Peeling off foregrounds with the constrained moment ILC method to unveil primordial CMB *B* modes

Mathieu Remazeilles ⬤,★ Aditya Rotti and Jens Chluba ⬤

Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

ABSTRACT

Galactic foregrounds are the main obstacle to observations of the cosmic microwave background (CMB) *B*-mode polarization. In addition to obscuring the inflationary *B*-mode signal by several orders of magnitude, Galactic foregrounds have non-trivial spectral signatures that are partially unknown and distorted by averaging effects along the line of sight, within the pixel/beam window, and by various analysis choices (e.g. spherical harmonic transforms and filters). Statistical moment expansion methods provide a powerful tool for modelling the effective Galactic foreground emission resulting from these averaging effects in CMB observations, while blind component separation treatments can handle unknown foregrounds. In this work, we combine these two approaches to develop a new *semiblind* component separation method at the intersection of parametric and blind methods, called *constrained moment ILC* (cMILC). This method adds several constraints to the standard ILC method to deproject the main statistical moments of the Galactic foreground emission. Applications to maps are performed in needlet space and when compared to the NILC method, this helps in significantly reducing residual foreground contamination (bias, variance, and skewness) in the reconstructed CMB *B*-mode map, power spectrum, and tensor-to-scalar ratio. We consider sky simulations for experimental settings similar to those of *LiteBIRD* and *PICO*, illustrating which trade-offs between residual foreground biases and degradation of the constraint on *r* can be expected within the new cMILC framework. We also outline several directions that require more work in preparation for the coming analysis challenges.

Key words: polarization – methods: analytical – methods: observational – cosmic background radiation – early Universe – inflation.

1 INTRODUCTION

The search for the primordial *B*-mode polarization (curl-like pattern) of the cosmic microwave background (CMB) radiation is recognized as one of the ultimate challenges in CMB cosmology. The primordial CMB *B*-mode signal would indeed be a clear signature of the primordial gravitational waves of quantum origin predicted by inflation (e.g. Starobinsky 1983), when the Universe underwent ultra-rapid accelerated expansion just about 10^{-35} s after the big bang. The amplitude of the primordial CMB *B*-mode power spectrum, parametrized using the tensor-to-scalar ratio *r*, would allow determining the energy scale of inflation: $V^{1/4} \simeq (r/0.01)^{1/4} \times 10^{16}$ GeV (Knox & Song 2002), and thereby distinguish main classes of early Universe models (e.g. Baumann et al. 2009; Martin, Ringeval & Vennin 2014).

One of the main goals for next-generation CMB experiments is a statistically significant detection of primordial CMB *B* modes down to $r \simeq 10^{-3}$. From space, this effort is led by *LiteBIRD* (Hazumi et al. 2019), whose launch is planned by JAXA for early 2028. Another space concept is *PICO* (Hanany et al. 2019), which is currently in NASA's concept study phase. In addition, ambitious

future missions with a broad range of capabilities including CMB spectroscopy are being discussed for the ESA Vogage 2050 program (Basu et al. 2019; Chluba et al. 2019; Delabrouille et al. 2019). From the ground, the Simons Observatory (SO; Ade et al. 2019), whose survey should start in 2021, and the CMB-S4 (Abazajian et al. 2016) project are relentlessly moving forward. For all these efforts, CMB foregrounds pose a real challenge because the *B*-mode signal is extremely faint (≤ 50 nK r.m.s fluctuations), with intense polarized Galactic emission (mostly synchrotron and thermal dust emission) obscuring it by several orders of magnitude. In addition, gravitational lensing effects on the CMB by the large-scale structure and several instrumental systematics create spurious contamination to CMB *B* modes. Component separation methods are thus critical, since the residual foreground contamination in the recovered CMB *B*-mode map will set the ultimate uncertainty limit with which *r* can be determined (Remazeilles et al. 2018).

Several component separation methods have been developed in the literature to disentangle the primary CMB signal from the foregrounds: Commander (Eriksen et al. 2008), SMICA (Delabrouille, Cardoso & Patanchon 2003; Cardoso et al. 2008), NILC (Delabrouille et al. 2009), SEVEM (Fernández-Cobos et al. 2012), and GNILC (Remazeilles, Delabrouille & Cardoso 2011b), to name only those that have been used for the data product release of the ESA's *Planck* satellite (Planck Collaboration IV 2020a). These

★ E-mail: mathieu.remazeilles@manchester.ac.uk

methods (which have been successfully applied to *Planck* temperature and *E*-mode data for which the signal-to-foregrounds ratio was relatively large) are now being recycled, along with most recent algorithms like xForecast (Errard et al. 2016; Stompor, Errard & Poletti 2016), BFoRe (Alonso et al. 2017), and tailored template fitting (Delta-map; Ichiki et al. 2019), for new faint signal-to-foregrounds regimes in *B*-mode forecasts (e.g. Remazeilles et al. 2018).

Some of these methods, such as the Bayesian fitting method Commander and maximum-likelihood fitting method xForecast, are *parametric* since they rely on astrophysical models of the foreground emission, while other methods, such as the variance-minimization method NILC, are *blind* in the sense that no explicit assumption is made on foregrounds. Each of them have their advantages and weaknesses, but consistency between blind and parametric methods on CMB *B*-mode reconstruction is desired for claiming a robust detection. In this respect, many lessons have been learned (Flauger, Hill & Spergel 2014; BICEP2/Keck Collaboration 2015), and we now understand the importance of further developing component separation methods for *B* modes that allow us to eliminate biases arising from residual foregrounds. In addition, the optimal method depends on the observable that is targeted, and there may be no *one-rules-it-all* approach.

Component separation methods rely on the distinct spectral signatures of the various components of emission (CMB, foregrounds) to disentangle them in multifrequency sky observations. However, Galactic foregrounds have non-trivial spectral energy distributions (SEDs), whose exact properties are still unknown at the sensitivity levels of $r \simeq 10^{-3}$. In parametric approaches, a slight mismodelling of the foregrounds may lead to large biases on r due to huge amplitude disparity between foregrounds and primordial *B* modes (Remazeilles et al. 2016; Hensley & Bull 2018). Similarly, the thermal dust emission from the interstellar medium is expected to be partially *decorrelated* across frequencies, i.e. the dust spectral parameters (spectral index and temperature) may not only vary across the sky but also across frequencies, because of the averaging of multiple cloud contributions of different spectral indices and temperatures along the line of sight (Tassis & Pavlidou 2015). Thus, while still sufficient at *Planck* sensitivity, the common modelling of the dust SED as a unique modified blackbody across frequencies may break at the targeted *B*-mode sensitivity, and lead to significant biases on r after component separation.

To minimize the effects of unknown foreground complexities, blind analysis methods provide a powerful remedy, while averaging effects of known SEDs can in principle be modelled parametrically through moment expansion methods (Stolyarov et al. 2005; Chluba, Hill & Abitbol 2017). Aside from line-of-sight averaging effects, similar spectral averaging effects arise from the limited beam/pixel resolution of the sky maps and various analysis choices, such as spherical harmonic decomposition, downgrading the resolution, or generally when applying filters (Chluba et al. 2017). The effective SED of the foreground emission in a given pixel/beam may therefore differ from the expected SED shape in each line of sight, since the former is the average over several line-of-sight SEDs.

As an illustration (see Fig. 1), let us assume that, in each single pixel of a high-resolution sky map, the synchrotron emission can be accurately described by a single power law, $f_{\rm sync}(\nu) \simeq \nu^{\beta}$, across frequencies ν, with a pixel-dependent spectral index $\beta(p)$ varying across the sky. While this SED model might be approximately valid at the given map resolution and sensitivity, it can actually break for maps of lower angular resolution, for which the effective SED is now
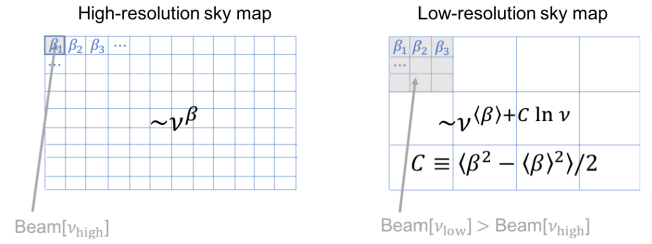

Figure 1. Mismodelling of the foregrounds across frequency due to beam averaging in multiresolution sky maps: a synchrotron power-law spectrum at high angular resolution becomes a *curved* power law at lower resolution, with the curvature given by the second-order moment of the spectral index.

the average of multiple power laws within a larger beam:

$$\langle f_{\rm sync}(\nu) \rangle = \langle \nu^{\beta(p)} \rangle$$
$$\simeq \left\langle e^{\langle\beta\rangle \ln \nu} + (\beta - \langle\beta\rangle) \frac{{\rm d}e^{\beta \ln \nu}}{{\rm d}\beta}|_{\beta=\langle\beta\rangle} + \frac{1}{2}(\beta - \langle\beta\rangle)^2 \frac{{\rm d}^2 e^{\beta \ln \nu}}{{\rm d}\beta^2}|_{\beta=\langle\beta\rangle} \right\rangle$$
$$\simeq e^{\langle\beta\rangle \ln \nu} \left\langle 1 + (\beta - \langle\beta\rangle) \ln \nu + \frac{1}{2}(\beta - \langle\beta\rangle)^2 (\ln \nu)^2 \right\rangle$$
$$\simeq e^{\langle\beta\rangle \ln \nu} \left(1 + \frac{1}{2}\langle\beta^2 - \langle\beta\rangle^2\rangle (\ln \nu)^2 \right) \simeq \nu^{\langle\beta\rangle + C \ln \nu}. \quad (1)$$

Here, we defined the beam-average $\langle Y(p, \nu) \rangle \equiv \langle Y(p, \nu) \rangle_{p \in {\rm beam}[\nu]}$. As equation (1) shows, the average SED is no longer a power law but a *curved* power law at lower angular resolution, with an effective curvature $C = \frac{1}{2}\langle\beta(p)^2 - \langle\beta\rangle^2\rangle$ given by the variance (second-order moment) of the spectral index within the beam (Chluba et al. 2017). In the last step, we resummed the series, which assumes that all higher order moment are Gaussian, which in general may not be valid. Therefore, while the spectral intensity of the synchrotron might be accurately modelled by a power law at one resolution, the power-law model is no longer valid at lower angular resolution. This results in an effective decorrelation of the foreground emission across frequencies. For similar reasons, omitting effective curvature due to beam averaging in the SED modelling of thermal dust can induce non-negligible biases on $r \lesssim 10^{-3}$ (Remazeilles et al. 2018). Modelling as many moments of the foreground emission as possible, given the sensitivity limits of an experiment, thus provides a natural way for capturing some of the additional complexity of the CMB foreground contamination.

