described our technique for measuring it from real and simulated CMB maps. By analyzing the resulting data associated with each feature individually, we have verified (where applicable) that our measurement of the SMICA map's  $p$ -values (single-tail probabilities) relative to the simulation ensembles are consistent with previous findings.

Fig. 3 shows a summary plot of these anomaly statistics measured from our ensembles of 100,000 synfast (left column), and 1,000 FFP (right column) simulations. In each panel, the grey histogram shows the distribution of simulation measurements, which are either made based on full-sky maps, or the cut UT78 sky, as indicated by the gray text in the lower right corner. These data are what will be used in subsequent sections to study the relationship between features in the context of isotropic  $\Lambda$ CDM. The vertical lines show feature measurements done using the Planck public QML  $C_\ell$ 's (blue), pseudo- $C_\ell$ 's extracted from the full-sky SMICA map (green), UT78 cut-sky pseudo- $C_\ell$ 's for the SMICA map (orange), and the theoretical expectation based on Planck's best-fit parameter values (dashed black). Note that the two statistics that depend on the phases ( $S_{\text{QO}}$  and  $A_{\text{LV}}$ ) do not have a corresponding measurement from either the published QML or the best-fit theory  $C_\ell$ , as they can not be related to the angular power spectrum. The QML and theory values for  $\sigma_{16}^2$  are computed using Eq. (12). The  $p$ -values for these measurements are shown in the same color on the right-hand side of each panel, with an arrow indicating which measurement we think is most relevant for the feature in question. These choices and the findings for each quantity are discussed in detail above.

To put the  $p$ -values in context, the 1, 2, and  $3\sigma$  error bars for a normal distribution correspond to single tail probabilities of  $p = 16\%$ ,  $2.2\%$ , and  $0.014\%$ , respectively. Thus, of the features we study, the SMICA values for  $S_{1/2}$ ,  $C_2$ ,  $\sigma_{16}^2$ ,  $S_{\text{QO}}$ , and  $A_{\text{LV}}$  are  $2-3\sigma$  unlikely compared

<sup>17</sup> Values taken from Fig. 27 ( $A_{\text{LV}}$  value) and Table 20 ( $p$ -value) of Ref. [7].

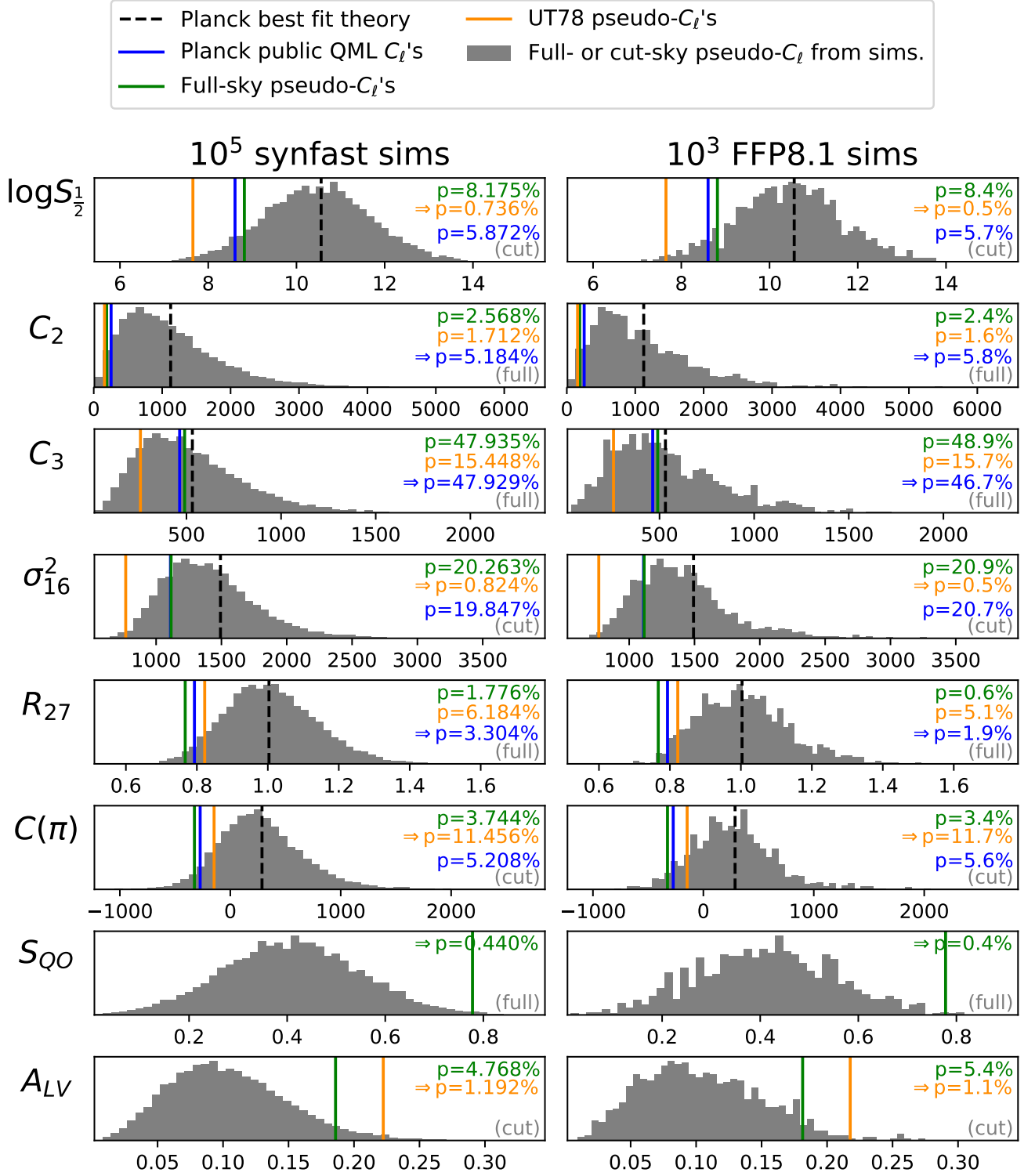


FIG. 3. Summary plot where each row shows the results for one of the features discussed in Sec. III. Grey histograms show statistics for the features of the CMB temperature map measured from our 100,000 synfast simulations (left column), and from 1,000 FFP 8.1 simulations (right column). The grey text in the lower right of each panel denotes whether simulations measurements are based on the full-sky or the cut-sky, where cut-sky measurements use the UT78 mask. Vertical lines show values for the Planck SMICA map, as well as theory predictions from Planck's best-fit cosmology. The  $p$ -values displayed are single-tail probabilities showing the percentage of simulations that are more extreme than the SMICA measurements. Arrows by the  $p$ -values indicate which measurement we think is most relevant for each feature.

to our simulations, while  $R_{27}$  and  $C(\pi)$  are between  $1\sigma$  and  $2\sigma$ , and  $C_3$  is not in tension. For the most part our measured  $p$ -values are consistent with previous findings. Where they are not, we can explain the discrepancies in terms of the power spectrum measurement technique (for  $R_{27}$ ) or the initial map resolution (for  $A_{LV}$ ).

We can also use the results in Fig. 3 to make some general observations about the impact of different measurement choices. Outside of small differences which are within the reasonable range of sampling error, the one-dimensional  $p$ -values comparing measurements of the real sky measurements relative to synfast simulations are in good agreement with those comparing the real sky to FFP simulations. We additionally note that in general the Planck public QML  $C_\ell$ 's give results that are very similar to full-sky  $C_\ell$  measurements. The fact that the cut-sky pseudo- $C_\ell$  SMICA measurements have comparably lower  $S_{1/2}$ ,  $C_2$ ,  $C_3$ , and  $\sigma_{16}^2$  are consistent with previous studies which have found that the observed lack of power in the CMB sky is more severe in regions further from the galactic mask.

