Delta-map method of removing CMB foregrounds with spatially varying spectra

Kiyotomo Ichiki[1,2,*], Hiroaki Kanai[3], Nobuhiko Katayama[4] Eiichiro Komatsu[4,5]

[1]*Graduate School of Science, Division of Particle and Astrophysical Science, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan*
[2]*Kobayashi-Maskawa Institute for the Origin of Particles and the Universe, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan*
[3]*Graduate School of Engineering, Yokohama National University, 79-5 Tokiwadai, Hodogaya-ku, Yokohama 240-8501, Japan*
[4]*Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU, WPI), Todai Institutes for Advanced Study, The University of Tokyo, Kashiwa 277-8583, Japan*
[5]*Max Planck Institute for Astrophysics, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany*
*E-mail: ichiki@a.phys.nagoya-u.ac.jp

....
We extend the internal template foreground removal method by accounting for spatially varying spectral parameters such as the spectral indices of synchrotron and dust emission and the dust temperature. As the previous algorithm had to assume that the spectral parameters are uniform over the full sky (or some significant fraction of the sky), it resulted in a bias in the tensor-to-scalar ratio parameter r estimated from foreground-cleaned polarization maps of the cosmic microwave background (CMB). The new algorithm, the "Delta-map method", accounts for spatially varying spectra to first order in perturbation. The only free parameters are the cosmological parameters such as r and the sky-averaged foreground parameters. We show that a cleaned CMB map is the maximum likelihood solution to first order in perturbation, and derive the posterior distribution of r and the sky-averaged foreground parameters using Bayesian statistics. Applying this to realistic simulated sky maps including polarized CMB, synchrotron, and thermal dust emission, we find that the new algorithm removes the bias in r down to an undetected level in our noiseless simulation ($r \lesssim 4 \times 10^{-4}$). We show that the frequency decorrelation of polarized foregrounds due to averaging of spatially varying spectra can be accounted for to first order in perturbation by using slightly different spectral parameters for the Stokes Q and U maps. Finally, we show that the effect of polarized anomalous microwave emission on foreground removal can be absorbed into the curvature parameter of the synchrotron spectrum.
....
Subject Index E60, E63, E65, F14

1. Introduction

The search for the signature of primordial gravitational waves from the early Universe in the B-mode polarization of the cosmic microwave background (CMB) [1] is already limited by the accuracy of the foreground removal procedure [2]. Among various methods for removing the foreground emission (see Refs. [3–5] and references therein), we focus on the so-called "internal template" method [6–10]. The basic idea behind this method is simple. Consider that we have maps at three frequencies: low, middle, and high; say, 40, 100, and 240 GHz. We know that the low- and high-frequency data are dominated by synchrotron and thermal dust emission, respectively [11]; thus, the

simplest way to remove them is to fit the low- and high-frequency maps to and subtract them from the map at the middle frequency. A cleaned map may be written as $x = ([Q, U]_{100} + \alpha_{40}[Q, U]_{40} + \alpha_{240}[Q, U]_{240})/(1 + \alpha_{40} + \alpha_{240})$, where α_{40} and α_{240} are fitting coefficients and Q and U are the Stokes parameters in thermodynamic temperature units. Fitting can be done in pixel, Fourier (spherical harmonics), or wavelet space. This method is called a *template* because the low- and high-frequency data are used as templates for synchrotron and thermal dust emission, respectively, and is called *internal* because no external data sets are used.

All of the previous implementations of this method assumed that the fitting coefficients are uniform over some fitted regions. In some studies the fitted region was taken to be the whole sky [6,7,9,10], and in others some large pixels [8]. In this paper, we build on our previous work [8], in which we find that the internal template method yields a small but non-negligible bias in the tensor-to-scalar ratio parameter r of order 10^{-3}, and that the origin of the bias is the spatial variation of the fitting coefficients, i.e., spatial variation of the spectral parameters of the foreground emission.

Motivated by this finding, we extend the internal template method by accounting for spatial variations of the spectra. If we have two high- (or low-)frequency channels as dust (or synchrotron) templates instead of one, we should be able to solve for both the mean and variation in one spectral parameter, e.g., the dust (or synchrotron) spectral index. We show that this is indeed the case. In general, we need $N + 1$ channels to account for spatial variations of N spectral parameters of one foreground component.

Armed with the new method, which we shall call the "Delta-map method" throughout this paper, we investigate two hotly debated topics of the polarized foreground: treatments of possible polarization in the anomalous microwave emission (AME) and the frequency decorrelation due to averaging of spatially varying foreground spectra.

This paper is organized as follows. We introduce the Delta-map method in Sect. 2, and explain how to implement it in Sect. 3. In Sect. 4, we interpret the Delta-map method within the Bayesian framework. In Sect. 5, we apply this to realistic foreground maps for validation. In Sect. 6, we show how to treat AME and frequency decorrelation. In Sect. 7, we study how instrumental noise affects the results. We conclude in Sect. 8.

2.   Delta-map method

2.1.   *General formula*

Without loss of generality, let us consider one foreground component with N spectral parameters, and model the observed Stokes parameters Q and U in thermodynamic temperature units at a given frequency ν as

$$[Q, U]_\nu(\hat{n}) = \mathrm{CMB}(\hat{n}) + g_\nu D_\nu(p^I(\hat{n}))[Q^f, U^f]_{\nu_*}(\hat{n}) + N_\nu(\hat{n}), \qquad (1)$$

where $\mathrm{CMB}(\hat{n})$ and $N_\nu(\hat{n})$ denote the CMB signal and instrumental noise towards a sky direction $\hat{n}$ in thermodynamic units, respectively, and $[Q^f, U^f]_{\nu_*}(\hat{n})$ denotes the foreground component at some reference frequency ν_* in brightness temperature units. As the dependence on ν_* cancels in the final result, $[Q^f, U^f]_{\nu_*}$ appears only as an auxiliary variable. For simplicity we have not written explicitly experiment-specific effects such as gain calibration factors, beam smearing, pixel window functions, or any other smoothing/filtering that may be applied to data. Such operations will be common to all sky signals at the same frequency channel.

The coefficient g_ν is given by

$$g_\nu \equiv \frac{(e^x - 1)^2}{e^x x^2} \quad \text{with} \quad x \equiv \frac{h\nu}{k_B T_{CMB}}, \tag{2}$$

which converts the units from brightness temperature to the CMB thermodynamic temperature ($T_{CMB} = 2.725$ K). While this expression does not take into account averaging over the bandpass of detectors, we shall include it in Sect. 2.5.

The function D_ν denotes the frequency dependence of the foreground emission in brightness temperature units, and $p^I(\hat{n})$ a set of parameters ($I = 1, 2, \ldots, N$) that characterize the emission law and can vary across the sky. For example, if we consider a power-law foreground emission component, then $N = 1$ and $p^1(\hat{n})$ is a power-law spectral index $\beta(\hat{n})$. Thus, $D_\nu = (\nu/\nu_*)^{\beta(\hat{n})}$. Decomposing the spectral parameters to the mean value and perturbation as $p^I(\hat{n}) = \bar{p}^I + \delta p^I(\hat{n})$, we obtain

$$D_\nu(p^I(\hat{n}))[Q^f, U^f]_{\nu_*}(\hat{n}) = D_\nu(\bar{p}^I)[Q^f, U^f]_{\nu_*}(\hat{n}) + \sum_{J=1}^{N} D_{\nu,J} \delta p^J(\hat{n})[Q^f, U^f]_{\nu_*}(\hat{n}) + \mathcal{O}(\delta p^2), \tag{3}$$

where we have written $D_{\nu,J} \equiv \left(\partial D_\nu/\partial p^J\right)_{p^J = \bar{p}^J}$.

The central assumption that we make throughout this paper is that we truncate the expansion at first order in δp^I [12,13]. Therefore, when a foreground component is parameterized by N parameters that vary across the sky only weakly, we can interpret the emission as a superposition of $(N + 1)$ template maps, namely $[Q^f, U^f]_{\nu_*}(\hat{n})$ and $\delta p^I(\hat{n})[Q^f, U^f]_{\nu_*}(\hat{n})$ ($I = 1, 2, \ldots, N$). If the $\mathcal{O}(\delta p^2)$ term is significant compared with the first two terms, it may cause a bias in r as foreground residuals. The bias becomes greater when the foreground channels are taken further away from the CMB channel. We find, however, that the bias is negligible for the band specification and realistic simulated sky maps used in this work, as we shall show later.

To subtract the foreground component in Eq. (1) with the help of Eq. (3), we form a linear combination of $(N + 2)$ maps at different frequencies as

$$[Q, U]_{\nu_{CMB}}(\hat{n}) + \sum_{j=1}^{N+1} \alpha_j [Q, U]_{\nu_j}(\hat{n}) = \left(1 + \sum_{j=1}^{N+1} \alpha_j\right) CMB(\hat{n})$$

$$+ \left(g_{\nu_{CMB}} D_{\nu_{CMB}}(\bar{p}^I) + \sum_{j=1}^{N+1} \alpha_j g_{\nu_j} D_{\nu_j}(\bar{p}^I)\right) [Q^f, U^f]_{\nu_*}(\hat{n})$$

$$+ \sum_{J=1}^{N} \left(g_{\nu_{CMB}} D_{\nu_{CMB},J} + \sum_{j=1}^{N+1} \alpha_j g_{\nu_j} D_{\nu_j,J}\right) \delta p^J(\hat{n})[Q^f, U^f]_{\nu_*}(\hat{n})$$

$$+ N_{\nu_{CMB}}(\hat{n}) + \sum_{j=1}^{N+1} \alpha_j N_{\nu_j}(\hat{n}), \tag{4}$$

where α_j are fitting coefficients, ν_{CMB} denotes a frequency employed as a CMB channel, and ν_j the foreground channels. The fitting coefficients are found by setting the foreground terms in the above equation (i.e., the second and third lines) to zero:

$$g_{\nu_{CMB}} D_{\nu_{CMB}}(\bar{p}^I) + \sum_{j=1}^{N+1} g_{\nu_j} D_{\nu_j}(\bar{p}^I) \alpha_j = 0, \tag{5}$$

$$g_{\nu_{\text{CMB}}} D_{\nu_{\text{CMB}},J} + \sum_{j=1}^{N+1} g_{\nu_j} D_{\nu_j,J} \alpha_j = 0, \quad \text{for } J = (1, 2, \ldots, N). \tag{6}$$

These equations can be cast into a matrix form as $\boldsymbol{A}\vec{\alpha} = -\vec{d}_{\nu_{\text{CMB}}}$, where

$$\vec{\alpha} \equiv (\alpha_1, \alpha_2, \ldots, \alpha_{N+1})^{\text{T}}, \tag{7}$$

$$\vec{d}_\nu \equiv g_\nu \left(D_\nu(\vec{p}^I), D_{\nu,1}, \ldots, D_{\nu,N} \right)^{\text{T}}, \tag{8}$$

$$\boldsymbol{A} \equiv (\vec{d}_{\nu_1}, \vec{d}_{\nu_2}, \ldots, \vec{d}_{\nu_{N+1}}). \tag{9}$$

The coefficients α_j can be obtained numerically by $-\boldsymbol{A}^{-1}\vec{d}_{\nu_{\text{CMB}}}$.[1] Once the coefficients are determined, we obtain a foreground-cleaned CMB map $\vec{x}$ as

$$\vec{x} = \frac{[Q, U]_{\nu_{\text{CMB}}}(\hat{n}) + \sum_{j=1}^{N+1} \alpha_j [Q, U]_{\nu_j}(\hat{n})}{1 + \sum_{j=1}^{N+1} \alpha_j} . \tag{10}$$

Since this is a linear combination of maps, we can construct a cleaned CMB in pixel, spherical harmonics, or wavelet space.

Multi-component foregrounds can be incorporated straightforwardly by expanding the entries of $\boldsymbol{A}$. For example, if we use D_ν for dust and S_ν for synchrotron, we may define a new vector for synchrotron as

$$\vec{s}_\nu \equiv g_\nu \left(S_\nu(\vec{p}^J), S_{\nu,1}, \ldots, S_{\nu,M} \right)^{\text{T}}, \tag{11}$$

where M is the number of parameters characterizing synchrotron, and write the matrix $\boldsymbol{A}$ as

$$\boldsymbol{A} = (\vec{d}_{\nu_1}, \vec{d}_{\nu_2}, \ldots, \vec{d}_{\nu_{N+1}}, \vec{s}_{\nu_1}, \vec{s}_{\nu_2}, \ldots, \vec{s}_{\nu_{M+1}}). \tag{12}$$

There are $N + M + 2$ coefficients to be determined, given by $\vec{\alpha} = -\boldsymbol{A}^{-1}(\vec{d}_{\nu_{\text{CMB}}}, \vec{s}_{\nu_{\text{CMB}}})^{\text{T}}$. Thus, obtaining one cleaned CMB map requires at least $N + M + 3$ frequencies. In Sect. 4, we derive the formula for a cleaned CMB map when we have more frequency channels than this minimum number.

2.2. Power law

To obtain a better feel for the new algorithm, let us work out three simpler, concrete examples.

First, we consider a single power-law spectrum with a spatially varying spectral index:

$$D_\nu(p^I(\hat{n})) = D_\nu(\beta(\hat{n})) = \left(\frac{\nu}{\nu_*} \right)^{\beta(\hat{n})} . \tag{13}$$

---

[1] The explicit form of $\boldsymbol{A}$ is

$$\boldsymbol{A} = \begin{pmatrix} g_{\nu_1} D_{\nu_1} & g_{\nu_2} D_{\nu_2} & \cdots & g_{\nu_{N+1}} D_{\nu_{N+1}} \\ g_{\nu_1} D_{\nu_1,1} & g_{\nu_2} D_{\nu_2,1} & \cdots & g_{\nu_{N+1}} D_{\nu_{N+1},1} \\ g_{\nu_1} D_{\nu_1,2} & g_{\nu_2} D_{\nu_2,2} & \cdots & g_{\nu_{N+1}} D_{\nu_{N+1},2} \\ \vdots & \vdots & \ddots & \vdots \\ g_{\nu_1} D_{\nu_1,N} & g_{\nu_2} D_{\nu_2,N} & \cdots & g_{\nu_{N+1}} D_{\nu_{N+1},N} \end{pmatrix}.$$

The Stokes parameters become [12,13]

$$[Q,U]_\nu(\hat{n}) = \mathrm{CMB}(\hat{n}) + N_\nu(\hat{n}) + g_\nu \left(\frac{\nu}{\nu_*}\right)^{\bar{\beta}} \left(1 + \ln\left(\frac{\nu}{\nu_*}\right)\delta\beta(\hat{n})\right)[Q^f,U^f]_{\nu_*}(\hat{n}) + \mathcal{O}(\delta p^2), \quad (14)$$

where $\bar{\beta}$ is the mean spectral index. Equation (8) becomes

$$\vec{d}_\nu = g_\nu \left(\frac{\nu}{\nu_*}\right)^{\bar{\beta}} \left(1, \ln\left(\frac{\nu}{\nu_*}\right)\right)^{\mathrm{T}}. \quad (15)$$

Therefore, the sum in Eq. (10) evaluates to

$$\sum_{j=1}^{2} \alpha_j [Q,U]_{\nu_j}(\hat{n}) = \frac{g_{\nu_{\mathrm{CMB}}}}{\ln(\nu_2/\nu_1)} \left\{ \frac{1}{g_{\nu_2}}\left(\frac{\nu_{\mathrm{CMB}}}{\nu_2}\right)^{\bar{\beta}} \ln(\nu_1/\nu_{\mathrm{CMB}})[Q,U]_{\nu_2}(\hat{n}) \right.$$
$$\left. - \frac{1}{g_{\nu_1}}\left(\frac{\nu_{\mathrm{CMB}}}{\nu_1}\right)^{\bar{\beta}} \ln(\nu_2/\nu_{\mathrm{CMB}})[Q,U]_{\nu_1}(\hat{n}) \right\}. \quad (16)$$

The reference frequency ν_* cancels in the final result, which is required from the physical ground: a cleaned CMB map result should be invariant under an arbitrary change in the reference frequency. Adding this to $[Q,U]_{\nu_{\mathrm{CMB}}}(\hat{n})$ removes the foreground emission including spatial variation of the spectral index to first order in $\delta\beta(\hat{n})$. Note that we do not fit for $\delta\beta$ because they are solved and eliminated. The only fitting parameter is the mean spectral index $\bar{\beta}$.