With this in mind, in this work we develop a *semiblind* component separation method, called *constrained moment ILC* (cMILC), which operates at the intersection of blind ILC methods and parametric foreground modelling methods. We extend the NILC method by adding several nulling constraints on the effective SEDs of the main moments of the dust and synchrotron emissions in order to remove the bulk of the residual foreground contamination in the recovered CMB *B*-mode map. This is achieved by combining the moment expansion technique (Chluba et al. 2017) with the constrained ILC (cILC) method (Remazeilles, Delabrouille & Cardoso 2011a). While here the focus is on *B*-mode signal, this type of semiblind approach has broader applicability and was recently introduced for the extraction of the relativistic Sunyaev–Zeldovich (SZ) effect (Remazeilles & Chluba 2020) and CMB spectral distortions (Rotti & Chluba 2021). In harmonic space, the power of moment methods was furthermore recently explored on *Planck* data (Mangilli et al. 2021). Moment expansion techniques have also been used in the modelling of SZ signals (Chluba et al. 2013). First-order moments have also been recently used to augment internal template fitting methods for *B*-mode foreground cleaning (Ichiki et al. 2019).

The paper is organized as follows. In Section 2, we review the basics of the ILC and cILC methods. In Section 3, we present our new semiblind component separation method cMILC, which plugs the foreground moment expansion technique described in Section 3.1 into the cILC method, as outlined in Section 3.2. We discuss the statistical properties of residual foregrounds and noise in Section 3.4. We then present the results of our *B*-mode analysis on sky simulations for experimental settings similar to *LiteBIRD* and *PICO* in Section 4. We discuss some directions for optimization of the method in Section 5, and draw our conclusions in Section 6.

2 BASICS OF ILC METHODS

2.1 The standard ILC

The standard internal linear combination (ILC; Bennett et al. 2003; Tegmark, de Oliveira-Costa & Hamilton 2003; Eriksen et al. 2004; Delabrouille et al. 2009; Basak & Delabrouille 2012) is a blind component separation method as there is no attempt at directly modelling the foregrounds. Hence, the sky observations $d_\nu(p)$, at each frequency ν and for each pixel p are written as

$$d_\nu(p) = a_\nu s(p) + n_\nu(p), \tag{2}$$

where a_ν is the known spectral response (SED)[1] of the CMB anisotropies $s(p)$ at frequency ν, while $n_\nu(p)$ is the unmodelled overall contamination from foregrounds and instrumental noise.

For convenience with the algebra, we recast equation (2) in a $n_f \times 1$ vector form, where n_f is the number of frequency channels:

$$\boldsymbol{d}(p) = \boldsymbol{a}s(p) + \boldsymbol{n}(p). \tag{3}$$

Here, $\boldsymbol{d}(p) = \{d_\nu(p)\}_\nu$ collects the set of available frequency maps, $\boldsymbol{a} = \{a_\nu\}_\nu$ is the known CMB SED vector, $s(p)$ corresponds to the unknown CMB anisotropies that we aim at extracting, and $\boldsymbol{n}(p) = \{n_\nu(p)\}_\nu$ collects the unmodelled foregrounds and noise that we aim at mitigating in the reconstructed CMB map.

The standard ILC estimate, $\hat{s}(p)$, of the CMB anisotropies is obtained by forming a weighted linear combination of the frequency maps

$$\hat{s} = \sum_\nu w_\nu d_\nu \equiv \boldsymbol{w}^{\mathrm{T}}\boldsymbol{d} \tag{4}$$

that is of minimum variance, i.e.

$$\frac{\partial}{\partial \boldsymbol{w}} \langle \hat{s}^2 \rangle = 0. \tag{5}$$

Here, the ILC weights $\boldsymbol{w} = \{w_\nu\}_\nu$ assigned to the frequency maps are constrained to provide unit response to the CMB SED:

$$\boldsymbol{w}^{\mathrm{T}}\boldsymbol{a} = 1 \tag{6}$$

in order to guarantee the full conservation of the CMB signal $s(p)$ in the variance minimization. This ensures that the ILC estimate

$$\hat{s} = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{d} = \boldsymbol{w}^{\mathrm{T}}(\boldsymbol{a}s + \boldsymbol{n}) = s + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{n} \tag{7}$$

does not alter the CMB signal s (i.e. no multiplicative error due to the constraint equation 6), while the variance of the residual foreground and noise $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{n}$ is minimized by equation (5).

The variance of the ILC map, $\langle \hat{s}^2 \rangle$, relates to the covariance of the data $\boldsymbol{d}$ as

$$\langle \hat{s}^2 \rangle = \boldsymbol{w}^{\mathrm{T}} \langle \boldsymbol{d}\boldsymbol{d}^{\mathrm{T}} \rangle \boldsymbol{w} = \boldsymbol{w}^{\mathrm{T}}\mathbf{C}\boldsymbol{w}, \tag{8}$$

[1]Modelled by the first temperature derivative of the blackbody spectrum.

where $\mathbf{C} = \langle \boldsymbol{d}\boldsymbol{d}^{\mathrm{T}} \rangle$ is the $n_f \times n_f$ covariance matrix of the data whose elements are $\mathbf{C}_{\nu\nu'} = \langle d_\nu d_{\nu'} \rangle$ for any pair of frequencies (ν, ν'). An empirical estimate of the data covariance matrix in each pixel p can be obtained by ergodicity:

$$\mathbf{C}(p) = \frac{1}{N_p} \sum_{p' \in \mathcal{D}(p)} \boldsymbol{d}(p')\boldsymbol{d}(p')^{\mathrm{T}}, \tag{9}$$

where $\mathcal{D}(p)$ is a domain of pixels surrounding pixel p, and N_p the number of pixels in this domain.

The ILC weights $\boldsymbol{w}$ are thus the solution of the constrained variance-minimization problem:

$$\begin{cases} \frac{\partial}{\partial \boldsymbol{w}}(\boldsymbol{w}^{\mathrm{T}}\mathbf{C}\boldsymbol{w}) = 0, \\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{a} = 1, \end{cases} \tag{10}$$

which can be addressed with a Lagrange multiplier λ by solving

$$\frac{\partial}{\partial \boldsymbol{w}}[\boldsymbol{w}^{\mathrm{T}}\mathbf{C}\boldsymbol{w} + \lambda(1 - \boldsymbol{w}^{\mathrm{T}}\boldsymbol{a})] = 0. \tag{11}$$

This yields $\boldsymbol{w}^T = \lambda \boldsymbol{a}^{\mathrm{T}}\mathbf{C}^{-1}$, while $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{a} = 1$ leads to $\lambda = (\boldsymbol{a}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{a})^{-1}$, so that the standard ILC weights are given by

$$\boldsymbol{w}^{\mathrm{T}} = (\boldsymbol{a}^{\mathrm{T}}\mathbf{C}^{-1}\boldsymbol{a})^{-1}\boldsymbol{a}^{\mathrm{T}}\mathbf{C}^{-1}. \tag{12}$$

2.2 The cILC

While the variance of the global contamination (i.e. foregrounds plus noise) is minimized by the standard ILC estimate, this is not necessarily the solution with minimal foreground variance. Instead of minimizing the variance of the global contamination, the cILC method (Remazeilles et al. 2011a) extends the standard ILC to fully eliminate selected foreground residuals by using prior knowledge on their SED and adding *orthogonality* constraints to the variance-minimization problem:

$$\begin{cases} \frac{\partial}{\partial \boldsymbol{w}}\left(\boldsymbol{w}^{\mathrm{T}}\mathbf{C}\boldsymbol{w}\right) = 0 \\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{a} = 1 \\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_1 = 0 \\ \vdots \\ \boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_m = 0, \end{cases} \tag{13}$$

where $\boldsymbol{b}_1, \cdots, \boldsymbol{b}_m$ are the SED vectors of m modelled foreground components, to which the cILC weights $\boldsymbol{w}$ are orthogonal. If we write the sky observations as

$$\boldsymbol{d}(p) = \boldsymbol{a}s(p) + \sum_{i=1}^{m} \boldsymbol{b}_i g_i(p) + \boldsymbol{n}(p), \tag{14}$$

where $\{g_i(p)\}_{1 \le i \le m}$ are the m foreground components whose SEDs $\{\boldsymbol{b}_i\}_{1 \le i \le m}$ are modelled, while $\boldsymbol{n}(p)$ includes unmodelled foregrounds and instrumental noise, then the orthogonality constraints $\boldsymbol{w}^T \boldsymbol{b}_i = 0$ of the cILC method in equation (13) guarantee the full cancellation of the residual foreground contamination from the m modelled components in the reconstructed CMB map:

$$\hat{s}(p) = (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{a})s(p) + \sum_{i=1}^{m} \left(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_i\right) g_i(p) + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{n}(p)$$

$$= s(p) + \boldsymbol{w}^{\mathrm{T}}\boldsymbol{n}(p). \tag{15}$$

This significantly differs from the ILC method (equation 10) which suffers from additional foreground residuals as $\sum_{i=1}^{m}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_i) g_i(p) \ne 0$ in this case. However, the penalty of adding nulling constraints on some foregrounds in the cILC is an increase of the noise residuals

$w^{\mathrm{T}}n(p)$ compared to the standard ILC. But there is a sweet spot between noise cost and reduction of foreground biases, as we show in this work. This optimization furthermore depends on the observable that is targeted, which in our case is r (see Section 4.2).