#### IV. RESULTS: ANOMALY COVARIANCES

We are now ready to tackle the main goal of this paper and study the relationships between the large-angle CMB temperature map properties which have been examined individually above. The features are: the integrated power of temperature fluctuations at angles  $\theta > 60^\circ$  ( $S_{1/2}$ ), the quadrupole amplitude ( $C_2$ ), the octopole amplitude ( $C_3$ ), the variance of the temperature map evaluated at resolution NSIDE = 16 ( $\sigma_{16}^2$ ), the parity statistic  $R$  with maximum multipole of  $\ell = 27$  ( $R_{27}$ ), the angular power spectrum at  $180^\circ$  ( $C(\pi)$ ), the quadrupole-octopole alignment ( $S_{QO}$ ), and the amplitude of the hemispherical power asymmetry ( $A_{LV}$ ). Using our measurements of these quantities from our synfast and FFP simulations, we will determine their covariances in order to build an understanding of how they are related under the assumption of isotropic  $\Lambda$ CDM. We will do so in three stages, first describing the relationship between pairs of features measured from the synfast simulations, then comparing the covariance matrices for the synfast and FFP simulations, and finally, further exploring the covariance structure by performing a principal component analysis.

We begin by inspecting how our fiducial set of 100,000 synfast simulations are distributed in the eight-dimensional space defined by the parameters  $S_{1/2}$ ,  $C_2$ ,  $C_3$ ,  $\sigma_{16}^2$ ,  $R_{27}$ ,  $C(\pi)$ ,  $S_{QO}$  and  $A_{LV}$  introduced in Sec. III. Fig. 4 shows the relationships between pairs of those quantities. The diagonal panels display the same one-dimensional statistical information as the left column of Fig. 3, with the gray histograms showing the distribution from simulations and the vertical lines showing the measurements of the SMICA map and theoretical expectations based on the Planck's best-fit cosmological parameters.

In the off-diagonal panels, the grey contours indicate the 1, 2, and  $3\sigma$  confidence regions based on simulation data. The contour locations are determined for each panel as follows. First, we use simulation data to make a two-dimensional histogram with fifty bins along each axis. Next, we smooth the histogram using a Gaussian filter with a width corresponding to one bin. The smoothed histogram is then thresholded at constant-count (constant-probability) surfaces so that 68% of the input realizations fall inside the  $1\sigma$  contours, etc. The number at the top of each panel displays the correlation coefficient  $R$  of the two quantities shown, computed based on the simulation samples. Measurements of the observed SMICA map and theory predictions are shown using colored crosses.

Examining Fig. 4, we can make a few general observations. First, there is structure and notable covariances in the relationships between most of the features that depend only on two-point functions, but not between  $C_2$  and  $C_3$ , nor between the  $a_{\ell m}$ -phase-dependent quantities ( $S_{QO}$  and  $A_{LV}$ ) and any of the other quantities. This is expected given the isotropic  $\Lambda$ CDM model used to generate the simulations. We additionally note that the distribution of simulation points in this eight-dimensional space is decidedly non-Gaussian; this is due to the asymmetric limits on the quantities measured, as well as the (in some cases) non-linear dependence of quantities on  $C_\ell$ .

More specifically, the covariances of the two-point-function-based quantities can be understood by how they depend on the power spectrum components  $C_\ell$ . In isotropic  $\Lambda$ CDM, we expect the power at different multipoles  $\ell$  to be independent. Correspondingly we see little covariance between  $C_2$  and  $C_3$ . The positive correlation between  $\sigma_{16}^2$  and either of these amplitudes is straightforward, given Eq. (12): all else being equal, adding power to low  $\ell$  increases the variance at large scales. Similarly, increasing  $C_2$  adds to even- $\ell$  power and increasing  $C_3$  adds to odd- $\ell$  power, so we expect and see that the parity measure  $R_{27}$  to be positively and negatively correlated with  $C_2$  and  $C_3$ , respectively.

Looking at  $C(\pi)$  allows us to clarify the relationship between the parity properties and  $S_{1/2}$ . We note that, given the parity properties of spherical harmonics and referencing Eq. (15), the contributions from even- $\ell$  modes to  $C(\pi)$  will be  $\propto (2\ell + 1)C_\ell$ , while odd- $\ell$  contributions are  $\propto -(2\ell + 1)C_\ell$ . Thus  $C(\pi)$  is effectively another way of characterizing the parity properties of the large-angle CMB. Accordingly we see that  $C(\pi)$  has a strong positive correlation with  $R_{27}$  and with  $C_2$  and a somewhat weaker negative correlation with  $C_3$ . Since  $C(\pi)$  is the measurement of the angular correlation at  $\theta = 180^\circ$ , small  $S_{1/2}$  values require that  $|C(\pi)|$  be close to zero. The triangular structure of the contours in the  $C(\pi)$ - $S_{1/2}$  plane reflect this behavior. (The fact that the triangle is asymmetric about  $C(\pi) = 0$  can be understood in terms of the fact that the simulations are based on the  $\Lambda$ CDM Planck best-fit power spectrum: it has  $C(\pi) > 0$ , because the

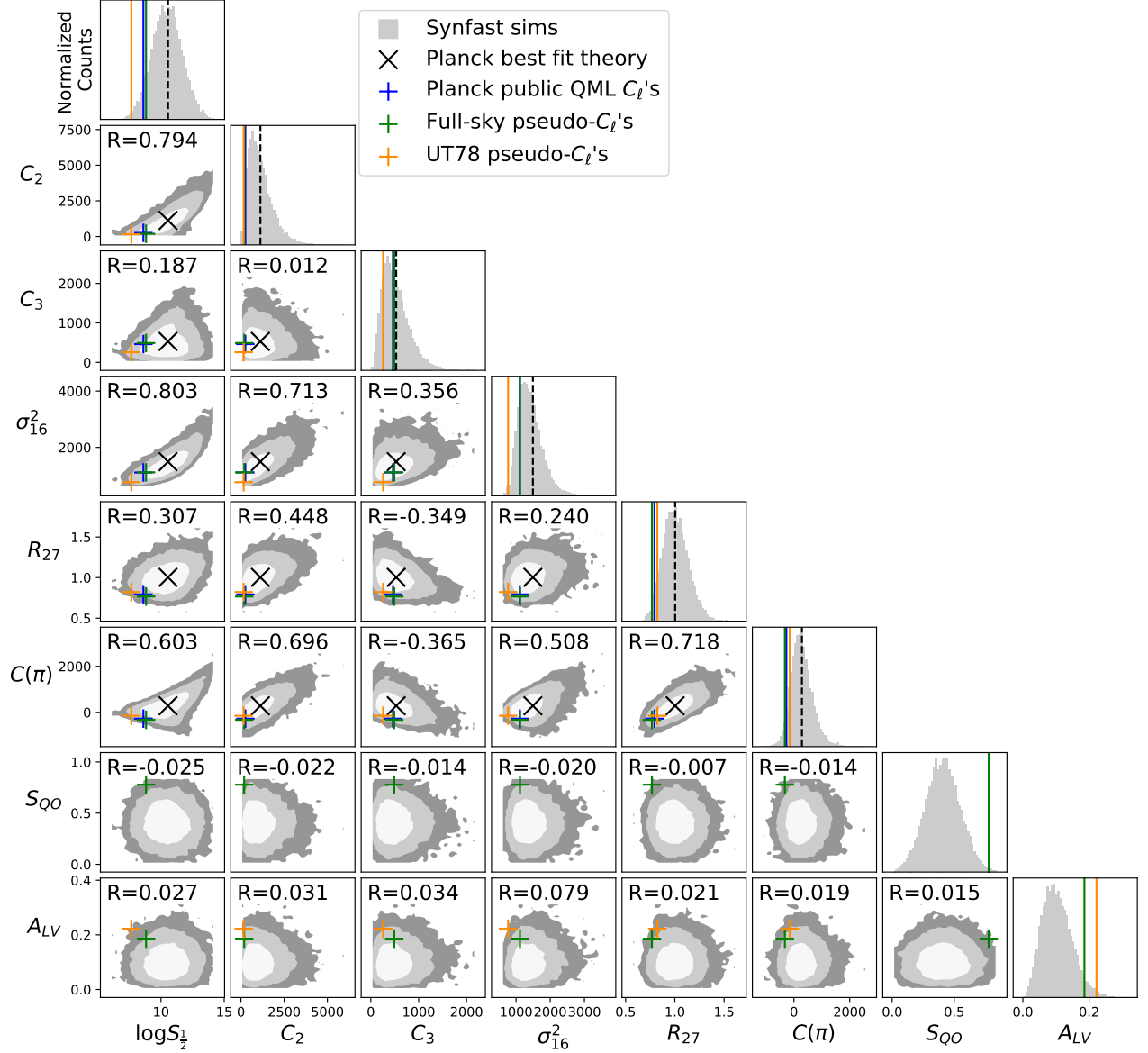


FIG. 4. Relationships between large-angle CMB features (see Table I for descriptions of these quantities). Gray contours show the 1, 2, and  $3\sigma$  confidence regions based on measurements from our ensemble of 100,000 synfast simulations, using the same measurement choices that produced the gray histograms in the left column of Fig. 3. The 1D histograms on the diagonal are the same as those in the left column of Fig. 3. The marked data points for SMICA measurements and theoretical expectations are equivalent to the vertical lines in Fig. 3. Note that the statistics based on the phases of the  $a_{\ell m}$ ,  $S_{QO}$  and  $A_{LV}$ , do not have the corresponding theoretical expectations because they cannot be computed analytically from  $C_\ell$ .

quadrupole  $\ell = 2$  mode generates a dominant positive term in Eq. (15).) Accordingly, the  $S_{1/2}$ - $R_{27}$  contours show an echo of that triangular structure. This provides an intuitive way to understand the result derived analytically in Ref. [24].

