Inference of the optical depth to reionization from low multipole temperature and polarization Planck data

Roger de Belsunce1, 2, Steven Gratton,1 William Coulton2 and George Efstathiou1

1Kavli Institute for Cosmology & Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 OHA, UK
2Kavli Institute for the Physics and Mathematics of the Universe (WPI), The University of Tokyo Institutes for Advanced Study (UTIAS), The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

Accepted 2021 July 27. Received 2021 June 29; in original form 2021 April 7

ABSTRACT

This paper explores methods for constructing low multipole temperature and polarization likelihoods from maps of the cosmic microwave background anisotropies that have complex noise properties and partial sky coverage. We use Planck 2018 High Frequency Instrument (HFI) and updated SRoll2 temperature and polarization maps to test our methods. We present three likelihood approximations based on quadratic cross spectrum estimators: (i) a variant of the simulation-based likelihood (SimBal) techniques used in the Planck legacy papers to produce a low multipole $EE$ likelihood; (ii) a semi-analytical likelihood approximation (momento) based on the principle of maximum entropy; (iii) a density-estimation ‘likelihood-free’ scheme (DELFi). Approaches (ii) and (iii) can be generalized to produce low multipole joint temperature-polarization ($TT$$EE$$EE$) likelihoods. We present extensive tests of these methods on simulations with realistic correlated noise. We then analyse the Planck data and confirm the robustness of our method and likelihoods on multiple inter- and intra-frequency detector set combinations of SRoll2 maps. The three likelihood techniques give consistent results and support a low value of the optical depth to reionization, $\tau$, from the HFI. Our best estimate of $\tau$ comes from combining the low multipole SRoll2 momento ($TT$$EE$$EE$) likelihood with the CamSpec high multipole likelihood and is $\tau = 0.0627^{+0.0050}_{-0.0058}$. This is consistent with the SRoll2 team’s determination of $\tau$, though slightly higher by $\sim 0.5\sigma$, mainly because of our joint treatment of temperature and polarization.

Key words: methods: data analysis – cosmic background radiation – cosmological parameters.

1 INTRODUCTION

Over the last decade, observations of the cosmic microwave background (CMB) (Hinshaw et al. 2013; Henning et al. 2018; Aiola et al. 2020; Planck Collaboration VI 2020), together with measurements of the baryon acoustic oscillation scale from large galaxy surveys (Gil-Marín et al. 2020; Bautista et al. 2021) and many other cosmological observations have transformed cosmology into a high-precision science. In cosmological data analysis, an accurate representation of the likelihood and the ability to model systematics are crucial in order to make reliable inferences from data. Exact likelihoods are often either unknown or computationally expensive to compute. In addition, systematics in the data may bias the results if they cannot be modelled with fidelity and included in the likelihood.

These issues are of particular importance for the measurement of the optical depth to reionization $\tau$ from Planck temperature and polarization CMB maps. Heuristic likelihood models for CMB data on a cut sky with idealized noise properties have been proposed by e.g. Hamimeche & Lewis (2008) and Mangilli, Plaszczynski & Tristram (2015). However, the accuracy of these models is difficult to quantify, especially for cross-correlations of Planck polarization maps which have complex noise correlations and systematics. For these reasons, the Planck collaboration adopted a simulation-based approach to construct a low multipole polarization likelihood from the HFI Planck maps (Planck Collaboration XLVI 2016, hereafter PSRoll1).

In this paper, we apply three likelihood approximations to measure the optical depth to reionization from Planck HFI maps. All three methods use Bayesian statistics to make inferences about models from data. Bayes’ theorem can be used to infer the posterior density $P(\theta|d_0, M)$ of a set of parameters $\theta$ describing a model $M$ from a realization of data $d_0$:

$$P(\theta|d_0, M) = \frac{P(d_0|\theta, M)P(\theta|M)}{P(d_0|M)} \propto P_\phi = \frac{L_\phi\pi}{Z_\phi},$$

(1)

where $P_\phi$ is the posterior, $L_\phi$ the likelihood, $\pi$ the prior, and $Z_\phi$ the evidence. The subscript $\phi$ denotes the dependence on the data set. We compare the simulation-based likelihood (SimBAL) method of PSRoll1, which was used in the Planck 2018 analysis of cosmological parameters (Planck Collaboration VI 2020, hereafter PCP18), with a flexible semi-analytic likelihood approximation (GLASS; Gratton 2017) and a density-estimation ‘likelihood-free’ method1 (DELFi; Alsing et al. 2019). GLASS can easily be adapted to produce a joint temperature-polarization likelihood at low multipoles. However, this

1 ‘Likelihood-free’ (LF) methods are clearly not likelihood-free. What is meant is that the likelihood $L$ is inferred by fitting to numerical simulations rather than being expressed as a simple functional form.

© 2021 The Author(s)

Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
is non-trivial for SIMBAL and DELFI when trying to achieve near optimal results.

The optical depth to reionization provides a measure of the time at which the intergalactic medium (IGM) was reionized by photons produced by early generations of stars and galaxies. Following recombination at \( z \approx 1000 \), the IGM remained almost neutral until reionization. Assuming abrupt reionization at \( z_{\text{re}} \), the Thomson optical depth \( \tau \) is

\[
\tau = \frac{2c\sigma_T (1 - Y_\text{p})}{m_p} \frac{H_0}{\Omega_\Lambda} \frac{\Omega_m}{8\pi G} \left( \Omega_m (1 + z_{\text{re}})^3 + \Omega_\Lambda - 1 \right),
\]

where \( \sigma_T \) is the Thomson cross-section and we have assumed the base Lambda cold dark matter (LCDM) model with a helium abundance by mass of \( Y_\text{p} \) (assuming that helium remains neutral). The Gunn–Peterson test (Gunn & Peterson 1965; Fan et al. 2006) provides strong astrophysical evidence that the IGM was highly ionized by a redshift of \( z = 6.5 \). Using the Planck 2018 base ΛCDM parameters, equation (2) yields a lower limit of \( \tau \approx 0.04 \) for \( z_{\text{re}} = 6.5 \).

In this paper, we have measured \( \tau \) from two sets of Planck maps: the Planck 2018 legacy release (Planck Collaboration II 2020; Planck Collaboration III 2020), which for HFI is based on the SRRoll1 map-making algorithm described in SRRoll1, and the improved map-making algorithm SRRoll2 (Delouis et al. 2019, hereafter SRRoll2). The values of \( \tau \) computed from the low multipole EE spectra alone from these maps are:

\[
\tau_{\text{SRRoll1}} = 0.0506 \pm 0.0086, \quad (\text{PCP18}),
\]

\[
\tau_{\text{SRRoll2}} = 0.0566 \pm 0.0062, \quad (\text{Pagano et al. 2020}).
\]

These estimates improve significantly on the result from WMAP (Hinshaw et al. 2013) of \( \tau = 0.089 \pm 0.014 \). Measurements of \( \tau \) using a pixel-based likelihood on Low Frequency Instrument (LFI) Planck and WMAP data have been presented in Lattanzi et al. (2017) and Natale et al. (2020), albeit with larger uncertainties than in equations (3a) and (3b). Note that the estimates from Planck are just above the Gunn–Peterson limit of \( \tau \approx 0.04 \) inferred from equation (2). This implies that reionization occurred late, i.e. \( z_{\text{re}} \) cannot be much greater than about 6.5 (see for example Kulkarni et al. 2019, and references therein).

Measuring \( \tau \) using the CMB is challenging. At high multipoles, the CMB power spectra are damped by a factor \( e^{-2\tau} \), leading to a degeneracy between the \( \tau \) parameter and the amplitude of the initial cosmological scalar perturbations \( A_s \) (which is partially broken by CMB lensing). Reionization also induces a polarization signal at superhorizon scales in the EE power spectrum leading to a ‘reionization bump’ at low multipoles (\( \ell \approx 20 \)) with an amplitude that scales approximately as \( \tau^2 \). The EE power spectrum at low multipoles can therefore be used to constrain \( \tau \), provided systematics can be kept under control.

An enormous effort has been made to improve the fidelity of the Planck HFI polarization maps. These improvements are presented in detail in SRRoll1 and SRRoll2. In the analysis described in SRRoll2, residual systematics at 100 and 143 GHz are reduced to levels below the notional detector noise levels at multipoles \( \ell \lesssim 10 \), as demonstrated by a number of null tests. SRRoll1 and SRRoll2 describe sets of end-to-end simulations of the two SRRoll pipelines. In this paper, we use these simulations to characterize large-scale systematic modes and correlated noise at low multipoles. We then generate a large number of realizations with realistic noise and systematics over a grid of \( \tau \) values, which are used to train, or calibrate, two of the likelihood models. All of the likelihoods are based on a quadratic cross spectrum (QCS) estimator, which we use to measure the foreground cleaned cross-spectra at low multipoles (\( 2 \leq \ell \leq 29 \)) from the SRRoll1 and SRRoll2 temperature and polarization maps.

This paper is organized as follows: in Section 2 we review the QCS power spectrum estimator. In Section 3, we discuss map compression and foreground cleaning procedures as well as residual systematics in the maps and their contribution to the power spectrum. In Section 3.2 we derive the pixel–pixel noise covariance matrices (NCMs) required for the QCS estimator and the likelihood computations. In Section 4 we present the different likelihood methods used to measure \( \tau \): the simulation-based likelihood (C-SimLow) in Section 4.1, the likelihood approximation scheme (momento) in Section 4.2, and the likelihood-free approach (pydelfi) in Section 4.3. In Section 5 we test our three likelihoods on simulations with realistic correlated noise. In Section 6 we analyse the Planck data and perform cross-checks to validate our results. Section 7 presents our conclusions.

2 QUADRATIC CROSS SPECTRUM ESTIMATOR

To make inferences from large data sets such as CMB maps, data compression is often required to reduce the size of the data vector to a manageable level. Here, we compress maps into summary statistics, namely the angular power spectrum \( C_\ell \), for each mode (\( \ell = TT, TE, EE, BB \ldots \)). Quadratic estimators can be constructed to measure the sky temperature and polarization power spectra on an incomplete sky which have lower variance than traditional pseudo-\( C_\ell \) (PCL) estimators (Tegmark & de Oliveira-Costa 2001; Efstathiou 2006) and are easily computable at low multipoles from low-resolution maps.

Ordering the \( T, Q, \) and \( U \) pixel values as a data vector \( x \), one can write down a quadratic power spectrum estimate \( y_\ell^\prime \) (Tegmark 1997a):

\[
y_\ell^\prime = x_i x_j E_{ij}^{\ell}, \quad (4)
\]

Most of our results are based on the 100 \times 143 full-mission cross-spectra, though we also investigate 100 \times 100 and 143 \times 143 ‘detector set’ cross-spectra.

Throughout the paper, we use the following notation: \( \hat{C}_\ell \) are the undeconvolved and \( \hat{C}_\ell \) the deconvolved power spectra of the data. Theory spectra are denoted by \( C_\ell \).

It is also possible to write down a pixel-based likelihood for low-resolution maps, provided the signal and noise are Gaussian and the noise covariance matrix \( N_\ell \) is known accurately, see Page et al. (2007), Lattanzi et al. (2017), and Natale et al. (2020).
where
\[
E^{\ell} = \frac{1}{2} C^{-1} \frac{\partial C}{\partial \epsilon} C^{-1},
\]
which we will assume is composed of a signal and a noise term, \(C = S + N\), evaluated for a fiducial model. In equation (5), \(C_{\epsilon}\) is the theoretical CMB power spectrum for mode \(\ell\). The Fisher matrix is invertible, unbiased estimates of the power spectra can be computed from
\[
\hat{C}_{\ell} = F_{\ell\ell}^{-1} \left( y_{\ell} - \text{tr} \left[ NE^{\ell} \right] \right),
\]
and the covariance matrix of equation (13) becomes \(\Delta \hat{C}_{\ell} \Delta \hat{C}_{\ell'} = F_{\ell\ell}^{-1} F_{\ell'\ell'}^{-1}\).

Note that if we had used the exact covariance matrix of equation (6), the Fisher matrix in equation (11) would take the more familiar form
\[
F_{\ell\ell} = \frac{1}{2} \left[ \frac{\partial C}{\partial \epsilon} C^{-1} \frac{\partial C}{\partial \epsilon} C^{-1} \right],
\]
and the covariance matrix of equation (13) becomes \(\Delta \hat{C}_{\ell} \Delta \hat{C}_{\ell'} = F_{\ell\ell}^{-1}\). For realistic sky cuts, using the reshaped covariance matrix in place of the exact covariance matrix leads to a negligible increase in variance compared to the optimal estimator, given in equation (4) (see Efstathiou 2006).