As the above expression uses a difference between two weighted maps at different frequencies, we started to call this map a "Delta-map", which became the name of our algorithm. However, as we show in Sect. 4, our algorithm is much more general than Eq. (16). A more appropriate (and general) reason for the name "Delta-map" would be the fact that we expand and truncate the spectral dependence of foregrounds at first order in perturbation. Henceforth we shall adopt this reasoning for the name of the algorithm.

2.3. Running spectral index

Second, we allow the spectral index to vary slowly with frequency:

$$D_\nu(p^I(\hat{n})) = D_\nu(\beta(\hat{n}), C(\hat{n})) = \left(\frac{\nu}{\nu_*}\right)^{\beta(\hat{n}) + C(\hat{n})\ln(\nu/\nu_s)}, \quad (17)$$

where C is the so-called curvature parameter or running of the spectral index, $\bar{C}$ is its mean value, and ν_s is a pivot frequency of the running index. There is no physical meaning for ν_s, but it can be chosen such that it minimizes a correlation between β and C.

The Stokes parameters become

$$[Q,U]_\nu(\hat{n}) = \mathrm{CMB}(\hat{n}) + N_\nu(\hat{n})$$
$$+ g_\nu\left(\frac{\nu}{\nu_*}\right)^{\bar{\beta}+\bar{C}\ln(\nu/\nu_s)}\left[1 + \ln\left(\frac{\nu}{\nu_*}\right)\left(\delta\beta(\hat{n}) + \ln\left(\frac{\nu}{\nu_s}\right)\delta C(\hat{n})\right)\right][Q^f,U^f]_{\nu_*}(\hat{n})$$
$$+ \mathcal{O}(\delta p^2), \quad (18)$$

and

$$\vec{d}_\nu = g_\nu\left(\frac{\nu}{\nu_*}\right)^{\bar{\beta}+\bar{C}\ln(\nu/\nu_s)}\left(1, \ln\left(\frac{\nu}{\nu_*}\right), \ln\left(\frac{\nu}{\nu_*}\right)\ln\left(\frac{\nu}{\nu_s}\right)\right)^{\mathrm{T}}. \quad (19)$$

In this case, we need maps at three different frequencies to remove the effects of spatially varying foreground parameters to first order in $\delta\beta(\hat{n})$ and $\delta C(\hat{n})$.

2.4. Modified blackbody

Finally, we consider a modified blackbody (MBB) spectrum for dust emission:

$$
D_\nu(p^I(\hat{n})) = D_\nu(\beta_{\rm d}(\hat{n}), T_{\rm d}(\hat{n})) = \left(\frac{\nu}{\nu_*}\right)^{\beta_{\rm d}(\hat{n})+1} \frac{e^{x_{{\rm d}*}(\hat{n})} - 1}{e^{x_{\rm d}(\hat{n})} - 1}, \tag{20}
$$

where $x_{\rm d}(\hat{n}) \equiv h\nu/k_{\rm B}T_{\rm d}(\hat{n})$, $x_{{\rm d}*}(\hat{n}) \equiv h\nu_*/k_{\rm B}T_{\rm d}(\hat{n})$, and $T_{\rm d}(\hat{n})$ is the dust temperature. In the low-frequency limit, $D_\nu \to (\nu/\nu_*)^{\beta_{\rm d}}$. The Stokes parameters become

$$
\begin{aligned}
[Q, U]_\nu(\hat{n}) = {} & {\rm CMB}(\hat{n}) + N_\nu(\hat{n}) + g_\nu \left(\frac{\nu}{\nu_*}\right)^{\bar{\beta}_{\rm d}+1} \frac{e^{\bar{x}_{{\rm d}*}} - 1}{e^{\bar{x}_{\rm d}} - 1} \\
& \times \left[1 + \ln\left(\frac{\nu}{\nu_*}\right)\delta\beta_{\rm d}(\hat{n}) + \left(\frac{\bar{x}_{\rm d}e^{\bar{x}_{\rm d}}}{e^{\bar{x}_{\rm d}} - 1} - \frac{\bar{x}_{{\rm d}*}e^{\bar{x}_{{\rm d}*}}}{e^{\bar{x}_{{\rm d}*}} - 1}\right)\frac{\delta T_{\rm d}(\hat{n})}{\bar{T}_{\rm d}}\right][Q^f, U^f]_{\nu_*}(\hat{n}) \\
& + \mathcal{O}(\delta p^2),
\end{aligned} \tag{21}
$$

where $\bar{T}_{\rm d}$ is the mean temperature, $\bar{x}_{\rm d} \equiv h\nu/k\bar{T}_{\rm d}$, and

$$
\vec{d}_\nu = g_\nu \left(\frac{\nu}{\nu_*}\right)^{\bar{\beta}_{\rm d}+1} \frac{e^{\bar{x}_{{\rm d}*}} - 1}{e^{\bar{x}_{\rm d}} - 1}\left(1, \ln\left(\frac{\nu}{\nu_*}\right), \left(\frac{\bar{x}_{\rm d}e^{\bar{x}_{\rm d}}}{e^{\bar{x}_{\rm d}} - 1} - \frac{\bar{x}_{{\rm d}*}e^{\bar{x}_{{\rm d}*}}}{e^{\bar{x}_{{\rm d}*}} - 1}\right)\frac{1}{\bar{T}_{\rm d}}\right)^{\rm T}. \tag{22}
$$

In this case, we need maps at three different frequencies to remove the effects of spatially varying foreground parameters to first order in $\delta\beta_{\rm d}(\hat{n})$ and $\delta T_{\rm d}(\hat{n})$.

2.5. Bandpass average

The above formalism applies when we have detectors with a delta-function response to a single frequency. In reality the sky signal must be averaged within a given bandpass of the detectors.

To incorporate the bandpass average, we first average an observed intensity as

$$
\langle I(\nu)\rangle \equiv \frac{\int W(\nu')I(\nu')d\nu'}{\int W(\nu')d\nu'} = \frac{1}{\Delta\nu}\int_{\nu-\Delta\nu/2}^{\nu+\Delta\nu/2} I(\nu')d\nu', \tag{23}
$$

where the last equality holds when the bandpass function $W(\nu)$ is a top-hat function with a width of $\Delta\nu$. We then relate $\langle I(\nu)\rangle$ to the thermodynamic CMB temperature fluctuation δT as

$$
\langle I(\nu)\rangle = \left\langle\frac{\partial B_\nu(T)}{\partial T}\right\rangle\delta T, \tag{24}
$$

where $B_\nu(T)$ is a blackbody intensity spectrum. The same relation holds for the Stokes parameters Q and U.

We use the brightness temperature as an intermediate step between the intensity and thermodynamic temperature. The brightness–thermodynamic temperature conversion factor g_ν (2) with bandpass averaging is then given by

$$
\langle g_\nu\rangle = \frac{2k_{\rm B}\nu^2/c^2}{\left\langle\frac{\partial B_\nu(T)}{\partial T}\right\rangle} = \frac{\nu^2}{\left\langle\nu^2\frac{e^x x^2}{(e^x-1)^2}\right\rangle}. \tag{25}
$$

There is no averaging for ν^2 in the numerator because it is the intensity that is averaged, and conversion from the intensity to brightness temperature is merely an intermediate step. Using these quantities, our sky model (1) is now expressed as

$$[Q, U]_\nu(\hat{n}) = \text{CMB}(\hat{n}) + \langle g_\nu \rangle \frac{\left\langle \nu^2 D_\nu(p^I(\hat{n})) \right\rangle}{\nu^2} [Q^f, U^f]_{\nu_*}(\hat{n}) + N_\nu(\hat{n})$$

$$= \text{CMB}(\hat{n}) + \frac{\left\langle \nu^2 D_\nu(p^I(\hat{n})) \right\rangle}{\left\langle \nu^2 \frac{e^x x^2}{(e^x-1)^2} \right\rangle} [Q^f, U^f]_{\nu_*}(\hat{n}) + N_\nu(\hat{n}). \tag{26}$$

As $[Q^f, U^f]_{\nu_*}$ is an auxiliary variable, there is no need to bandpass-average it. The average is done only for variables depending on ν. Therefore, to incorporate bandpass averaging, all we need is to replace $g_\nu D_\nu$ with $\langle g_\nu \rangle \langle \nu^2 D_\nu \rangle / \nu^2$.

Incorporating bandpass averaging in the foreground removal algorithm is important because it changes the effective frequency at which intensity of foregrounds should be evaluated. Using an incorrect bandpass leads to a bias in r. How precisely we should know the bandpass needs to be determined by the requirement for science, e.g., a target value of r, coupled with a careful analysis including foreground removal and systematic errors in our knowledge of the bandpass. This is work in progress. For the rest of this paper we shall ignore bandpass averaging in the simulations and leave the detailed study of its effect for future work.

3. Implementation

3.1. Cleaning one map

Our likelihood for a cleaned CMB map $\vec{x}$ (10) is

$$-2 \ln \mathcal{L} = \vec{x}(\bar{p}^I)^{\mathrm{T}} C^{-1}(r, s, \bar{p}^I) \vec{x}(\bar{p}^I) + \ln |2\pi C|, \tag{27}$$

where the covariance matrix C is given by [8]

$$C(r, s, \bar{p}^I) = C^{\text{tens}}(r) + C^{\text{scal}}(s) + \frac{N_{\nu_{\text{CMB}}} + \sum_{j=1}^{N+1} \alpha_j^2 N_{\nu_j}}{\left(1 + \sum_{j=1}^{N+1} \alpha_j\right)^2} + E_{\text{diag}}, \tag{28}$$

and C^{tens} and C^{scal} are the tensor and scalar CMB signal covariance matrices (including beam smearing, pixel window function, and any other extra smoothing/filtering applied to data), respectively, calculated from theoretical power spectra following Appendix A of Ref. [8]. The noise covariance matrices N_ν and E_{diag} will be explained after Eq. (31).

Theoretical power spectra are given by

$$C_\ell^{\text{EE}} = r c_\ell^{\text{tens,EE}} + s c_\ell^{\text{scal,EE}}, \tag{29}$$

$$C_\ell^{\text{BB}} = r c_\ell^{\text{tens,BB}} + s c_\ell^{\text{lens,BB}}, \tag{30}$$

where $c_\ell^{\text{tens,EE}}$ and $c_\ell^{\text{tens,BB}}$ are the E- and B-mode polarization power spectra from the tensor mode with $r = 1$, respectively, and $c_\ell^{\text{scal,EE}}$ and $c_\ell^{\text{lens,BB}}$ are the E mode from the scalar mode and the B mode from the gravitational-lensing effect, respectively.

The parameter s rescales the amplitude of the scalar perturbation, and the fiducial value is $s = 1$. In Ref. [8], we found that marginalization over the scalar amplitude s was able to remove a chance correlation between the template coefficients and the tensor-to-scalar ratio r.

To compute the fiducial theoretical power spectra, we use the Planck 2015 cosmological parameters for "TT+LowP+lensing" [14]: $\Omega_b h^2 = 0.022\,26$, $\Omega_c h^2 = 0.1186$, $h = 0.6781$, $\tau = 0.066$, $A_s = 2.137 \times 10^{-9}$, and $n_s = 0.9677$.

In this paper, we work with low-resolution maps at a Healpix [15] resolution of $N_{\text{side}} = 4$. At this resolution and with the very low instrumental noise that we assume in this paper (anticipating future space missions such as LiteBIRD [16]), the signal covariance matrix becomes ill conditioned due to a strong scalar E-mode signal that is not band-limited. To solve this numerical issue, we follow the procedure described in Appendix A of Ref. [17]; specifically, we apply a Gaussian smoothing to a signal-plus-noise map with 2200 arcmin (FWHM), which is 2.5 times the pixel size at $N_{\text{side}} = 4$, and resample the smoothed map to $N_{\text{side}} = 4$. As the smoothed map at $N_{\text{side}} = 4$ is dominated by the scalar E-mode signal at all angular scales supported by the map resolution, the covariance matrix of this map is singular. In order to regularize the covariance matrix, we add an artificial, homogeneous white noise of $0.2\,\mu$K arcmin to the smoothed map, such that the map becomes noise dominated at the Nyquist frequency of the map:

$$\vec{x} = \frac{[Q, U]_{\nu_{\text{CMB}}}(\hat{n}) + \sum_{j=1}^{N+1} \alpha_j [Q, U]_{\nu_j}(\hat{n})}{1 + \sum_{j=1}^{N+1} \alpha_j} + \text{artificial noise}(\hat{n}). \tag{31}$$

The noise covariance matrices $N_{\nu_{\text{CMB}}}$ and N_{ν_j} in Eq. (28) become non-diagonal because of smoothing, and E_{diag} is a diagonal matrix for the artificial noise used to regularize the covariance matrix.

We now wish to estimate $\bar{p}^I$, r, and s from the likelihood given in Eq. (27). We follow the following procedure:

1. **Initialization:** Set initial values for s, r, and $\bar{p}^I$.
2. **Iteration step I:** Minimize the χ^2 part of the likelihood (the first term in Eq. (27)),[2]

$$\bar{p}^I = \underset{\bar{p}}{\arg\min}\, \vec{x}(\bar{p}^I)^{\text{T}} C^{-1}(r, s, \bar{p}^I) \vec{x}(\bar{p}^I), \tag{32}$$

   with s and r fixed. During minimization, the covariance matrix C and its inverse C^{-1} are updated at each evaluation of $\bar{p}^I$.
3. **Iteration step II:** Maximize the likelihood given in Eq. (27) for s and r with the foreground parameters $\bar{p}^I$ fixed to the values found during iteration step I. During maximization, C, C^{-1}, and the determinant $|C|$ are updated at each evaluation of s and r.
4. Return to step 2 until a convergence criterion is satisfied.

We find that ten iterations are typically sufficient to meet the convergence criterion that we set as $|r_{\text{new}} - r_{\text{old}}| < 10^{-3} r_{\text{old}}$. To maximize (minimize) the functions we have used the *nlopt* library [18].

Finally, we include the mask in the covariance matrix following Ref. [6] (see also Appendix A). We use the "P06 mask" defined by the WMAP polarization analysis, which retains 73% of sky for the analysis. We degrade the original P06 mask to $N_{\text{side}} = 4$.

For $N_{\text{side}} = 4$ simulations, analyzing 1000 maps roughly takes 1000 seconds on a laptop (four cores, 16 GB memory). The most time-consuming part is matrix inversion, the computational cost of which scales as the number of matrix elements cubed; thus, the computational cost scales as $\propto N_{\text{side}}^6$.

---

[2] One may wonder why we do not maximize the likelihood for $\bar{p}^I$, r, and s simultaneously using Eq. (27). We provide a reason for this in Sect. 4.

3.2.   *Cleaning multiple maps*

We may use the Delta-map method to clean two (or more) maps at two (or more) CMB frequencies. Here we consider a case in which we clean two maps at different CMB frequencies (e.g., 100 GHz and 140 GHz) using lower- and higher-frequency maps as the common template maps for cleaning. Having two frequency channels cleaned separately (instead of combining them into one cleaned map) is useful for null tests, as we describe in Sect. 5.3.