We note that the panel showing the cross correlation between  $C_2$  and  $\log S_{1/2}$  is comparable to that studied

in Ref. [25], though we find a looser relation between the two quantities. The reason for this difference is that we mix masking choices for our fiducial synfast map statistics, using full-sky measurements for the  $C_2$  and cut-sky measurements for  $S_{1/2}$ . We verify that if we either measure both quantities from full-sky maps or both on cut-sky maps, the shape of our contours closely resemble the

distribution in Ref. [25]. (Ref. [25] uses UT78 cut-sky measurements of both features.)

The results from FFP simulations are visually similar to those in Fig. 4, so we do not show the scatter plot for that ensemble. Instead, below we examine the quantitative difference between the feature covariances based on the synfast and FFP sets of simulations.

### A. Covariance structure comparison

Here we measure the covariance between large-angle features measured from our simulated CMB temperature maps, and compare the covariances extracted from the synfast and FFP simulations. By making this comparison, we can gauge whether survey properties modeled in the FFP but not in the synfast simulations affect the relationship between the features studied. This in turn can potentially provide insight into whether those survey properties influence observed anomalies in the SMICA map — though the fact that no one has yet found a convincing systematics-based explanation for any of them makes this unlikely.

For a given ensemble of  $n$  simulations, we represent each realization as a  $d$ -dimensional vector  $x$ , where  $d = 8$  is the number of large-angle quantities measured ( $S_{1/2}$ ,  $C_2$ ,  $C_3$ , etc.). Before measuring the covariance matrix, we center and normalize the data so the  $j$ th vector component of realization  $i$  becomes

$$\tilde{x}_i^{(j)} = (x_i^{(j)} - \bar{x}^{(j)})/\sigma^{(j)}, \quad (22)$$

where  $\bar{x}^{(j)}$  and  $\sigma^{(j)}$  are the mean and standard deviation over realizations of the  $j$ th quantity being measured, respectively. This ensures that the covariance structure is not dominated by differences in the characteristic size of some of the quantities we study (e.g.  $\sigma_{16}^2$  compared to  $A_{LV}$ ). It will also mean that the covariance we measure will be equivalent to the correlation coefficients appearing in the subplots of Fig. 5.

Once the data are preprocessed, we define a  $d \times n$ -dimensional matrix  $X = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ , where each column corresponds to one realization. This allows us to concisely write the measured  $d \times d$ -dimensional feature covariance matrix as

$$S = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i \otimes \tilde{x}_i = \frac{1}{n} X X^T. \quad (23)$$

We show the covariance matrix  $S_{\text{syn}}$  for our ensemble of 100,000 synfast simulations on the left side of Fig. 5.

We would like to study how the covariance matrix derived from the ensemble of 1,000 FFP simulations,  $S_{\text{FFP}}$ , differs from that measured from our 100,000 synfast simulations,  $S_{\text{syn}}$ . To make that comparison meaningful, we must ensure that differences we see are not an artifact of sample variance due to the smaller number of FFP simulations. On the right side of Fig. 5 we therefore show

the relative difference for covariance matrix entries  $S_{ij}$ ,

$$(\Delta S)_{ij} = \frac{[(S_{\text{FFP}})_{ij} - (S_{\text{syn}})_{ij}]}{[\sigma_{\text{syn}}^{(1000)}]_{ij}}, \quad (24)$$

where the denominator  $[\sigma_{\text{syn}}^{(1000)}]_{ij}$  is sampling error for when  $S_{ij}$  is measured from a set of 1,000 synfast simulations.

We estimate  $\sigma_{\text{syn}}^{(1000)}$  based on  $N = 100$  subdivisions of the 100,000 synfast simulations. This allows us to measure the covariance matrix  $S_{\text{syn}}^{(1000, \alpha)}$  for each subsample  $\alpha \in \{1, \dots, N\}$ . For each entry  $ij$  of the matrix, we can then compute the mean over subsamples  $[\bar{S}_{\text{syn}}^{(1000)}]_{ij}$ , as well as the sample variance,

$$[\sigma_{\text{syn}}^{(1000)}]_{ij}^2 = \frac{1}{N-1} \sum_{\alpha=1}^N \left( [S_{\text{syn}}^{(1000, \alpha)}]_{ij} - (\bar{S}_{\text{syn}}^{(1000)})_{ij} \right)^2. \quad (25)$$

Thus, assuming the errors on the covariance matrix entries are Gaussian, the values plotted in the right panel of Fig. 5 show the difference between the FFP and synfast covariances in units of their 1000-synfast-realization-based standard deviation. Plots of the absolute difference  $S_{\text{FFP}} - S_{\text{syn}}$  and the sampling error  $\sigma_{\text{syn}}^{(1000)}$  are shown for completeness in Appendix A.

We see that there are several moderately significant differences between the feature covariances derived using the FFP and synfast simulations. The most prominent of these are between the quadrupole-octopole alignment  $S_{\text{QO}}$  and  $\log S_{1/2}$ ,  $C_2$ ,  $C_3$ , and  $\sigma_{16}^2$ , with differences ranging from  $(2-3.6)\sigma$ . There is also a  $2.3\sigma$  difference in the  $S_{1/2}-C_2$  entry, and several other less significant differences in the range  $(1-2)\sigma$ .

Noting that the largest  $\Delta S$  entries involve the quadrupole, we hypothesized that these differences might be driven by the kinematic quadrupole, which is partially simulated in the FFP maps but not in the synfast simulations. To test this idea, we created an alternative version of the synfast ensemble where the Doppler quadrupole (DQ) correction  $a_{2m}^{\text{DQ}}$  given in Table 3 of Ref. [57] is added to each map<sup>18</sup>. When we use this synfast+DQ ensemble to reproduce Fig. 5 there are no significant changes in the feature covariance matrix or its differences from the FFP feature covariance. If the  $S_{\text{FFP}} - S_{\text{syn}}$  differences were mainly driven by the kinematic quadrupole present in the FFP maps, we would expect that adding a DQ correction to the synfast simulations would significantly change the structure of  $\Delta S$ . Because it does not, we conclude that modeled foregrounds or survey properties other than the kinematic quadrupole are driving the differences between  $S_{\text{FFP}}$  and  $S_{\text{syn}}$ .

<sup>18</sup> This DQ correction is slightly different than that included in the FFP simulations, which model only the residual frequency-dependent portion of the kinematic quadrupole that is not removed during the Planck map processing.