In our application, the Planck noise properties in polarization are complex and so it is dangerous to estimate power spectra using equation (12) since this requires the subtraction of a noise term. We therefore modified the quadratic estimator by applying it to cross-spectra of maps \((a)\) and \((b)\) on the assumption that the noise between these maps is uncorrelated (Efstathiou & Gratton 2014). The QCS estimator is
\[
\hat{y}^{(a,b)}_{\ell} = y_{\ell}^{(a)} y_{\ell}^{(b)} E_{ij},
\]
and compute the PCL cross spectrum \(C^{(a,b)}_{\ell}\) of the maps \(x^{(a)}\) and \(x^{(b)}\). The estimator in equation (15) is then
\[
\hat{y}^{(a,b)}_{\ell} = \frac{(2\ell + 1)}{2\Omega_{\ell}} \hat{C}_{\ell}^{(a+b)},
\]
where \(\Omega_{\ell}\) is the solid angle of a single map pixel, each assumed to be of the same size. The QCS estimates can therefore be computed very rapidly for large numbers of simulations, since the matrices \((\hat{C}^{(a)}_{\ell})^{-1}\) and \((\hat{C}^{(b)}_{\ell})^{-1}\) need only be computed once.

The expectation value of equation (15) is
\[
\langle y^{(a,b)}_{\ell} \rangle = F^{(a,b)}_{\ell\ell} \hat{C}_{\ell}^{(a+b)},
\]
and estimates of the power spectra \(\hat{C}_{\ell}^{(a+b)}\) can be recovered by inversion of equation (19) as in equation (12).

The QCS estimator was used, together with the simulation-based likelihood (SimBAl), to analyse the Planck HFI maps in PSRoll1, PCP18, and Pagano et al. (2020). Although the QCS estimator is not ‘optimal’ in any formal sense, it has a significantly lower variance than a PCL estimator applied to the Planck polarization maps. However, in addition to lower variance, the QCS estimator produces estimates of the EE power spectrum with a covariance matrix that is effectively diagonal. It is because the QCS estimates \(\hat{C}_{\ell}^{(a+b)}\) for each multipole are effectively independent that the SimBAl likelihood approach is feasible.

The variance of the QCS estimates is somewhat more complicated than equation (11):
\[
\langle y^{(a,b)}_{\ell} \rangle \langle y^{(a,b)}_{\ell'} \rangle = \langle y_{\ell}^{(a)} \rangle \langle y_{\ell'}^{(b)} \rangle = \left[ 2 S_{1q} S_{ip} + (N_{1p}^{(a)} + N_{1p}^{(b)}) S_{1q} + N_{1p}^{(a)} N_{1q}^{(b)} E_{ij}^{(a,b)} E_{ij}^{(a,b)'} \right].
\]

The error bars shown in plots of the power spectra below (Figs 4 and 5) are computed from equation (20), though this expression is not used in the likelihoods. In this paper, we assume a base \(\Lambda\)CDM model with \(\tau = 0.06\) to compute the signal matrix \(S\). The construction of the NCIs \(N\) (including realistic correlated noise) is described in Section 3.2 and is based on the end-to-end simulations described in Planck Collaboration III (2020) and PSRoll2, respectively, for the analysis of SRoll1 and SRoll2 maps. In contrast, the analysis of Pagano et al. (2020) used a simplified noise model based on the Planck FFP8 simulations (Planck Collaboration XII 2016) for the QCS computations.

3 DATA

The SRoll1 and SRoll2 map-making algorithms are described in detail in PSRoll1, Planck Collaboration III (2020), and PSRoll2. Briefly, the algorithms find global solutions minimizing the variance in the response of each polarized bolometer within a given frequency band with respect to a number of instrumental parameters. Analogue to digital converter non-linearity (ADCNL) introduced.
large polarization systematics at low multipoles in the 2015 Planck HFI maps (Planck Collaboration VIII 2016). These systematics were substantially reduced with the SRoll1 processing used to produce the Planck 2018 HFI legacy maps. Although SRoll1 reduced systematics arising from first-order ADCNL, second-order ADCNL caused temperature to polarization dipole leakage in the maps (Planck Collaboration VI 2020). The SRoll2 map-making algorithm reduced these large-scale polarization systematics for 100 and 143 GHz still further via the following refinements:

(i) the revised ADCNL corrections in SRoll2 obviate the need for fitting an effective gain variation of the bolometers;

(ii) the polarization angle and efficiency for each bolometer were treated as marginalized parameters;

(iii) the thermal dust and CO templates were updated.

These improvements reduced significantly large-scale systematics in the polarization data. The SRoll1 and SRoll2 100 and 143 GHz Q and U maps are compared in figs 4 and 5 of Pagano et al. (2020). In this paper, we have analysed both the SRoll1 and SRoll2 maps, together with their respective sets of end-to-end simulations, so that the reader can assess the impact of changes in the HFI data processing.

3.1 Map compression and foreground cleaning

To apply the QCS estimator, we degrade the high-resolution Planck T, Q, and U maps following a procedure similar to that described in PSRoll1 and PSRoll2. We first apply an apodized mask with \( f_{\text{sky}} = 0.85 \) (as plotted in Fig. 1) to suppress the Galactic plane region.\(^7\)

The maps were then smoothed using the harmonic-space smoothing operator:

\[
w(\ell) = \begin{cases} 
1 & , \ell \leq \ell_1 \\
\frac{1}{2} \left[ 1 + \cos \pi \frac{\ell - \ell_1}{\ell_2 - \ell_1} \right] & , \ell_1 < \ell \leq \ell_2 , \\
0 & , \ell > \ell_2
\end{cases}
\]  

with \( \ell_1 = N_{\text{side}}^U \) and \( \ell_2 = 3N_{\text{side}}^U \), and degraded from \( N_{\text{side}} = 2048 \) (5.03 \( \times \) 10\(^4\) HealPix pixels, Górski et al. 2005) to \( N_{\text{side}}^U = 16 \) (3072 pixels) in the low-resolution maps. We apply the smoothing operator given in equation (21) and a HealPix pixel window function at the map level to match the \( N_{\text{side}}^U = 16 \) low-resolution covariance matrices, \( N_0 \), discussed in Section 3.2.

\(^7\)This procedure is unnecessary for foreground-subtracted simulations, but required for the real data to avoid smearing high-amplitude foreground emission in the Galactic plane to high Galactic latitudes.

Table 1. Cleaning coefficients at low multipoles for temperature and polarization maps. The polarization cleaning coefficients are listed in pairs, depending on whether Planck 30 GHz (\( a_{30}^P \)) or WMAP 22 GHz (\( a_{22}^P \)) polarization maps were used as synchrotron templates. In temperature, we cleaned only for dust using the 353 cleaning coefficients listed in bold face from table 7 of EG19.

| Data set | \( \nu [\text{GHz}] \) | \( a_{353}^T \) | \( a_{353}^P \) | \( a_{30}^T \) | \( a_{30}^P \) | \( a_{22}^T \) | \( a_{22}^P \) |
|----------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| SRoll1   | 100             | 0.0237          | 0.0190          | 0.0126          | 0.0179          | 0.0104          |                  |
|          | 143             | 0.0398          | 0.0402          | 0.0096          | 0.0396          | 0.0071          |                  |
| SRoll2   | 100             | 0.0237          | 0.0193          | 0.0191          | 0.0193          | 0.0189          | 0.0094          |
|          | 143             | 0.0398          | 0.0396          | 0.0102          | 0.0391          | 0.0062          |                  |

The low-resolution maps are foreground cleaned by fitting high- and low-frequency templates. Specifically, we use the 353 GHz maps as a dust template and (in polarization only) either the Planck LFI 30 GHz or WMAP K-band (22 GHz) maps as synchrotron templates. Following Efstathiou & Gratton (2019, hereafter EG19, see sections 7.1 and 8) we minimize cleaned map residuals

\[
\sigma^2 = \sum_i \left( (1 + \alpha_1 + \alpha_2)m_i - \alpha_1 m_i^T - \alpha_2 m_i^P \right)^2 ,
\]  

with respect to the coefficients \( \alpha_1 \) and \( \alpha_2 \) for the two map templates \( m_i^T \) and \( m_i^P \). In polarization, the sum in equation (22) extends over the unmasked pixels in the \( Q \) and \( U \) maps defined by the \( f_{\text{sky}} = 0.70 \) mask plotted in Fig. 1. We therefore determine two sets of coefficients for polarization, which we denote \( \alpha_1^P \) and \( \alpha_2^P \). For temperature, we have applied dust template subtraction with a coefficient \( \alpha_1^T \) and ignored synchrotron, for reasons discussed below.

The template coefficients used in this paper are listed in Table 1. The polarization coefficients are listed in pairs, one pair for each of the SRoll1 and SRoll2 maps to give an impression of the sensitivity of the dust coefficient on the choice of low-frequency template. Polarized dust emission dominates the 100 and 143 GHz \( Q \) and \( U \) maps at low resolution, with synchrotron making a small (but non-negligible) contribution at 100 GHz. As Table 1 shows, the polarization dust coefficients are stable. However, for 100 GHz the amplitude of the 30 GHz coefficient differs between SRoll1 and SRoll2. The SRoll2 polarization cleaning coefficients are in excellent agreement with the coefficients determined by Pagano et al. (2020) (\( \alpha_{30}^P = 0.0186 \), \( \alpha_{22}^P = 0.0095 \) for 100 GHz). As we will show below, ignoring the synchrotron correction in polarization causes shifts in \( \tau \) of a fraction of a standard deviation. Maps cleaned with WMAP K-band are nearly indistinguishable to the ones cleaned with Planck 30 GHz maps. This similarity gives us confidence in our synchrotron cleaning coefficients.

Dust cleaning in temperature using 353 GHz or higher frequencies removes almost all of the foreground emission at low multipoles at 143 GHz leaving noise-free CMB signal over most of the sky (and indistinguishable from the Planck component separated maps), as discussed in detail in EG19. At 100 GHz, the main contaminant, following 353 cleaning, is CO line emission which makes a small but easily detectable contribution to the signal. Since the residual foregrounds are small at low multipoles, we subtract only dust emission in temperature using 353 GHz maps and the cleaning coefficients determined by EG19 (as listed in Table 1). We use the same temperature cleaning coefficients for SRoll1 and SRoll2 since these coefficients are insensitive to the map-making algorithm.

To avoid introducing correlated noise into the QCS spectra, we foreground clean the frequency maps using pairs of half-mission...
Figure 2. Full-mission SRoll2 Q and U maps at 100 and 143 GHz degraded to $N_{\text{side}} = 128$ smoothed with a $\sigma = 2^\circ$ Gaussian beam. The $f_{\text{sky}} = 0.70$ mask of Fig. 1 has been applied. The upper set of maps at each frequency shows the $Q$ and $U$ maps before foreground subtraction. The lower set of maps at each frequency shows the $Q$ and $U$ maps after subtraction of 353 and 30 GHz Planck maps with the template cleaning coefficients listed in Table 1. The colour scale is in units of $\mu$K.

Figure 3. Polarization maps used for the main cosmological results presented in this paper. The plots show the $N_{\text{side}}^U = 16$ foreground cleaned full-mission 100 and 143 GHz $Q$ and $U$ maps. SRoll1 and SRoll2 maps are shown in the upper and lower panels, respectively. The $f_{\text{sky}} = 0.54$ mask of Fig. 1 has been applied. The colour scale is in units of $\mu$K.

3.2 Noise covariance matrices

In this section, we describe how we fit a parametric model to estimates of the pixel–pixel NCMs computed from end-to-end simulations. Once we have a model for an NCM, we can generate large numbers of simulations on the assumption that the noise is Gaussian, allowing us to construct likelihoods as discussed in Section 4. We compute two sets of NCMs, one each for SRoll1 and SRoll2, from the end-to-end simulations of the map-making pipelines. In each case, the fiducial CMB and the input foreground model was subtracted, leaving maps containing noise and map-making systematics. We then construct empirical NCMs, $\hat{N}$,

$$\hat{N} = \frac{1}{n_s} \sum_{i=1}^{n_s} (\mathbf{n}_i - \bar{n})(\mathbf{n}_i - \bar{n})^\top,$$

where $\mathbf{n}_i$ are the simulated sky maps for the Stokes parameters $T$, $Q$, and $U$, $\bar{n}$ a smoothed template of the mean of the maps (see equation

8For HFI, the half-mission maps are constructed by splitting the available rings from the full-mission frequency maps into two halves, i.e. each frequency channel has two half-mission maps: HM1 and HM2.