In this case, the map vector becomes

$$\vec{x}' = \begin{pmatrix} \vec{x}_{\nu_{\mathrm{CMB1}}} \\ \vec{x}_{\nu_{\mathrm{CMB2}}} \end{pmatrix}, \tag{33}$$

where $\vec{x}_{\nu_{\mathrm{CMB1,2}}}$ are cleaned CMB maps at frequencies ν_{CMB1} and ν_{CMB2}, given by

$$\vec{x}_{\nu_{\mathrm{CMB1,2}}} = \frac{[Q,U]_{\nu_{\mathrm{CMB1,2}}}(\hat{n}) + \sum_{j=1}^{N+1} \alpha_j^{\nu_{\mathrm{CMB1,2}}}[Q,U]_{\nu_j}(\hat{n})}{1 + \sum_{j=1}^{N+1} \alpha_j^{\nu_{\mathrm{CMB1,2}}}} + \text{artificial noise}_{1,2}(\hat{n}). \tag{34}$$

The noise covariance matrix for this map vector is given by

$$\boldsymbol{N}(\hat{n},\hat{n}') = \begin{pmatrix} \boldsymbol{n}_{Q_1 Q_1} & \boldsymbol{n}_{Q_1 U_1} & \boldsymbol{n}_{Q_1 Q_2} & \boldsymbol{n}_{Q_1 U_2} \\ \boldsymbol{n}_{U_1 Q_1} & \boldsymbol{n}_{U_1 U_1} & \boldsymbol{n}_{U_1 Q_2} & \boldsymbol{n}_{U_1 U_2} \\ \boldsymbol{n}_{Q_2 Q_1} & \boldsymbol{n}_{Q_2 U_1} & \boldsymbol{n}_{Q_2 Q_2} & \boldsymbol{n}_{Q_2 U_2} \\ \boldsymbol{n}_{U_2 Q_1} & \boldsymbol{n}_{U_2 U_1} & \boldsymbol{n}_{U_2 Q_2} & \boldsymbol{n}_{U_2 U_2} \end{pmatrix}, \tag{35}$$

where the components of the noise covariance matrix are, e.g.,

$$\boldsymbol{n}_{Q1Q1} = \frac{\boldsymbol{n}_{\nu_{\mathrm{CMB1}}}^{QQ} + \sum_{j=1}^{N+1} \left(\alpha_j^{\nu_{\mathrm{CMB1}}}\right)^2 \boldsymbol{n}_{\nu_j}^{QQ}}{\left(1 + \sum_{j=1}^{N+1} \alpha_j^{\nu_{\mathrm{CMB1}}}\right)^2}, \tag{36}$$

$$\boldsymbol{n}_{Q1U2} = \frac{\sum_{j=1}^{N+1} \alpha_j^{\nu_{\mathrm{CMB1}}} \alpha_j^{\nu_{\mathrm{CMB2}}} \boldsymbol{n}_{\nu_j}^{QU}}{\left(1 + \sum_{j=1}^{N+1} \alpha_j^{\nu_{\mathrm{CMB1}}}\right) \left(1 + \sum_{j=1}^{N+1} \alpha_j^{\nu_{\mathrm{CMB2}}}\right)}. \tag{37}$$

Here $\boldsymbol{n}_\nu^{QQ}, \boldsymbol{n}_\nu^{QU}$, and $\boldsymbol{n}_\nu^{UU}$ are the noise covariance matrices for the Stokes Q and U maps at frequency ν. The signal covariance is given more simply by

$$\boldsymbol{S}(\hat{n},\hat{n}') = \begin{pmatrix} \boldsymbol{s}_{QQ} & \boldsymbol{s}_{QU} & \boldsymbol{s}_{QQ} & \boldsymbol{s}_{QU} \\ \boldsymbol{s}_{UQ} & \boldsymbol{s}_{UU} & \boldsymbol{s}_{UQ} & \boldsymbol{s}_{UU} \\ \boldsymbol{s}_{QQ} & \boldsymbol{s}_{QU} & \boldsymbol{s}_{QQ} & \boldsymbol{s}_{QU} \\ \boldsymbol{s}_{UQ} & \boldsymbol{s}_{UU} & \boldsymbol{s}_{UQ} & \boldsymbol{s}_{UU} \end{pmatrix}, \tag{38}$$

where $\boldsymbol{s}_{QQ}, \boldsymbol{s}_{QU}, \boldsymbol{s}_{UQ}$, and $\boldsymbol{s}_{UU}$ are the signal covariance matrices for the Stokes Q and U of the CMB. Here, the same signal matrix elements appear for different frequency channels; however, it should be understood that they are different in practice due to different beam smearing and filtering depending on channels. We have not written this effect explicitly for simplicity throughout this paper.

3.3.   *Test*

Before moving on to simulations with realistic foreground models, we test the algorithm using simplified ones. Specifically, we generate foreground maps exactly matching the form assumed in
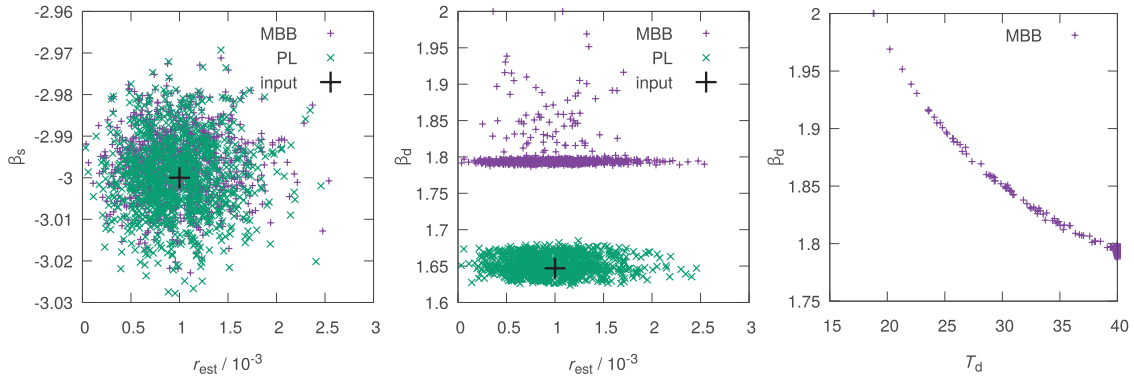
Fig. 1. Values of r and the mean foreground parameters obtained from 1000 random realizations of the CMB. No instrumental noise is included. The simulation also includes power-law synchrotron and power-law dust with spatially varying indices, and the mean indices are $\bar{\beta}_s = -3.0$ and $\bar{\beta}_d = 1.647$. While we always use the Delta-map method for a power-law spectrum (Sect. 2.2) for synchrotron, we compare two Delta-map methods for dust: power-law and MBB (Sect. 2.4). The input map for dust is always a power-law (only in this test). The green (denoted as "PL") and purple (denoted as "MBB") points show the results from the Delta-map method for power-law dust and MBB dust, respectively. We use flat priors with $r = [0, 1]$, $\bar{\beta}_s = [-4, -2]$, $\bar{\beta}_d = [1, 2]$, and $T_d = [10\,\mathrm{K}, 40\,\mathrm{K}]$.

Eq. (14), i.e., power-law emission expanded and truncated at first order in $\delta\beta$. For Q and U maps and power-law indices of synchrotron and dust, we use the same maps as in Ref. [8], in which the mean synchrotron and dust indices are $\bar{\beta}_s = -3.00$ and $\bar{\beta}_d = 1.647$, respectively.

We add scalar and tensor modes of CMB with a tensor-to-scalar ratio of $r = 0.001$. We do not add instrumental noise but add artificial noise to regularize the covariance matrix. We then apply the Delta-map method to estimate r, $\bar{\beta}_d$, and $\bar{\beta}_s$.

For this test we use CMB+foreground maps generated at 40, 60, 140, 230, and 340 GHz, and use the two lowest and two highest frequencies to clean a map at 140 GHz. We do not apply bandpass averaging. Figure 1 shows the values of r, $\bar{\beta}_s$, and $\bar{\beta}_d$ obtained from 1000 random realizations of the CMB. We successfully recover the input values.

Next, we apply the Delta-map method for an MBB for dust (Sect. 2.4), while the simulated maps always have a power-law dust (only in this section). For this case we generate another map at 400 GHz so that we have the three highest frequencies for dust. We find that we recover the input values of r and $\bar{\beta}_s$. This is because, in the limited frequency range considered here, an MBB can mimic a power law with a spectral index of 1.647 by appropriately tuning $\bar{\beta}_d$ and $\bar{T}_d$. (MBB becomes a power law in the Rayleigh–Jeans limit, i.e., $k_B T_d / h\nu \to \infty$.) The clear correlation seen in the right panel of Fig. 1 supports this interpretation.

4. Bayesian derivation of the Delta-map method

4.1. Likelihood

So far we have formulated the Delta-map method in a "bottom-up approach"; namely, we started with a plausible form of a foreground-cleaned CMB map given in Eq. (10) and constructed a likelihood as in Eq. (27). In this section we formulate the Delta-map method in a "top-down approach" by starting with the likelihood of data $\vec{m}$ given a general sky signal $\vec{s}$, mixing matrix $\boldsymbol{D}$, and noise covariance matrix $\boldsymbol{N}$:

$$-2\ln\mathcal{L}(\vec{m}|\vec{s}, \boldsymbol{D}) = (\vec{m} - \boldsymbol{D}\vec{s})^\mathrm{T}\,\boldsymbol{N}^{-1}\,(\vec{m} - \boldsymbol{D}\vec{s}) + \ln|2\pi\boldsymbol{N}|\,. \tag{39}$$

We have concatenated the Stokes parameter data at all pixels and frequencies as

$$\vec{m} \equiv \left(\{ [Q, U]_{\nu_1}(\hat{n}_i)] \}, \ldots, \{ [Q, U]_{\nu_{N_{\mathrm{freq}}}}(\hat{n}_i) \} \right)^{\mathrm{T}}. \tag{40}$$

We similarly concatenate the noise covariance matrix $\boldsymbol{N}$, which now becomes a $2N_{\mathrm{pix}}N_{\mathrm{freq}}$-by-$2N_{\mathrm{pix}}N_{\mathrm{freq}}$ matrix. (The factor of 2 is for Q and U.) Here, N_{pix} is the number of pixels, N_{freq} is the number of frequency channels, and $\hat{n}_i$ is a sky direction unit vector toward the ith pixel. We have also defined the sky signal vector as

$$\vec{s} \equiv \left(\{ \mathrm{CMB}(\hat{n}_i) \}, \{ \mathrm{Dust}(\hat{n}_i) \}, \{ \mathrm{Synch}(\hat{n}_i) \}, \ldots \right)^{\mathrm{T}}, \tag{41}$$

whose number of elements is $2N_{\mathrm{pix}}(N_{\mathrm{temp}}+1)$, where N_{temp} is the number of foreground ("template") components. When we have synchrotron and dust, $N_{\mathrm{temp}} = 2$.

The number of rows of the mixing matrix $\boldsymbol{D}$ is $2N_{\mathrm{pix}}N_{\mathrm{freq}}$, and the number of columns is $2N_{\mathrm{pix}}(N_{\mathrm{temp}} + 1)$. We may write $\boldsymbol{D}$ using $2N_{\mathrm{pix}}N_{\mathrm{freq}}$-by-$2N_{\mathrm{pix}}$ sub-matrices as

$$\boldsymbol{D} = \left(\boldsymbol{D}^{\mathrm{CMB}} \quad \boldsymbol{D}^{\mathrm{dust}} \quad \boldsymbol{D}^{\mathrm{synch}} \quad \cdots \right), \tag{42}$$

where

$$\boldsymbol{D}^{\mathrm{CMB}} \equiv \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \tag{43}$$

$$\boldsymbol{D}^{\mathrm{dust}} \equiv \begin{pmatrix} g_{\nu_1} D_{\nu_1}^{\mathrm{dust}}(\beta_{\mathrm{d},1}, T_{\mathrm{d},1}, \ldots) & 0 & \cdots & 0 \\ 0 & g_{\nu_1} D_{\nu_1}^{\mathrm{dust}}(\beta_{\mathrm{d},2}, T_{\mathrm{d},2}, \ldots) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{\nu_1} D_{\nu_1}^{\mathrm{dust}}(\beta_{\mathrm{d},N_{\mathrm{pix}}}, T_{\mathrm{d},N_{\mathrm{pix}}}, \ldots) \\ g_{\nu_2} D_{\nu_2}^{\mathrm{dust}}(\beta_{\mathrm{d},1}, T_{\mathrm{d},1}, \ldots) & 0 & \cdots & 0 \\ 0 & g_{\nu_2} D_{\nu_2}^{\mathrm{dust}}(\beta_{\mathrm{d},2}, T_{\mathrm{d},2}, \ldots) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{\nu_2} D_{\nu_2}^{\mathrm{dust}}(\beta_{\mathrm{d},N_{\mathrm{pix}}}, T_{\mathrm{d},N_{\mathrm{pix}}}, \ldots) \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}, \tag{44}$$

and similarly for $\boldsymbol{D}^{\mathrm{synch}}$. Here, $\beta_{\mathrm{d},i} \equiv \beta_{\mathrm{d}}(\hat{n}_i)$ etc., and each block in $\boldsymbol{D}^{\mathrm{dust}}$ should include entries for both Q and U, i.e., each block is a $2N_{\mathrm{pix}}$-by-$2N_{\mathrm{pix}}$ matrix.

The effect that we have not written explicitly here is the effect of gain calibration, beam smearing, pixel window functions, or any extra smoothing/filtering applied to data. These effects will modify

all elements of the mixing matrix; e.g., "1" in the CMB matrix becomes a product of the calibration factor, beam transfer, and pixel window function operators, etc.

The next step is to expand the mixing matrix to first order in spatial variations of the spectral parameters. We redefine the signal vectors as

$$\vec{s} \rightarrow \Big(\{\text{CMB}(\hat{n}_i)\}, \{\text{Dust}(\hat{n}_i)\}, \{\text{Dust}(\hat{n}_i)\delta\beta_{\text{d},i}\}, \{\text{Dust}(\hat{n}_i)\delta T_{\text{d},i}\}, \dots,$$

$$\{\text{Synch}(\hat{n}_i)\}, \{\text{Synch}(\hat{n}_i)\delta\beta_{\text{s},i}\}, \{\text{Synch}(\hat{n}_i)\delta C_{\text{s},i}\}, \dots \Big)^{\text{T}}, \quad (45)$$

as well as the mixing matrix as

$$\boldsymbol{D} \rightarrow \Big(\boldsymbol{D}^{\text{CMB}} \quad \boldsymbol{D}^{\text{dust}}(\bar{p}^I) \quad \boldsymbol{D}^{\text{dust}}_{,1} \quad \cdots \quad \boldsymbol{D}^{\text{dust}}_{,N} \quad \boldsymbol{D}^{\text{synch}}(\bar{p}^J) \quad \boldsymbol{D}^{\text{synch}}_{,1} \quad \cdots \quad \boldsymbol{D}^{\text{synch}}_{,M} \quad \cdots \Big), \quad (46)$$

where dust and synchrotron have N ($I = 1, \dots, N$) and M ($J = 1, \dots, M$) spectral parameters, respectively. The subscript ",I" denotes derivatives with respect to the Ith foreground parameter (e.g., β_{d} and T_{d} for dust and β_{s} and C_{s} for synchrotron). All elements in this new mixing matrix are evaluated at the mean spectral parameters. The form of the likelihood given in Eq. (39) remains unchanged.