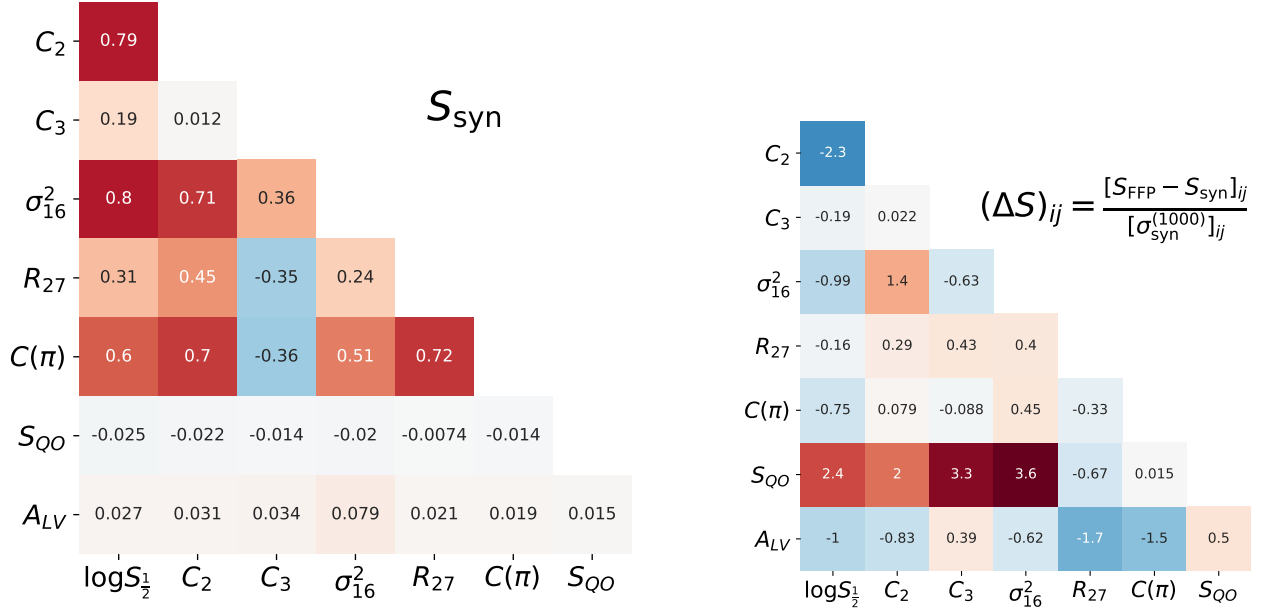


FIG. 5. Left: The feature covariance matrix  $S_{\text{syn}}$  measured from the synfast simulation ensemble. Right: The difference  $\Delta S$  (given in Eq. (24)), between the covariances measured from the FFP and synfast simulations in units of its sampling error  $\sigma_{\text{syn}}^{(1000)}$  estimated from sets 1,000 synfast realizations (defined in Eq. (25)).

### B. Principal component analysis

We next use a principal component analysis (PCA) to investigate whether large-angle CMB anomalies can be reduced to a few fundamental “building blocks” — features, or combinations thereof, which explain the ways that the observed CMB sky is unusual compared to our ensembles of simulations. Ref. [8] conjectured that there are three such building blocks in the CMB maps observed by WMAP and Planck: missing large-angle power, alignments between the low multipoles, and dipolar modulation of the CMB (which is roughly equivalent to the hemispherical asymmetry studied in this paper). We now have an opportunity to quantitatively test this conjecture by using our simulation measurements. By finding the simulation data’s principal components (PCs) in the eight-dimensional feature space we consider, we can determine which linear combinations of features explain most of the covariance structure discussed in Sec. IV A. It is our hope that studying the position of the SMICA in this PC basis will allow us to further characterize the large-angle properties of the observed CMB temperature map.

PCA is a dimensionality reduction technique which works by identifying the directions in a  $d$ -dimensional parameter space along which a set of data points have the maximum variance. The principal components are defined sequentially: the first PC corresponds to the direction in which our simulation realizations have the most

variance; the second PC corresponds to the direction of maximum variance after the components of the data in the direction of PC 1 are projected out; and so on. In practice, the PCs are obtained by finding the eigenvectors of the data’s covariance matrix. Therefore, in our analysis, we determine the eigenvectors of the covariance matrices  $S$  derived above in Eq. (23) to obtain PCs which are unit-length vectors in the basis defined by the quantities  $S_{1/2}$ ,  $C_2$ ,  $C_3$ ,  $\sigma_{16}^2$ ,  $R_{27}$ ,  $C(\pi)$ ,  $S_{QO}$ , and  $A_{LV}$ . The first principal component is the eigenvector with the largest eigenvalue, the second PC has the second largest eigenvalue, and so on.

PCA works as a dimensionality reduction technique because we can capture much of the information about the input data’s variance by projecting it onto the first  $d' \leq d$  PC directions. Heuristically, the fraction of the information that is retained in this projection is equal to ratio between the sum of the first  $d'$  covariance matrix eigenvalues to the sum of all  $d$  eigenvalues. To quantify the relative importance of the various PCs, we adopt the complement of this quantity, the fractional residual variance (FRV), the fraction of the variance that is *not* captured by the first  $d'$  PCs. It is given by the expression

$$\text{FRV} \equiv 1 - \frac{\sum_{i=1}^{d'} \lambda_i}{\sum_{i=1}^d \lambda_i}, \quad (26)$$

where the eigenvalues have been ordered so that  $\lambda_i \geq \lambda_{i+1}$ .



Fig 6 shows the properties of the PCs derived from our ensemble of 100,000 synfast simulations. The left panel show PCs (the eigenvectors of  $S_{\text{syn}}$ ), with each column corresponding to one PC. The rows correspond to contributions of each of the original eight quantities to the PCs. The right panel shows the fractional residual variance as a function of the number of PCs retained as well as the individual contribution of each PC to its sum.

Studying the eigenvectors themselves, we find that interpretation of the first four PCs is fairly straightforward because they can be associated with input quantities which are largely independent of one another. Together, they capture about 90% of the simulations' variation in the space of our eight measured features.

The first PC, which captures 42% of the simulations' variance, roughly corresponds to missing large-angle correlations, as it is dominated by  $C_2$ ,  $S_{1/2}$ ,  $\sigma_{16}^2$  and  $C(\pi)$  (and, to an extent,  $R_{27}$ ). These particular features are positively correlated, with correlation coefficients  $R$  varying between about 0.5 and 0.8 (see Fig. 4), so it is mathematically expected that they would form a principal component whose eigenvector components have same signs and comparable amplitudes, as we observe in PC 1. One can understand the relationship between these features by noting how their quantities will change if we change the quadrupole amplitude, all else being equal. Raising  $C_2$  will increase large-angle power, and will increase the relative power in even-parity modes compared to odd-parity modes, so it make sense that  $S_{1/2}$ ,  $\sigma_{16}^2$ ,  $C(\pi)$ , and  $R_{27}$  will all increase. Thus, one interpretation of the first PC is that it picks out direction in our feature space similar to that associated with variations in the quadrupole amplitude.

Next, PC 2 accounts for 20% of the simulations' variance and is dominated by the octopole. Given the correlations of  $C_3$  and the other statistics in Fig. 4, it is also unsurprising that PC 2 receives moderate contributions from  $S_{1/2}$  and  $\sigma_{16}^2$  with the same sign as  $C_3$ , and from  $R_{27}$  and  $C(\pi)$  with the opposite sign. As with PC 1, we can understand this in terms of how other quantities will respond if we raise or lower  $C_3$  without changing power at other multipoles. More octopole power will generally add to large-angle power, increasing  $S_{1/2}$  and  $\sigma_{16}^2$ , but specifically through odd parity contributions which correspond to lower  $R_{27}$  and  $C(\pi)$ .

The third and fourth PCs are associated with the  $a_{\ell m}$ -phase dependent quantities. PC 3 is associated with the sum of the quadrupole-octopole alignment  $S_{\text{QO}}$  and the hemispherical asymmetry statistic  $A_{\text{LV}}$ , and captures 13% of the simulations' variance, while PC 4 is associated with their difference and captures 12% of the variance. We note that because their associated covariance eigenvalues  $\lambda_3$  and  $\lambda_4$  are nearly equal, the ordering of PC 3 and PC 4 is somewhat arbitrary. This reflects the fact that the correlation between  $S_{\text{QO}}$  and  $A_{\text{LV}}$  is very small, and means that using PC 3 and PC 4 together is basically equivalent to defining two unit vectors in the  $A_{\text{LV}}$  and  $S_{\text{QO}}$  directions.

The structure of the fifth through eighth PC, which account for the remaining 10% of the simulations' variance, resists simple interpretation. This is because they are determined by the relationships between the non-independent quantities, after the variation of the data in the direction of the first four PCs (roughly  $C_2$ ,  $C_3$ ,  $S_{\text{QO}}$  and  $A_{\text{LV}}$ ) are projected out. One could infer, for example, that because PC 5 has  $C_2$  and  $C_3$  components with different signs, that it might capture some information about whether the power from the quadrupole and octopole cancel one another, but this is far from clear. PCs 6-8 all have small eigenvalues with  $\lambda_6 \sim \lambda_7 \sim \lambda_8$ , so their order and the way that they divide up whatever degrees of freedom are left after the first five PCs are removed are somewhat arbitrary.