9SRoll1 simulations (labelled FFP10) available at https://pla.esac.esa.int; SRoll2 simulations available at http://sroll20.ias.u-psud.fr.
Inference of $\tau$ using Planck cross-spectra

3.3 Quadratic temperature and polarization power spectra

As an illustration of our methods, Fig. 4 shows the $TT$, $TE$, $EE$, and $BB$ spectra, computed using the QCS estimator technique presented in Section 2, for the cross-correlation of the 100 and 143 GHz full-mission maps for both S$\text{Roll1}$ and S$\text{Roll2}$ for $2 \leq \ell \leq 31$. The $T$, $Q$, and $U$ maps have been foreground cleaned as discussed in Section 3.1. We apply a mask with $f_{\text{sky}} = 0.54$, shown in Fig. 1, to the $Q$ and $U$ maps. For the $TT$ cross-spectra we compare our results with the $\text{COMMANDER}$ 2018 spectrum (PCP18) and observe good agreement between them, even though the $\text{COMMANDER}$ $TT$ spectrum is computed using a mask with larger sky fraction ($f_{\text{sky}} = 0.86$). In particular, we observe the same behaviour in the spectra: first, at $\ell = 2$ there is a very low value, and, second, there is a ‘dip’ at around $\ell \simeq 20–25$. The error bars on the $\text{COMMANDER}$ spectrum are asymmetric for posterior widths.

To guide the eye, the solid and dashed black lines in Fig. 4 show the $TT$, $TE$, and $EE$ spectra for base $\Lambda$CDM model with $\tau = 0.055$, close to the best-fitting value for the S$\text{Roll1}$ and S$\text{Roll2}$ maps cited in equations (3a) and (3b), and for $\tau = 0.070$, which is disfavoured at about the 2–3$\sigma$ level. The $TT$ and $TE$ spectra have large cosmic variance and so do not provide strong constraints on $\tau$ given the low values of $\tau$ inferred from the EE spectra. The S$\text{Roll1}$ and S$\text{Roll2}$ $TT$, $TE$, and $EE$ spectra are very similar and mainly show differences at low multipoles. The $BB$ spectra are approximately consistent with zero, providing an important null-test for S$\text{Roll1}$ and S$\text{Roll2}$. The $\chi^2$ values divided by the degrees of freedom (29 multipoles) are 1.28 for S$\text{Roll1}$ and 1.13 for S$\text{Roll2}$. Therefore, we use 100 simulations from each set to allow validation tests of the likelihoods as discussed in Section 5.

We approach the problem of fitting a model to $\hat{N}$ as a maximum likelihood inference problem. We assume a Gaussian probability distribution for each noise realization

$$
\mathcal{L} = \mathcal{P}(\hat{N}|M) \equiv \prod_{i=1}^{n_s} \frac{1}{\sqrt{2\pi}|M|} e^{-\frac{1}{2} (n_i - a_i M^{-1} n_i)^T}
$$

where $M$ is the model for the NCM. We assume that $M$ consists of three terms

$$
M = \alpha N_0 + \beta N_1 + \Psi \Psi^T,
$$

(25)

where $\alpha$ and $\beta$ are scaling parameters for two matrices $N_0$ and $N_1$ and the term $\Psi \Psi^T$ models large-scale modes with parameters $\Psi$ as described more fully below. Since we are dealing with $Q$ and $U$ maps, the NCMs $M$ are of size $(2N_{\text{max}}^Q, 2N_{\text{max}}^U)$. We neglect noise in temperature and only fit polarization noise since, the low-resolution $T$ maps are signal dominated to high accuracy. Moreover, we assume that the noise for 100 and 143 GHz maps is uncorrelated. The matrices $N_0$ are the $N_{\text{large}}$ 16 low-resolution map-making covariance matrices (see e.g. Tristram et al. 2011) computed for the parameters of the FFP8 simulations (Planck Collaboration VIII 2016; Planck Collaboration XII 2016). These covariance matrices contain structure representing the scanning strategy, detector white noise, and ‘1f’-type noise, but do not include complexities associated with corrections for ADCNL. Since these matrices were designed to match the 2015 Planck half-mission maps, they do not necessarily match the noise levels of the S$\text{Roll1}$ and S$\text{Roll2}$ maps at high multipoles. The matrices $N_1$ were constructed from the diagonal components $(\sigma^2)^T_j$, $(\sigma^2)^Q_j$, $(\sigma^2)^U_j$ of the $3 \times 3 T$, $Q$, $U$ high-resolution pixel noise estimates produced by the map-making algorithms. These noise estimates were degraded in resolution to $N_{\text{large}}$ 16 (appropriate for a low-resolution map $X$) by computing:

$$
N_1 \equiv \langle X, X \rangle = \sum_{i \in \ell_2} \sum_{p \in q} \sigma_{p q}^2 \frac{(2\ell_2 + 1)(2\ell_1 + 1)}{4\pi} \Omega_p \Omega_q \times P_l(\cos \theta_{pq}) P_l(\cos \theta_{pq}) f_{\ell_1} f_{\ell_2},
$$

(26)

for each $Q$ and $U$ (see App. A of Efstathiou, Gratton & Paci 2009). We assume that the noise is diagonal $(x_i, x_j) = \sigma_{pq}^2 \delta_{pq}$, $\Omega_q$ is the solid angle of a high-resolution map pixel, $P_l$ denotes the Legendre polynomials, and $f_{\ell_1}$ is the smoothing operator applied to the high-resolution maps given in equation (21). The sum

$$
N = \alpha N_0 + \beta N_1
$$

(27)

therefore allows us to model destriping noise correlations with adjustable ‘white noise’ levels, using the matrices $N_0$ and $N_1$ as templates.

The final term in equation (25) models additional large-scale noise correlations in the S$\text{Roll1}$ and S$\text{Roll2}$ end-to-end simulations. Terms up to $\ell_{\text{max}} = 4$ are added to the noise model via $Y \Psi Y^T$, where each column of the matrix $Y$ is a map of the spherical harmonic functions $Y_{lm}(\theta, \phi)$. The square matrix $\Psi$, with dimensions equal to the number of large-scale modes that we wish to fit (42 modes to model $QQ$, $QU$, and $UU$ correlations), controls the covariance of the modes. In the noise fitting procedure we apply the binary polarization mask with $f_{\text{sky}} = 0.54$, shown in Fig. 1, to the $Q$ and $U$ maps. This avoids fitting modes behind the Galactic plane.

To solve our inference problem, we minimize the ‘action’ $S = -\ln(\mathcal{P}(\hat{N}|M))$

$$
S = \frac{n_s}{2} \left( u M^{-1} \hat{N} + \ln |M| \right)
$$

(28)

with respect to the free parameters in our model for $M$. We solve numerically for the scalar parameters $\alpha$ and $\beta$, solving analytically for the $\Psi$ at each step. The matrix $\Psi$ is given by (for the derivation see App. A)

$$
\Psi = (Y^T N_0^{-1} Y)^{-1} \left( Y^T N_0^{-1} \left[ \hat{N} - M \right] N_0^{-1} Y \right) (Y^T N_0^{-1} Y)^{-1}.
$$

(29)

When fitting the NCM $M$ using equation (25) the dominating components are the $1/f$-type noise matrix ($\alpha \approx 1.8$) as well as the large-scale modes up to $\ell_{\text{max}} = 4$. The smoothed low-resolution covariance matrix $\tilde{N}_0$ only subtracted a small amount of power from the diagonal with $\beta \approx -0.1$.

As summarized in the start of this section, uncorrected ADCNL leads to ‘stripy’ residuals in the Planck $Q$ and $U$ maps at 100, 143, and 217 GHz, which are substantially reduced in S$\text{Roll1}$ compared to S$\text{Roll2}$ (see, for example, fig. 4 in Pagano et al. 2020). The end-to-end simulations provide templates for these residuals. Instead of subtracting the mean map averaged over the simulations, $\tilde{x}$, we subtract a smoothed template

$$
\tilde{n} = Y (Y^T N_0^{-1} Y)^{-1} Y^T N_0^{-1} \tilde{x},
$$

(30)

with $\ell_{\text{max}} = 4$ (this is the maximum likelihood solution of the map-making equation, as discussed in e.g. Tagmark 1997b). We have approximated the NCM in equation (30) by the term $\tilde{N}_0$ to avoid having to iterate to obtain a solution for $M$. 

Downloaded from https://academic.oup.com/mnras/article/507/1/1072/6348120 by guest on 24 August 2021
Fig. 4. Low-ℓ QCS estimates for 100 × 143 temperature (TT, TE) and polarization (EE, BB) full-mission cross-spectra of SRoll12 (red) and Planck 2018 (blue) maps with \( D_\ell = (\ell + 1)C_\ell / (2\pi) \). The COMMANDER 2018 TT cross-spectrum (PCP18) (dotted lined) with the associated error bar as a grey shaded region is shown for comparison. The maps are foreground cleaned with the 30 and 353 GHz channel for synchrotron and dust emission, respectively. Theoretical spectra are shown for \( \tau = 0.055 \) (solid line) and \( \tau = 0.070 \) (dashed line).

Fig. 5 shows TT, TE, EE, and BB spectra computed using the QCS scheme for intra- and inter-frequency cross-spectra for combinations of the publicly available SRoll12 detector set\(^{10}\) maps. The T, Q, and U maps are foreground cleaned, following the procedure discussed in Section 3, i.e. the ds1 maps are cleaned using 353 GHz HM1 and 30 GHz HM1 maps and the ds2 maps are cleaned with 353 GHz HM2 and 30 GHz HM2 maps. The temperature and polarization masks are the same as those used as for the previously discussed full-mission cross-spectra and shown in Fig. 1. For comparison, the empty black circles are the 100 full-mission cross-spectra and shown in Fig. 1. For comparison, the black circles show the 100 × 143 full-mission QCS estimates with corresponding error bars.

Fig. 5. Low-ℓ QCS estimates for TT, TE, EE, and BB autospectra and cross-spectra of detector set combinations for 100 and 143 GHz low-resolution SRoll12 maps with \( D_\ell = (\ell + 1)C_\ell / (2\pi) \). The maps are foreground cleaned with the 30 and 353 GHz channels for synchrotron and dust emission, respectively. Theoretical spectra are shown for \( \tau = 0.055 \) (solid line) and \( \tau = 0.070 \) (dashed line). For comparison, the black circles show the 100 × 143 full-mission QCS estimates with corresponding error bars.

set combinations as expected. In the low multipole EE polarization regime that is most constraining for \( \tau \) (2 ≤ \( \ell \) ≤ 12), the estimates scatter between the theoretical curves with \( \tau = 0.055 \) and \( \tau = 0.070 \). This indicates that higher values of \( \tau \) may be favoured for some detector set combinations. At intermediate scales, 10 ≤ \( \ell \) ≤ 25, the spectra scatter around the theoretical curves and no clear trend is visible. However, the 100 ds1 × 100 ds2 intra-frequency cross spectrum shows a few outliers in polarization (purple line). This potentially suggests that the 100 GHz maps are more affected by noise and unresolved systematics than the 143 GHz ones. The illustrated BB spectra are all approximately consistent with the null hypothesis of zero signal power.

4 LIKELIHOODS

In this section we present and compare a simulation-based likelihood, a likelihood-approximation scheme, and a likelihood-free approach.

\(^{10}\)A detector set denotes a combination of bolometers chosen able to fully determine the polarization state of (non-circularly polarized) incoming light: At 100 GHz, ds1 consists of the 100-1a/b and 100-4a/b bolometer pairs and ds2 of the 100-2a/b and 100-3a/b pairs. At 143 GHz, ds1 consists of the 143-1a/b and 143-3a/b pairs, and ds2 of the 143-2a/b and 143-4a/b pairs (PSRoll11).
The methods are called SimBAL, GLASS, and DELFI and their respective likelihoods are named C-SimLow,11 momento, and pydelfi.

4.1 Simulation-based likelihood

The simulation-based likelihood (SimBAL), originally presented in PSRolI, uses low-ℓ QCS estimates of only the EE polarization spectrum to measure the optical depth.

The joint sampling distribution for all the power spectrum elements is in general a function of all the power spectra components defining the model. However, the QCS procedure with reshaping, see equation (7), does a good job of approximately factorizing this distribution by multipole. Then the distribution of the power spectrum elements at a given multipole depends mainly on the theory elements at that multipole. By only considering a single spectrum, EE in this case, the requirement to handle intra-multipole correlations is avoided, motivating an approximate likelihood form:

$$
\mathcal{L}(C|\hat{C}) = \prod_{\ell=\ell_{\text{min}}}^{\ell_{\text{max}}} \mathcal{L}_\ell \left( C_\ell | \hat{C}_\ell \right),
$$

(31)

with \( \ell_{\text{min}} = 2 \) and \( \ell_{\text{max}} = 29 \), a product of one-dimensional functions.