If we introduce the $n_f \times (m + 1)$ SED matrix $\mathbf{A} = [a \, b_1 \cdots b_m]$, whose columns collect the SED vectors of the modelled components, and the $1 \times (m + 1)$ transposed vector $e^{\mathrm{T}} = [1 \, 0 \, \cdots \, 0]$, then the system of equations (13) can be rewritten in the compact form:

$$\begin{cases} \frac{\partial}{\partial w} \left(w^{\mathrm{T}} \mathbf{C} \, w \right) = 0 \\ w^{\mathrm{T}} \mathbf{A} = e^{\mathrm{T}}. \end{cases} \tag{16}$$

Hence, the problem is equivalent to solving

$$\frac{\partial}{\partial w} [w^{\mathrm{T}} \mathbf{C} \, w + \Lambda^{\mathrm{T}} (e - \mathbf{A}^{\mathrm{T}} w)] = 0, \tag{17}$$

where $\Lambda^{\mathrm{T}} = [\lambda \, \mu_1 \cdots \mu_m]$ is a vector of $(m + 1)$ Lagrange multipliers. This yields to $w^{\mathrm{T}} = \Lambda^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1}$, while the constraints $w^{\mathrm{T}} \mathbf{A} = e^{\mathrm{T}}$ give $\Lambda^{\mathrm{T}} = e^{\mathrm{T}} (\mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{A})^{-1}$. The cILC weights are then given by

$$w^{\mathrm{T}} = e^{\mathrm{T}} (\mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{A})^{-1} \mathbf{A}^{\mathrm{T}} \mathbf{C}^{-1}, \tag{18}$$

which is a generalization of equation (12). Equation (18) has been previously derived in the literature for the extraction of the relativistic SZ effect (Remazeilles & Chluba 2020) and CMB spectral distortions (Rotti & Chluba 2021). Similar multidimensional ILC filters have been derived for other component separation methods like GNILC (Remazeilles et al. 2011b) and MILCA (Hurier, Macías-Pérez & Hildebrandt 2013), the main difference between them lying in the form of the mixing matrix $\mathbf{A}$.

It should be noted that for a single orthogonality constraint, i.e. $\mathbf{A} = [a \, b]$ and $e^{\mathrm{T}} = [1 \, 0]$, the general expression equation (18) reduces to

$$w^{\mathrm{T}} = \frac{(b^{\mathrm{T}} \mathbf{C}^{-1} b) a^{\mathrm{T}} \mathbf{C}^{-1} - (a^{\mathrm{T}} \mathbf{C}^{-1} b) b^{\mathrm{T}} \mathbf{C}^{-1}}{(a^{\mathrm{T}} \mathbf{C}^{-1} a)(b^{\mathrm{T}} \mathbf{C}^{-1} b) - (a^{\mathrm{T}} \mathbf{C}^{-1} b)^2}, \tag{19}$$

which was introduced by Remazeilles et al. (2011a) to null SZ cluster residuals in primary CMB maps, when a and b are the SEDs of CMB and thermal SZ, respectively. Finally, in the absence of any orthogonality constraint, i.e. $\mathbf{A} \equiv a$ and $e \equiv 1$, the expression equation (18) reduces to the standard ILC given in equation (12).

3 SEMIBLIND COMPONENT SEPARATION METHOD

We now work out our new *semiblind* component separation method for B modes, the cMILC, which combines foreground moment expansion (Chluba et al. 2017) and cILC method (Remazeilles et al. 2011a).

3.1 Moment expansion of the foreground emission

The spectral intensity[2] of the Galactic foreground emission depends on several spectral parameters $\boldsymbol{\beta}(p) \equiv \{\beta_i(p)\}_{1 \le i \le n}$ (e.g. spectral indices, temperatures) that vary across the sky (hence the p dependence for 'pixel') and along the line of sight:

$$I(\nu, p) = A_{\nu_0}(p) f(\nu, \boldsymbol{\beta}(p)). \tag{20}$$

$A_{\nu_0}(p)$ is the amplitude of the foreground emission at some pivot frequency ν_0, and $f(\nu, \boldsymbol{\beta}(p))$ is the SED which varies both across

frequencies ν and across the sky, depending on the local spectral parameters $\boldsymbol{\beta}(p) \equiv \{\beta_i(p)\}_{1 \le i \le n}$.

As stressed by Chluba et al. (2017), we can expand the foreground SED, $f(\nu, \boldsymbol{\beta}(p))$, in terms of *statistical moments* around some fixed pivot parameters $\overline{\boldsymbol{\beta}} \equiv \{\overline{\beta}_i\}_i$:

$$f(\nu, \boldsymbol{\beta}(p)) = \sum_k \sum_{\alpha_1 + \cdots + \alpha_n = k}$$
$$\times \frac{(\beta_1(p) - \overline{\beta}_1)^{\alpha_1} \cdots (\beta_n(p) - \overline{\beta}_n)^{\alpha_n}}{\alpha_1! \cdots \alpha_n!} \frac{\partial^k f(\nu, \overline{\boldsymbol{\beta}})}{\partial \overline{\beta}_1^{\alpha_1} \cdots \partial \overline{\beta}_n^{\alpha_n}}. \tag{21}$$

This is nothing but a multidimensional Taylor series in all the spectral parameters, however, it gives the flexibility to describe various physical and observational averaging processes. The moment expansion highlights that the foreground emission can be captured by a set of *moment components* in the sky:

$$m_{\alpha_1, \cdots, \alpha_n}(p) = A_{\nu_0}(p) \frac{(\beta_1(p) - \overline{\beta}_1)^{\alpha_1} \cdots (\beta_n(p) - \overline{\beta}_n)^{\alpha_n}}{\alpha_1! \cdots \alpha_n!}, \tag{22}$$

each of them having a *uniform* SED across the sky given by

$$\frac{\partial^{\alpha_1 + \cdots + \alpha_n} f}{\partial \beta_1^{\alpha_1} \cdots \partial \beta_n^{\alpha_n}} (\nu, \boldsymbol{\beta}(p) = \overline{\boldsymbol{\beta}}). \tag{23}$$

Next, we will consider thermal dust and synchrotron as the main polarized Galactic foregrounds for B modes. The SED of the synchrotron emission is known to be close to a power law

$$f_{\mathrm{sync}}(\nu, \beta_s(p)) = \left(\frac{\nu}{\nu_s} \right)^{\beta_s(p)}, \tag{24}$$

for data in Rayleigh–Jeans brightness temperature units, with an average spectral index of $\langle \beta_s(p) \rangle_p \simeq -3$ over the sky (Kogut et al. 2007; Miville-Deschênes et al. 2008; Krachmalnicoff et al. 2018; Planck Collaboration IV 2020a). Conversely, the SED of the thermal dust emission is known to be close to a modified blackbody

$$f_{\mathrm{dust}}(\nu, \beta_d(p), T_d(p)) = \left(\frac{\nu}{\nu_d} \right)^{\beta_d(p) + 1} \frac{1}{\exp \left(\frac{h\nu}{k T_d(p)} \right) - 1}, \tag{25}$$

for data in Rayleigh–Jeans brightness temperature units, with an average spectral index of $\langle \beta_d(p) \rangle_p \simeq 1.5$ and an average temperature of $\langle T_d(p) \rangle_p \simeq 20 \, \mathrm{K}$ (Planck Collaboration XI 2014; Planck Collaboration X 2016a; Planck Collaboration Int. XLVIII 2016c).

Hence, the moment expansion of the synchrotron emission up to second order yields[3]

$$\begin{aligned} I_{\mathrm{sync}}(\nu, p) = \, & A_{\nu_s}(p) f_{\mathrm{sync}}(\nu, \overline{\beta}_s) \\ & + A_{\nu_s}(p) \Delta\beta_s(p) \partial_{\overline{\beta}_s} f_{\mathrm{sync}}(\nu, \overline{\beta}_s) \\ & + \frac{1}{2} A_{\nu_s}(p) \Delta\beta_s^2(p) \partial_{\overline{\beta}_s}^2 f_{\mathrm{sync}}(\nu, \overline{\beta}_s) + \mathcal{O}\left(\beta_s^3\right), \end{aligned} \tag{26}$$

where $\Delta\beta_s(p) = \beta_s(p) - \overline{\beta}_s$ and

$$f_{\mathrm{sync}}(\nu, \overline{\beta}_s) = \left(\frac{\nu}{\nu_s} \right)^{\overline{\beta}_s}, \tag{27a}$$

$$\partial_{\overline{\beta}_s} f_{\mathrm{sync}}(\nu, \overline{\beta}_s) = \ln \left(\frac{\nu}{\nu_s} \right) f_{\mathrm{sync}}(\nu, \overline{\beta}_s), \tag{27b}$$

$$\partial_{\overline{\beta}_s}^2 f_{\mathrm{sync}}(\nu, \overline{\beta}_s) = \left[\ln \left(\frac{\nu}{\nu_s} \right) \right]^2 f_{\mathrm{sync}}(\nu, \overline{\beta}_s) \tag{27c}$$

---

[2]By 'intensity', we mean either I, Q, U, E, or B, without loss of generality.