Specifically,

$$\Big(\boldsymbol{D}^{\text{dust}}(\bar{p}^I) \quad \boldsymbol{D}^{\text{dust}}_{,1} \quad \cdots \quad \boldsymbol{D}^{\text{dust}}_{,N} \Big)$$

$$= \begin{pmatrix} g_{\nu_1} D^{\text{dust}}_{\nu_1}(\bar{p}^I)\boldsymbol{I} & g_{\nu_1} D^{\text{dust}}_{\nu_1,1}\boldsymbol{I} & \cdots & g_{\nu_1} D^{\text{dust}}_{\nu_1,N}\boldsymbol{I} \\ g_{\nu_2} D^{\text{dust}}_{\nu_2}(\bar{p}^I)\boldsymbol{I} & g_{\nu_2} D^{\text{dust}}_{\nu_2,1}\boldsymbol{I} & \cdots & g_{\nu_2} D^{\text{dust}}_{\nu_2,N}\boldsymbol{I} \\ \vdots & \vdots & \ddots & \vdots \\ g_{\nu_{\text{Nfreq}}} D^{\text{dust}}_{\nu_{N_{\text{freq}}}}(\bar{p}^I)\boldsymbol{I} & g_{\nu_{N_{\text{freq}}}} D^{\text{dust}}_{\nu_{N_{\text{freq}}},1}\boldsymbol{I} & \cdots & g_{\nu_{N_{\text{freq}}}} D^{\text{dust}}_{\nu_{N_{\text{freq}}},N}\boldsymbol{I} \end{pmatrix}, \quad (47)$$

and similarly for synchrotron etc. Here, $\boldsymbol{I}$ is a $2N_{\text{pix}}$-by-$2N_{\text{pix}}$ identity matrix.

When we have only CMB, dust, and synchrotron in the sky, the number of elements of $\vec{s}$ is $2N_{\text{pix}}(N+M+3)$ and $\boldsymbol{D}$ becomes a $2N_{\text{pix}}N_{\text{freq}}$-by-$2N_{\text{pix}}(N+M+3)$ matrix. When $N_{\text{freq}} = N+M+3$, $\boldsymbol{D}$ becomes a square matrix and invertible. This is a special case in which the number of frequency channels is just enough to solve for a cleaned CMB map.

Now that the mixing matrix $\boldsymbol{D}$ does not depend on pixels, Eq. (39) can be written trivially in pixel, spherical harmonics, or wavelet space. This is an important advantage of linearization used by the Delta-map method.

4.2. Delta-map as the maximum likelihood solution

Maximizing the likelihood with respect to the CMB signal, we obtain [19]

$$\text{CMB}^{\text{ML}}(\hat{n}) = \big[(\boldsymbol{D}^{\text{T}}\boldsymbol{N}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}^{\text{T}}\boldsymbol{N}^{-1}\vec{m} \big]_{\text{CMB}}, \quad (48)$$

where the subscript "CMB" indicates that we take the elements relevant to the CMB, and the superscript "ML" stands for the maximum likelihood value. We can evaluate this in pixel, spherical harmonics, or wavelet space.

When the number of frequency channels is just enough to solve for a cleaned CMB map, $\boldsymbol{D}$ is invertible and we obtain $\text{CMB}^{\text{ML}} = \big[\boldsymbol{D}^{-1}\vec{m} \big]_{\text{CMB}}$. For example, when we have a spatially uniform

power-law component as foreground, the solution for a cleaned CMB map from two maps at two frequencies is

$$
\mathrm{CMB}^{\mathrm{ML}}(\hat{n}) = \frac{g_{\nu_2} D_{\nu_2} m_{\nu_1}(\hat{n}) - g_{\nu_1} D_{\nu_1} m_{\nu_2}(\hat{n})}{g_{\nu_2} D_{\nu_2} - g_{\nu_1} D_{\nu_1}},
\tag{49}
$$

where $D_\nu = (\nu/\nu_*)^{\bar{\beta}}$. Note that ν_* cancels. If we identify ν_2 with the reference frequency ν_*, the expression further simplifies to $\mathrm{CMB}^{\mathrm{ML}}(\hat{n}) = [m_{\nu_1}(\hat{n}) - (g_{\nu_1}/g_{\nu_*}) D_{\nu_1} m_{\nu_2}(\hat{n})]/[1 - (g_{\nu_1}/g_{\nu_*}) D_{\nu_1}]$, which coincides with the usual internal template cleaning solution with $\alpha = (g_{\nu_1}/g_{\nu_*}) D_{\nu_1}$ (see, e.g., Eq. (6) of Ref. [8]). We can obtain the same result from Eq. (10) with α given by $-A^{-1}\vec{d}_{\nu_{\mathrm{CMB}}}$ and $\nu_1 = \nu_{\mathrm{CMB}}$.

When we have a spatially varying power-law component as foreground, the solution for a cleaned CMB map from three maps at three frequencies is

$$
\mathrm{CMB}^{\mathrm{ML}}(\hat{n}) = \frac{\xi_{12} m_{\nu_3}(\hat{n}) + \xi_{31} m_{\nu_2}(\hat{n}) - \xi_{23} m_{\nu_1}(\hat{n})}{\xi_{12} + \xi_{31} - \xi_{23}},
\tag{50}
$$

where $\xi_{ij} \equiv g_{\nu_i} g_{\nu_j} \left(D_{\nu_i,\beta} D_{\nu_j} - D_{\nu_i} D_{\nu_j,\beta} \right) = g_{\nu_i} g_{\nu_j} \left(\nu_i \nu_j / \nu_*^2 \right)^{\bar{\beta}} \ln(\nu_i/\nu_j)$. The numerator evaluates to

$$
\begin{aligned}
& g_{\nu_1} g_{\nu_2} (\nu_1 \nu_2)^{\bar{\beta}} \ln\left(\frac{\nu_1}{\nu_2}\right) m_{\nu_3}(\hat{n}) \\
& + g_{\nu_3} g_{\nu_1} (\nu_3 \nu_1)^{\bar{\beta}} \ln\left(\frac{\nu_3}{\nu_1}\right) m_{\nu_2}(\hat{n}) - g_{\nu_2} g_{\nu_3} (\nu_2 \nu_3)^{\bar{\beta}} \ln\left(\frac{\nu_2}{\nu_3}\right) m_{\nu_1}(\hat{n}),
\end{aligned}
\tag{51}
$$

where we have canceled ν^* with that in the denominator. We find that this result reproduces the numerator of Eq. (16), $[Q, U]_{\nu_{\mathrm{CMB}}} + \sum_{j=1}^{2}[Q, U]_{\nu_j}$, up to a multiplicative factor if we identify ν_3 with ν_{CMB}.

We thus conclude that Eq. (10) gives the maximum likelihood solution for a cleaned CMB map to first order in perturbation, provided that the number of frequency channels is just enough to solve for a cleaned CMB map. We need to use Eq. (48) when we have more channels, which would give a less noisy CMB map.

4.3. Should we cancel CMB in a template map?

SEVEM (see Appendix A.6 of Ref. [3]) is an internal template method used by the Planck Collaboration. In this method one constructs a template foreground map by first taking the difference between two frequency channels (so that CMB cancels), fitting this difference map to a CMB channel, and subtracting the foreground. Specifically, when we have one foreground component and three frequency channels, SEVEM constructs

$$
\mathrm{CMB}^{\mathrm{SEVEM}}(\hat{n}) = m_{\nu_1}(\hat{n}) - \alpha \left[m_{\nu_2}(\hat{n}) - m_{\nu_3}(\hat{n}) \right],
\tag{52}
$$

where α is a fitting coefficient. (For this case, $\nu_{\mathrm{CMB}} = \nu_1$.) When the foreground has a spatially uniform spectrum, the coefficient is $\alpha = g_{\nu_1} D_{\nu_1} / (g_{\nu_2} D_{\nu_2} - g_{\nu_3} D_{\nu_3})$, which gives an unbiased estimate of CMB.

We find that this is not the maximum likelihood solution given in Eq. (48). Assuming that the noise covariance matrix is diagonal and homogeneous over pixels, $N = \mathrm{diag}(\sigma_{\nu_1}^2 I, \sigma_{\nu_2}^2 I, \sigma_{\nu_3}^2 I)$, where I is

a $2N_{\text{pix}}$-by-$2N_{\text{pix}}$ identity matrix, Eq. (48) becomes

$$\text{CMB}^{\text{ML}}(\hat{n}) = \frac{(g_{\nu_1}D_{\nu_1} - g_{\nu_2}D_{\nu_2})\left[g_{\nu_1}D_{\nu_1}m_{\nu_2}(\hat{n}) - g_{\nu_2}D_{\nu_2}m_{\nu_1}(\hat{n})\right]\sigma_{\nu_3}^2 + (2\text{ perm.})}{(g_{\nu_1}D_{\nu_1} - g_{\nu_2}D_{\nu_2})^2\sigma_{\nu_3}^2 + (2\text{ perm.})}, \tag{53}$$

where "2 perm." indicates two permutation terms. While this may look complicated, there is a clear interpretation. To see this, let us define

$$\text{CMB}_{ij}^{\text{ML,2band}}(\hat{n}) \equiv \frac{g_{\nu_i}D_{\nu_i}m_{\nu_j}(\hat{n}) - g_{\nu_j}D_{\nu_j}m_{\nu_i}(\hat{n})}{g_{\nu_i}D_{\nu_i} - g_{\nu_j}D_{\nu_j}}, \tag{54}$$

which is the maximum likelihood solution when we have only two frequency channels ν_i and ν_j (49). We find

$$\text{CMB}^{\text{ML}}(\hat{n}) = \frac{(g_{\nu_1}D_{\nu_1} - g_{\nu_2}D_{\nu_2})^2\text{CMB}_{12}^{\text{ML,2band}}(\hat{n})\sigma_{\nu_3}^2 + (2\text{ perm.})}{(g_{\nu_1}D_{\nu_1} - g_{\nu_2}D_{\nu_2})^2\sigma_{\nu_3}^2 + (2\text{ perm.})}. \tag{55}$$

Therefore, the maximum likelihood solution is equal to three cleaned CMB maps combined optimally, but is not equal to the SEVEM construction.

4.4. Posterior of the parameters

The next task is to derive a posterior distribution of the mean spectral parameters $\bar{p}^I$ and the cosmological parameters such as the tensor-to-scalar ratio r. We start by using Bayes's theorem to relate the posterior distribution to the likelihood given in Eq. (39) as

$$P(\bar{p}^I, \vec{s}, \boldsymbol{S}|\vec{m}) = \frac{\mathcal{L}(\vec{m}|\vec{s}, \boldsymbol{D})P(\boldsymbol{D}, \vec{s}, \boldsymbol{S})}{P(\vec{m})}, \tag{56}$$

where $\boldsymbol{S}$ is a covariance matrix of the sky signal $\vec{s} = \vec{s}_{\text{CMB}} + \vec{s}_f$ (CMB $\vec{s}_{\text{CMB}}$ and foregrounds $\vec{s}_f$), $P(\boldsymbol{D}, \vec{s}, \boldsymbol{S})$ is the prior distribution, and $P(\vec{m})$ is Bayes's (normalization) factor. Note that $\bar{p}^I$ is a vector of the parameters that determine the mixing matrix $\boldsymbol{D}$. We shall assume a flat prior on $\boldsymbol{D}$, i.e., $P(\boldsymbol{D}, \vec{s}, \boldsymbol{S}) \propto P(\vec{s}, \boldsymbol{S})$.

We first marginalize Eq. (56) over the CMB signal. To this end we assume a Gaussian probability density distribution for the CMB signal part of the prior: $-2\ln P(\vec{s}, \boldsymbol{S}) = \vec{s}_{\text{CMB}}(\boldsymbol{S}_0^{\text{CMB}})^{-1}\vec{s}_{\text{CMB}} + \ln|2\pi\boldsymbol{S}_0^{\text{CMB}}| + (\text{other terms})$, where $\boldsymbol{S}_0^{\text{CMB}}$ is a $2N_{\text{pix}}$-by-$2N_{\text{pix}}$ CMB signal covariance matrix with no beam smearing, pixel window function, extra filtering, or calibration factor.

Integrating both sides of Eq. (56) over $\vec{s}_{\text{CMB}}$, we obtain

$$-2\ln P(\bar{p}^I, \vec{s}_f, \boldsymbol{S}|\vec{m}) = \left(\vec{m} - \tilde{\boldsymbol{D}}\vec{s}_f\right)^{\text{T}}\left(\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N}\right)^{-1}\left(\vec{m} - \tilde{\boldsymbol{D}}\vec{s}_f\right)$$
$$+ \ln\left|2\pi(\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N})\right| - 2\ln P(\vec{s}_f, \boldsymbol{S}^f) + \text{constant}, \tag{57}$$

where $\boldsymbol{S}^{\text{CMB}} = \boldsymbol{D}^{\text{CMB}}\boldsymbol{S}_0^{\text{CMB}}(\boldsymbol{D}^{\text{CMB}})^{\text{T}}$, $\vec{s}_f$ is the foreground sky signal vector with $P(\vec{s}_f, \boldsymbol{S}^f)$ being its prior, and $\tilde{\boldsymbol{D}}$ is a new mixing matrix eliminating CMB, i.e.,

$$\tilde{\boldsymbol{D}} \equiv \left(\boldsymbol{D}^{\text{dust}}(\bar{p}^I) \quad \boldsymbol{D}_{,1}^{\text{dust}} \quad \cdots \quad \boldsymbol{D}_{,N}^{\text{dust}} \quad \boldsymbol{D}^{\text{synch}}(\bar{p}^J) \quad \boldsymbol{D}_{,1}^{\text{synch}} \quad \cdots \quad \boldsymbol{D}_{,M}^{\text{synch}} \quad \cdots\right). \tag{58}$$

Note that $\tilde{\boldsymbol{D}}$ is *not* a square matrix and thus not invertible. For example, when we have only dust and synchrotron as foregrounds, and dust and synchrotron have N and M spectral parameters,

respectively, $\tilde{\boldsymbol{D}}$ becomes a $2N_{\text{pix}}N_{\text{freq}}$-by-$2N_{\text{pix}}(N + M + 2)$ matrix; however, to be able to solve for a cleaned CMB map we need $N_{\text{freq}} \geq N + M + 3$.

We maximize the marginalized likelihood part of Eq. (57) (the first line on the right-hand side) with respect to the foreground sky signal to obtain the maximum likelihood value of $\vec{s}_f$ as

$$\vec{s}_f^{\text{ML}} = \left[\tilde{\boldsymbol{D}}^{\text{T}} (\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N})^{-1} \tilde{\boldsymbol{D}} \right]^{-1} \tilde{\boldsymbol{D}}^{\text{T}} (\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N})^{-1} \vec{m} \,. \tag{59}$$

Note that $\vec{s}_f^{\text{ML}}$ is *not* the maximum likelihood value of Eq. (39) but is the maximum likelihood value of the CMB-marginalized likelihood, Eq. (57), ignoring the prior term.

Inserting $\vec{s}_f^{\text{ML}}$ back into Eq. (57), we obtain the final expression for the posterior distribution:

$$\begin{aligned}
-2 \ln P(\vec{p}^I, \vec{s}_f^{\text{ML}}, \boldsymbol{S} | \vec{m}) = {} & \vec{m}^{\text{T}} \left(\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N} \right)^{-1} \vec{m} \\
& - \left[\tilde{\boldsymbol{D}}^{\text{T}} (\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N})^{-1} \vec{m} \right]^{\text{T}} \left[\tilde{\boldsymbol{D}}^{\text{T}} (\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N})^{-1} \tilde{\boldsymbol{D}} \right]^{-1} \tilde{\boldsymbol{D}}^{\text{T}} (\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N})^{-1} \vec{m} \\
& + \ln \left| 2\pi (\boldsymbol{S}^{\text{CMB}} + \boldsymbol{N}) \right| - 2 \ln P(\vec{s}_f^{\text{ML}}, \boldsymbol{S}^f) + \text{constant} \,.
\end{aligned} \tag{60}$$

This formula is the foundation of the Delta-map method. Once again, as $\tilde{\boldsymbol{D}}$ no longer depends on sky directions $\hat{n}$, we can evaluate the posterior in pixel, spherical harmonics, or wavelet space by using $\vec{m}$, $\boldsymbol{S}^{\text{CMB}}$, and $\boldsymbol{N}$ in the appropriate basis.