One then uses realizations generated according to a set of theory models to fit parametric forms to each of these one-dimensional sampling distributions. This is done for each ℓ at the input theory \( C_\ell \) values that happen to occur in the models considered. These fits are then evaluated at a realized \( C_\ell \) value set equal to that of the real data \( \hat{C}_\ell \). Finally, a further fit to these numbers, now as a function of the theory \( C_\ell \), gives the effective likelihood function at that multipole.

As the mask and noise do correlate the multipoles even with QCS, one may ask where such effects manifest themselves in the above procedure. At a given ℓ = ℓ₀ say, one can imagine computing the marginalized one-dimensional sampling distribution for \( \hat{C}_\ell \) for a specific theory power spectrum. Because of the couplings, in general this will differ between models even if they happen to share the same value of \( C_\ell \). The fitting procedure above then effectively averages over these distributions. One expects the variations to be relatively modest for plausible models, and hopes that the theory models used to generate the realizations are close enough to reality not to lead to significant errors in the effective averaging.

We therefore need to generate full maps of the observed CMB on which to measure the spectra. Considering the theory \( C_\ell \)'s for each mode over a large region of the parameter space would be computationally costly, since one would typically spend much time exploring low-probability regions. We thus only explore the region of the power spectrum around the theory \( C_\ell \) of interest.

Our implementation of the simulation-based likelihood largely follows that of Planck Collaboration V (2020c). The main difference is that we use Gaussian realizations of our noise fit to the Planck simulations, rather than using the outputs of the noise simulations directly, given the limited number of the latter. This allows us to use many more independent noise realizations throughout the procedure.

We generate 191 theoretical power spectra, \( C_\ell(\tau, \theta) \), uniformly sampled over a range of \( \tau \) values from 0.01 to 0.2 inclusive with a step size of \( \Delta \tau = 0.001 \), where \( \theta \) denotes all the cosmological model parameters defined in Section 6. Only \( A_\ell \) is varied along with \( \tau \), to keep the product \( 10^9 A_\ell e^{-2\tau} \) fixed at 1.870, consistent with a high-ℓ likelihood constraint. For each of these theoretical

11 We implement our own version of the publicly available likelihood SimLow (PSRolI) and call it C-SimLow; differences in implementation will be highlighted below.

Inference of τ using Planck cross-spectra

The General Likelihood Approximate Solution Scheme (GLASS; Gratton 2017) was developed to allow a principled Bayesian analysis of data even in situations where the sampling distribution for the data is not fully known. Our low-ℓ analysis of the CMB polarization is a case in point – because of difficulties in quantifying the noise in the maps, we choose to use quadratic cross-spectra for robustness. However, the joint distribution of the multipoles of such spectra, computed on a masked sky, does not have a simple analytic form. Instead, GLASS assumes one can compute certain moments of functions of the data, here the spectral multipoles, as a function of the parameters of the model under investigation. One can then imagine GLASS uses a maximum entropy construction to compute a least-presumptive sampling distribution consistent with these moments. We now introduce the method, briefly summarizing Gratton (2017), to which the reader is referred for a fuller presentation.

For an initial illustration, consider a situation in which one can obtain the mean \( \hat{x}(q) \) and variance \( \sigma^2(q) \) of some function of the data \( x \) in terms of a model parameter \( q \), ideally analytically but also potentially via forward simulations. Then, one maximizes the entropy of the system

$$
H(P) = -\int dx \ P(x) \ln \frac{P(x)}{\pi(x)}
$$

(33)

subject to the constraints on \( \hat{x}(q) \), \( \sigma^2(q) \), and normalization of \( P(x) \), imposed via Lagrange multipliers. The required values of the multipliers, \( \lambda_{x}(q) \) and \( \lambda_{\sigma^{2}}(q) \), may be solved for, in general numerically, as a function of \( q \). The sampling distribution is then
given as
\[ P(x) = \frac{\pi(x)e^{-\lambda x^2}}{\int dx \pi(x)e^{-\lambda x^2}} = \frac{1}{Z} \pi(x)e^{-\lambda x^2}, \]
where \( Z \) is a theory-dependent normalization constant (also called the evidence or partition function).

Equivalently, this can be understood as solving for the action (or negative log-likelihood) \( S \),
\[ S(x, q) = -\ln \pi(x) + \lambda_x(q)x + \lambda_{xx}(q)x^2 + \ln Z(\lambda(q)). \]

The method naturally extends to multiple statistics \( x_i, i \in [1, \ldots, n] \) fitted to models with multiple parameters \( q^a, a \in [1, \ldots, m] \), and using higher moments:
\[ S(x, q) = -\ln \pi(x) + \lambda_{x}x^T + \lambda_{x|x}^T x + \lambda_{x|x|^2} + \lambda_{x|x|^2|y} + \sum \ln Z(\lambda(q^a)), \]
where summation is implied over multiple indices.

However, the evaluation of the Lagrange multipliers quickly becomes very expensive, requiring the numerical computation of many multidimensional integrals. We can avoid this cost by instead computing more moments. To simplify notation, we introduce a 'meta-index' \( I \) to first run over all indices \( i \), then all pairs of indices \( ij \), and so on. \( X_i \) then runs over the \( x_i \), then the \( x_i x_i \) and so on, and similarly \( \lambda \) runs over the \( \lambda_i \), then the \( \lambda_{ij} \) and so on. We can express the moments of the \( X \) as derivatives of the evidence \( Z \) with respect to the Lagrange multipliers:
\[ \langle X^T \rangle(\lambda) = -\frac{\partial \ln Z}{\partial \lambda_I}, \]
\[ \langle X^T X \rangle(\lambda) = \frac{\partial^2 \ln Z}{\partial \lambda_I \partial \lambda_J}. \]

Next, we differentiate equation (37) with respect to \( q^a \)
\[ \langle X^T \rangle_a = -\frac{\partial^2 \ln Z}{\partial \lambda_I \partial \lambda_J} = -\langle X_i X_j \rangle \lambda_{ij}. \]

Now, differentiating the action in equation (36) yields
\[ S_a = \langle X_i - \langle X_i \rangle \rangle \lambda_{ij}. \]

We can solve equation (39) for the \( \lambda_{ij} \) in terms of the derivatives of the first moments and second-order cumulants. Substituting into (40), and adopting a vector/matrix notation to avoid explicitly writing meta-indices, we obtain
\[ S_a = -\left( X - \langle X \rangle \right) \left( \langle X^T X \rangle \right)^{-1} \left( X \right)_a, \]
which does not depend explicitly on the prior and the Lagrange multipliers. So, obtaining the moments by calculation or simulations, we can compute the gradient of \( S \). This gradient can then subsequently be integrated between two points in parameter space in order to find the difference in approximate log-likelihood between the two models.

For an instructive if overly simple example, consider applying the above procedure in a one-dimensional problem in which the prior \( \pi(x) \) is uniform and the first two moments of \( x \) happen to be calculable as \( \langle x \rangle = \mu \) and \( \langle x^2 \rangle = \mu^2 + \sigma^2 \), where \( \mu \) is a variable parameter of the model and \( \sigma^2 \) is fixed. Assume we will work to linear order in \( \mu \) for \( S \). Then ranges over just one element, with \( X \) simply being \( x \). Knowing that higher orders in \( x \), and thus \( X \), is sufficient to evaluate the single component of \( \langle XX^T \rangle \), which is \( \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2 \). Equation (41) simply reads
\[ S_{\mu} = -\langle x - \mu \rangle \frac{1}{\sigma^2} \]
with the restriction that we can integrate by inspection to find \( S = (x - \mu)^2/\sigma^2 \) up to a constant. For inference of \( \mu \) of this course may be rewritten as \( S = (x - \mu)^2/\sigma^2 + const \), the exact Gaussian result that we might have anticipated from the form of the moments.

For a more complicated example, consider observing a number of vectors \( y \) of Gaussian-distributed components, with the vectors being independent of each other but allowing components within a single vector to be correlated with each other according to a covariance matrix \( C \). The sampling distribution is then
\[ P(y|C) \propto e^{-\frac{1}{2} \sum y_i^2 C^{-1} y_i}, \]
and we see that the components of the observed covariance matrix \( \hat{C} \)
\[ \hat{C} = \frac{1}{2\ell + 1} \sum y_i y_i^T, \]
serve as sufficient statistics for learning about the components of \( C \). The GLASS scheme recovers the posterior associated with (43) simply by working to linear order in the components of \( \hat{C} \) for \( S \); after calculating their first and second moments, with some work equation (41) recovers the optimal result
\[ S_{true} = \left( \ell + 1 \right)^2 \left( \text{tr} C^{-1} \hat{C} + \ln \frac{|C|}{|\hat{C}|} - 1 \right) \]

Hence we see how to use GLASS to compute an approximate likelihood, which we name \( \text{momento} \), for the cross-spectra in our problem. One can in principle compute all of the intra- and inter- \( \ell \) cumulants between all of the (TT, TE and EE) cross-spectra up to some required degree of approximation, assuming the underlying maps are Gaussianly distributed (around some offset noise template). In practice this is relatively easily manageable up to fourth order in the spectra for moderate \( \ell_{max} \). Such moments are the natural generalizations of the following for multiple of a single cross spectrum \( \tilde{C}_{12} \) on the full sky with isotropic noise (and no unsubtracted map mean noise template):
\[ \langle \tilde{C}_{12} \rangle = C, \]
\[ (2l+1)^2 \langle \tilde{C}_{12}^3 \rangle = C^2 + (C + N_{11})(C + N_{22}), \]
\[ (2l+1)^2 \langle \tilde{C}_{12}^3 \rangle = 2C^2 + 6(C + N_{11})(C + N_{22}), \]
\[ (2l+1)^3 \langle \tilde{C}_{12}^4 \rangle = 6(C^4 + 2(C + N_{11})^2(C + N_{22})^2 + 6C^2(C + N_{11})(C + N_{22})). \]

Taking the theory power components themselves as parameters of the theory, using such cumulants we can numerically integrate \( S_{true} \) up along a path from a fiducial model to a model in question, for a selection of degrees of approximation (linear and quadratic, requiring from quadratic up to fourth-order moments of the spectra). We find that even the linear approximation performs well and so typically use this in our work.

To summarize, \( \text{momento} \) uses QCS power spectra, which use a reasonable fiducial model with power spectrum \( C_{\ell}^{fid} \) to construct the appropriate cumulants. Then, for each likelihood evaluation, \( \text{momento} \):

(i) takes as input a set of theory \( C_{\ell} \)’s,
(ii) computes the difference \( \Delta C_{\ell} = C_{\ell} - C_{\ell}^{fid} \) between the theory and the fiducial model;
(iii) uses Romberg integration to compute the change in \( S \) going from the fiducial \( C_{\ell}^{fid} \)’s to the theory \( C_{\ell} \)’s along the line \( C_{\ell}^{fid} + a \Delta C_{\ell} \)

MNRAS 507, 1072–1091 (2021)
in power spectrum space, with \( \alpha \) being a parameter ranging from zero to one. We choose to use a step size in \( \alpha \) of 0.25.

(iv) This requires the computation of gradients of \( S \) with respect to \( \alpha \) at four new positions in power spectrum space for every new likelihood evaluation (at \( \alpha = 0.25, 0.5, 0.75, 1 \)), as those computed at the fiducial model (\( \alpha = 0 \)) can be reused.

(v) The gradients of \( S \) with respect to \( \alpha \) are linear combinations of those of \( S \) with respect to the associated \( C_{\ell} \)'s, and

(vi) these \( \delta S / \delta C_{\ell} \)'s are computed via equation (41), which require both the QCS power spectra of the data and

(vii) the multidimensional moments/cumulants of the QCS power spectra evaluated at the theory model corresponding to each \( \alpha \), via the appropriate multidimensional generalizations of equations (46)–(47) in quick ‘linear’ mode or equations (46)–(49) in the fuller ‘quadratic’ mode.

In general, GLASS is a very flexible scheme to compute principled posteriors where likelihoods are challenging to compute either for computational efficiency or more fundamental reasons. Since the posteriors where likelihoods are challenging to compute either for speed or at the fiducial model (\( \alpha = 0 \)). We choose to use a step size in \( \alpha \) of 0.25.

12 We remind the reader that for the SIMBAL approach an explicit functional form for the distribution of spectra was assumed.