[3]We will be using the notation $\partial_{\overline{\beta}} f(\overline{\beta}) \equiv \frac{\partial f(\beta)}{\partial \beta} \big|_{\beta = \overline{\beta}}$ throughout the paper.
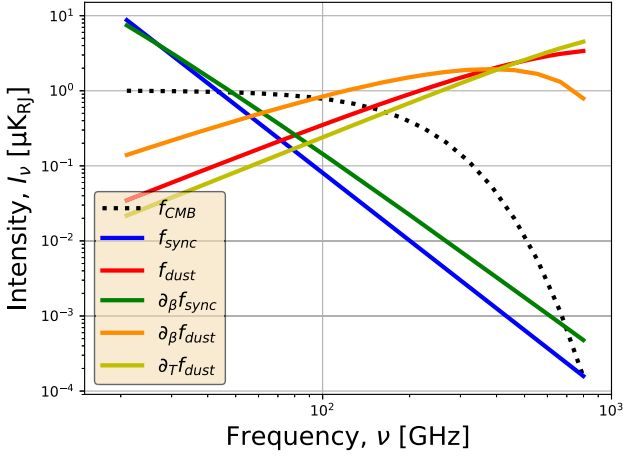
Figure 2. SED of the zeroth- and first-order moments of the dust and synchrotron (*coloured solid lines*), along with the SED of the CMB (*black dotted line*), across frequencies. The overall normalization of the SEDs is arbitrary.

are the SEDs of zeroth-order (equation 27a), first-order (equation 27b), and second-order (equation 27c) moment components of the synchrotron.

Similarly, the moment expansion of the thermal dust emission up to second order yields

$$
\begin{aligned}
I_{\text{dust}}(\nu, p) = {} & A_{\nu_d}(p) f_{\text{dust}}(\nu, \overline{\beta}_d) \\
& + A_{\nu_d}(p) \Delta\beta_d(p)\, \partial_{\overline{\beta}_d} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) \\
& + A_{\nu_d}(p) \Delta T_d(p)\, \partial_{\overline{T}_d} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) \\
& + \frac{1}{2} A_{\nu_d}(p) \Delta\beta_d^2(p)\, \partial_{\overline{\beta}_d}^2 f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) \\
& + A_{\nu_d}(p) \Delta\beta_d(p)\Delta T_d(p)\, \partial_{\overline{\beta}_d} \partial_{\overline{T}_d} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) \\
& + \frac{1}{2} A_{\nu_d}(p) \Delta T_d^2(p)\, \partial_{\overline{T}_d}^2 f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) \\
& + \mathcal{O}\left(\Delta\beta_d^3, \Delta T_d^3\right),
\end{aligned}
\tag{28}
$$

where $\Delta\beta_d(p) = \beta_d(p) - \overline{\beta}_d$, $\Delta T_d(p) = T_d(p) - \overline{T}_d$ and

$$
f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) = \left(\frac{\nu}{\nu_d}\right)^{\overline{\beta}_d + 1} \frac{1}{\exp(\bar{x}) - 1},
\tag{29a}
$$

$$
\partial_{\overline{\beta}_d} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) = \ln\left(\frac{\nu}{\nu_d}\right) f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d),
\tag{29b}
$$

$$
\partial_{\overline{T}_d} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) = \frac{\bar{x}}{\overline{T}_d} \frac{\exp(\bar{x})}{\exp(\bar{x}) - 1} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d),
\tag{29c}
$$

$$
\partial_{\overline{\beta}_d}^2 f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) = \left[\ln\left(\frac{\nu}{\nu_d}\right)\right]^2 f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d),
\tag{29d}
$$

$$
\partial_{\overline{T}_d}^2 f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) = \left[\bar{x} \coth\left(\frac{\bar{x}}{2}\right) - 2\right] \frac{1}{\overline{T}_d} \partial_{\overline{T}_d} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d),
\tag{29e}
$$

$$
\partial_{\overline{\beta}_d} \partial_{\overline{T}_d} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d) = \ln\left(\frac{\nu}{\nu_d}\right) \partial_{\overline{T}_d} f_{\text{dust}}(\nu, \overline{\beta}_d, \overline{T}_d)
\tag{29f}
$$

are the SEDs of zeroth-order (equation 29a), first-order (equations 29b and 29c), and second-order (equations 29d–29f) moment components of the dust, and $\bar{x} \equiv h\nu/k\overline{T}_d$. The SED shapes of zeroth- and first-order moments of the dust and synchrotron are plotted in Fig. 2.

By choosing pivot parameters that are close to the expected mean of the spectral parameters over the sky, $\overline{\beta}_s = \langle \beta_s(p) \rangle_p \simeq -3$, $\overline{\beta}_d = \langle \beta_d(p) \rangle_p \simeq 1.5$, and $\overline{T}_d = \langle T_d(p) \rangle_p \simeq 20\,\text{K}$, we ensure that the bulk of the Galactic foreground emission projects on to the zeroth-, first-, and second-order moments. However, a detailed optimization is more complicated as we briefly discuss in Section 5.1.

3.2 The cMILC

Our *semiblind* component separation method cMILC aims at recovering the CMB *B*-mode polarization signal from sky observations by adding nulling constraints on the main foreground moments in the cILC framework in order to improve the removal of the Galactic foreground contamination compared to standard ILC methods, which rely only on simple variance minimization. We thus build the cMILC method by enforcing the following constraints:

$$
\begin{cases}
\frac{\partial}{\partial \boldsymbol{w}} \left(\boldsymbol{w}^{\text{T}} \mathbf{C}\, \boldsymbol{w}\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \boldsymbol{a} = 1 \\
\boldsymbol{w}^{\text{T}} \cdot \boldsymbol{f}_{\text{sync}}\left(\overline{\beta}_s\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \partial_{\overline{\beta}_s} \boldsymbol{f}_{\text{sync}}\left(\overline{\beta}_s\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \partial_{\overline{\beta}_s}^2 \boldsymbol{f}_{\text{sync}}\left(\overline{\beta}_s\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \boldsymbol{f}_{\text{dust}}\left(\overline{\beta}_d, \overline{T}_d\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \partial_{\overline{\beta}_d} \boldsymbol{f}_{\text{dust}}\left(\overline{\beta}_d, \overline{T}_d\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \partial_{\overline{T}_d} \boldsymbol{f}_{\text{dust}}\left(\overline{\beta}_d, \overline{T}_d\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \partial_{\overline{\beta}_d}^2 \boldsymbol{f}_{\text{dust}}\left(\overline{\beta}_d, \overline{T}_d\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \partial_{\overline{T}_d}^2 \boldsymbol{f}_{\text{dust}}\left(\overline{\beta}_d, \overline{T}_d\right) = 0 \\
\boldsymbol{w}^{\text{T}} \cdot \partial_{\overline{\beta}_d} \partial_{\overline{T}_d} \boldsymbol{f}_{\text{dust}}\left(\overline{\beta}_d, \overline{T}_d\right) = 0.
\end{cases}
\tag{30}
$$

It is important to stress that cMILC does not attempt to fit the foregrounds, which are spatially correlated and uncertain components, but instead aims at *deprojecting* the dominant foreground moments in sky maps without altering the CMB component. Thus, bulk of the Galactic foreground contamination that is spatially correlated with the deprojected moments is eliminated from the recovered CMB *B*-mode power spectrum. Any remaining unconstrained foreground contribution is just variance-minimized, like in blind methods.[4]

The cMILC weights are thus again given by equation (18):

$$
\boldsymbol{w}_{\text{cMILC}}^{\text{T}} = \boldsymbol{e}^{\text{T}} (\mathbf{A}^{\text{T}} \mathbf{C}^{-1} \mathbf{A})^{-1} \mathbf{A}^{\text{T}} \mathbf{C}^{-1},
\tag{31}
$$

where, when limiting constraints to zeroth- and first-order moments as an example, the SED matrix $\mathbf{A}$ is a $n_f \times 6$ matrix given by

$$
\mathbf{A} = \left(\boldsymbol{a}\ \ \boldsymbol{f}_{\text{sync}}\ \ \boldsymbol{f}_{\text{dust}}\ \ \partial_{\overline{\beta}_s} \boldsymbol{f}_{\text{sync}}\ \ \partial_{\overline{\beta}_d} \boldsymbol{f}_{\text{dust}}\ \ \partial_{\overline{T}_d} \boldsymbol{f}_{\text{dust}}\right),
\tag{32}
$$

and the vector $\boldsymbol{e}$ is given by

$$
\boldsymbol{e}^{\text{T}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.
\tag{33}
$$

The SED matrix $\mathbf{A}$ and vector $\boldsymbol{e}$ can be augmented by including higher order moments, depending on the sensitivity floor of the experiment. We stress that by applying the Woodbury formula to equation (41), the cILC/cMILC filter equation (31) reduces to $\boldsymbol{e}^{\text{T}} (\mathbf{A}^{\text{T}} \mathbf{N}^{-1} \mathbf{A})^{-1} \mathbf{A}^{\text{T}} \mathbf{N}^{-1}$, where, unlike maximum-likelihood parametric fitting methods (e.g. Stompor et al. 2016), the covariance matrix $\mathbf{N}$ here does not only account for the noise but also for the *unconstrained* foregrounds (e.g. unmodelled foregrounds or higher order moments) that may be omitted in the mixing matrix $\mathbf{A}$. The data covariance

---

[4]Hence, the exact pivots parameters do not need to be absolutely known (Fig. A1), although pivots can help optimizing the method (see Section 5.1).