We also performed a PCA on the FFP simulation data. The results are very similar to those for the synfast simulations, and any differences that exist mainly just reflect the differences between the structure of the synfast and FFP covariance matrices discussed in Sec. IV A. Given this, we do not show the FFP-based PCs.

We next calculate the probability of CMB maps projected to the PC basis. This can inform whether linear combinations of the features that tend to “come together” in simulated skies are particularly anomalous (or not) when observed on our CMB sky. We proceed as follows. The measurement of each simulation corresponds to a vector in our eight-dimensional feature space. By taking the dot product of that vector with each PC, we find its components in the new PC basis. The grey histograms in Fig. 7 show the resulting distributions of the simulations projected to the PC basis, where as before we show results from the synfast simulations in the left column, and from the FFP simulations in the right column. The red vertical lines correspond to the observed CMB sky, using the fiducial SMICA map measurements (indicated in Fig. 3 by arrows next to their  $p$ -values).

Fig. 7 shows some instructive trends. Since the first principal component (PC 1) is a linear combination of the features that encode the missing angular correlations (low  $C_2$ ,  $S_{1/2}$ ,  $\sigma_{16}^2$  and  $C(\pi)$ ), it makes sense that the probability observed sky's PC 1 coefficient is low. However the fact that this probability is lower than that for any of the individual features ( $p$ -value = 0.064%) further indicates that, even given the lowness of one of its constituent statistics (for example, the quadrupole), the other features that make up PC 1 are still lower than expected in  $\Lambda$ CDM. Next, the probability of PC 2 is not anomalous ( $p = 47\%$ ), which is unsurprising given that it largely reflects the observed sky's rather average  $C_3$ . The PC 3 probability, however, is surprisingly high ( $p = 0.032\%$ ), which is the smallest  $p$ -value among all PCs. The extremely high value of the statistics projected to PC 3 comes from the fact that this principal component is largely a sum of  $S_{\text{QO}}$  and  $A_{\text{LV}}$ , which are both high on our sky but uncorrelated ( $R = 0.015$ ) in  $\Lambda$ CDM. Hence, PC 3 is the sum of two high-valued statistics, and is therefore very high itself. In contrast, PC 4 is mainly

a difference between the same two high-valued statistics ( $S_{\text{QO}}$  and  $A_{\text{LV}}$ ), and so is itself average ( $p = 39\%$ ). The higher PCs do not shed significant further light on the statistical properties of the features we study.

In concluding this section, we caution that PCA as a method is only able to capture linear structures in the dataset. Because the relationship between many of the quantities we measure are nonlinear by definition, the PCAs will therefore not capture all of the structure in the simulations' distribution.

## V. CONCLUSIONS

In this paper we have studied the relationships expected in  $\Lambda$ CDM between a set of large-angle CMB features, with the goal of better understanding the interdependence of large-angle temperature anomalies observed in WMAP and Planck data. In particular, we have studied eight features measured via the quantities defined in Sec. III: the integrated power of temperature fluctuations at angles  $\theta > 60^\circ$  ( $S_{1/2}$ ), the quadrupole amplitude ( $C_2$ ), the octopole amplitude ( $C_3$ ), the variance of the temperature map evaluated at resolution  $N_{\text{SIDE}} = 16$  ( $\sigma_{16}^2$ ), the parity statistic  $R$  with maximum multipole of  $\ell = 27$  ( $R_{27}$ ), the angular correlation function at  $180^\circ$  ( $C(\pi)$ ), the quadrupole-octopole alignment ( $S_{\text{QO}}$ ), and the amplitude of the hemispherical asymmetry ( $A_{\text{LV}}$ ). The first six of these features depend on the angular power spectrum and quantify various aspects of isotropic angular clustering at large scales, while the last two depend on phases of the  $a_{\ell m}$  and quantify large-angle alignments and power asymmetry observed in CMB maps. In addition to several commonly studied anomalous features, this list includes a few features which have not been individually reported to be anomalous in the CMB data (the octopole is a notable example), but which have been flagged in previous work as potentially interesting in relation to the other features.

Our analysis was based on on measurements of two ensembles of  $\Lambda$ CDM simulations: 100,000 noiseless Gaussian CMB temperature maps generated using the `synfast` function in `healpy`, and 1000 full focal plane (FFP8.1) simulations provided by the Planck team that contain astrophysical foregrounds and other physical artifacts expected in the observed sky. We began by using these ensembles to find the probability of each feature in  $\Lambda$ CDM. This allowed us to study the impact of analysis choices on the features' statistics and to make sure we could recover results from previous work. We found generally excellent agreement between the statistics measured from our two sets of synthetic maps, and summarized the results in Fig. 3.

Then, selecting a fiducial set of analysis choices, in Sec. IV we used those same simulation measurements to fulfill the principal goal of this paper by calculating the correlation between the eight features studied. Fig. 4 shows, for the first time, a complete covariance of the

features associated with the most commonly discussed large-angle CMB anomalies. Our results confirm and quantify various aspects of the features that were previously either merely conjectured or calculated in isolation. For example, the quadrupole  $C_2$ , the missing large-angle correlations statistic  $S_{1/2}$ , and the variance  $\sigma_{16}^2$  are all positively correlated and largely uncorrelated to the phase-dependent features. The phase-dependent features — the quadrupole-octopole alignment  $S_{\text{QO}}$  and the hemispherical asymmetry statistic  $A_{\text{LV}}$  — are uncorrelated both with each other and with all other features studied. Less trivially, the observed low  $S_{1/2}$  essentially guarantees (in  $\Lambda$ CDM) that the configuration-space clustering amplitude at the largest observable scale,  $C(\pi)$ , is very close to zero, which is in fact observed in the data. Furthermore, we find that the covariance between  $S_{\text{QO}}$  and several other features, though still low compared to the covariances between other features, is significantly higher when measured from FFP simulations than from `synfast`. Introducing a kinetic quadrupole correction to the `synfast` simulations has little impact on that difference.

We then diagonalized the measured covariance matrix to obtain the principal components of features' expected distribution in  $\Lambda$ CDM. This allowed us to quantify whether most of the simulations' variation in our eight-dimensional feature space is retained in some smaller number of PCs. We find that 42% of the simulations' variance is in the direction of the first PC, which quantifies the missing large-angle correlations and has comparable coefficients of the same sign in  $C_2$ ,  $S_{1/2}$ ,  $\sigma_{16}^2$  and  $C(\pi)$ . Another 20% is the PC 2 direction, which largely lies in the direction of  $C_3$ , along with less dominant contributions from features correlated with the octopole. The next two PCs quantify the sum and difference of the quadrupole-octopole alignment  $S_{\text{QO}}$  and the hemispherical asymmetry statistic  $A_{\text{LV}}$ , capturing 13% and 12% of the data's variance, respectively. These first four PCs together explain about 90% of variation in the space of the (eight) features.

It is important to remind ourselves that apart from the few (generally  $2\text{-}3\sigma$ ) anomalies discussed here and elsewhere, the  $\Lambda$ CDM model describes most of the current cosmological observations with immense success. Given the significant cosmic variance inherent in the largest angular scales of the CMB, as well as the absence of concrete models that are competitive with  $\Lambda$ CDM, we should be wary of putting too much weight on these anomalies as motivations for new physics. However, given the success of  $\Lambda$ CDM, any observational clues as to how to build a more fundamental description of, for example, the physics of inflation or dark energy will (initially at least) take the form of small deviations from its predictions [66]. Given this, we should certainly take a careful look at reported tensions and anomalies, making sure we understand how assumptions related to modeling and analysis affect their significance.