To understand this rather complicated expression, let us use an example of a spatially uniform power-law component as foreground and work with two maps at two frequencies. Then the first two lines yield

$$\text{CMB}^{\text{ML}} \left[\boldsymbol{S}_0^{\text{CMB}} + \frac{g_{\nu_2}^2 D_{\nu_2}^2 \boldsymbol{N}_{\nu_1} + g_{\nu_1}^2 D_{\nu_1}^2 \boldsymbol{N}_{\nu_2}}{\left(g_{\nu_2} D_{\nu_2} - g_{\nu_1} D_{\nu_1} \right)^2} \right]^{-1} \text{CMB}^{\text{ML}} \,, \tag{61}$$

where CMB^{ML} is given by Eq. (49). We find that this expression agrees with the likelihood that we constructed in Eq. (27) with appropriate coefficients α_j for this example, up to the diagonal noise (which we have not added here yet). Therefore, we conclude that Eq. (27) gives the CMB-marginalized posterior distribution of $\vec{p}^I$ and the cosmological parameters, with the maximum likelihood estimate of foregrounds for the CMB-marginalized likelihood.

As the determinant term in Eq. (60) does not depend on the mixing matrix (hence $\vec{p}^I$), maximizing the posterior with respect to $\vec{p}^I$ with a flat prior, $P(\vec{s}_f, \boldsymbol{S}^f) = \text{constant}$, amounts to minimizing χ^2 as in Eq. (32). This justifies our algorithm presented in Sect. 3.

We could have marginalized Eq. (57) over $\vec{s}_f$. Then it would also modify the determinant, which, in turn, would depend on $\vec{p}^I$. This would be the situation in which we maximize Eq. (27) including the determinant for $\vec{p}^I$, r, and s simultaneously. However, we find that this procedure produces a biased result for $\vec{p}^I$. The same behavior was observed in Ref. [19] for a flat prior on $\vec{s}_f$. An intuitive reason for this is that, when foregrounds are allowed to vary to unreasonably large values by a flat prior, the marginalized distribution has $\vec{p}^I$ that is far away from the maximum likelihood values. Therefore, a better approach would be either to use $\vec{s}_f^{\text{ML}}$ instead of marginalizing over it (which we have adopted here), to marginalize over $\vec{s}_f$ with a reasonable choice of $\boldsymbol{S}^f$ (see Refs. [20,21] for this approach), or to marginalize over $\boldsymbol{S}^f$.

Note that the posterior distribution given in Eq. (60) does not make a distinction between "CMB channels", "synchrotron channels", and "dust channels". This is a nice feature; while a heuristic

construction of cleaned CMB maps based on identifying low- and high-frequency maps as foreground channels (discussed in Sect. 2) as well as its likelihood given in Eq. (27) is intuitive and easy to understand/interpret, it might give readers the impression that there is some arbitrariness in choosing which channels are CMB etc. Consider the case where we have far more channels than the number of foreground components. Then, how should we decide which ones are CMB and foreground channels? The posterior distribution that we have derived in this section tells us that this arbitrariness does not exist: we can use Eq. (60) to combine all the data optimally to estimate r and the mean foreground parameters.

For the rest of this paper, we shall explore the simplest cases in which the number of frequency channels is just enough to solve for one CMB map and the necessary number of mean foreground parameters. Then Eq. (27) is the appropriate likelihood. That is to say, $N_{\mathrm{freq}} = N + M + 3$. For more general cases in which N_{freq} is greater, Eq. (60) should be used.

5. Validation with realistic foreground maps

5.1. Sky model

Our fiducial sky model consists of three components: polarized CMB, synchrotron, and thermal dust emission. Later, in Sect. 6.2, we also explore the effect of AME on cleaning synchrotron.

For synchrotron, we use "SynchrtronPol-commander_0256_R2.00" at 30 GHz as a template, which was derived from Planck maps using the Commander algorithm [22]. We extrapolate this map to other frequencies assuming power-law emission with spatially varying spectral indices β_{s}. The spatial variation is taken from Ref. [23], which was derived from WMAP 23 GHz and Haslam 408 MHz maps.

For the dust component, we explore two possibilities: one-component and two-component MBB dust. The latter is to test the robustness of the Delta-map method for more complex sky models; namely, we build the Delta-map method assuming one component, and apply it to a sky map containing two-component dust. Our starting point is the dust polarization template "DustPol-commander_1024_R2.00" at 353 GHz, which was derived from Planck maps using the Commander algorithm [22]. Extrapolation to other frequencies is done by either one- or two-component MBB.

The one-component MBB is given by Eq. (20), with maps of $\beta_{\mathrm{d}}(\hat{n})$ and $T_{\mathrm{d}}(\hat{n})$ taken from the Commander analysis given in Ref. [22].

A two-component MBB model is motivated by an observation that the one-component model does not explain simultaneously the Rayleigh–Jeans (microwave) and Wien (infrared) regions of the dust spectrum observed by the FIRAS and DIRBE instruments on board COBE, respectively [24]. A two-component MBB dust model has been proposed to explain the observations. Physically, two components might represent different dust grains (e.g., silicate- and carbon-dominated grains) with different temperatures. Our two-component model is based on Ref. [25]. The model assumes the following frequency dependence in brightness temperature:

$$F_\nu = a_1 \left(\frac{\nu}{\nu_0} \right)^{\beta_1+1} \frac{1}{e^{h\nu/k_{\mathrm{B}}T_1(\hat{n})} - 1} + a_2 \left(\frac{\nu}{\nu_0} \right)^{\beta_2+1} \frac{1}{e^{h\nu/k_{\mathrm{B}}T_2(\hat{n})} - 1}, \tag{62}$$

with $\nu_0 = c/\lambda_0$ and $\lambda_0 = 100\,\mu\mathrm{m}$. This function then gives $D_\nu = F_\nu/F_{\nu_*}$. Here, a_1, a_2, β_1, and β_2 are constant fitting parameters, whereas $T_2(\hat{n})$ (hot dust temperature) is a spatially varying parameter. The remaining parameter, $T_1(\hat{n})$ (cold dust temperature), is derived from T_2 assuming

thermal equilibrium with the same interstellar-radiation field [24], i.e.,

$$[T_1(\hat{n})]^{4+\beta_1} = \frac{1}{q_r} \frac{\zeta(4+\beta_2)}{\zeta(4+\beta_1)} \frac{\Gamma(4+\beta_2)}{\Gamma(4+\beta_1)} \left(\frac{h\nu_0}{k_B}\right)^{\beta_1-\beta_2} [T_2(\hat{n})]^{4+\beta_2}. \tag{63}$$

Here we have set the constant parameters as $\beta_2 = 2.82$, $\beta_1 = 1.63$, $a_1 = f_1 q_r$, $a_2 = (1 - f_1)$ with $f_1 = 0.0485$, and $q_r = 8.219$ [25]. The spatially varying parameters $T_2(\hat{n})$ and $[Q^f, U^f]_{\nu_*}(\hat{n})$ with $\nu_* = 353$ GHz are set according to the Planck results [25].

5.2. Results

We apply the Delta-map method to estimate r from 1000 random realizations of the CMB at a Healpix resolution of $N_{\text{side}} = 4$. We do not include instrumental noise here but investigate the effect of noise in Sect. 7. We prepare two sets of simulations: one with no tensor mode ($r_{\text{input}} = 0$) and another with $r_{\text{input}} = 10^{-3}$.

We smooth foreground maps and resample them at low-resolution pixels of $N_{\text{side}} = 4$ following the procedure described in Sect. 3. We use the P06 mask defined by the WMAP polarization analysis [6] and implement it in the likelihood evaluation following Appendix A.

We compare two combinations of frequencies: $(40, 60, 140, 230, 280, 340)$ GHz (Case A) and $(40, 60, 140, 230, 340, 400)$ GHz (Case B). We show the results for $\nu_{\text{CMB}} = 140$ GHz in this paper but one can choose other ν_{CMB}.[3] We do not apply a bandpass average. In this and following sections, we explore the likelihood in the parameter ranges of $r = [0, 1]$, $\beta_s = [-4.0, -2.0]$, $\beta_d = [1.0, 2.0]$, and $T_d = [5 \text{ K}, 40 \text{ K}]$ to find the best-fitting parameters for each realization. We find that all the best-fitting parameters are well within these parameter ranges listed above for all the 1000 realizations.

In Fig. 2, we show the results from applying the Delta-map method for power-law synchrotron (Sect. 2.2) and one-component MBB (Sect. 2.4) to CMB+power-law synchrotron+one-component MBB simulations, i.e., the simulations and assumptions match up to first order in spatial variation of the foreground parameters. For $r_{\text{input}} = 0$ (left panel), we find no significant bias, with a slightly larger uncertainty in r for Case B. The uncertainties are dominated by cosmic variance of the gravitational-lensing effect [8] and are found to be $r < 0.34 \times 10^{-3}$ for Case A and $r < 0.48 \times 10^{-3}$ for Case B (95% CL). These results are much better than our previous study, which found a significant bias for the P06 mask, $\delta r = 2 \times 10^{-3}$ [8], because our previous algorithm had to assume that the spectral parameters are uniform across the sky. The Delta-map method takes care of the spatial variation of the spectral parameters, as promised. Note that, for $r_{\text{input}} = 0$, a large number of samples fall into the first bin because foregrounds are successfully removed and we have put a hard prior of $r \geq 0$.

For $r_{\text{input}} = 10^{-3}$ (right panel), we find $r = (1.12 \pm 0.45) \times 10^{-3}$ for Case A and $(1.22 \pm 0.48) \times 10^{-3}$ for Case B (68% CL). The error bar here is dominated by both CMB lensing and cosmic variance of the input tensor mode.

Next, we apply the Delta-map method for power-law synchrotron and one-component MBB to CMB+power-law synchrotron+*two*-component MBB simulations, i.e., the simulations and assumptions no longer match for dust. Can we still recover the input r without a significant bias? In Fig. 3 we show the recovered r. For $r_{\text{input}} = 0$ (left panel) we find $r < 0.26 \times 10^{-3}$ for Case A and $r < 0.31 \times 10^{-3}$ for Case B (95% CL). For $r_{\text{input}} = 10^{-3}$ (right panel) we find

---

[3] As stressed at the end of Sect. 4.4, choosing a "CMB channel" makes sense only when we have just enough frequency channels to solve for one CMB map. For more general cases in which we have redundant frequencies, Eq. (60) should be used.
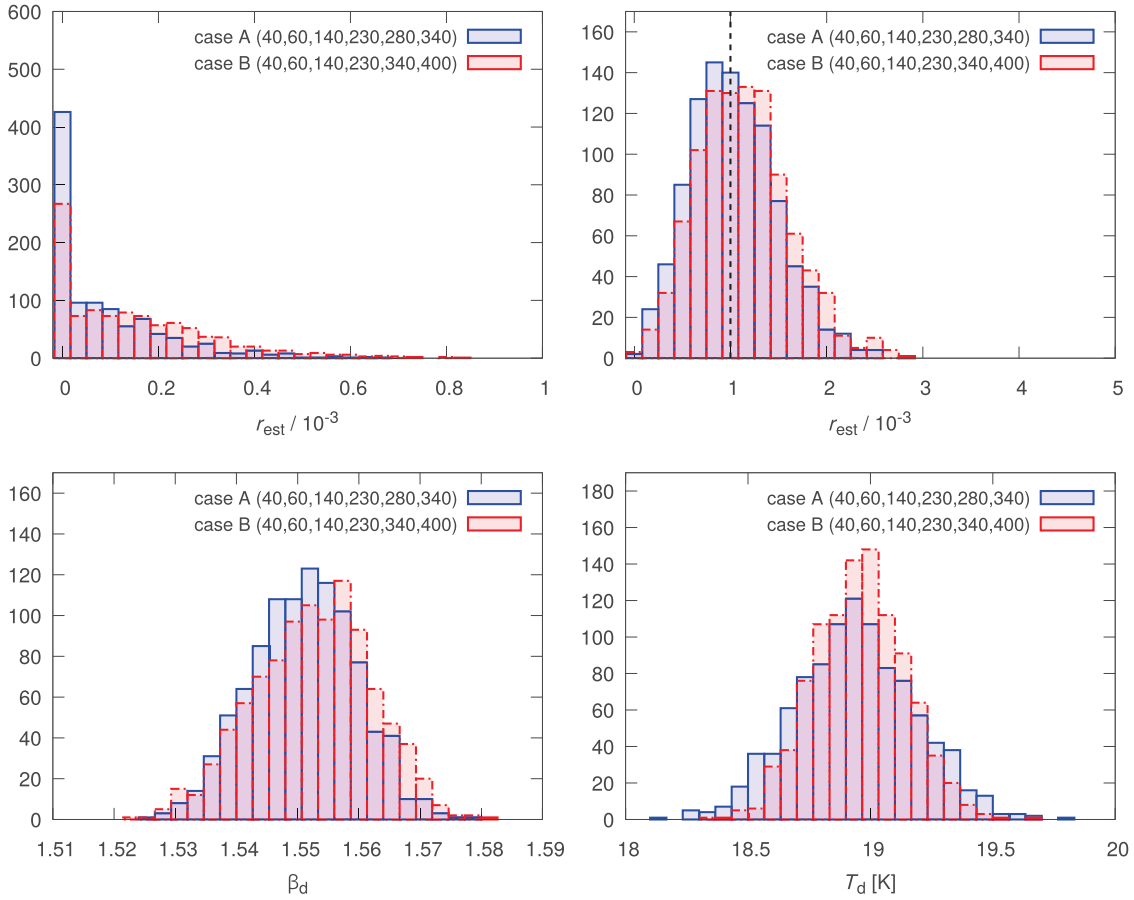
Fig. 2. Performance of the Delta-map method for power-law synchrotron and one-component MBB dust. (Upper panels) We show the distribution of r estimated from 1000 random realizations of the CMB. No instrumental noise is included. The simulation also includes power-law synchrotron and one-component MBB dust. (Left) The input value of r is zero. (Right) The input value of r is 10^{-3}. We show the results for $\nu_{\rm CMB} = 140\,{\rm GHz}$. The other channels are indicated in the figure. (Bottom panels) The distribution of $\bar{\beta}_{\rm d}$ and $\bar{T}_{\rm d}$ estimated from 1000 random realizations, for the input value of $r = 0$.

$r = (1.02 \pm 0.42) \times 10^{-3}$ for Case A and $(1.05 \pm 0.43) \times 10^{-3}$ for Case B (68% CL). Interestingly we obtain a slightly better result for this case, but the difference with respect to the previous case is well within the error bar.

We thus conclude that the Delta-map method recovers the input value of r without a significant bias in the presence of realistic spatially varying foreground spectral parameters. The slight difference in the uncertainties for Cases A and B arises from foreground residuals. Generally, in the absence of instrumental noise we expect smaller uncertainties if the foreground channels are closer to the CMB channel, because a mismatch of foreground emission between the CMB and foreground channels decreases as the CMB and foreground channels become closer to one another.