Table 1. Nomenclature of the NILC and cMILC methods for several combinations of constraints on various moments of the foreground emission.

| Case	Moments (SED)	Parameters
NILC	$a_{\rm CMB}$	1
cMILC01	$a_{\rm CMB}$; $f_{\rm sync}$	2
cMILC02	$a_{\rm CMB}$; $f_{\rm dust}$	2
cMILC03	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$	3
cMILC04	$a_{\rm CMB}$; $f_{\rm dust}$; $\partial_\beta f_{\rm dust}$	3
cMILC05	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm sync}$	4
cMILC06	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm dust}$	4
cMILC07	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm sync}$; $\partial_\beta f_{\rm dust}$	5
cMILC08	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm sync}$; $\partial_\beta f_{\rm dust}$; $\partial_T f_{\rm dust}$	6
cMILC09	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm sync}$; $\partial_\beta f_{\rm dust}$; $\partial_T f_{\rm dust}$; $\partial_T^2 f_{\rm dust}$	7
cMILC10	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm sync}$; $\partial_\beta f_{\rm dust}$; $\partial_T f_{\rm dust}$; $\partial_\beta^2 f_{\rm sync}$; $\partial_T^2 f_{\rm dust}$	8
cMILC11	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm sync}$; $\partial_\beta f_{\rm dust}$; $\partial_T f_{\rm dust}$; $\partial_\beta^2 f_{\rm sync}$; $\partial_T^2 f_{\rm dust}$; $\partial_\beta \partial_T f_{\rm dust}$	9
cMILC12	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm sync}$; $\partial_T f_{\rm dust}$ at low ℓ (hybrid case; see Section 5)	5
	$a_{\rm CMB}$; $f_{\rm sync}$; $f_{\rm dust}$; $\partial_\beta f_{\rm dust}$; $\partial_T f_{\rm dust}$ at high ℓ	

matrix **C** in equation (31) thus allows for blind variance minimization of any remaining unparametrized foregrounds but also of possible residuals that would arise from suboptimal pivot choices (see Fig. A1 in Appendix A).

For our analysis, we will be applying several cMILC filters (see Table 1) to the sky simulations by considering more and more columns in the SED matrix **A**, starting from the standard ILC method without nulling constraints (i.e. $\mathbf{A} \equiv \boldsymbol{a}$), and ending with the most constrained version of cMILC with all the nulling constraints on zeroth-, first-, and second-order foreground moments (i.e. with a $n_f \times 10$ SED matrix **A**). This progressive approach will help to highlight how increasing the number of moment constraints affects the trade-offs between foreground and noise residuals (bias, variance, and higher order statistics) in the recovered CMB *B*-mode map, power spectrum, and tensor-to-scalar ratio. Below we will see that the optimal number of moments is a function of the experimental sensitivity and spectral coverage.

3.3 Needlet implementation

ILC filtering can be implemented directly in pixel space on the sky maps $d_\nu(p)$ (e.g. Bennett et al. 2003; Eriksen et al. 2004) or in harmonic space on spherical harmonic coefficients $a_\nu(\ell, m)$ of the sky maps (e.g. Tegmark et al. 2003). Since pixel-space ILC methods are fully local in pixel space, they are non-local in harmonic space because of uncertainty principle, hence all multipoles are given the same weights. Similarly, harmonic-space ILC methods being fully local in multipole space are non-local in pixel space and thus provide same weights to all pixels of a sky map. Non-local weights in either pixel or multipole space are not optimal for component separation, since the local conditions of contamination by foregrounds and noise vary both across the sky and across multipoles.

Because of their localization properties both in pixel space and harmonic space, spherical wavelets such as *needlets* (Narcowich, Petrushev & Ward 2006; Guilloux, Faÿ & Cardoso 2009) provide a powerful alternative to allow the ILC weights to adjust themselves depending on local conditions of contamination both across the sky and across angular scales (Delabrouille et al. 2009). Therefore, in this work we implement both ILC and cMILC methods on the same needlet frame, so that hereafter the standard ILC method will be referred to as NILC for needlet ILC (Delabrouille et al. 2009), while our needlet-based cMILC method is still referred to as cMILC.
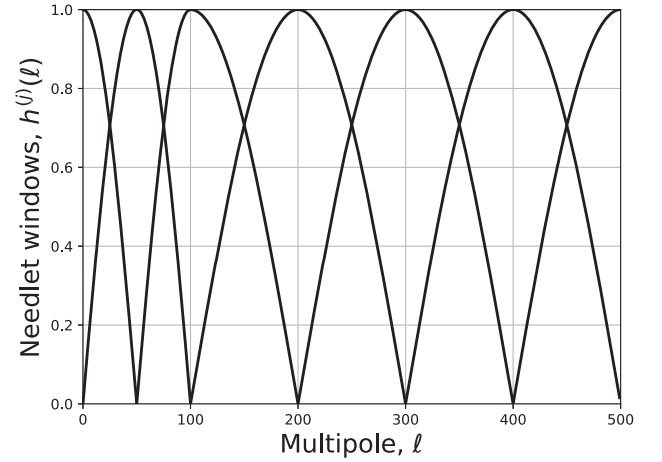
Figure 3. Needlet windows in harmonic space that are used in the analysis.

Needlet decomposition is performed on *B*-mode sky maps at each frequency as follows. We first define a set of seven bandpass windows $\{h^{(j)}(\ell)\}_{1 \leq j \leq 7}$ of cosine shape in harmonic space (Fig. 3), following the prescription of Basak & Delabrouille (2013). Each needlet window isolates a subrange of multipoles or angular scales, while the whole set satisfies the property $\sum_{j=1}^{7}(h^{(j)}(\ell))^2 = 1$ in order to not lose any power in the data processing. The number and width of needlet windows were chosen to allow decent localization in harmonic space while conserving sufficient localization in pixel space for efficient foreground cleaning, by following a dyadic scheme with wider windows at high multipoles to ensure more localization in pixel space for the cleaning of small-scale fluctuations in sky maps. The number, width, and shape of the needlet windows are free parameters and different prescriptions have been used in the literature depending on the observable of interest (e.g. Planck Collaboration XXII 2016b), thus leaving room for further optimization (see Section 5). The spherical harmonic coefficients $a_\nu(\ell, m)$ of the frequency maps are then bandpass filtered by the needlet windows as $\tilde{a}_\nu^{(j)}(\ell, m) = h^{(j)}(\ell) a_\nu(\ell, m)$. The inverse spherical harmonic transform of the bandpass-filtered harmonic coefficients $\tilde{a}_\nu^{(j)}(\ell, m)$ then yields seven needlet maps $\tilde{d}_\nu^{(j)}(p)$ for each frequency. Each needlet map thus contains typical anisotropies of specific angular scales selected by the needlet window. The ILC weights at all frequencies

are then computed for each needlet scale (j) independently, so that we obtain seven ILC CMB maps $\widehat{s}^{(j)}(p) = \sum_\nu w_\nu^{(j)}(p)\widetilde{d}_\nu^{(j)}(p)$ for each needlet frame. The spherical harmonic coefficients, $\widehat{s}^{(j)}(\ell, m)$, of the seven needlet ILC CMB maps are then bandpass-filtered again as $\widehat{z}^{(j)}(\ell, m) = h^{(j)}(\ell)\widehat{s}^{(j)}(\ell, m)$, and inverse spherical-harmonic transformed into maps $\widehat{z}^{(j)}(p)$, which are finally co-added to form the complete NILC or cMILC CMB map $\widehat{s}(p) = \sum_{j=1}^{7} \widehat{z}^{(j)}(p)$.

3.4 Statistics

In this section, we first compare the statistics of the constrained ILC (cILC, cMILC) and standard ILC (NILC) approaches, by generalizing the formulas of the ILC variance (e.g. Tegmark et al. 2003; Delabrouille et al. 2009) to our cMILC filter. We then illustrate our findings on simulations in Section 4.2.2. We stress that, while the standard ILC is the solution of minimum *overall* variance (foregrounds plus noise), cMILC leads to the solution of minimum *foreground* variance. In particular, cMILC beats the standard ILC in terms of residual foreground contamination (bias, variance, and skewness) due to nulling constraints, but at the cost of larger residual noise contamination due to the increased volume of the parameter space. However, there is an optimal trade-off point for cMILC, where the noise penalty is largely compensated by the significant reduction of the residual foreground bias. Aside from experimental parameters, this sweet spot depends on the observable and here we consider r as figure of merit (see Figs 8 and 9 in Section 4.2), illustrating the gains for both the variance and the bias on r through the cumulative systematic and statistical error $\delta r = \sqrt{(r - r^{\text{true}})^2 + \sigma^2(r)}$.