Inference of \( \tau \) using Planck cross-spectra

with the only modification that equation (50) is changed to a product of Jacobians. The intuition is that we decompose the complex mapping into a series of simple transformations that slowly deform the probability density into a distribution that approaches that of the complex mapping.13 We consider a particularly simple series of mappings. For the \( i \)th mapping we perform the following transformation:

\[
f_i(x|\theta) = \frac{x - \mu_j(x^{1-i}, \theta)}{\sigma_j(x^{1-i}, \theta)} ,
\]

i.e. that the \( i \)th output of the mapping is obtained by subtracting and scaling the \( i \)th input by a function of all the previous \((i - 1) \) inputs. This has the nice property that the Jacobian in equation (50) for each transformation has the trivial form of the product of the functions \( \sigma_j(x^{1-i}, \theta) \). Thus, the desired distribution can be written as

\[
P(\hat{C}_\ell|\theta) d^N \hat{C}_\ell = P(u_\ell|\theta) \prod_{j=1}^{D} \prod_{i=1}^{N} \sigma_j(u_\ell, \theta) d^N u_\ell .
\]

This expression of the problem has shifted the complexity from fitting a functional form for the likelihood to identifying a suitable series of mapping functions \( \sigma_j(u_\ell|\theta) \).

To proceed we consider a family of functions \( \sigma_j(u_\ell|\theta; \mathbf{w}) \), parametrized by \( \mathbf{w} \), and optimize the parameters to find the appropriate mapping. Equivalently stated, we solve a variational inference problem: we have a parametrized form for the likelihood \( P(\hat{C}_\ell|\theta; \mathbf{w}) \) and we wish to optimize the parameters so that we can approximate the true likelihood as accurately as possible.

To find the best-fitting weights of the NN, we minimize the Kullback–Leibler divergence \( D_{KL}(P^*|P) \), between the parametric distribution \( P(\hat{C}_\ell|\theta; \mathbf{w}) \) and the true distribution \( P^*(\hat{C}_\ell|\theta) \), which is defined as

\[
D_{KL}(P^*|P) = -\int P^*(\hat{C}_\ell|\theta) \ln \frac{P(\hat{C}_\ell|\theta; \mathbf{w})}{P^*(\hat{C}_\ell|\theta)} d^N \hat{C}_\ell .
\]

The Kullback–Leibler divergence is a measure of the difference between probability distributions; it is a non-negative function that is zero only when the two distributions are identical. By minimizing this function, we minimize the mismatch between our parametric conditional distribution and the true conditional distribution.

As we do not have access to the true distribution, only samples, we perform a Monte Carlo approximation of the Kullback–Leibler divergence using

\[
D_{\text{KL}}(P^*|P) = -\sum_{i=1}^{N_{\text{samples}}} \ln P(\hat{C}_\ell|\theta; \mathbf{w}) ,
\]

which, as the number of samples tends to infinity, approaches equation (54) up to an additive constant.

13 It has been shown that for some sufficiently flexible mappings arbitrary distributions can be modelled via such series of transformations (Huang et al. 2018; Jaini, Selby & Yu 2019).
given sufficient data they can represent any function meaning we have a sufficiently flexible class of functions.

We use the pydelfi implementation of this method to construct a polarization-only and a joint temperature-polarization likelihood, with the precise configuration shown in App. C. When dealing with a high-dimensional problem \( D \geq 30 \) in likelihood-free-inference, score compression is required both to reduce computational cost and give stable results. The greater the degree of compression, the more sub-optimal the likelihood. As a consequence, the results of our \( TTEE \) pydelfi likelihood have larger uncertainties, as discussed in detail in App. C3 (for a more general discussion, see Alsing, Wandelt & Feeney 2018).

5 LIKELIHOOD VALIDATION ON SIMULATIONS

In this section we present tests of the three likelihood methods discussed in Section 4 against simulations. The aim is to investigate whether there are biases or significant differences in their performance. We also test what happens if the input value of \( \tau \) in the simulations is changed to be higher or lower than the fiducial value of 0.06 used to construct the covariance matrices required for the QCS estimator.

Since end-to-end simulations are used to model noise, we have taken care to use an independent subset of the simulations for the tests described in this section. Thus, the noise estimation steps used all but 100 end-to-end simulations for each SRoll1 and SRoll2, leaving the remaining simulations available for likelihood validation. We also generated 100 new Gaussian CMB realizations for each of three values of \( \tau: \tau = 0.05, 0.06, \) and 0.07, using the same CMB realizations for SRoll1 and SRoll2. We use the same masks for the tests as for the data analysis (illustrated in Fig. 1).

Fig. 6 compares the performance of the three likelihood methods for polarization-only \((EE)\) inferences on \( \tau \) for the SRoll2 simulations (the SRoll1 case is very similar). The figure shows posteriors for each of the 100 test simulations using the \( \tau = 0.06 \) CMB realizations. All three likelihoods perform satisfactorily, without significant bias. The mean maximum likelihood values \( \tau_{\text{ML}} \), mean of the posterior widths \( \sigma(\tau) \), and the standard deviation of the maximum likelihood values \( \sigma(\tau_{\text{ML}}) \) found for each likelihood are given in Table 2 (which also lists values for the \( \tau = 0.05 \) and 0.07 CMB realizations).

While investigating the average behaviour of the likelihoods, we have also compared them realization by realization. Fig. 7 shows scatter plots of the maximum likelihood values for each pair of likelihoods for SRoll2 with an input \( \tau = 0.06 \). To guide the eye, the grey shaded area shows the range expected for a \( \pm 1\sigma \) error of \( \delta \tau = 0.006 \). There is a high degree of correlation between all three likelihoods, especially between C-SimLow and pydelfi, as discussed in Section 4.3. This behaviour is expected since the pydelfi approach is effectively a generalization of the simulation-based likelihood C-SimLow (fitting a set of Gaussian’s to the conditional distributions, instead of using a pre-defined functional form). The scatter between C-SimLow and pydelfi is about a sixth of a sigma, and between either C-SimLow or pydelfi and momento it is roughly half a sigma.\(^{15}\) Applied to the same simulations, methodological differences in the likelihood implementation lead to differences in the maximum likelihood value of \( \tau \) of less than a standard deviation.

Table 2 also gives results for simulations in which the CMB realizations are generated from models with both lower (0.05) and higher (0.07) values of \( \tau \) than the fiducial value \( \tau = 0.06 \) used to compute the QCS estimates. No bias is seen for any of the likelihoods confirming that the methods are insensitive to the choice of fiducial cosmology.

We note that the posteriors on \( \tau \) determined from momento are about 10 per cent tighter than those determined from either C-SimLow or pydelfi. The distribution of peak maximum likelihood values of \( \tau \), shown in Fig. 6, is also tighter for momento. We have therefore chosen to use momento as our default low-\( \ell \) likelihood in Section 6 when combining with the high-\( \ell \) \( TTEE \) likelihood. Finally, all methods give average posterior widths that are slightly less than the scatter of their maximum likelihood values. The distribution of maximum likelihood values of \( \tau \) should be closely related to the width of the posterior distribution but is not guaranteed to be the same. The agreement is, however, close enough to demonstrate that the widths of the posterior distributions are not seriously in error.

\(^{15}\)The quoted \( \sigma \)-shifts are calculated as a fraction of the scatter between the \( \tau_{\text{ML}} \) measurements, in other words of \( \sigma(\tau_{\text{ML}}) \).

---

Figure 6. A test of the three likelihoods (C-SimLow, momento, and pydelfi) on 100 simulated signal maps with an input \( \tau = 0.060 \) and 100 noise and systematics maps from the SRoll2 end-to-end simulations. Each posterior per simulation is shown in black and the dashed black line is the mean of the maximum likelihood values for each method. The \( \tau \) value in each top right corner shows the mean of the maximum likelihood values for \( \tau \) over the simulations and the mean of the posterior widths.
0.96 (PCP18). In Section 6.3, we relax the constraint on 10^9 \tau and 6.2 we perform one-dimensional parameter scans in \tau to derive constraints on \tau to \tau (max. likelihood value model 1) and 6.2 we perform one-dimensional parameter scans in \tau to derive the results quoted in equations (3a) and (3b). Section 6.2 presents a more extensive investigation of the parameter space using momento in conjunction with the high-\ell CamSpec v12.SHM likelihood (EG19).

6.1 Constraints using 100 × 143 full-mission QCS

Table 3 summarizes the results from foreground-cleaned 100 × 143 full-mission maps. These results are plotted in Fig. 8 and compared with the results of equations (3a) and (3b) (shown as the blue and

\begin{table}
\centering
\caption{Summary of likelihood tests performed using Gaussian realizations of CMB signal maps and 100 end-to-end SRoll2 simulations for the simulation-based likelihood (C-SimLow), the likelihood approximation scheme (momento), and the density-estimation likelihood-free (pydelfi) method. Here, we make inferences on \tau using low-\ell polarization data only. For each likelihood the mean maximum likelihood values \overline{\tau}_{ML}, mean of the posterior widths \sigma(\overline{\tau}), and the standard deviation of the maximum likelihood values \sigma(\tau_{ML}) are computed. The input \tau for the CMB realizations is denoted by \tau_{in}. No evidence for any bias in the recovered \tau_{ML}'s, even when the CMB is drawn from a distribution that does match the fiducial model with \tau = 0.06, is seen.}
\begin{tabular}{|c|c|c|c|c|c|c|c|}
\hline
Likelihoods & \tau_{in} = 0.050 & \tau_{in} = 0.060 & \tau_{in} = 0.070 \\
& \overline{\tau}_{ML} & \sigma(\overline{\tau}) & \sigma(\tau_{ML}) & \overline{\tau}_{ML} & \sigma(\overline{\tau}) & \sigma(\tau_{ML}) & \overline{\tau}_{ML} & \sigma(\overline{\tau}) & \sigma(\tau_{ML}) \\
\hline
C-SimLow & 0.0503 & 0.0064 & 0.0077 & 0.0603 & 0.0063 & 0.0069 & 0.0703 & 0.0603 & 0.0063 & 0.0066 \\
momento & 0.0498 & 0.0056 & 0.0063 & 0.0600 & 0.0057 & 0.0063 & 0.0704 & 0.0600 & 0.0060 & 0.0065 \\
pydelfi & 0.0496 & 0.0064 & 0.0077 & 0.0597 & 0.0060 & 0.0068 & 0.0697 & 0.058 & 0.0065 \\
\hline
\end{tabular}
\end{table}

\begin{table}
\centering
\caption{Summary of \tau constraints for 100 × 143 full-mission cross-spectra obtained using a simulation-based likelihood (C-SimLow), a likelihood approximation scheme (momento), and a density-estimation likelihood-free (pydelfi) approach. For \tau (EE) we measure \tau only using the low multipole polarization data and for \tau (TT/EE) we compute a joint likelihood for temperature and polarization data.}
\begin{tabular}{|c|c|c|c|}
\hline
Data set & Likelihood & \tau (EE) & \tau (TT/EE) \\
\hline
\text{Planck} 2018 & C-SimLow & 0.0530 ± 0.0071 & – \\
& momento & 0.0507 ± 0.0063 & 0.0527 ± 0.0058 \\
& pydelfi & 0.0517 ± 0.0070 & 0.0513 ± 0.0078 \\
SRoll2 & C-SimLow & 0.0582 ± 0.0057 & – \\
& momento & 0.0581 ± 0.0055 & 0.0604 ± 0.0052 \\
& pydelfi & 0.0588 ± 0.0054 & 0.0580 ± 0.0064 \\
\hline
\end{tabular}
\end{table}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig7}
\caption{Scatter plots between the maximum likelihood values corresponding to the posteriors shown in Fig. 6 for the three pairs of the likelihood methods (C-SimLow, momento, and pydelfi). The grey shaded area shows the typical posterior width for the optical depth \sigma(\overline{\tau}). The black, red, and green dashed lines show linear fits to each set of points. Their correlation coefficients (slopes of the linear fit) are given in the legend. There is a high degree of correlation between all three methods, especially between C-SimLow and pydelfi.}
\end{figure}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig8}
\caption{Summary plot of \tau posteriors obtained from 100 × 143 full-mission Planck 2018 and SRoll2 maps using a simulation-based likelihood (C-SimLow), a likelihood approximation scheme (momento), and a likelihood-free inference (pydelfi) approach. pydelfi and momento measure \tau by, first, using only polarization (EE) and, second, using temperature and polarization data combined (TT/EE). The vertical dashed lines and corresponding shaded regions indicate maximum likelihood values only using EE polarization data, for Planck 2018 (SRoll1) (PCP18) and SRoll2 (Pagano et al. 2020).}
\end{figure}
red dashed lines, respectively, together with 1σ errors shown by the red and blue shaded areas). For all three likelihood approximations, our results reproduce the upward movement in τ between SRoll1 and SRoll2 (as noted in Delouis et al. 2019). Furthermore, the results from all three likelihoods are consistent with each other. There are, however, some interesting features that are worth noting:

(i) The TTTEEE pydelfi results have larger error bars than the EE results alone, even though additional data are included in the TTTEE activation (though the best-fitting value of τ hardly changes). As outlined in Section 4.3 one of the drawbacks of likelihood-free inference is that higher dimensional problems (D ≥ 30) require additional data compression. In our TTTEE EEE implementation we compressed the 84 power spectrum components for 2 ≤ ℓ ≤ 29 into three summary statistics, one each for TT, TE, and EE, whereas for the EE-only implementation we were able to avoid compression entirely, using all 29 EE power spectrum multipole. The larger TTTEE pydelfi error is a consequence of lossy compression which actually degrades the EE block. In fact, we found that a score-compressed EE posterior from pydelfi gives a maximum likelihood value for τ that is lower by ~0.25σ compared to the results without compression. With compression the TE data does actually pull the posterior upwards by ~0.3σ, largely cancelling the effects of compression on EE. This is discussed further in App. C3.