5.3. Null test

To test our algorithm more precisely, we perform a null test; namely, we clean two maps (100 and 140 GHz) following the procedure in Sect. 3.2, estimate r, and take the difference to see if there is any residual bias in the difference. This is a powerful null test (see Ref. [26] for application to the optical depth parameter τ), as it is free from cosmic variance, allowing for a precise test with
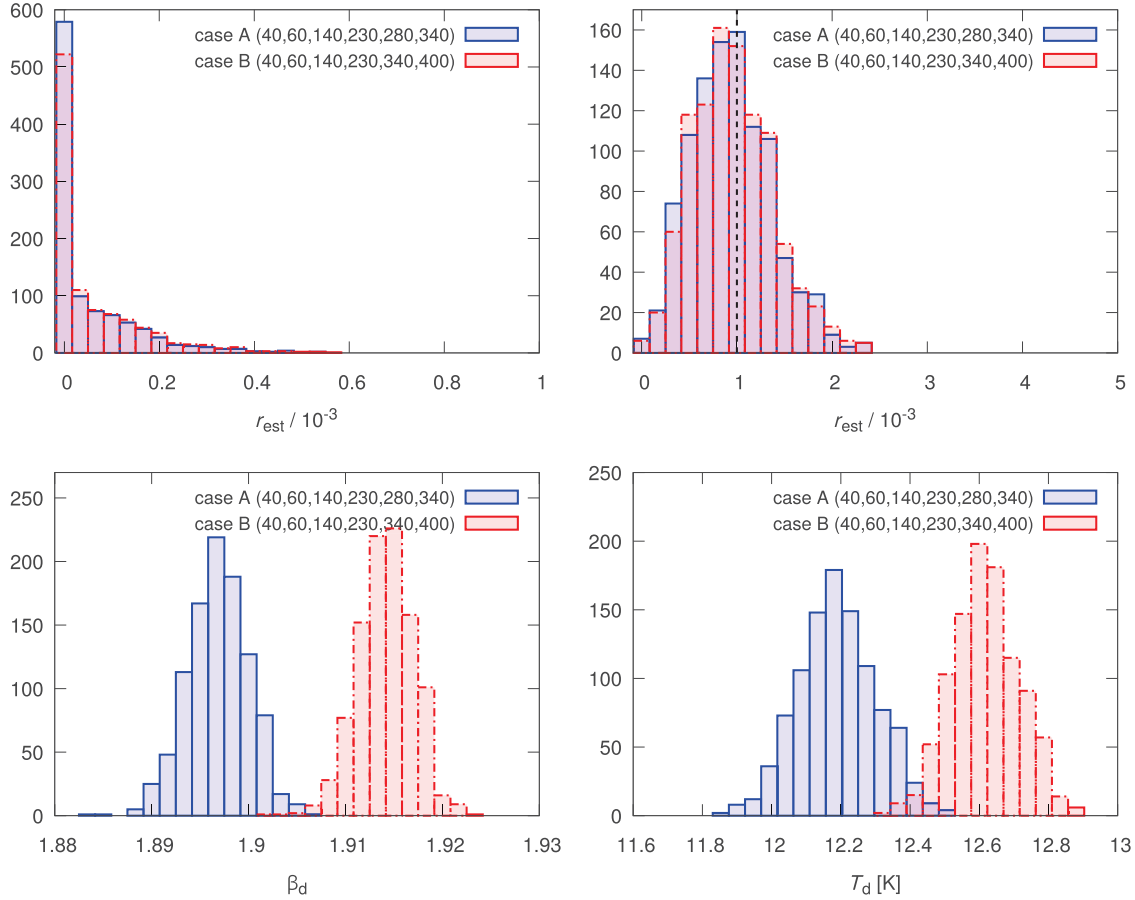
Fig. 3. Same as Fig. 2, but the Delta-map method is applied to simulations with CMB+power-law synchrotron+two-component MBB dust. Note that the Delta-map method still assumes that dust is one-component MBB. We show the results for $\nu_{CMB} = 140\,\text{GHz}$. The other channels are indicated in the figure.

a smaller error bar. We must perform this kind of test on the real data to make sure that a future detection of r is not due to a residual foreground. As we do not include instrumental noise yet, the estimates of r at $100\,\text{GHz}$ (r_{100}) and $140\,\text{GHz}$ (r_{140}) should be identical if the foreground emission is removed adequately.

We perform the null test for the simulations with $r_{input} = 10^{-3}$, power-law synchrotron, and two-component MBB dust (while the Delta-map method assumes one-component MBB). Frequency channels are $(40, 60, 100, 140, 230, 280, 340)\,\text{GHz}$.

In the left panel of Fig. 4, we show that r_{100} and r_{140} are indeed highly correlated. In the right panel, we show the distribution of the difference, $r_{100} - r_{140}$, as well as those of r_{100} and r_{140}. We find that the distribution is consistent with the null hypothesis and that the small spread of the distribution of $r_{100} - r_{140}$ is due to foreground residuals. We find $r_{100} - r_{140} = (0.72 \pm 2.98) \times 10^{-5}$ (68% CL).

6. Applications

6.1. Decorrelation

Internal template methods assume that a map of $[Q, U]$ at one frequency can be used to predict $[Q, U]$ at other frequencies as a function of foreground parameters. However, this assumption breaks down and leads to a phenomenon called "decorrelation", when we have a superposition of multiple
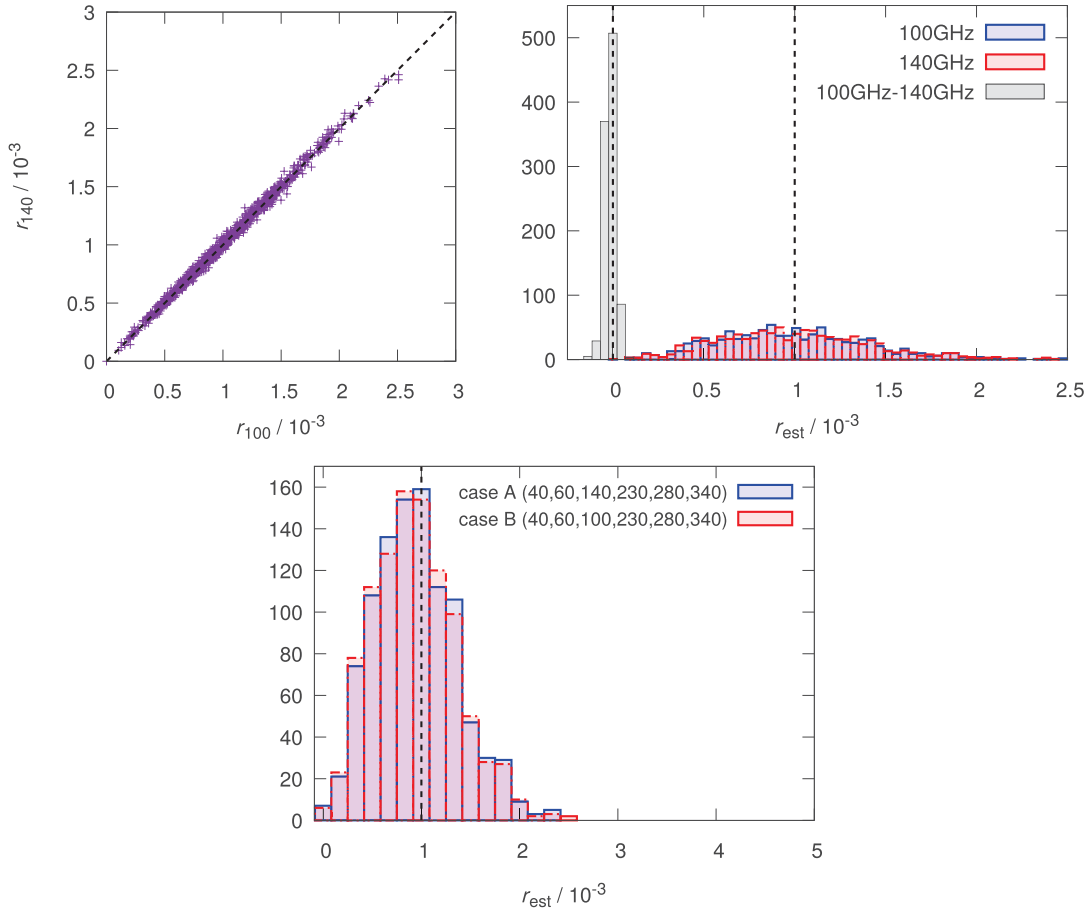
Fig. 4. Null test for r measured from cleaned 100 GHz (r_{100}) and 140 GHz (r_{140}) data. (Upper left) Scatter plot of r_{100} and r_{140}. (Upper right) Distribution of r_{100} (purple), r_{140} (green), and $r_{100} - r_{140}$ (red). (Bottom left) A zoom-in on the distributions of r_{100} and r_{140} in the upper-right panel, with the same bin size as in the top-right panel of Fig. 3. The input value of r is 10^{-3}.

components with different foreground parameters, e.g., superposition of multiple dust components with different values of β_d and T_d along the line of sight [27,28]. The same effect arises when we pixelize (or coarse-grain) a map with big pixels containing many smaller pixels with different values of β_d and T_d. Currently there is no evidence for the former, intrinsic decorrelation effect [29], whereas the latter effect can be modeled given the distribution of foreground parameters on smaller pixels. As we shall show in this section, one way to deal with this decorrelation effect is to use the foreground parameters that are slightly different for Stokes Q and U.

6.1.1. Theory

Let us take dust as an example, though the results can be generalized straightforwardly to other foreground components. Consider that we have multiple dust components along the line of sight or within a big pixel. Each ith component has a temperature $T_{d,i}$ and a polarization direction γ_i. We then observe

$$Q_\nu = \sum_i Q_{\nu,i} = g_\nu \sum_i I_{d,\nu}(T_{d,i}) \Pi_i \cos(2\gamma_i), \tag{64}$$

$$U_\nu = \sum_i U_{\nu,i} = g_\nu \sum_i I_{d,\nu}(T_{d,i}) \Pi_i \sin(2\gamma_i), \tag{65}$$

where I_d is an unpolarized dust amplitude in brightness temperature units and Π_i is a polarization fraction. We have omitted the sky direction vector $\hat{n}$ to simplify notation. Here we assume that it is only the dust temperature $T_{d,i}$ that varies and we fix the dust emissivity β_d, but it is straightforward to include it in the final result.

We consider Q first. Writing dust temperatures as $T_{d,i} = \bar{T}_d(1+\delta_i)$, we obtain

$$Q_\nu = g_\nu I_{d,\nu}(\bar{T}_d) \sum_i \left(1 + \frac{x_d e^{x_d} \delta_i}{e^{x_d}-1} + \cdots\right) \Pi_i \cos(2\gamma_i), \tag{66}$$

where $x_d \equiv h\nu/(k_B \bar{T}_d)$. Writing $A \equiv \sum_i \Pi_i \cos(2\gamma_i)$, we obtain

$$Q_\nu = g_\nu I_{d,\nu}(\bar{T}_d) A \left[1 + \frac{x_d e^{x_d}}{e^{x_d}-1} \frac{\sum_i \delta_i \Pi_i \cos(2\gamma_i)}{\sum_i \Pi_i \cos(2\gamma_i)} + \cdots\right]. \tag{67}$$

If we define a new dust temperature fluctuation variable

$$\delta_Q \equiv \frac{\sum_i \delta_i \Pi_i \cos(2\gamma_i)}{\sum_i \Pi_i \cos(2\gamma_i)}, \tag{68}$$

and truncate expansion at first order in δ_i, we find $Q_\nu = g_\nu I_{d,\nu}(T_Q) A$, where $T_Q \equiv \bar{T}_d(1 + \delta_Q)$. Performing similar calculations for U, we find $U_\nu = g_\nu I_{d,\nu}(T_U) B$, where $B \equiv \sum_i \Pi_i \sin(2\gamma_i)$, $T_U \equiv \bar{T}_d(1 + \delta_U)$, and

$$\delta_U \equiv \frac{\sum_i \delta_i \Pi_i \sin(2\gamma_i)}{\sum_i \Pi_i \sin(2\gamma_i)}. \tag{69}$$

Now, we can always write A and B as $A = \Pi \cos(2\gamma)$ and $B = \Pi \sin(2\gamma)$ by defining Π and γ by $\Pi \equiv \sqrt{A^2 + B^2}$ and $\tan(2\gamma) \equiv B/A$. We obtain the final result:

$$Q_\nu = g_\nu I_{d,\nu}(T_Q) \Pi \cos(2\gamma), \tag{70}$$

$$U_\nu = g_\nu I_{d,\nu}(T_U) \Pi \sin(2\gamma). \tag{71}$$

Therefore, to first order in temperature perturbation, Q and U behave as if they had slightly different dust temperatures. This makes an observed polarization direction, γ_{obs}, different from γ because

$$2\gamma_{obs} = \tan^{-1}\left[\frac{U(\nu)}{Q(\nu)}\right] = \tan^{-1}\left[\frac{I_{d,\nu}(T_U)}{I_{d,\nu}(T_Q)} \tan(2\gamma)\right]. \tag{72}$$

Importantly, *the observed polarization direction changes with frequency*, i.e., $\gamma_{obs} = \gamma_{obs}(\nu)$. This is the origin of frequency decorrelation.

We can account for variations in any other foreground parameters (e.g., β_d, β_s, C_s) by using slightly different values for Q and U.

6.1.2. Impact on r

We include decorrelation in our two-component MBB dust simulation as follows. We perturb the temperature of hot dust, T_2, in Q and U as

$$T_2^Q(\hat{n}) \equiv T_2(\hat{n}) + \epsilon_Q \sigma_T, \tag{73}$$
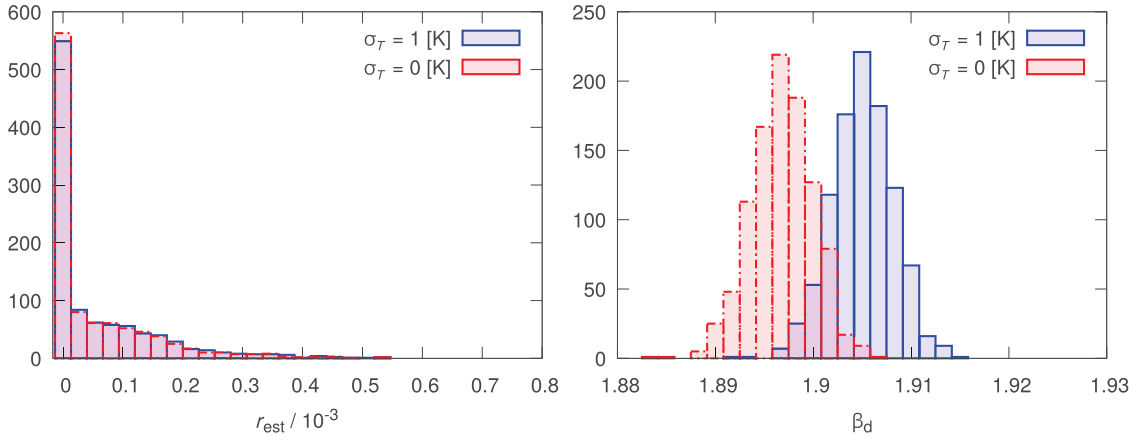
Fig. 5. Distribution of r and $\bar{\beta}_{\rm d}$ obtained from 1000 random realizations of the CMB with (purple) and without (green) decorrelation of dust polarization in the simulation. The Delta-map method assumes that there is no decorrelation. We used (40,60,230,280,340) GHz channels to clean the 140 GHz channel. The simulation also includes power-law synchrotron and two-component MBB dust.

$$T_2^U(\hat{n}) \equiv T_2(\hat{n}) + \epsilon_U \sigma_T, \tag{74}$$

where ϵ_Q and ϵ_U are Gaussian random numbers with unit variance and σ_T is a parameter that controls the effect of decorrelation. The temperatures of cold dust, T_1^Q and T_1^U, are related to T_2^Q and T_2^U via Eq. (63). Using these temperatures, we extrapolate the Planck dust polarization maps in brightness temperature units at 353 GHz to other frequencies using

$$Q_\nu(\hat{n}) = g_\nu \frac{I_{{\rm d},\nu}(T_2^Q(\hat{n}))}{I_{{\rm d},353}(T_2(\hat{n}))} Q_{353}^f(\hat{n}), \tag{75}$$

$$U_\nu(\hat{n}) = g_\nu \frac{I_{{\rm d},\nu}(T_2^U(\hat{n}))}{I_{{\rm d},353}(T_2(\hat{n}))} U_{353}^f(\hat{n}), \tag{76}$$

where $I_\nu(T_2(\hat{n}))$ is the frequency spectrum in the two-component model given by Eq. (62).