3.4.1 Overall variance

Using equations (8) and (12), we recover the standard expression (e.g. Tegmark et al. 2003) for the variance of the ILC map[5]:

$$\sigma_{\text{ILC}}^2 \equiv \langle \widehat{s}_{\text{ILC}}^2 \rangle = \boldsymbol{w}_{\text{ILC}}^{\text{T}}\mathbf{C}\,\boldsymbol{w}_{\text{ILC}} = \frac{1}{\boldsymbol{a}^{\text{T}}\mathbf{C}^{-1}\boldsymbol{a}}, \tag{34}$$

where $\mathbf{C} = \langle \boldsymbol{dd}^T \rangle$ is the $n_f \times n_f$ covariance matrix of the data, and $\boldsymbol{a}$ is the SED for CMB. Similarly, by using equation (31) we can derive the expression for the variance of the cMILC map, $\sigma_{\text{cMILC}}^2 \equiv \langle \widehat{s}_{\text{cMILC}}^2 \rangle$:

$$\sigma_{\text{cMILC}}^2 = \boldsymbol{w}_{\text{cMILC}}^{\text{T}}\mathbf{C}\,\boldsymbol{w}_{\text{cMILC}} = \boldsymbol{e}^T(\mathbf{A}^{\text{T}}\mathbf{C}^{-1}\mathbf{A})^{-1}\boldsymbol{e}, \tag{35}$$

where $\mathbf{A} = [\boldsymbol{a}\,\boldsymbol{b}_1 \cdots \boldsymbol{b}_m]$ is the $n_f \times (m + 1)$ SED matrix for the CMB and the constrained foreground moments (e.g. equation 32), and $\boldsymbol{e}^{\text{T}} = [1\,0 \cdots 0]$.

By defining the submatrix $\mathbf{B} = [\boldsymbol{b}_1 \cdots \boldsymbol{b}_m]$ as the $n_f \times m$ SED matrix for the modelled foreground moments only, the full SED matrix $\mathbf{A}$ can be written in the block matrix form:

$$\mathbf{A} = (\boldsymbol{a}\,\mathbf{B}), \tag{36}$$

so that the variance of the cMILC map equation (35) reads as

$$\sigma_{\text{cMILC}}^2 = \boldsymbol{e}^{\text{T}}\begin{pmatrix} \boldsymbol{a}^{\text{T}}\mathbf{C}^{-1}\boldsymbol{a} & \boldsymbol{a}^{\text{T}}\mathbf{C}^{-1}\mathbf{B} \\ \mathbf{B}^{\text{T}}\mathbf{C}^{-1}\boldsymbol{a} & \mathbf{B}^{\text{T}}\mathbf{C}^{-1}\mathbf{B} \end{pmatrix}^{-1}\boldsymbol{e}. \tag{37}$$

Using block matrix inversion, the variance of the cMILC map reduces to

$$\sigma_{\text{cMILC}}^2 = \frac{\sigma_{\text{ILC}}^2}{1 - \sigma_{\text{ILC}}^2\,\boldsymbol{a}^{\text{T}}\mathbf{C}^{-1}\mathbf{B}\,\left(\mathbf{B}^{\text{T}}\mathbf{C}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^{\text{T}}\mathbf{C}^{-1}\boldsymbol{a}}. \tag{38}$$

---

[5]Note that in the derivation of the standard formulas, equations (34) and (35), for the variance we make the assumption that empirical correlations between the weights $\boldsymbol{w}(p)$ and the data $d(p)$ can be neglected to first order, and thus the weights can be pulled out of $\langle \cdots \rangle$ in equation (8).

Therefore, the overall variance of the cMILC map is larger than the overall variance of the standard ILC map, with an amplification factor of

$$\frac{\sigma_{\text{cMILC}}^2}{\sigma_{\text{ILC}}^2} = \frac{1}{1 - \sigma_{\text{ILC}}^2\,\boldsymbol{a}^{\text{T}}\mathbf{C}^{-1}\mathbf{B}\,\left(\mathbf{B}^{\text{T}}\mathbf{C}^{-1}\mathbf{B}\right)^{-1}\mathbf{B}^{\text{T}}\mathbf{C}^{-1}\boldsymbol{a}} \geq 1. \tag{39}$$

This increase of the overall variance is expected, since the volume of the parameter space (number of modelled components) is larger for the cMILC than for the standard ILC. However, while the standard ILC is the solution of minimum overall variance, it is not the solution of minimum foreground variance. As we show below, cMILC yields lower foreground contamination than standard ILC.

3.4.2 Foregrounds residual variance

Let us model the data with independent contributions from the various components:

$$\boldsymbol{d}(p) = \boldsymbol{a}\,s(p) + \sum_{i=1}^{m} \boldsymbol{b}_i\,g_i(p) + \boldsymbol{n}(p), \tag{40}$$

where $s(p)$ is the CMB signal and $\boldsymbol{a}$ its SED, $\{g_i(p)\}_{1 \leq i \leq m}$ the m foreground components whose SEDs $\{\boldsymbol{b}_i\}_{1 \leq i \leq m}$ are constrained in cMILC, while not in the ILC, and $\boldsymbol{n}(p)$ the contamination from both instrumental noise and unmodelled foregrounds. The covariance matrix of the data thus reads as

$$\mathbf{C} = \boldsymbol{a}\,\sigma_s^2\,\boldsymbol{a}^{\text{T}} + \sum_{i=1}^{m} \boldsymbol{b}_i\,\sigma_{g_i}^2\,\boldsymbol{b}_i^{\text{T}} + \mathbf{N}, \tag{41}$$

where $\sigma_s^2 = \langle s(p)^2 \rangle$ is the variance of intrinsic CMB anisotropies, $\sigma_{g_i}^2 = \langle g_i(p)^2 \rangle$ the variance of the modelled foregrounds, and $\mathbf{N} = \langle \boldsymbol{n}(p)\boldsymbol{n}(p)^T \rangle$ the covariance matrix of the noise and unmodelled foregrounds.

By inserting equation (41) into equation (34), the contribution from the different components to the overall variance of the standard ILC map becomes explicit:

$$\sigma_{\text{ILC}}^2 = \boldsymbol{w}_{\text{ILC}}^{\text{T}}\left(\boldsymbol{a}\,\sigma_s^2\,\boldsymbol{a}^{\text{T}} + \sum_{i=1}^{m} \boldsymbol{b}_i\,\sigma_{g_i}^2\,\boldsymbol{b}_i^{\text{T}} + \mathbf{N}\right)\boldsymbol{w}_{\text{ILC}}$$

$$= \sigma_s^2 + \sum_{i=1}^{m} \left(\boldsymbol{w}_{\text{ILC}}^{\text{T}}\boldsymbol{b}_i\right)^2\sigma_{g_i}^2 + \boldsymbol{w}_{\text{ILC}}^{\text{T}}\mathbf{N}\,\boldsymbol{w}_{\text{ILC}}, \tag{42}$$

since $\boldsymbol{w}_{\text{ILC}}^{\text{T}}\boldsymbol{a} = 1$, but $\boldsymbol{w}_{\text{ILC}}^{\text{T}}\boldsymbol{b}_i \neq 0$. In contrast to the ILC solution, the variance of the cMILC map equation (35) reduces to

$$\sigma_{\text{cMILC}}^2 = \boldsymbol{w}_{\text{cMILC}}^{\text{T}}\left(\boldsymbol{a}\,\sigma_s^2\,\boldsymbol{a}^{\text{T}} + \sum_{i=1}^{m} \boldsymbol{b}_i\,\sigma_{g_i}^2\,\boldsymbol{b}_i^{\text{T}} + \mathbf{N}\right)\boldsymbol{w}_{\text{cMILC}}$$

$$= \sigma_s^2 + \boldsymbol{w}_{\text{cMILC}}^{\text{T}}\mathbf{N}\,\boldsymbol{w}_{\text{cMILC}}, \tag{43}$$

since $\boldsymbol{w}_{\text{cMILC}}^{\text{T}}\boldsymbol{a} = 1$ and $\boldsymbol{w}_{\text{cMILC}}^{\text{T}}\boldsymbol{b}_i = 0$. Therefore, the variance of the cMILC map is cleared of any contribution from the constrained foregrounds (equation 43), while some residual variance from those foregrounds still contributes to the variance of the standard ILC map (equation 42). If we now consider that the unconstrained foreground moments are those which are below the noise of the instrument, then $\boldsymbol{w}_{\text{ILC}}^{\text{T}}\mathbf{N}\,\boldsymbol{w}_{\text{ILC}}$ and $\boldsymbol{w}_{\text{cMILC}}^{\text{T}}\mathbf{N}\,\boldsymbol{w}_{\text{cMILC}}$ are just the noise variance contributions to the standard ILC and cMILC maps, and hence cMILC moves towards the solution with minimum residual foreground contamination, unlike the standard ILC. Consequently, the inequality equation (39) on overall variance implies that the residual noise variance must be larger in the cMILC map than in the standard ILC map, i.e. $\boldsymbol{w}_{\text{cMILC}}^{\text{T}}\mathbf{N}\,\boldsymbol{w}_{\text{cMILC}} \geq \boldsymbol{w}_{\text{ILC}}^{\text{T}}\mathbf{N}\,\boldsymbol{w}_{\text{ILC}}$.

To conclude, cMILC reduces the variance and bias from residual foreground contamination, while it has larger noise variance than the standard ILC. By including more and more foreground moments in cMILC, the residual noise variance increases in the CMB map, but in contrast the residual foreground variance decreases towards the minimum foreground variance solution, thus largely beating the standard ILC. These theoretical expectations are confirmed by our analysis on sky simulations (see Fig. 6).

While foreground residuals and noise residuals both cause biases on the CMB *B*-mode power spectrum, and hence on the inferred tensor-to-scalar ratio *r*, the noise bias can be corrected for (i.e. by Jackknife), such that any bias on a detection arises from foreground residuals, not from instrumental noise residuals. In this respect, the cMILC method focuses on nulling foreground biases at the expense of increased noise variance.