(ii) For momento, adding TT and TE spectra to EE causes upward shifts in τ of approximately 0.002 (~0.4σ) for both SRoll1 and SRoll2 τ values. The posterior width is reduced by a modest ~5 per cent. This behaviour is consistent with the parameter-shift criteria developed by Gratton & Challinor (2020).

(iii) The SRoll2 likelihoods consistently yield higher values of τ, and with slightly tighter errors, than the corresponding SRoll1 likelihoods.

(iv) The errors on τ from our application of C-SimLow to SRoll1 are about 20 per cent smaller than those quoted in equations (3a). There are two reasons for this: (a) the results of PPC18 used a sub-optimal noise model based on Planck FFP8 end-to-end simulations for the QCS computations; (b) we subtracted a smoothed noise template at the map level (see equation 30), which reduces the size of the posterior widths by ~5–10 per cent for SRoll1, as explained in more detail in App. B.

Recently, a new set of Planck maps for the LFI and HFI frequency bands have been developed (Planck Collaboration LVII 2020, hereafter NPIPE). As with SRoll1 and SRoll2, a set of systematic templates are fitted as part of the map-making stage. However, amongst other differences, NPIPE retains the CMB Solar dipole in each map and uses foreground polarization priors at 30, 217, and 353 GHz to break parameter degeneracies. The use of polarization priors leads to a suppression of the polarization signal at low multipoles, necessitating the calibration of EE power spectrum transfer functions from end-to-end numerical simulations. The transfer functions corrections are quite large for the EE multipole ℓ = 2–7 that contain most of the information on τ. The analysis of the NPIPE 100 × 143 EE spectrum presented in Planck Collaboration LVII (2020) gives a value for τ that is lower by 1.2σ–1.6σ compared to the results of Table 3. The τ results from NPIPE are therefore broadly in agreement with those from SRoll1 and SRoll2.

6.2 The optical depth from inter- and intra-frequency detector set combinations of SRoll2 maps

To assess the robustness of the results of the previous section, we have analysed the inter- and intra-frequency spectra computed from SRoll2 detector set maps. The SRoll2 cross-spectra used in this subsection are shown in Fig. 5.

We stress that the likelihoods for each map pair have been computed/trained afresh using the appropriate detector set NCMs constructed from the relevant simulations. The results of the cross-checks, for each of the three likelihoods, are presented in Table 4. Each column lists the mean value for τ and the associated posterior width for the indicated spectrum combination.

Focusing on the results from momento, the posterior widths for 143ds1 × 143ds2 are about 5 per cent smaller than for 100ds1 × 100ds2, which is expected because the 143 GHz maps are less noisy than the 100 GHz maps. The reduction in errors is, however, smaller in the other two likelihoods. More significantly, at multipoles ℓ = 3–5 the 143ds1 × 143ds2 EE spectrum lies above the τ = 0.055 theoretical line. As a consequence, for all likelihoods, the 143ds1 × 143ds2 value of τ is higher than that for the 100 × 143 spectra by about 1.4–2σ. The 100ds1 × 100ds2 also shows a preference for higher values of τ compared to the 100 × 143 spectra, but to a lesser extent. This is suggestive of correlated residual systematics in the 100 GHz detset and 143 GHz set maps which partially cancel when 100 GHz set maps are cross-correlated against 143 GHz set maps (as discussed by Pagano et al. 2020). The effects are relatively small, but in agreement with the conclusions of Pagano et al. (2020), our results suggest that the 100 × 143 spectra are likely to provide the most reliable constraints on τ.

The last row in Table 4 shows the mean of τ values for all of the inter- and intra-frequency cross detector set spectra constraints (ignoring correlations). The 100ds1 × 100ds2 and 143ds1 × 143ds2 results pull the means to slightly higher values of τ compared to the 100 × 143 full-mission results, but only by about 0.5σ. Thus, while there is some evidence of small systematic-related biases in the 100ds1 × 100ds2 and 143ds1 × 143ds2 τ values, the net effect of residual systematics on the 100 × 143 full-mission results are probably at the level of a standard deviation or less. This statement depends on the fidelity of the noise and systematics simulations.

Columns 3 and 4 of Table 4 illustrate the impact of polarized synchrotron cleaning on τ. (This test was done only for the C-SimLow likelihood). As expected, the effect on τ is most pronounced for 100ds1 × 100ds2, with synchrotron cleaning lowering τ by about 1σ and bringing it into closer agreement with the 100 × 143 full-mission result. However, the effects of synchrotron cleaning on the 100 × 143 and 143ds1 × 143ds2 spectra are significantly smaller, leading to changes in τ of ~0.3σ. Synchrotron cleaning, while non-negligible, is not a critical factor in the τ constraints from the 100 × 143 spectra.

6.3 Full Monte Carlo Markov Chain parameter exploration

We explore the full ΛCDM cosmological parameter space by combining our full-mission 100 × 143 momento EE and TTTEE likelihoods at low multipoles for both SRoll1 and SRoll2 with the high-ℓ Camspec v12.5HM (TTTEE) likelihood (which uses SRoll1 maps). The EE momento likelihood uses the multipole range 2 ≤ ℓ ≤ 29 and is utilized with the Planck 2018 low-ℓ TT likelihood over the same multipole range. The momento TTTEE likelihood uses the multipole range 2 ≤ ℓ ≤ 10 to speed up the low multipole likelihood evaluations. The multipoles 11–29 in TE and EE
have very little constraining power on \( \tau \) and so little information on \( \tau \) is lost by truncating the \( \momento\ TTTEEE\ ) likelihood used with the \( \text{Planck} \) 2018 low-\( \ell \) \( TT\ ) likelihood over the multipole range \( 11 \leq \ell \leq 29 \), so that there are no multipole gaps in the \( TT\ ) likelihood.

Table 5 lists the results of full MCMC exploration of the parameters of the base \( \Lambda CDM\ ) cosmology. These results are similar to those summarized in Table 3 for the one-dimensional \( \tau \) scans. The \( \SRoll2\ ) results for \( \tau \) are about 1\( \sigma \) higher than those from \( \SRoll1\ ), and the \( \momento\ TTTEEE\ ) \( \tau \) results give values for \( \tau \) that are about 0.3\( \sigma \) higher than those using the \( EE\ ) likelihoods. The values for \( 10^9 A_{e^{-2r}} \) are also within about 1\( \sigma \) of the value assumed for the one-dimensional \( \tau \) scans.

We can compare the results from Section 5 with those of full parameter analyses using the \( \text{SIMBAL} \ ) likelihood combined with the low multipole \( TT\ ) likelihood and the \( \text{Plik} \ ) high multipole \( TTTEEE\ ) likelihood reported in PCP18 and Pagano et al. (2020):

\[
\tau = 0.0544^{+0.0070}_{-0.0081}, \quad \SRoll1, \quad (56a)
\]

\[
\tau = 0.0591^{+0.0054}_{-0.0068}, \quad \SRoll2. \quad (56b)
\]

Our results using \( \tau \) are higher by about 0.3–0.5\( \sigma \) and (formally) have slightly smaller error bars.

Fig. 9 illustrates the changes to the \( \tau \) constraints caused by switching from \( \text{SimLow} \ ) to \( \text{momento} \ ) EE likelihoods and then to \( \text{momento}\ TTTEEE\ ). This figure shows contours in the \( \ln 10^9 A_{e^{-2r}}-\tau \) plane for base \( \Lambda CDM\ ). In each case we use the same high-\( \ell \) \( \text{CamSpec} \ ) v12.5HM \( TTTEEE\ ) likelihood for \( 30 \leq \ell \leq 2500 \) and so only the low-\( \ell \) likelihoods change: grey contours for \( \text{SimLow} \ ), red contours for \( \SRoll1 \ ) \( \momento\ ) likelihoods, and blue for \( \SRoll2\ ) \( \momento\ ) likelihoods. Interestingly, the \( \momento\ TTTEEE\ ) likelihood disfavours values of \( \tau \leq 0.04 \) that are already excluded by the Gunn–Peterson test (see equation 2). In other words, the posteriors of the \( \momento\ ) likelihood are consistent with what we know about the IGM.

The main results of this section are summarized in Fig. 10. The plot to the left shows the marginalized posterior distribution for \( \tau \) for \( \text{CamSpec} \ ) combined with the \( \SRoll1\ ) likelihoods as described above. This shows the small shifts in \( \tau \) when we use the \( \SRoll1\ ) \( \momento\ ) likelihoods in place of the \( \text{Planck} \ ) 2018 low-\( \ell \) \( EE\ ) likelihood. The right-hand plot shows how the posteriors change if we use the \( \SRoll2\ ) \( \momento\ ) likelihoods. The \( \momento\ ) constraints shift to higher values of \( \tau \) as a consequence of the changes to the HFI map-making. We take as our ‘best’ estimate of \( \tau \) and redshift of reionization, \( z_{re} \), the results from the combined \( \text{CamSpec} + \SRoll2\ ) \( \momento\ TTTEEE\ ) likelihood:

\[
\tau = 0.0627^{+0.0059}_{-0.0058}, \quad z_{re} = 8.51 \pm 0.52. \quad (57)
\]
In this paper, we have developed and compared three likelihood techniques on the Planck SRoll1 and SRoll2 maps. The first is a variant of the SimBAL scheme described in PSRoll1 and PSRoll2, but using more accurate simulation-based NCMs to construct quadratic cross-spectra and to generate a large number of independent noise realizations. The second (momento) is based on the GLASS maximum entropy approach developed by Gratton (2017) and the third (pydefi) is a density-estimation ‘likelihood free’ scheme that follows closely the implementation described by Alsing et al. (2019). The momento and pydefi approaches can be generalized to construct low multipole TT/EE likelihoods. (Though not explored in this paper, it is straightforward to adapt these schemes to develop likelihoods incorporating other low multipole spectra, e.g. BB, ET.)

Our main conclusion is that all three likelihood methods are in good agreement and support the conclusions on \( \tau \) reported in PCP18 and Pagano et al. (2020); we do, however, see small differences between the likelihoods as summarized in Tables 3 and 4. Using only the spectra at low multipoles, our results tend to give higher values of \( \tau \) than those using SimBAL (equations 3a) and 3b) by up to \( \sim 0.8\sigma \). However, if we include the high multipole CamSpec \( \text{TT/EE} \) likelihood, the results for \( \tau \) using the SRoll2 momento EE likelihood is very close to that given in equation (56b) though with a smaller formal error.

We constructed low multipole TT/EE likelihoods using momento and pydefi. For momento, using a TetEE likelihood leads to smaller errors on \( \tau \) than using EE alone, as expected. However, we had to apply data compression to produce a pydefi TetEE likelihood that was numerically fast and robust enough for likelihood analysis. This resulted in a loss of information and to \( \tau \) constraints that had slightly larger errors using pydefi TetEE compared to using pydefi EE, though with no evidence of any bias; see App. C3 for a detailed discussion.

We also made a thorough analysis of different detector set data splits at 100 and 143 GHz, as summarized in Table 4. For all likelihoods, the 100ds1 × 100ds2 and 143ds1 × 143ds2 spectra give \( \tau \) values that are higher than those from the baseline 100 × 143 full-mission analysis by between 0.8 and 2σ. There is therefore evidence that within a frequency band there remain correlated systematic effects (as is apparent visually from Fig. 3) that bias \( \tau \) high by \( \sim 0.01 \). The series of null tests described in PSRoll2, together with the absence of any statistically significant B-mode signal at low multipoles, suggests that the 100 × 143 full-mission cross-spectra should provide unbiased estimates of \( \tau \). The changes between PSRoll1 and PSRoll2 suggest an upper bound of about 1σ to biases in \( \tau \) caused by residual systematics.