It is in this spirit that this work contributes to the dis-

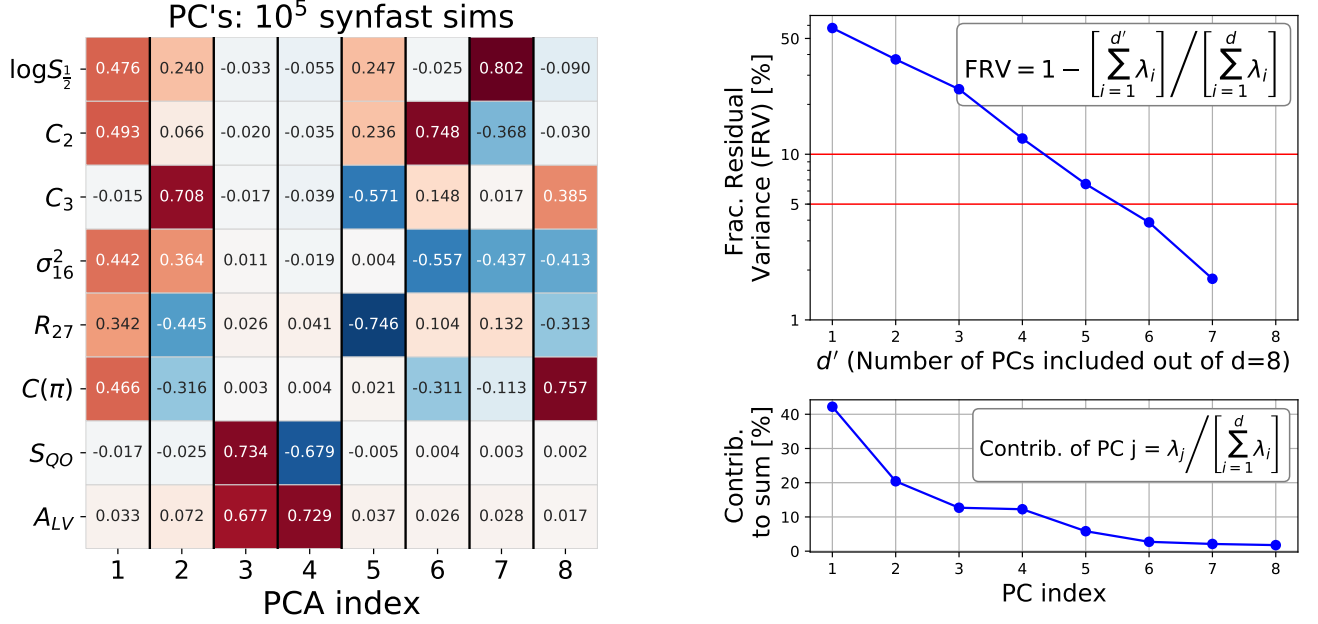


FIG. 6. Left: Eigenvectors of the anomaly covariance for our fiducial set of  $10^5$  synfast simulations. Each column is one PC; the one with index 1 points in the direction of the data's maximum variance. The rows indicate the features' contributions to each PC. Right: Fractional residual variance for these PCs and the eigenvalues associated with each PC as a fraction of the total sum of all eigenvalues.

cussion of large-angle CMB anomalies: by understanding in detail how observed features are related in  $\Lambda$ CDM, we can better assess the independent ways in which our observed CMB sky is unusual, and thus whether they might provide clues about beyond- $\Lambda$ CDM physics. An interesting potential avenue for future work could be to study how the covariance between the anomalies changes when assuming underlying models that are extensions of or alternatives to  $\Lambda$ CDM.

## ACKNOWLEDGMENTS

JM has been supported by the Rackham Graduate School through a Predoctoral Fellowship. DH and SA have been supported by NASA under contract 14-ATP14-0005. DH and JM have also been supported by DOE under contract DE-FG02-95ER40899. Many thanks to Diego Molinari for answering our questions about the Planck parity symmetry analysis. The analysis in this paper was done using an Anaconda [67] (version 4.3.22) installation of Python (version 2.7.13), and made use of the packages SciPy [68] (version 0.19.1), NumPy [69] (version 1.12.1), Matplotlib [70] (version 2.0.2), and Healpy (version 1.10.3), the Python implementation of Healpix [32]. Jupyter [71] (version 5.0.0) notebooks were used for parts of the analysis. Additionally, some of the results in this paper have been derived using the multipole vector code available

at <http://www.phys.cwru.edu/projects/mpvectors/>, which was developed with support from the US Department of Energy.

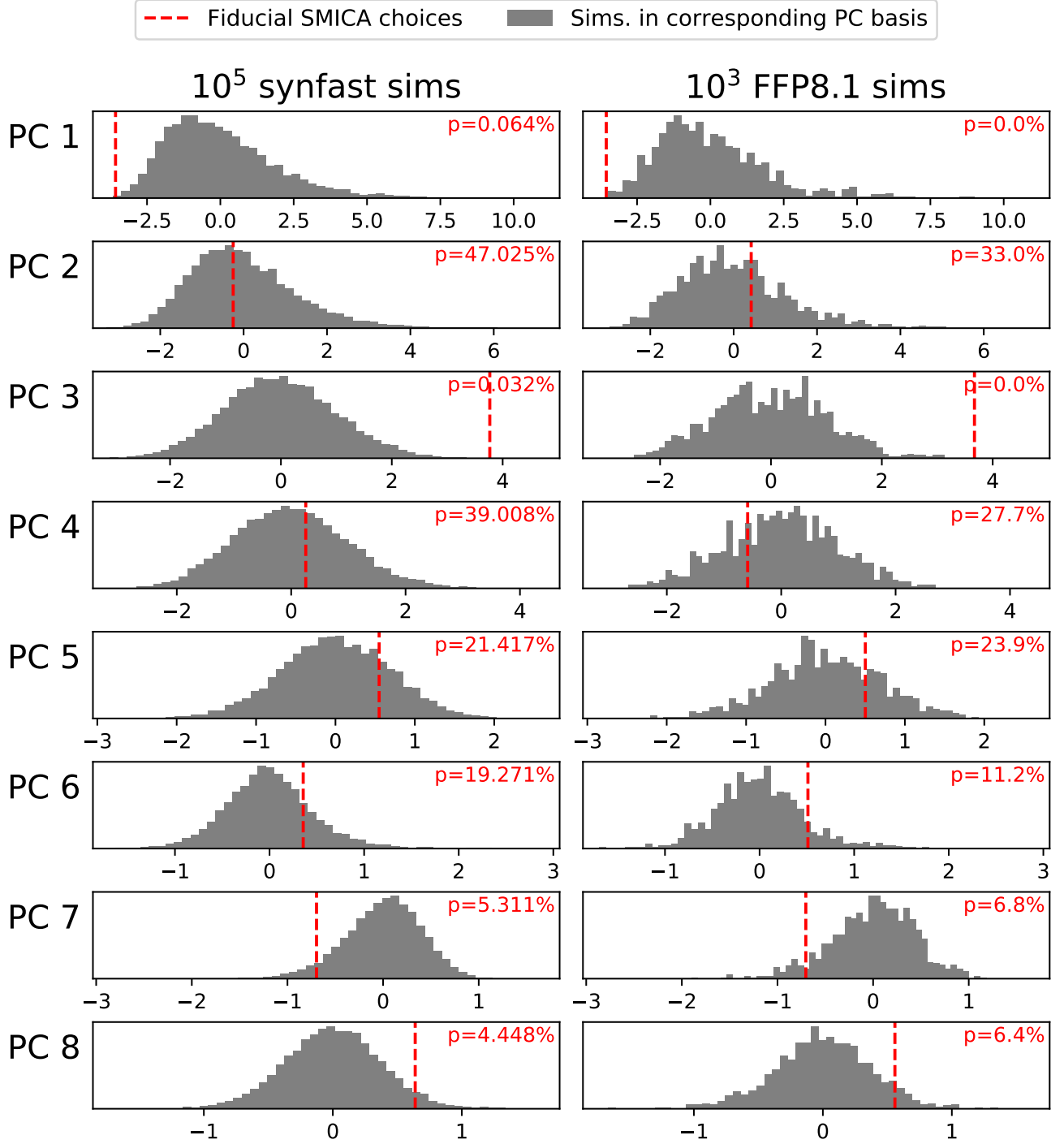


FIG. 7. Summary of feature statistics projected into our PCA basis. The row labels indicate the basis vectors of the PC basis, which are equivalent to the unit-eigenvectors of the synfast (left) and FFP (right) simulations' feature covariance matrix. The gray histograms show the distribution of the components of simulation realizations in the direction of each PC. The red lines show the same projection of our fiducial SMICA map measurements (using the measurement methods whose  $p$ -values in Fig. 3 are denoted by an arrow). The red numbers in the top right corner of each panel show the percentage of simulations that are more extreme than the corresponding SMICA measurement.