3.4.3 Higher order statistics

By realizing that CMB and instrumental noise are mostly Gaussian fields compared to Galactic foregrounds which are highly non-Gaussian, it becomes clear that cMILC will get rid of most non-Gaussian residuals in the CMB map through nulling constraints, while the standard ILC will not. As an illustration, the skewness of the reconstructed CMB map reads

$$\langle \hat{s}^3 \rangle = \langle s^3 \rangle + \sum_{i,j,k=1}^{m} (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_i)(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_j)(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_k)\langle g_i\, g_j\, g_k \rangle + \langle (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{n})^3 \rangle$$

$$\simeq \sum_{i,j,k=1}^{m} (\boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_i)(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_j)(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{b}_k)\langle g_i\, g_j\, g_k \rangle, \quad (44)$$

where in the last line we assumed that CMB and noise are mostly Gaussian and have negligible skewness, i.e. $\langle s^3 \rangle \simeq 0$ and $\langle \boldsymbol{n}^3 \rangle \simeq 0$. By construction, the cMILC weights give $\boldsymbol{w}_{\mathrm{cMILC}}^{\mathrm{T}}\boldsymbol{b}_i = 0$, such that the skewness due to foreground residuals in equation (44) should reduce to zero for the cMILC map, while this is not the case for the standard ILC map. These expectations are also confirmed by our analysis on sky simulations (see Fig. 6). While this work focuses on *B* modes, we refer to Section 5 for a brief discussion of the potential benefits of reducing non-Gaussian residuals for primordial non-Gaussianity and CMB lensing analyses.

4 ANALYSIS

4.1 Sky simulations

In our sky simulations, we consider different models of the foreground emission, capitalizing on existing tools. For *PICO*, we use the public *PICO* sky simulations,[6] which have been delivered on NERSC as an open data analysis challenge. These include both the PySM d1s1 model (Thorne et al. 2017), which is consistent with *Planck* observations, and a more complicated dust model derived from MHD simulations (Kritsuk, Flauger & Ustyugov 2018) that has non-trivial spectral dependence and accounts for line-of-sight effects, as we discuss in Section 4.2.5. For *LiteBIRD*, since there is no available public simulation, we replicated the PySM d1s1 model of the public *PICO* simulation with specifications given in Table B1.

We first evaluate the performance of our semiblind component separation method cMILC on the PySM sky simulations d1s1 for

---

[6]https://zzz.physics.umn.edu/ipsig/20180424_dc_maps (models 91 and 96).

both *LiteBIRD* and *PICO* experiments, and compare the results of cMILC with those obtained with the standard NILC method. The simulated sky maps in polarization for the model d1s1 include several components of emission: CMB anisotropies with a tensor-to-scalar ratio of $r = 0$, an optical depth to reionization of $\tau = 0.054$, and full lensing contamination ($A_{\mathrm{L}} = 1$); Galactic thermal dust emission as a modified blackbody component with varying spectral index and temperature across the sky, based on the *Planck* Commander maps (Planck Collaboration X 2016a); and Galactic synchrotron emission as a power-law component with varying spectral index across the sky, based on the template maps from Miville-Deschênes et al. (2008). For more details regarding PySM and its implementation we refer to Thorne et al. (2017).

The sky emission is integrated over either *LiteBIRD* (Table B1) or *PICO* frequency bands (Table B2), for which current simulations assume δ-function bandpasses. Each frequency map has a HEALPix $N_{\mathrm{side}} = 512$ pixelization scheme (Górski et al. 2005), and is smoothed with a Gaussian beam of full width at half-maximum (FWHM) values listed in Tables B1 and B2. Gaussian white noise map realizations of typical r.m.s. values listed in the aforementioned tables were then co-added to the sky maps. Therefore, we have 15 sky maps ranging from 40 to 402 GHz for the *LiteBIRD*-like simulation, and 21 sky maps ranging from 20 to 800 GHz for the *PICO* simulation.

We perform foreground removal by applying the NILC and cMILC methods to these two sets of sky maps, and study the impact on CMB *B*-mode reconstruction (map, power spectrum, and tensor-to-scalar ratio) of deprojecting more and more foreground moments with cMILC, with combinations as listed in Table 1. For our analysis, we first transform the Stokes Q, U full-sky maps into full-sky *B*-mode maps at each frequency, then apply the NILC and cMILC methods on the set of full-sky *B*-mode observations, thus avoiding *E*-to-*B* leakage inherent to masking procedures. To obtain numerically stable results, it was important to carefully consider the precision settings of the linear algebra routines. The CMB *B*-mode map reconstruction is performed at 40 arcmin angular resolution for both *LiteBIRD* and *PICO*. As a main figure of merit we use r, but we expect other observables such as $\sum m_\nu$ and N_{eff} to also benefit from cMILC (see discussion in Section 5).

4.2 Results

4.2.1 Visual inspection of recovered maps in the BICEP2 region

While the reconstruction of the CMB *B*-mode map from the *LiteBIRD* and *PICO* sky maps by the NILC and cMILC methods is performed on the entire sky, for illustration we first inspect the quality of the reconstructed maps and their residual foreground and noise contamination in the BICEP2 region (BICEP2 Collaboration 2014). This allows us to better appreciate the improvement in foreground removal with cMILC in comparison to NILC. To evaluate residual foreground and noise contamination in the reconstructed CMB *B*-mode map after component separation, we applied the same NILC/cMILC weights that were assigned in the reconstruction of the CMB *B*-mode map to the input foreground and noise maps of the simulations.

Fig. 4 summarizes the results for *LiteBIRD* in the BICEP2 region. The first row shows the NILC results, while the following rows show the cMILC results when including progressively more constraints on foreground moments, starting in the second row with cMILC03 which deprojects zeroth-order moments (f_{sync}, f_{dust}), and ending in the bottom row with cMILC08 which deprojects both zeroth- and first-order moments (f_{sync}, f_{dust}, $\partial_{\beta_s} f_{\mathrm{sync}}$, $\partial_{\beta_d} f_{\mathrm{dust}}$, $\partial_{T_d} f_{\mathrm{dust}}$). No further
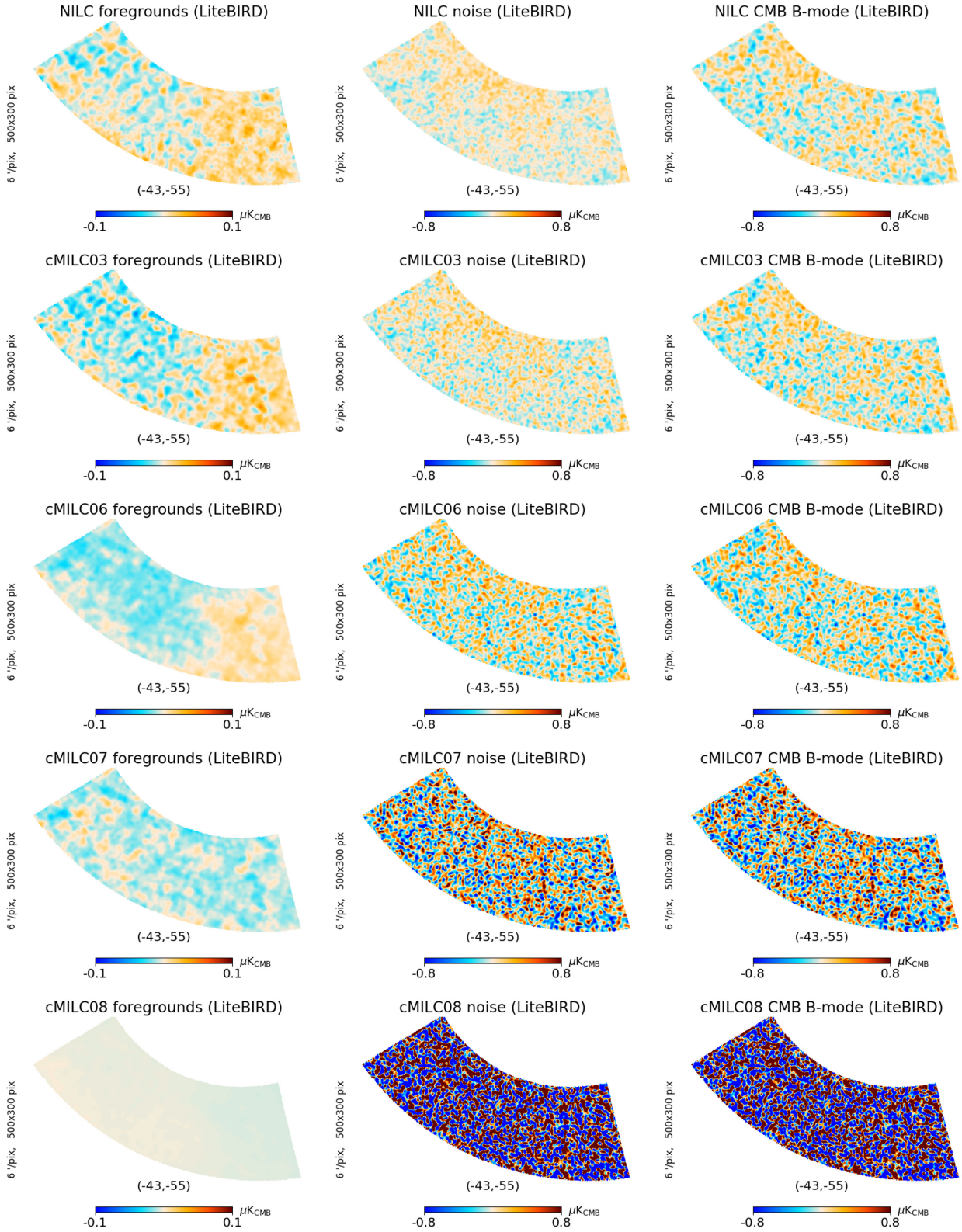
Figure 4. *Top to bottom rows*: cMILC results for *LiteBIRD* in the BICEP2 region when deprojecting more and more foreground moments. *Left column*: residual foregrounds. *Middle column*: residual noise. *Right column*: reconstructed CMB *B*-mode map. Deprojecting moments with cMILC significantly reduces the residual foreground contamination in the recovered CMB *B*-mode map, although this comes along with a noise penalty. Among these maps, CMILC06 is the optimal choice in terms of trade-off between residual foreground bias and noise penalty on the tensor-to-scalar ratio *r* (see Section 4.2.4).

constraints on second-order moments were imposed in cMILC because the noise degradation prevented us from probing higher order moments of the foreground emission.