As noted above, the likelihood techniques explored here have wider applications and can be adapted to other problems involving low multipole polarization maps, particularly if the maps have complex noise properties. An obvious example is the measurement of the tensor-to-scalar ratio \( r \), in addition to \( \tau \), from the forthcoming CMB satellite LiteBIRD (Sugai et al. 2020).

ACKNOWLEDGEMENTS

RdB is grateful to Pablo Lemos, Oliver Friedrich, and Will Handley for fruitful discussions regarding likelihood-free inference, to Erik Rosenberg for help with CamSpec, and to Paul Murdin for suggestions on an early version of the manuscript. We are indebted to the members of the Planck collaboration for their enormous efforts in producing such a wonderful set of data and especially thank the Bware team for their SRoll2 products.
References

Aiola S. et al., 2020, J. Cosmol. Astropart. Phys., 2020, 047
Alsing J., Wandelt B., Feeney S., 2018, MNRAS, 477, 2874
Alsing J., Charnock T., Feeney S., Wandelt B., 2019, MNRAS, 488, 4440
Bautista J. E. et al., 2021, MNRAS, 500, 736
Carassou S., de Lapparent V., Bertin E., Le Borgne D., 2017, A&A, 605, A9
Davies F. B., Hennawi J. F., Eilers A.-C., Lukić Z., 2018, ApJ, 855, 106
Delouis J. M., Pagano L., Mottet S., Puget J. L., Vibert L., 2019, A&A, 629, A38 (PSRoll1)
Efstathiou G., 2006, MNRAS, 370, 343
Efstathiou G., Gratton S., 2014, Planck Internal Communication.
Efstathiou G., Gratton S., 2019, preprint (arXiv:2010.00483) (EG19)
Efstathiou G., Gratton S., Paci F., 2009, MNRAS, 397, 1355
Fan X. et al., 2006, AJ, 132, 137
Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, PASP, 125, 306
Germain M., Gregor K., Murray I., Larochelle H., 2015, preprint (arXiv:1502.03509)
Gillet-Marlin H. et al., 2020, MNRAS, 498, 2492
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Gratton S., 2017, preprint (arXiv:1708.08479)
Gratton S., Challinor A., 2020, MNRAS, 499, 3410
Gun J. E., Peterson B. A., 1965, ApJ, 142, 1633
Hahn C., Vakili M., Walsh K., Hearin A. P., Hogg D. W., Campbell D., 2017, MNRAS, 469, 2791
Hamimeche S., Lewis A., 2008, Phys. Rev. D, 77, 103013
Handley W. J., Hobson M. P., Lasenby A. N., 2015a, MNRAS, 450, L61
Handley W. J., Hobson M. P., Lasenby A. N., 2015b, MNRAS, 453, 4384
Henning J. W. et al., 2018, ApJS, 852, 97
Hinsdill G. et al., 2013, ApJS, 208, 19
Hinton S. R., 2016, J. Open Source Softw., 1, 00045
Huang C.-W., Krueger D., Lacoste A., Courville A., 2018, preprint (arXiv:1804.00779)
Jain P., Selby K. A., Yu Y., 2019, preprint (arXiv:1905.02325)
Jeffrey N., Alsing J., Lanusse F., 2021, MNRAS, 501, 954
Kacprzak T., Herbel J., Amara A., Réfrégier A., 2018, J. Cosmol. Astropart. Phys., 2018, 042
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kulkarni G., Keating L. C., Haeffel M. G., Bosman S. E. F., Puchewein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Lattanzi M. et al., 2017, J. Cosmol. Astropart. Phys., 2017, 041
Leclercq F., 2018, Phys. Rev. D, 98, 063511
Lemos P., Jeffrey N., Whiteway L., Lahav O., Lubsken N., Hoffman Y., 2021, Phys. Rev. D, 103, 023009
Lin C.- A., Kilbinger M., 2015, A&A, 583, A70
Lueckmann J.-M., Bassetto G., Karalesos T., Macke J. H., 2018, preprint (arXiv:1805.09294)
Mangilli A., Plasszczynski S., Tristram M., 2015, MNRAS, 453, 3174
Natale U., Pagano L., Lattanzi M., Migliaccio M., Colombo L. P., Gruppuso A., Natali P., Polenta G., 2020, A&A, 644, A32
Pagano L., Delouis J. M., Mottet S., Puget J. L., Vibert L., 2020, A&A, 635, A99
Page L. et al., 2007, ApJS, 170, 335
Papamakarios G., Murray I., 2016, NIPS 2016
Papamakarios G., Sterratt D. C., Murray I., 2018, preprint (arXiv:1805.07226)
Pearson K., Jeffery G. B., Elderton E. M., 1929, Biometrika, 21, 164
Planck Collaboration VIII, 2016, A&A, 594, A8
Planck Collaboration XII, 2016, A&A, 594, A12
Planck Collaboration XI, 2016, A&A, 596, A107 (PSRoll1)
Planck Collaboration II, 2020, A&A, 641, A2
Planck Collaboration III, 2020, A&A, 641, A3
Planck Collaboration V, 2020, A&A, 641, A5
Planck Collaboration VI, 2020, A&A, 641, A6 (PCP18)
Planck Collaboration VII, 2020, A&A, 643, A42
Smyth P., Wolpert D., 1998, NIPS 1997, 668
Smyth P., Wolpert D., 1999, Mach. Learn., 36, 59
Sugai H. et al., 2020, J. Low Temp. Phys., 199, 1107
Tegmark M., 1997a, Phys. Rev. D, 55, 5895
Tegmark M., 1997b, Phys. Rev. D, 56, 4514
Tegmark M., de Oliveira-Costa A., 2001, Phys. Rev. D, 64, 063001
Torrado J., Lewis A., 2021, J. Cosmol. Astropart. Phys., 2021, 057
Tristram M., Filliard C., Perdereau O., Plasszczynski S., Stompor R., Touze A., 2011, A&A, 534, A88
Uria B., Côté M.-A., Gregor K., Murray I., Larochelle H., 2016, preprint (arXiv:1605.02226)
Weyant A., Schafer C., Wood-Vasey W. M., 2013, ApJ, 764, 116

APPENDIX A: ANALYTIC SOLUTION TO THE NOISE INFERENCE PROBLEM

In order to minimize equation (28) analytically for $\Psi$, we first consider the variation in $\delta S$ caused by small changes in $M$. Omitting the $n_{\perp}/2$ pre-factor, we have:

$$\delta S = \text{tr}(\delta M) \delta N + \delta \ln |M|$$

(A1)

$$= \text{tr}[-M^{-1}(\delta M)M^{-1}\delta N + M^{-1}(\delta M)]$$

(A2)

$$= \text{tr}[-M^{-1}(\delta \Psi)M^{-1}\delta N + M^{-1}(\delta \Psi)M^{-1}]$$

(A3)

using the standard results for matrices that $\delta M^{-1} = -M^{-1}(\delta M)M^{-1}$ and $\delta \ln |M| = \text{tr}^{-1}(\delta M)$, and $\delta M = Y \delta \Psi Y^T$. Next, we use the cyclic property of the trace to obtain:

$$\delta S = \text{tr}[\delta \Psi (Y^{-1}M^{-1}(M - \delta N)M^{-1})Y]$$

(A4)

At the minimum we require $\delta S = 0$ for arbitrary $\delta \Psi$ and thus need

$$Y^{-1}M^{-1}(M - \delta N)M^{-1}Y = 0.$$

(A5)
We now use the generalized Sherman–Morrison–Woodbury formula
\[ M^{-1} = N^{-1} - N^{-1}Y(\Psi^{-1} + Y^TN^{-1}Y)^{-1}Y^TN^{-1} \]  
(A6)
(and adding and subtracting \( \Psi^{-1} \)) to rewrite \( Y^TM^{-1} \) as
\[ Y^TM^{-1} = Y^T(N^{-1} - N^{-1}Y(\Psi^{-1} + Y^TN^{-1}Y)^{-1}Y^TN^{-1}) \]
(A7)
\[ = Y^TN^{-1} - (Y^TN^{-1}Y + \Psi^{-1} - \Psi^{-1}) \]
\[ \times(\Psi^{-1} + Y^TN^{-1}Y)^{-1}Y^TN^{-1} \]
(A8)
\[ = \Psi^{-1}(\Psi^{-1} + Y^TN^{-1}Y)^{-1}Y^TN^{-1}. \]
(A9)
Substituting this and its transpose into Equation (A5) yields
\[ Y^TN^{-1}[N + YY^T - \tilde{N}]N^TY = 0. \]  
(A10)
Rearranging this for \( \Psi \) then gives the desired Equation (29).

APPENDIX B: SMOOTHED TEMPLATE SUBTRACTION

This appendix discusses the effect of smoothed template subtraction. The ADCNL effectively leads to CMB-independent offsets in Planck HFI maps. The CMB signal can then suffer chance correlations with these offsets, leading to additional scatter in power spectra. This effect is seen in simulations and, as discussed in Section 3.2, can potentially be mitigated by computing a smoothed estimate of the offset for each map and then subtracting the appropriate estimate from each of the QCS input maps. Applying this prescription to SRoll1 spectra leads to a \( \sim 10 \) per cent reduction in the posterior width for \( \tau \) from SRoll1 and to a \( \sim 5 \) per cent reduction from SRoll2. This is consistent with the hypothesis that SRoll2 better reduces large-scale residuals than SRoll1 and so sees less of an improvement. Note that the procedure of removing smoothed templates from real data only leads one closer to the truth if the simulations from which the smoothed templates are computed are relatively accurate not only in the power spectrum domain but also in the map domain.

Fig. B1 shows the effect of smoothed template subtraction on the \( \tau \) posterior for Planck 2018 data using C-SimLow (EE). The solid line yields a posterior of \( \tau = 0.0521 \pm 0.0077 \) where no smoothed templates have been subtracted. The dashed line yields a posterior \( \tau = 0.0530 \pm 0.0071 \).

**Figure B1.** Effect of smoothed noise template subtraction on the \( \tau \) posterior computed from 100 \( \times \) 143 full-mission cross-spectra of Planck 2018 maps. This test is performed using the simulation-based likelihood C-SimLow. The solid (dashed) line shows the posterior obtained when the subtraction is (is not) performed.

**Figure B2.** QCS estimates for \( TE, EE \) and \( BB \) 100 \( \times \) 143 full-mission foreground-cleaned cross-spectra computed for the difference between SRoll2 and SRoll1 maps. The black points show the spectra computed before any smoothed templates are removed, while the red points show the spectra computed after the maps have their corresponding templates subtracted. The error bars for the cross-spectra are taken from the \( BB \) spectrum as the signal has been removed. The \( TT \) residual spectrum is not shown as it is consistent with zero. Note that the \( BB \) errors plotted on the \( TE \) spectrum are smaller than the points at low \( \ell \); there are significant changes (relative to the \( BB \) noise) between SRoll1 and SRoll2 in \( TE \).

**B1 Comparison between SRoll1 and SRoll2**

Power spectra of the difference maps computed between SRoll1 and SRoll2 are shown in Fig. B2. The error bars are based on the error estimates obtained from the \( BB \) spectra since no signal should be left in the difference maps. Both maps have been foreground cleaned with the respective maps, as detailed in Section 3.1. The red points show the spectra computed from maps that have corresponding smoothed templates (denoted \( n_1 \) for SRoll1 and \( n_2 \) for SRoll2) subtracted. The black points show the difference at the map level without subtracting the templates. We note that the subtractions only have large effects for \( \ell \leq 4 \) (as one would expect given that this is the maximum multipole retained in the templates) and, except for the quadrupole, reduce the residuals. With or without the template subtractions, the residual cross-spectra show differences for the very low multipoles below \( \ell \sim 10 \) for \( TE, EE \), and \( BB \). We do not show the \( TT \) spectrum since it is signal dominated and, as shown in Fig. 4, the \( TT \) spectra are almost identical for SRoll1 and SRoll2. This test confirms that the SRoll2 map-making algorithm indeed alters large-scale polarization features relative to SRoll1.
In the present Appendix, we will elaborate on the technical details of the likelihood-free inference code pydelfi (Alsing et al. 2019), based on theoretical ground work by Leclercq (2018), Alsing et al. (2018), Lueckmann et al. (2018), and Papamakarios, Sterratt & Murray (2018). These methods provide a means of performing Bayesian inference when it is possible to produce high-accuracy simulations of the observables in question, even when it is not possible to write down an explicit analytic expression for the likelihood. Examples of such problems include measuring cosmological parameters from type Ia supernovae (Weyant, Schafer & Wood-Vasey 2013), weak lensing peak counts (Lin & Kilbinger 2015; Jeffrey, Alsing & Lanusse 2021), analysing the galaxy-halo connection (Hahn et al. 2017), inferring photometric redshifts and size evolution of galaxies (Carassou et al. 2017), measuring cosmological redshift distributions (Kacprzak et al. 2018), estimating the ionizing background from the Lyman-α and Lyman-β forests (Davies et al. 2018), and inferring the sum of the masses of Andromeda and the Milky Way (Lemos et al. 2021).