- 
- [1] P. A. R. Ade et al. (Planck), *Astron. Astrophys.* **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO].
- [2] A. de Oliveira-Costa, M. Tegmark, M. Zaldarriaga, and A. Hamilton, *Phys. Rev.* **D69**, 063516 (2004), arXiv:astro-ph/0307282 [astro-ph].
- [3] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, *Phys. Rev.* **D75**, 023507 (2007), arXiv:astro-ph/0605135 [astro-ph].
- [4] H. K. Eriksen, F. K. Hansen, A. J. Banday, K. M. Gorski, and P. B. Lilje, *Astrophys. J.* **605**, 14 (2004), [Erratum: *Astrophys. J.* 609, 1198 (2004)], arXiv:astro-ph/0307507 [astro-ph].
- [5] K. Land and J. Magueijo, *Phys. Rev.* **D72**, 101302 (2005), arXiv:astro-ph/0507289 [astro-ph].
- [6] P. Vielva, E. Martinez-Gonzalez, R. B. Barreiro, J. L. Sanz, and L. Cayon, *Astrophys. J.* **609**, 22 (2004), arXiv:astro-ph/0310273 [astro-ph].
- [7] P. A. R. Ade et al. (Planck), *Astron. Astrophys.* **594**, A16 (2016), arXiv:1506.07135 [astro-ph.CO].
- [8] D. J. Schwarz, C. J. Copi, D. Huterer, and G. D. Starkman, *Class. Quant. Grav.* **33**, 184001 (2016), arXiv:1510.07929 [astro-ph.CO].
- [9] C. L. Bennett et al., *Astrophys. J. Suppl.* **192**, 17 (2011), arXiv:1001.4758 [astro-ph.CO].
- [10] C. R. Contaldi, M. Peloso, L. Kofman, and A. D. Linde, *JCAP* **0307**, 002 (2003), arXiv:astro-ph/0303636 [astro-ph].
- [11] C. Gordon, W. Hu, D. Huterer, and T. M. Crawford, *Phys. Rev.* **D72**, 103002 (2005), arXiv:astro-ph/0509301 [astro-ph].
- [12] E. F. Bunn and A. Bourdon, *Phys. Rev.* **D78**, 123509 (2008), arXiv:0808.0341 [astro-ph].
- [13] S. Adhikari, S. Shandera, and A. L. Erickcek, *Physical Review D* **93**, 023524 (2016), arXiv:1508.06489.
- [14] C. Dvorkin, H. V. Peiris, and W. Hu, *Phys. Rev.* **D77**, 063008 (2008), arXiv:0711.2321 [astro-ph].
- [15] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, *Mon. Not. Roy. Astron. Soc.* **434**, 3590 (2013), arXiv:1303.4786 [astro-ph.CO].
- [16] A. Yoho, S. Aiola, C. J. Copi, A. Kosowsky, and G. D. Starkman, *Phys. Rev.* **D91**, 123504 (2015), arXiv:1503.05928 [astro-ph.CO].
- [17] M. O'Dwyer, C. J. Copi, L. Knox, and G. D. Starkman, *Mon. Not. Roy. Astron. Soc.* **470**, 372 (2017), arXiv:1608.02234 [astro-ph.CO].
- [18] D. Contreras, J. P. Zibin, D. Scott, A. J. Banday, and K. M. Gorski, *Phys. Rev.* **D96**, 123522 (2017), arXiv:1704.03143 [astro-ph.CO].
- [19] A. Yoho, C. J. Copi, G. D. Starkman, and A. Kosowsky, *Mon. Not. Roy. Astron. Soc.* **442**, 2392 (2014), arXiv:1310.7603 [astro-ph.CO].
- [20] C. Gibelyou, D. Huterer, and W. Fang, *Phys. Rev.* **D82**, 123009 (2010), arXiv:1007.0757 [astro-ph.CO].
- [21] D. Sarkar, D. Huterer, C. J. Copi, G. D. Starkman, and D. J. Schwarz, *Astropart. Phys.* **34**, 591 (2011), arXiv:1004.3784 [astro-ph.CO].
- [22] A. Rakic and D. J. Schwarz, *Phys. Rev.* **D75**, 103002 (2007), arXiv:astro-ph/0703266 [astro-ph].
- [23] L. Polastri, A. Gruppuso, and P. Natoli, *JCAP* **1504**, 018 (2015), arXiv:1503.01611 [astro-ph.CO].
- [24] J. Kim and P. Naselsky, *Astrophys. J.* **739**, 79 (2011), arXiv:1011.0377 [astro-ph.CO].
- [25] R. Knight and L. Knox, (2017), arXiv:1705.01178 [astro-ph.CO].
- [26] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, *Mon. Not. Roy. Astron. Soc.* **399**, 295 (2009), arXiv:0808.3767 [astro-ph].
- [27] A. Hajian, (2007), arXiv:astro-ph/0702723 [ASTRO-PH].
- [28] P. A. R. Ade et al. (Planck), *Astron. Astrophys.* **594**, A12 (2016), arXiv:1509.06348 [astro-ph.CO].
- [29] R. Adam et al. (Planck), *Astron. Astrophys.* **594**, A9 (2016), arXiv:1502.05956 [astro-ph.CO].
- [30] P. A. R. Ade et al. (Planck), *Astron. Astrophys.* **571**, A23 (2014), arXiv:1303.5083 [astro-ph.CO].
- [31] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, *Mon. Not. Roy. Astron. Soc.* **451**, 2978 (2015), arXiv:1310.3831 [astro-ph.CO].
- [32] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman, *Astrophys. J.* **622**, 759 (2005), arXiv:astro-ph/0409513 [astro-ph].
- [33] M. Kamionkowski and L. Knox, *Phys. Rev.* **D67**, 063001 (2003), arXiv:astro-ph/0210165 [astro-ph].
- [34] G. Chon, A. Challinor, S. Prunet, E. Hivon, and I. Szapudi, *Mon. Not. Roy. Astron. Soc.* **350**, 914 (2004), arXiv:astro-ph/0303414 [astro-ph].
- [35] D. Molinari, A. Gruppuso, G. Polenta, C. Burigana, A. De Rosa, P. Natoli, F. Finelli, and F. Paci, *Mon. Not. Roy. Astron. Soc.* **440**, 957 (2014), arXiv:1403.1089 [astro-ph.CO].
- [36] N. Aghanim et al. (Planck), *Astron. Astrophys.* **594**, A11 (2016), arXiv:1507.02704 [astro-ph.CO].
- [37] M. Chu, H. K. Eriksen, L. Knox, K. M. Gorski, J. B. Jewell, D. L. Larson, I. J. O'Dwyer, and B. D. Wandelt, *Phys. Rev.* **D71**, 103002 (2005), arXiv:astro-ph/0411737 [astro-ph].
- [38] P. K. Aluri, J. P. Ralston, and A. Weltman, *Mon. Not. Roy. Astron. Soc.* **472**, 2410 (2017), arXiv:1703.07070 [astro-ph.CO].
- [39] P. Vielva, *Adv. Astron.* **2010**, 592094 (2010), arXiv:1008.3051 [astro-ph.CO].
- [40] D. N. Spergel et al. (WMAP), *Astrophys. J. Suppl.* **148**, 175 (2003), arXiv:astro-ph/0302209 [astro-ph].
- [41] G. Hinshaw, A. J. Banday, C. L. Bennett, K. M. Gorski, A. Kogut, C. H. Lineweaver, G. F. Smoot, and E. L. Wright, *Astrophys. J.* **464**, L25 (1996), arXiv:astro-ph/9601061 [astro-ph].
- [42] A. Gruppuso, *Mon. Not. Roy. Astron. Soc.* **437**, 2076 (2014), arXiv:1310.2822 [astro-ph.CO].
- [43] C. L. Bennett, A. Banday, K. M. Gorski, G. Hinshaw, P. Jackson, P. Keegstra, A. Kogut, G. F. Smoot, D. T. Wilkinson, and E. L. Wright, *Astrophys. J.* **464**, L1 (1996), arXiv:astro-ph/9601067 [astro-ph].
- [44] C. L. Bennett et al. (WMAP), *Astrophys. J. Suppl.* **148**, 1 (2003), arXiv:astro-ph/0302207 [astro-ph].
- [45] I. J. O'Dwyer, H. K. Eriksen, B. D. Wandelt, J. B. Jewell, D. L. Larson, K. M. Gorski, A. J. Banday, S. Levin, and P. B. Lilje, *Astrophys. J.* **617**, L99 (2004), arXiv:astro-ph/0407027 [astro-ph].