We can implement T_2^Q and T_2^U directly into the Delta-map method by increasing the number of mean temperatures from one $\bar{T}_{\rm d}$ to two $\bar{T}_2^Q$ and $\bar{T}_2^U$. However, in this section we instead apply the Delta-map method for one temperature $\bar{T}_{\rm d}$ and see how much impact the decorrelation has on our estimation of r. To this end, we generate temperature fluctuations with an r.m.s. of 1 K, i.e., $\sigma_T = 1$ K, and include them in the hot dust temperature of two-component MBB dust. To avoid negative temperature by fluctuations, we do not perturb it if $T_2 < 5$ K.

In Fig. 5 we find that decorrelation with $\sigma_T = 1$ K shifts the best-fitting value of $\bar{\beta}_{\rm d}$ (right panel); a superposition of varying foreground parameters changes the average foreground spectrum as shown by Ref. [30]. Fortunately, this has little impact on r (left panel). This is good news for estimation of r, though the appropriate magnitudes of σ_T in real sky and r.m.s. of other foreground parameters are yet to be determined. At least our method provides a framework within which frequency decorrelation due to any foreground parameters can be treated consistently to first order in perturbation. In Table 1, we summarize the mean and variance of the sky-averaged foreground parameters obtained from 1000 random realizations of the CMB with the assumed foreground maps and the frequency channels used in the analysis.

Table 1. Mean and variance of the reconstructed sky-averaged foreground parameters

dust model ($r = 0$)	$\bar{\beta}_s$	$\bar{\beta}_d$	$\bar{T}_d$
one-component (Case A)	-2.920 ± 0.020	1.5523 ± 0.009	18.97 ± 0.24
one-component (Case B)	-2.906 ± 0.024	1.5540 ± 0.010	18.99 ± 0.19
two-component (Case A)	-2.940 ± 0.026	1.8977 ± 0.003	12.21 ± 0.11
two-component (Case B)	-2.934 ± 0.029	1.9150 ± 0.003	12.64 ± 0.10
two-component (Case A with $\sigma_T = 1$ [K])	-2.939 ± 0.026	1.9059 ± 0.003	12.06 ± 0.11

6.2. AME

While it is widely believed that Galactic synchrotron and thermal dust emissions are the most dominant sources of polarized foreground emission on large angular scales [11], there might be a third polarized component, AME, whose most plausible candidate includes tiny PAH particles spinning with dipole moments [31,32]. While the AME has been shown to contribute to μK-level fluctuations in unpolarized intensity at 40 GHz and even smaller fluctuations at higher frequencies [33], little is known about its polarization properties.

The Planck Collaboration has put an upper bound on the polarization fraction of AME of $\Pi \lesssim 1.6\%$ in the Perseus region [34]. The QUIJOTE Collaboration has placed a limit of $\Pi \lesssim 0.39\%$ in the molecular complexes W43 [35]. The polarization degree in more diffuse areas that are suitable for cosmological analyses is unknown owing to the strong contamination from Galactic synchrotron emission. Even though the polarization contribution of AME to the microwave data near the foreground minimum, $\nu \approx 80$–100 GHz, is expected to be negligible, it affects component separation by changing the interpretation of low-frequency data from pure synchrotron to synchrotron plus AME. In this sense, AME affects the estimation of r indirectly. In this section we show that this indirect effect can be absorbed into the synchrotron curvature parameter C_s.

To estimate the effect of AME, we model it following Ref. [36]. Specifically, we take the intensity map of thermal dust emission at 353 GHz and rescale it by a factor of 0.9 that is found in the Planck 2015 results [34] to predict the AME intensity at 22.8 GHz. We then assign a 1% polarization fraction with the same polarization angles as those of the thermal dust component, with a frequency spectrum given by the cold-medium model described in Ref. [37].

We first apply the Delta-map method to the simulation with CMB+power-law synchrotron+one-component MBB dust+AME, using the Delta-map method for power-law synchrotron and one-component MBB dust. We use $(40, 60, 140, 230, 280, 340)$ GHz. We find a significant bias in the estimated r: $r_{est} = 0.068 \pm 0.011$ (68% CL) for $r_{input} = 0$. At first this result was surprising because the amplitude of AME at 140 GHz is too small to be relevant. We then find that AME affects r by changing the best-fitting synchrotron. We therefore added curvature to the synchrotron spectrum and implemented it in the Delta-map method following the procedure in Sect. 2.3. Including C_s requires one more frequency channel for synchrotron, so we add one more frequency, which is to be determined. When adding C_s, we need to specify the pivot frequency ν_s; we write a curved power-law synchrotron spectrum as $\nu^{\beta_s + C_s \ln(\nu/\nu_s)}$. We tried $\nu_s = 40$ and 50 GHz. The choice of ν_s does not affect the CMB result because it cancels in the cleaned CMB map, but it would change the meaning of the synchrotron parameters.

In the top panels of Fig. 6, we show scatter plots of $\bar{C}_s$ and $\bar{\beta}_s$ from 1000 simulations of CMB+AME+power-law synchrotron+one-component MBB dust emission. We study four configurations: to our Case A configuration $(40, 60, 140, 230, 280, 340)$ GHz we add 78 GHz (left panel)
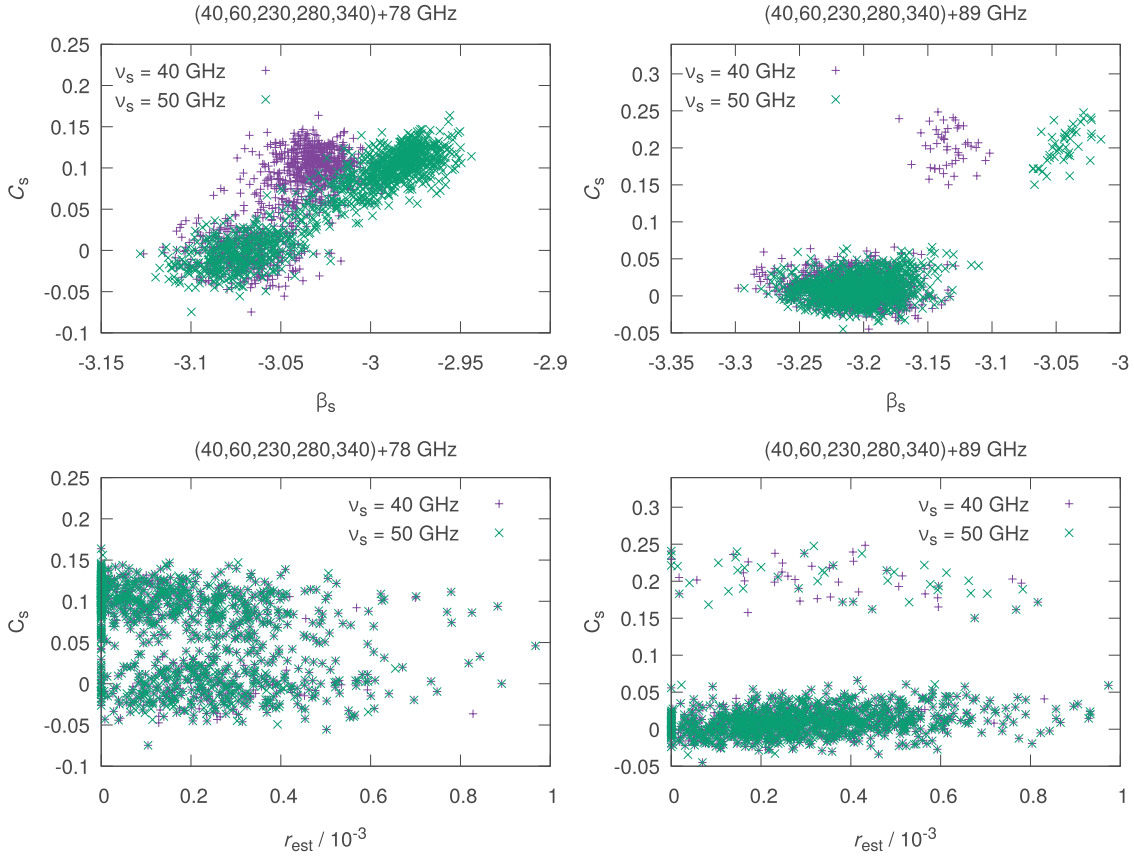
Fig. 6. Scatter plots of $\bar{\beta}_s$ and $\bar{C}_s$ (top panels) and r and $\bar{C}_s$ (bottom panels) obtained from 1000 random realizations of the CMB with $r = 0$. The simulation also includes AME, power-law synchrotron, and one-component MBB dust. We apply the Delta-map method for curved power-law synchrotron and one-component MBB dust. We show the results for $\nu_{\mathrm{CMB}} = 140\,\mathrm{GHz}$. The other frequencies used are shown above each panel. We consider two pivot frequencies (40 and 50 GHz), as shown in the figure.

or 89 GHz (right panel) for $\bar{C}_s$, and for the pivot scale we choose $\nu_s = 40\,\mathrm{GHz}$ or $50\,\mathrm{GHz}$. We find that the distribution is split into two clusters. One clusters around $\bar{C}_s \simeq 0$ and $\bar{\beta}_s \simeq -3.07$ and -3.20 in the left and right panels, respectively, and the other around $\bar{C}_s = 0.1$ and 0.2 in the left and right panels, respectively. In the former cluster, because $|\bar{\beta}_s| \gg |\bar{C}_s \ln(\nu/\nu_s)|$, most of the low-frequency foreground emission can be removed as a power-law emission, and AME is absorbed by the Delta-map terms proportional to $\delta\beta_s(\hat{n})$ and $\delta C_s(\hat{n})$ (see Eq. (18)). This interpretation is supported by the fact that the distribution in this cluster is not very sensitive to the choice of ν_s. In some realizations, however, non-zero $\bar{C}_s$ is needed by chance; in this case the results fall into the latter cluster.

How about estimates of r? In the bottom panels of Fig. 6, we find that the mean curvature parameter $\bar{C}_s$ is not correlated with r. In the left panel of Fig. 7, we show the distribution of r for two frequency configurations with $r_{\mathrm{input}} = 0$. We find $r < 5.2 \times 10^{-4}$ and $r < 6.4 \times 10^{-4}$ (95% CL) for the additional 78 and 89 GHz channels, respectively; thus, we find no evidence for bias, unlike when we did not include $C_s(\hat{n})$ to absorb the effect of AME. In fact, the different number of realizations in the first bin in the left panel of Fig. 7 is due to foreground residuals. Because the polarization intensity of the AME rapidly decreases above 40 GHz, an additional channel at 78 GHz does a better job than 89 GHz in order to capture the spectrum of the AME component, and relatively larger foreground residuals remain for an additional channel at 89 GHz.
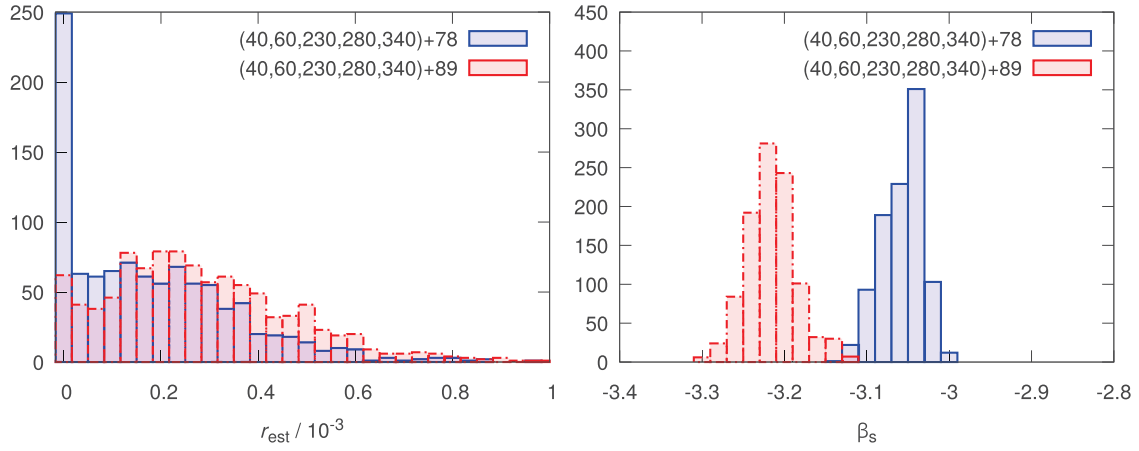
Fig. 7. Distribution of r (left) and $\bar{\beta}_s$ (right) estimated from 1000 random realizations of the CMB with $r = 0$. The simulation also includes AME, power-law synchrotron, and one-component MBB dust. We use the Delta-map method for one-component MBB dust and curved power-law synchrotron with the pivot frequency of $\nu_s = 40$ GHz. We show the results for $\nu_{\mathrm{CMB}} = 140$ GHz.

On the other hand, the distribution of $\bar{\beta}_s$ shifts significantly between the cases of adding 78 or 89 GHz. The peak positions in the distribution of $\bar{\beta}_s$ depend upon the frequency channels used to absorb AME and the frequency spectrum of AME. Because of the fact that the AME spectrum is generally characterized by a negative spectral index and a negative running index, the Delta-map method returns smaller $\bar{\beta}_s$ if we choose a higher-frequency channel to absorb AME (farther away from the other two low-frequency channels at 40 and 60 GHz). Even though the results for $\bar{\beta}_s$ are different (indicating that the physical meaning of $\bar{\beta}_s$ is no longer a simple synchrotron index), we find no bias in the estimated r in either case. (We find a slightly better result for an additional channel at 78 GHz as described above.) This is good news for the estimation of r.

7. Effect of instrumental noise

All the results that we have presented so far were derived without instrumental noise. Here, we shall include instrumental noise in the simulation. As with other template-cleaning methods, the Delta-map method also suffers from a boosted noise; because the method is based on a linear combination of maps at different frequencies having the CMB signal in common, we cannot avoid removing some portion of the CMB when removing foreground components from the CMB channel. A linear combination of Gaussian random noises, on the other hand, will simply result in a larger Gaussian random noise; thus, the net effect is to boost the noise relative to the CMB signal in the map after the Delta-map method is applied.

The effective noise level in a cleaned CMB map can be readily estimated once the noise level at each observing frequency channel is specified. When we use two channels for synchrotron and three channels for dust, the noise covariance matrix for each cleaned CMB map is given by the noise terms in Eq. (28) as $\boldsymbol{n}^{\nu_{\mathrm{CMB}},\mathrm{cleaned}} = \left(\boldsymbol{N}_{\nu_{\mathrm{CMB}}} + \sum_{j=1}^{5} \alpha_j^2 \boldsymbol{N}_{\nu_j} \right) / \left(1 + \sum_{j=1}^{5} \alpha_j \right)^2$, where α_i are the coefficients given by solving Eqs. (5) and (6), and $\boldsymbol{N}_\nu$ is the noise covariance matrix at a frequency channel ν.

In Table 2, we show some examples of noise specifications and derived constraints on r using them. The effective noise levels that we find are 3.0, 4.0, 3.9, 3.7, 3.4, 3.6, 3.8, 3.9, 4.7, 4.5, and

Table 2. Some example noise specifications at 40, 60, 100, 140, 230, 280, 340, and 400 GHz, and derived upper limits on r from cleaned CMB maps after the Delta-map method. They are in units of thermodynamic μK arcmin. The lowest values are shown in boldface.