The key principle behind LF methods is the understanding that a simulated observable, d, and the input parameters, θ, are a sample from the joint probability distribution P(d, θ|M) conditioned on the theoretical model. For compactness we will hereafter suppress the conditional aspect of the distribution and make the assumption that our simulations accurately represent the true physical process.17 With a sufficiently large number of samples from P(d, θ) we can obtain a posterior on the parameters by selecting the samples that have ddata = dim. In many situations, it is essentially impossible to obtain ddata = dim and thus |ddata − dim| < ε is used (for some suitable metric).

An immediate challenge to the above idea is that, for many situations including cases in cosmology, it can require an incredibly large number of simulations to obtain samples from the posterior, especially if one requires a small ε parameter. This is particularly true when the dimension of the observable is large. This challenge can be addressed with density estimation likelihood-free inference (DELFi). The conceptually simplest DELFi methods use a density estimator to obtain a parametric form for the joint distribution P(d, θ|w), where w are the parameters of the density estimator. These methods have been found to require many orders of magnitude fewer simulations. In this work we use a slight variation: we use a density estimator to model the conditional distribution P(d|θ, w), utilizing the fact that, when conditioned on the parameters, dim is a sample from P(d|θ). This variation means that one can be agnostic about the properties of the chosen simulation points; see Alsing et al. (2019) for a more detailed discussion. The likelihood is then obtained by evaluating this conditional distribution at the observed data d0. Given the likelihood and a prior, we then use MCMC sampling to obtain posterior samples.

C1 Masked Autoregressive Flows

We require the density estimation to be flexible enough to allow us to accurately approximate the true conditional distribution and also to be computationally tractable in order to both fit and use. Recent work by e.g. Papamakarios & Murray (2016) and Uria et al. (2016) have shown that Masked Autoregressive Flows (MAFs) provide one such method.

MAFs rely on two steps to achieve these goals. First, they utilize the chain rule to express the multidimensional conditional distribution, P(d|θ), as a series of one-dimensional conditionals

\[ P(d|θ) = \prod_{i=1}^{n} P(d_i|d_{i-1}, θ). \]  

(C1)

A model utilizing this decomposition is known as an autoregressive model. Next, a form is chosen for the one-dimensional conditional distributions. We assume the conditionals are one-dimensional Gaussian distributions whose means, μ, and standard deviations, σ, depend on di−1 and θ. To achieve the desired flexibility we use neural networks, with parameters w, to parametrize these functions as μ(di−1, θ|w) and σ(di−1, θ|w). As was shown in Germain et al. (2015) this set-up can equivalently be formulated as developing a mapping from d to the variable uit = xi − μi(di−1, θ|w)/σi(di−1, θ|w), where u are independent zero mean, unit variance Gaussian random variables. This means the density estimator has the simple form

\[ P(d|θ, w) = \mathcal{N}[u(d, θ; w)|0, 1] \times \prod_{i=1}^{n} \sigma^2_i(d, θ; w). \]  

(C2)

where the product over the standard deviations is the Jacobian from transforming from u to d. We use an efficient implementation of this set-up called the Gaussian Masked Autoencoders for Density Estimation and hereafter refer to this method as MADE.

There are two main limitations for using MADE: first, they depend sensitively on the order of factorization and, second, the assumption of Gaussian conditionals may be overly restrictive. To mitigate these shortcomings we stack a series of MADEs to make a MAF. The output u of each MADE is the input of the next one. The conditional density estimator is thus given by

\[ P(d|θ, w) = \prod_{i} P(d_i|d_{i-1}, θ) \]

\[ = \mathcal{N}[u(d, θ; w)|0, 1] \times \prod_{n} \prod_{i=1}^{N_{\text{MADE}}} \sigma^2_i(d, θ; w). \]  

(C3)

Thus, we obtain a conditional distribution that is both analytically tractable (a simple product of Gaussians) and highly flexible.

Finally, we fit the weights of the neural network as detailed in Section 4.3 around equation (54). By minimizing the negative loss function in equation (55) we train the neural density estimators (NDEs) with respect to the network weights. The problem of overfitting is mitigated by applying the standard machine learning procedures of early stopping, dropouts, and the random selection of training and testing subsets.

C2 Architecture of pydelfi

For the density estimation, we select an ensemble of NDEs, all of them MA斐, with different numbers of MADEs, hidden layers, and neurons per layer, as shown in Table C1. All the NDEs use a tanh activation function, and are trained using the stochastic gradient optimizer ADAM (Kingma & Ba 2014). To shorten training times of the NN, we trained the model on graphics processing units. To avoid overfitting, we use one-tenth of the training set at each training cycle.
Table C1. Details of the eight MAFs making up the NDEs of the pydelfi architecture used in this paper, along with the $\tau$ posteriors from each NDE for the SRoll1 100 × 143 EE analysis. (The individual posteriors are shown in Fig. C1.).

| No. | $N_{MADE}$ | Hidden layers | Units | Validation loss | $\tau$ Posterior |
|-----|------------|---------------|-------|----------------|-----------------|
| 1   | 20         | 2             | 50    | 26.0           | 0.0519 ± 0.0071 |
| 2   | 10         | 2             | 50    | 25.9           | 0.0512 ± 0.0071 |
| 3   | 5          | 2             | 50    | 26.0           | 0.0526 ± 0.0073 |
| 4   | 3          | 2             | 75    | 26.0           | 0.0526 ± 0.0069 |
| 5   | 5          | 2             | 75    | 25.8           | 0.0529 ± 0.0073 |
| 6   | 10         | 2             | 75    | 25.9           | 0.0514 ± 0.0070 |
| 7   | 20         | 2             | 75    | 26.0           | 0.0519 ± 0.0068 |
| 8   | 5          | 3             | 35    | 26.0           | 0.0521 ± 0.0067 |

Figure C1. Planck 2018 posteriors obtained from our eight NDEs in our pydelfi architecture. The black lines show the individual constraints, and the blue solid line shows the weighted average that makes our pydelfi result. (The means and standard deviations of the individual posteriors are given in Table C1.).

Figure C2. Score-compressed pydelfi posteriors from SRoll1 100 × 143 spectra. The individual $\tau$ posteriors for the EE (dashed black line), TE (dash–dotted black line), and TT (dotted black line) statistics are shown, along with the result from the three-statistic $TTTEEE$ likelihood (solid red line).

Figure C3. Comparison of $\tau$ posteriors from 100 × 143 SRoll1 $TTTEEE$ spectra using pydelfi with (dotted line) and without compression (dashed line) and using momento (solid line). For this test all likelihoods use the same multipole range of $2 \leq \ell \leq 10$. We note that avoiding score compression for pydelfi results in a shift upwards by ~0.5σ of the $\tau$ maximum likelihood value compared to that from the score compressed likelihood, leading to a similar result to momento.

C3 Score and data compression

To compute a joint likelihood using low multipole temperature and polarization QCS spectra ($TT$, $TE$, $EE$) an additional compression step is required to reduce the dimensionality of the problem. Therefore, the full $N$-dimensional data set $D \in \mathbb{R}^N$ is first compressed to quadratic cross-spectra, i.e. summary statistics $d \in \mathbb{R}^M$, with $M < N$. Then, we score compress the vector of $M$ power spectrum measurements $d$ into a vector of $n$ components $t \in \mathbb{R}^n$, with $n < M < N$. In order to inform the choice of statistic for the score compression, an approximate form of the log-likelihood function $L$ is assumed. The statistic $t$ is then the gradient of the approximate log-likelihood, evaluated at some fiducial parameter values $\theta_0$, i.e. $d \mapsto t = \nabla_\theta \mathcal{L}_0$.

In the main body of the paper we compressed the $TT$, $TE$, and $EE$ cross-spectra separately for computational simplicity. This is suboptimal, as power spectrum elements between spectra are correlated. However, this does not bias our cosmological constraints and pydelfi does capture residual correlations between the individual $TT$, $TE$, and $EE$ statistics. The approximate log-likelihood we use for...
the score compression step is the analytic result that can be derived for cross-spectra computed on the full sky with Gaussian isotropic noise, namely the variance-gamma distribution (Pearson, Jeffery & Elderton 1929):

\[
P(\hat{C}_l^{XY}) = \frac{(2\ell + 1) |\hat{C}_l^{XY}|^\ell}{\Gamma\left(\frac{2\ell + 1}{2}\right)\sqrt{2^{2\ell}\pi(1 - \rho^2_x)} C^{XX}_\ell C^{YY}_\ell} \times K_\ell \left(\frac{(2\ell + 1) |\hat{C}_l^{XY}|}{(1 - \rho^2_x) \sqrt{C^{XX}_\ell C^{YY}_\ell}}\right) \exp \left(\frac{(2\ell + 1)\rho_x |\hat{C}_l^{XY}|}{(1 - \rho^2_x) \sqrt{C^{XX}_\ell C^{YY}_\ell}}\right).
\]

(C5)

Here, \(\Gamma(n)\) is the gamma function, \(K_\ell(x)\) is the modified Bessel function of the second kind, and \(\rho_x \equiv C^{XY} / \sqrt{C^{XX} C^{YY}}\) is the correlation coefficient. Using a fiducial model with \(\tau = 0.06\), we differentiate equation (C5) to obtain

\[
t = \nabla \ln P(\hat{C}_l^{XY}) |_{\hat{\tau}},
\]

one each for \(TT\), \(TE\), and \(EE\).

We then fit the parameters of the \texttt{pydelfi} likelihood using the compressed statistics of the simulations and the compressed data vector, yielding the pair \(\{\tau, t\}\). For the case where we only consider \(EE\), the compressed statistic is related to the maximum likelihood estimate for \(\tau\); see Alsing et al. (2018) for a more detailed discussion. The limitations of the compression steps are discussed further in Alsing et al. (2018).

### C4 Effect of score compression on the maximum likelihood value of \(\tau\) and comparison between \texttt{pydelfi} and \texttt{momento}

This section considers in more detail the properties of the score-compressed \texttt{pydelfi} likelihoods, both amongst themselves and then in comparison with \texttt{momento}.

First, in Fig. C2 we illustrate the score-compressed posteriors obtained with \(100 \times 143\) spectra from \texttt{SRoll2} frequency maps using \texttt{pydelfi}. The different posteriors from the \(EE\) (dashed line), \(TE\) (dash--dotted line), and \(TT\) (dotted line) statistics are shown, along with that from the combined three-statistic \(TTTEE\) likelihood (solid red line). These all used the full multipole range \((2 \leq \ell \leq 29)\) for the construction of the statistic(s). The \(EE\) posterior is most constraining, the \(TT\) posterior somewhat so, whereas the \(TE\) one hardly varies over the \(\tau\) scan. In the three-statistic likelihood, the addition of \(TE\) information moves the mean \(\tau\) value upwards by \(\sim 0.25\sigma\) from that from the \(EE\)-statistic likelihood. This upwards shift is also seen for \texttt{momento} in Table 3 but is not apparent there for \texttt{pydelfi}. This is because the table shows results from the \textit{non-score-compressed} \texttt{pydelfi} \(EE\) likelihood and from the \textit{score-compressed} three-statistic \texttt{pydelfi} \(TTTEE\) likelihood. The upwards shift from adding \(TE\) partially cancels the downward movement caused in passing from a full \(EE\) likelihood to a score-compressed one.

Second, to further understand how \texttt{pydelfi} performs in comparison to \texttt{momento} in the joint temperature-polarization case, we produce a version of \texttt{pydelfi} that does not need score compression. By limiting the power spectrum multipoles to those that are most constraining for \(\tau\), i.e. \(2 \leq \ell \leq 10\), rather than \(2 \leq \ell \leq 29\), we have a computationally manageable 27-dimensional problem, rather than a difficult 84-dimensional one. Comparing the posteriors with and without score compression on \texttt{SRoll2} maps for the three-statistic \(TTTEE\) likelihoods, we obtain:

\[
\tau = 0.0578 \pm 0.0063, \quad \text{(\texttt{pydelfi}, \ell \leq 10 + comp.). (7a)}
\]

\[
\tau = 0.0612 \pm 0.0060, \quad \text{(\texttt{pydelfi}, \ell \leq 10 + no comp.). (7b)}
\]

Fig. C3 illustrates these posteriors and compares them to the \texttt{momento} posterior. This confirms the role of the score compression discussed above in affecting constraints on \(\tau\).

This paper has been typeset from a TEX/LATEX file prepared by the author.