- [46] A. Slosar and U. Seljak, Phys. Rev. **D70**, 083002 (2004), arXiv:astro-ph/0404567 [astro-ph].
- [47] M. Cruz, P. Vielva, E. Martinez-Gonzalez, and R. B. Barreiro, Mon. Not. Roy. Astron. Soc. **412**, 2383 (2011), arXiv:1005.1264 [astro-ph.CO].
- [48] C. Monteserin, R. B. B. Barreiro, P. Vielva, E. Martinez-Gonzalez, M. P. Hobson, and A. N. Lasenby, Mon. Not. Roy. Astron. Soc. **387**, 209 (2008), arXiv:0706.4289 [astro-ph].
- [49] J. Kim and P. Naselsky, Phys. Rev. **D82**, 063002 (2010), arXiv:1002.0148 [astro-ph.CO].
- [50] J. Kim and P. Naselsky, Astrophys. J. **714**, L265 (2010), arXiv:1001.4613 [astro-ph.CO].
- [51] A. Gruppuso, F. Finelli, P. Natoli, F. Paci, P. Cabella, A. De Rosa, and N. Mandolesi, Monthly Notices of the Royal Astronomical Society **411**, 1445 (2011).
- [52] D. Molinari, private communication (2018).
- [53] R. Adam et al. (Planck), Astron. Astrophys. **594**, A10 (2016), arXiv:1502.01588 [astro-ph.CO].
- [54] C. J. Copi, D. Huterer, and G. D. Starkman, Phys. Rev. **D70**, 043515 (2004), arXiv:astro-ph/0310511 [astro-ph].
- [55] D. J. Schwarz, G. D. Starkman, D. Huterer, and C. J. Copi, Phys. Rev. Lett. **93**, 221301 (2004), arXiv:astro-ph/0403353 [astro-ph].
- [56] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, Mon. Not. Roy. Astron. Soc. **367**, 79 (2006), arXiv:astro-ph/0508047 [astro-ph].
- [57] C. J. Copi, D. Huterer, D. J. Schwarz, and G. D. Starkman, Mon. Not. Roy. Astron. Soc. **449**, 3458 (2015), arXiv:1311.4562 [astro-ph.CO].
- [58] P. A. R. Ade et al. (Planck), Astron. Astrophys. **571**, A1 (2014), arXiv:1303.5062 [astro-ph.CO].
- [59] F. K. Hansen, P. Cabella, D. Marinucci, and N. Vittorio, Astrophys. J. **607**, L67 (2004), arXiv:astro-ph/0402396 [astro-ph].
- [60] H. K. Eriksen, A. J. Banday, K. M. Gorski, F. K. Hansen, and P. B. Lilje, Astrophys. J. **660**, L81 (2007), arXiv:astro-ph/0701089 [astro-ph].
- [61] J. Hoftuft, H. K. Eriksen, A. J. Banday, K. M. Gorski, F. K. Hansen, and P. B. Lilje, Astrophys. J. **699**, 985 (2009), arXiv:0903.1229 [astro-ph.CO].
- [62] Y. Akrami, Y. Fantaye, A. Shafieloo, H. K. Eriksen, F. K. Hansen, A. J. Banday, and K. M. Gorski, Astrophys. J. **784**, L42 (2014), arXiv:1402.0870 [astro-ph.CO].
- [63] S. Adhikari, Mon. Not. Roy. Astron. Soc. **446**, 4232 (2015), arXiv:1408.5396 [astro-ph.CO].
- [64] N. Aghanim et al. (Planck), Astron. Astrophys. **571**, A27 (2014), arXiv:1303.5087 [astro-ph.CO].
- [65] M. Quartin and A. Notari, JCAP **1501**, 008 (2015), arXiv:1408.5792 [astro-ph.CO].
- [66] D. Scott (2018) arXiv:1804.01318 [astro-ph.CO].
- [67] Anaconda, Inc., “Anaconda software distribution.” (2018), [Computer Software. Version 4.3.22].
- [68] E. Jones, T. Oliphant, P. Peterson, et al., “SciPy: Open source scientific tools for Python,” (2001–).
- [69] T. E. Oliphant, *A guide to NumPy* (Trelgol Publishing, USA, 2006).
- [70] J. D. Hunter, Computing in Science & Engineering **9**, 90 (2007).
- [71] F. Pérez and Granger, Computing in Science & Engineering **9**, 21 (2007).

#### Appendix A: Covariance difference and sampling errors

Here we complement the results in Fig. 5 in Sec. IV A to show more information about the differences in the features’ covariance matrix calculated using the synfast simulations and those using the FFP simulations.

The left panel of Fig. 8 shows the *absolute* differences in the coefficients calculated in the two sets of simulations. It provides additional information to the relative differences shown in the right panel of Fig. 5 because the overall size of the correlation coefficients (shown in the left panel of Fig. 5) varies by two orders of magnitude.

The right panel of Fig. 8 shows the sampling error in the correlation coefficients in the 1000 FFP simulations, which we measure by splitting the 100,000 synfast simulations into 100 subsamples. Note that right panel of Fig. 5 shows the ratio between the two panels of Fig. 8.

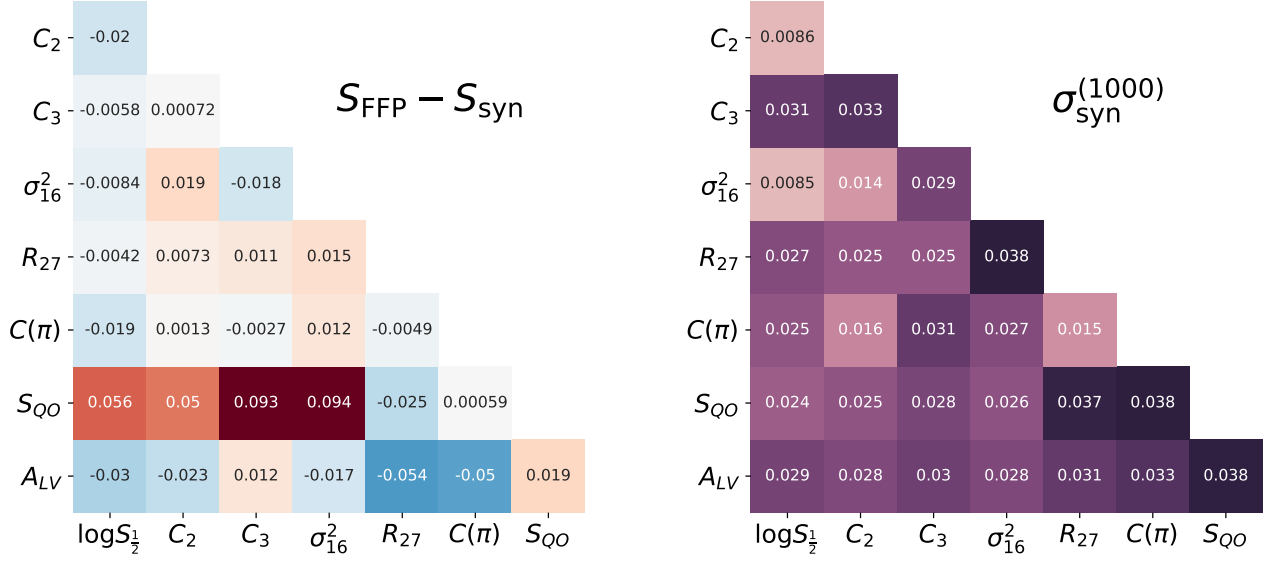


FIG. 8. Left: Absolute difference between the covariance matrix measured from 1000 FFP simulations and from 100,000 synfast simulations. Right: Sample variance error bars for covariance matrix entries measured from 100 subsamples of the 100,000 synfast realizations, each containing 1000 realizations. The right panel of Fig. 5 shows the ratio between the two panels of this figure.