Frequency [GHz]	synchrotron		CMB		dust				limits on r
	40	60	100	140	230	280	340	400	(95% CL)
	20.0	6.0	–	**1.0**	**1.0**	–	5.0	18.0	$< \mathbf{1.4 \times 10^{-3}}$
	20.0	6.0	–	2.0	2.0	–	5.0	18.0	$< 2.1 \times 10^{-3}$
	10.0	6.0	–	2.0	2.0	–	5.0	18.0	$< 2.0 \times 10^{-3}$
	20.0	**3.0**	–	2.0	2.0	–	5.0	18.0	$< 1.9 \times 10^{-3}$
	20.0	6.0	–	**1.0**	2.0	–	5.0	18.0	$< 1.6 \times 10^{-3}$
Noise	20.0	6.0	–	2.0	**1.0**	–	5.0	18.0	$< 1.9 \times 10^{-3}$
[μK arcmin]	20.0	6.0	–	2.0	2.0	–	**2.5**	18.0	$< 1.9 \times 10^{-3}$
	20.0	6.0	–	2.0	2.0	–	5.0	**9.0**	$< 2.0 \times 10^{-3}$
	20.0	6.0	2.0	–	2.0	–	5.0	18.0	$< 2.9 \times 10^{-3}$
	20.0	6.0	–	2.0	–	2.0	5.0	18.0	$< 2.8 \times 10^{-3}$
	20.0	6.0	–	2.0	2.0	2.0	5.0	–	$< 5.7 \times 10^{-3}$

7.0 μK arcmin from the top to bottom specifications in the table.[4] To derive these values, we used the Delta-map method with a power-law synchrotron (Sect. 2.2) with $\bar{\beta}_s = -3.0$ and one-component MBB dust (Sect. 2.4) with $\bar{\beta}_d = 1.65$ and $\bar{T}_d = 19.15$ K. The total effective noise slightly depends upon the foreground parameters. In fact, recent results from S-PASS have shown that the spectral index of Galactic synchrotron emission is slightly steeper than this value ($\bar{\beta}_s^{S-PASS} = -3.22$) [38]; in that case, the total effective noise for the specification described in the top row in Table 2 would change from 3.0 to 2.65 μK arcmin.

If we combine several cleaned CMB maps (e.g., 100 and 140 GHz), the effective noise will be reduced to some extent and its inverse noise covariance matrix will be given by the inverse weighted sum of the noise matrices of the cleaned CMB maps as

$$\left(\boldsymbol{n}^{\text{tot}}\right)^{-1} = \sum_{\nu_{\text{CMB}}} \left(\boldsymbol{n}^{\nu_{\text{CMB}},\text{cleaned}}\right)^{-1}. \tag{77}$$

Note that the noise cannot be reduced by as much as the inverse square of the number of maps ($n_{\text{map}}^{-1/2}$), because the noises in the cleaned maps are correlated with one another. This might give readers the impression that the Delta-map method as presented in this paper is unbiased, but it is not necessarily minimum variance. However, this is not the case; as we discussed at the end of Sect. 4.4, the posterior distribution given in Eq. (60) can be used to combine all maps optimally to estimate r and the mean foreground parameters.

8.   Conclusions

In this paper we have extended the internal template method by accounting for spatially varying foreground parameters such as β_s, C_s, β_d, and T_d. This was achieved by expanding the frequency spectra of foregrounds to first order in perturbations (spatial variations) [12,13]. We show that a

---

[4] Note that these values are derived with a Gaussian smoothing of 2220 arcmin (FWHM). To obtain the equivalent non-smoothed white noise amplitudes in units of μK arcmin, we need to multiply these numbers by a factor of 3.76.

foreground-cleaned CMB map obtained by our method ("Delta-map method") is the maximum likelihood solution (48), and its posterior distribution for the CMB parameters (e.g., tensor-to-scalar ratio parameter r) and the sky-averaged, mean foreground parameters (e.g., $\bar{\beta}_\mathrm{s}$, $\bar{C}_\mathrm{s}$, $\bar{\beta}_\mathrm{d}$, and $\bar{T}_\mathrm{d}$) are obtained by marginalizing over CMB and using the maximum likelihood solution to the foreground maps (60). We have found that the Delta-map method removes a bias in r, which was found in the previous study assuming uniform spectral indices [8], successfully. We have checked this against realistic sky simulations including polarized CMB, power-law synchrotron, and one- or two-component MBB dust emission.

While we have formulated and validated the Delta-map method in pixel space in this paper, it is valid in any other space such as spherical harmonics and wavelet space.

Armed with this new tool, we have presented physically motivated ways to address the frequency decorrelation due to averaging of different foreground parameters along the line of sight (or over pixels) and the effect of AME on foreground cleaning. We have shown that, to first order in perturbation, the decorrelation effect can be accounted for by using slightly different foreground parameters for Stokes Q and U. As an application we simulated a dust map with the r.m.s. temperature fluctuation along the line of sight of 1 K. We show that the best-fitting dust index $\bar{\beta}_\mathrm{d}$ is affected by this but the estimation of r is hardly affected.

While AME has a negligible contribution to the foreground minimum, 80–100 GHz, it affects the estimation of r via its effect on synchrotron cleaning. If not accounted for, it can result in a bias in r as large as $\delta r \simeq 0.06$ (for 1% polarization in AME). However, we have found that the effect of AME on synchrotron cleaning can be absorbed by the curvature parameter in the synchrotron spectrum, C_s, within the Delta-map method. We find that the Delta-map method can reduce the bias in r to an undetected level in our simulation.

The remaining question is how the performance of the Delta-map method, especially the derived uncertainty in r, compares with those of other foreground cleaning methods in the literature. We leave this for future work.

Appendix A.   Mask

One advantage of pixel-based internal template methods is that one can easily and exactly account for the effect of a mask through the noise covariance matrix [6]. To implement it, we rewrite the likelihood function as

$$\mathcal{L} = \frac{\exp[-\frac{1}{2}(\boldsymbol{N}^{-1}\vec{x})^\mathrm{T}(\boldsymbol{N}^{-1}\boldsymbol{S}\boldsymbol{N}^{-1} + \boldsymbol{N}^{-1})^{-1}(\boldsymbol{N}^{-1}\vec{x})]|\boldsymbol{N}^{-1}|}{|\boldsymbol{N}^{-1}\boldsymbol{S}\boldsymbol{N}^{-1} + \boldsymbol{N}^{-1}|^{1/2}}, \qquad (A.1)$$

where $\boldsymbol{S} = \boldsymbol{C}^\mathrm{tens}(r) + \boldsymbol{C}^\mathrm{scal}(s)$ and $\boldsymbol{N}$ are the CMB signal and noise covariance matrices, respectively.

Fig. A.1. Conservative (left), fiducial (middle), and aggressive (right) masks used in this work, the sky coverages of which are $f_{\text{fsky}} = 0.22$, 0.59, and 0.69, respectively.

Table A.1. Limits on r with different masks against the foreground model of one-component MBB dust and power-law synchrotron emissions with our Case A configuration (40, 60, 140, 230, 280, 340 GHz). The input tensor-to-scalar ratio is $r = 0$.

Mask	f_{sky}	mean	limit (95% CL)
full-sky	1.0	3.4×10^{-4}	$< 0.72 \times 10^{-3}$
aggressive	0.69	1.7×10^{-4}	$< 0.48 \times 10^{-3}$
fiducial	0.59	9.9×10^{-5}	$< 0.34 \times 10^{-3}$
conservative	0.22	7.7×10^{-5}	$< 0.52 \times 10^{-3}$

We rearrange the data vector and the covariance matrix such that unmasked pixels appear first and masked pixels afterward. Suppose that the structure of N is given by

$$N = \begin{pmatrix} T & U \\ U^{\text{T}} & W \end{pmatrix}, \tag{A.2}$$

where T is the noise covariance matrix for unmasked pixels, W is that for masked pixels, and U is that for their correlation. Then the matrix inverse is given by

$$N^{-1} = \begin{pmatrix} T & U \\ U^{\text{T}} & W \end{pmatrix}^{-1} = \begin{pmatrix} T^{-1} + T^{-1}UQ^{-1}U^{\text{T}}T^{-1} & -T^{-1}UQ^{-1} \\ -Q^{-1}U^{\text{T}}T^{-1} & Q^{-1} \end{pmatrix} \equiv \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}, \tag{A.3}$$

where $Q \equiv W - U^{\text{T}}T^{-1}U$ and $T^{-1} = B_{11} - B_{12}B_{22}^{-1}B_{21}$ (Schur complement). We then assign infinite noise to the masked pixels such that $N \rightarrow N + \lambda(I - M)$, where I is the unit matrix and M is the diagonal matrix whose elements are zero for masked pixels and unity otherwise. In the limit of $\lambda \rightarrow \infty$, $Q \rightarrow \infty$, and the inverse of N is given by

$$N^{-1} \rightarrow \begin{pmatrix} T^{-1} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} B_{11} - B_{12}B_{22}^{-1}B_{21} & 0 \\ 0 & 0 \end{pmatrix}. \tag{A.4}$$

In this work, we have utilized the "P06 mask" defined by the WMAP polarization analysis, degrading it to $N_{\text{side}} = 4$. The sky coverage of our default P06 mask at $N_{\text{side}} = 4$ is $f_{\text{sky}} = 0.56$. Whilst we find that this mask is sufficient to suppress the residual $\mathcal{O}(\delta p^2)$ term in Eq. (3), it is interesting to investigate how the results change if one uses different masks. We apply the Delta-map method to the simulations with two different masks, namely more aggressive and conservative ones, whose sky coverages are $f_{\text{sky}} = 0.22$ and 0.69, respectively, as shown in Fig. A.1. The results are summarized in Table A.1. We find that the aggressive mask leads to a larger bias and a weaker constraint because of foreground residuals, while the conservative mask leads to a smaller bias and a weaker constraint because of the smaller sky coverage. We caution, however, that one should not accept the results for

the aggressive (and full-sky) case at face value because our foreground model may not be accurate enough close to the Galactic plane.

References

[1] M. Kamionkowski and E. D. Kovetz, Ann. Rev. Astron. Astrophys. **54**, 227 (2016) [arXiv:1510.06042 [astro-ph.CO]] [Search INSPIRE].

[2] P. A. R. Ade et al. BICEP2/Keck and Planck Collaborations, Phys. Rev. Lett. **114**, 101301 (2015) [arXiv:1502.00612 [astro-ph.CO]] [Search INSPIRE].

[3] S. M. Leach et al., Astron. Astrophys. **491**, 597 (2008) [arXiv:0805.0269 [astro-ph]] [Search INSPIRE].

[4] M. Remazeilles et al., J. Cosmol. Astropart. Phys. **4**, 023 (2018) [arXiv:1704.04501 [astro-ph.CO]] [Search INSPIRE].

[5] Y. Akrami et al. [Planck Collaboration IV], arXiv:1807.06208 [astro-ph.CO] [Search INSPIRE].

[6] L. Page et al., Astrophys. J. Suppl. **170**, 335 (2007) [arXiv:astro-ph/0603450] [Search INSPIRE].

[7] G. Efstathiou, S. Gratton, and F. Paci, Mon. Not. R. Astron. Soc. **397**, 1355 (2009) [arXiv:0902.4803 [astro-ph.CO]] [Search INSPIRE].

[8] N. Katayama and E. Komatsu, Astrophys. J. **737**, 78 (2011) [arXiv:1101.5210 [astro-ph.CO]] [Search INSPIRE].

[9] R. Fernández-Cobos, P. Vielva, R. B. Barreiro, and E. Martínez-González, Mon. Not. R. Astron. Soc. **420**, 2162 (2012) [arXiv:1106.2016 [astro-ph.CO]] [Search INSPIRE].

[10] R. Fernández-Cobos, A. Marcos-Caballero, P. Vielva, E. Martínez-González, and R. B. Barreiro, Mon. Not. R. Astron. Soc. **459**, 441 (2016) [arXiv:1601.01515 [astro-ph.CO]] [Search INSPIRE].

[11] K. Ichiki, Prog. Theor. Exp. Phys. **2014**, 06B109 (2014).

[12] F. R. Bouchet and R. Gispert, New Astron. **4**, 443 (1999) [arXiv:astro-ph/9903176] [Search INSPIRE].

[13] R. Saha and P. K. Aluri, Astrophys. J. **829**, 113 (2016) [arXiv:1509.03656 [astro-ph.CO]] [Search INSPIRE].

[14] P. A. R. Ade et al. [Planck Collaboration XIII], Astron. Astrophys. **594**, A13 (2016) [arXiv:1502.01589 [astro-ph.CO]] [Search INSPIRE].

[15] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, Astrophys. J. **622**, 759 (2005) [arXiv:astro-ph/0409513] [Search INSPIRE].

[16] H. Ishino et al., Proc. SPIE. **9904**, 9904 (2016).

[17] H. K. Eriksen et al., Astrophys. J. **656**, 641 (2007) [arXiv:astro-ph/0606088] [Search INSPIRE].

[18] Steven G. Johnson, The NLopt nonlinear-optimization package, http://ab-initio.mit.edu/nlopt, date last accessed March 1, 2019.

[19] R. Stompor, S. Leach, F. Stivoli, and C. Baccigalupi, Mon. Not. R. Astron. Soc. **392**, 216 (2009) [arXiv:0804.2645 [astro-ph]] [Search INSPIRE].

[20] F. Vansyngel, B. D. Wandelt, J.-F. Cardoso, and K. Benabed, Astron. Astrophys. **588**, A113 (2016) [arXiv:1409.0858 [astro-ph.CO]] [Search INSPIRE].

[21] F. Vansyngel, The blind Bayesian approach to cosmic microwave background data analysis, *Ph.D. Thesis*, Université Pierre et Marie Curie—Paris VI (2014).

[22] R. Adam et al. [Planck Collaboration X], Astron. Astrophys. **594**, A10 (2016) [arXiv:1502.01588 [astro-ph.CO]] [Search INSPIRE].

[23] M.-A. Miville-Deschênes, N. Ysard, A. Lavabre, N. Ponthieu, J. F. Macías-Pérez, J. Aumont, and J. P. Bernard, Astron. Astrophys. **490**, 1093 (2008) [arXiv:0802.3345 [astro-ph]] [Search INSPIRE].

[24] D. P. Finkbeiner, M. Davis, and D. J. Schlegel, Astrophys. J. **524**, 867 (1999) [arXiv:astro-ph/9905128] [Search INSPIRE].

[25] A. M. Meisner and D. P. Finkbeiner, Astrophys. J. **798**, 88 (2015) [arXiv:1410.7523 [astro-ph.GA]].

[26] G. Hinshaw et al., Astrophys. J. Suppl. **180**, 225 (2009) [arXiv:0803.0732 [astro-ph]] [Search INSPIRE].

[27] K. Tassis and V. Pavlidou, Mon. Not. R. Astron. Soc. **451**, L90 (2015) [arXiv:1410.8136 [astro-ph.CO]] [Search INSPIRE].

[28] J. Poh and S. Dodelson, Phys. Rev. D **95**, 103511 (2017).

[29] Y. Akrami et al. [Planck Collaboration XI], arXiv:1801.04945 [astro-ph.GA] [Search INSPIRE].

[30] J. Chluba, J. C. Hill, and M. H. Abitbol, Mon. Not. R. Astron. Soc. **472**, 1195 (2017).

[31] E. M. Leitch and A. C. R. Readhead, Adv. Astron. **2013**, 352407 (2013).

[32] C. Dickinson et al., New Astron. Rev. **80**, 1 (2018).

[33] C. L. Bennett et al., Astrophys. J. Suppl. **208**, 20 (2013) [arXiv:1212.5225 [astro-ph.CO]] [Search INSPIRE].

[34] P. A. R. Ade et al. [Planck Collaboration XXV], Astron. Astrophys. **594**, A25 (2016) [arXiv:1506.06660 [astro-ph.GA]] [Search INSPIRE].

[35] R. Génova-Santos et al., Mon. Not. R. Astron. Soc. **464**, 4107 (2017) [arXiv:1605.04741 [astro-ph.GA]] [Search INSPIRE].

[36] M. Remazeilles, C. Dickinson, H. K. K. Eriksen, and I. K. Wehus, Mon. Not. R. Astron. Soc. **458**, 2032 (2016) [arXiv:1509.04714 [astro-ph.CO]] [Search INSPIRE].

[37] Y. Ali-Haïmoud, C. M. Hirata, and C. Dickinson, Mon. Not. R. Astron. Soc. **395**, 1055 (2009) [arXiv:0812.2904 [astro-ph]] [Search INSPIRE].

[38] N. Krachmalnicoff et al., Astron. Astrophys. **618**, A166 (2018) [arXiv:1802.01145 [astro-ph.GA]] [Search INSPIRE].