The residual foreground contamination left by each method is shown in the first column: clearly, the level of residual foregrounds is significantly reduced in the cMILC CMB *B*-mode maps when including more and more moments in the component separation. The second and third columns show, respectively, the residual noise fluctuations and the recovered CMB *B*-mode fluctuations for each method. We see that the improvement on foreground removal due to moment constraints is at the expense of increasing noise in the reconstructed CMB *B*-mode map, as expected. Hence, there is a clear trade-off between mitigation of residual foreground contamination and noise degradation, but also a sweet spot where the noise penalty is more than compensated by the reduction of residual foreground contamination, depending on the observable of interest. As we will see in Section 4.2.4, for *r* the optimal solution is given by cMILC06 for *LiteBIRD*.

Fig. 5 summarizes the results for *PICO* in the BICEP2 region. As expected, the level of residual foregrounds and noise in the CMB *B*-mode map reconstruction for *PICO* is further reduced compared to *LiteBIRD* due to higher sensitivity and broader spectral coverage (20–800 GHz). *PICO* also has the capability to probe second-order moments of the foreground emission. In particular, adding constraints on first- and second-order dust temperature moments, $\boldsymbol{w}^{\mathrm{T}} \cdot \partial_{T_d} f_{\mathrm{dust}} = 0$ (cMILC08; fourth row) and $\boldsymbol{w}^{\mathrm{T}} \cdot \partial_{T_d}^2 f_{\mathrm{dust}} = 0$ (cMILC09; bottom row), significantly reduces the residual foreground contamination at small scales. Again, the left column highlights the improvement in terms of residual foreground contamination when including more and more moments in cMILC, with best results obtained in the bottom row (cMILC09), where several moments of dust and synchrotron up to second order (f_{sync}, f_{dust}, $\partial_{\beta_s} f_{\mathrm{sync}}$, $\partial_{\beta_d} f_{\mathrm{dust}}$, $\partial_{T_d} f_{\mathrm{dust}}$, $\partial_{T_d}^2 f_{\mathrm{dust}}$) were deprojected. As expected, additional moment constraints are at the expense of an increase of residual noise contamination (middle column), but remain at a reasonable level compared to the r.m.s of CMB *B*-mode fluctuations (right column) due to the high sensitivity of *PICO*. The best trade-off for *r* among those maps is actually given by cMILC08 for *PICO*, as we show in Section 4.2.4.

4.2.2 *Statistical properties of residual foregrounds and noise*

Besides map visualization, it is instructive to look at the statistical properties of the residual foreground and noise contaminations in the recovered CMB *B*-mode maps, which we present in Fig. 6 for *LiteBIRD* (upper panels) and *PICO* (lower panels).

In the left-hand panels of Fig. 6, we computed the one-point statistics (probability distribution function – PDF) of the residual foreground contamination on $f_{\mathrm{sky}} = 50$ per cent of the sky for the NILC CMB *B*-mode map (red line) and for several cMILC CMB *B*-mode maps, with more and more moments being deprojected. Similarly, in right-hand panels of Fig. 6, we show the PDF of the residual noise contamination for the same NILC and cMILC maps. Overall, the residual foreground and noise contamination is lower for *PICO* than *LiteBIRD* due to higher sensitivity and broader spectral coverage. As we already stressed in Section 3.4, the variance of the residual noise contamination (right-hand panels) increases when adding constraints on moments in cMILC because of the increasing volume of the parameter space with respect to NILC. In stark contrast, the variance of the residual foreground contamination (left-hand panels) significantly decreases by deprojecting moments with cMILC.

Clearly, the PDF of foreground residuals (left-hand panels) shows larger variance and skewness for NILC (red line), while adding nulling constraints on zeroth- and first-order moments of dust and synchrotron (f_{sync}, f_{dust}, $\partial_{\beta_s} f_{\mathrm{sync}}$, $\partial_{\beta_d} f_{\mathrm{dust}}$, $\partial_{T_d} f_{\mathrm{dust}}$) with cMILC significantly reducing the variance and skewness of the residual foreground contamination (cMILC08; blue line), therefore leaving only negligible Gaussian residuals in the recovered CMB *B*-mode map. More quantitatively, for *LiteBIRD* the variance of residual foregrounds is $\sigma_{\mathrm{FG}}^2 = 0.05\,\sigma_{\mathrm{CMB}}^2$ for NILC and $\sigma_{\mathrm{FG}}^2 = 0.003\,\sigma_{\mathrm{CMB}}^2$ for cMILC08, hence a reduction of the foreground variance by 94 per cent with cMILC. For *PICO*, the variance of residual foregrounds is $\sigma_{\mathrm{FG}}^2 = 0.0025\,\sigma_{\mathrm{CMB}}^2$ for NILC and $\sigma_{\mathrm{FG}}^2 = 0.0012\,\sigma_{\mathrm{CMB}}^2$ for cMILC10, hence a reduction of the foreground variance by 52 per cent with cMILC. Similarly, the skewness of residual foregrounds on $f_{\mathrm{sky}} = 50$ per cent of the sky is $\langle s_{\mathrm{FG}}^3 \rangle / \sigma_{\mathrm{FG}}^3 = -0.3$ for NILC and $\langle s_{\mathrm{FG}}^3 \rangle / \sigma_{\mathrm{FG}}^3 = 0.18$ for cMILC08, hence a reduction of the residual skewness by 40 per cent with cMILC (with respect to the CMB r.m.s., the residual skewness is $\langle s_{\mathrm{FG}}^3 \rangle / \sigma_{\mathrm{CMB}}^3 = -3 \times 10^{-3}$ for NILC and $\langle s_{\mathrm{FG}}^3 \rangle / \sigma_{\mathrm{CMB}}^3 = 4 \times 10^{-5}$ for cMILC08). For *PICO*, the skewness of residual foregrounds is $\langle s_{\mathrm{FG}}^3 \rangle / \sigma_{\mathrm{FG}}^3 = -0.3$ for NILC and $\langle s_{\mathrm{FG}}^3 \rangle / \sigma_{\mathrm{FG}}^3 = -0.11$ for cMILC10, hence a reduction of the residual skewness by 60 per cent with cMILC (with respect to the CMB r.m.s., the residual skewness is $\langle s_{\mathrm{FG}}^3 \rangle / \sigma_{\mathrm{CMB}}^3 = -4 \times 10^{-5}$ for NILC and $\langle s_{\mathrm{FG}}^3 \rangle / \sigma_{\mathrm{CMB}}^3 = -5 \times 10^{-6}$ for cMILC10). Therefore, cMILC leads to a solution that minimizes foreground variance and skewness, largely beating NILC on that front. The trends observed in our numerical results are consistent with those predicted by the analytical expressions in Section 3.4.

4.2.3 B-*mode power spectrum*

In Fig. 7, we compute the angular power spectrum of the residual foreground contamination (C_ℓ^{fgds}; left-hand panels) in the recovered CMB *B*-mode maps on $f_{\mathrm{sky}} = 50$ per cent of the sky, along with the standard deviation of the residual noise and lensing power spectra ($\sqrt{2/(2\ell + 1) f_{\mathrm{sky}}}\,C_\ell^{\mathrm{noise/lensing}}$; right-hand panels). The results from NILC and cMILC are shown for *LiteBIRD* in the upper panels and for *PICO* in the lower panels. The power spectra are deconvolved from the Galactic mask and the beam window function using MASTER (Hivon et al. 2002), and binned across multipoles with $\Delta\ell = 16$. The same Galactic mask was used for the NILC and cMILC maps for fair comparison, but in principle the shape of the Galactic mask can be optimized for each variant of cMILC, since the morphology of the residual foreground contamination obviously differs, as is also suggested by the studies presented in Sections 4.2.1 and 4.2.2.

It is clear from the left-hand panels of Fig. 7 that adding constraints on dust and synchrotron moments in cMILC significantly reduces the residual foreground contamination across a broad range of multipoles compared to NILC, and thus potentially prevents biases due to residual foregrounds in the recovered primordial CMB *B*-mode power spectrum. In contrast, while the residual noise bias on the power spectrum can be corrected for using Jackknife, the noise error (standard deviation of residual noise power spectrum) increases when increasing the volume of the constrained parameter space in cMILC, as it is shown in the right-hand panels of Fig. 7.

For *LiteBIRD* (upper left panel), while the power spectrum of residual foreground contamination for NILC (red line) is reduced well below the lensing *B*-mode signal, it is still at a level equivalent to $r \gtrsim 3 \times 10^{-3}$ across the full range of multipoles. Therefore, after delensing, NILC would not guarantee proper recovery of the
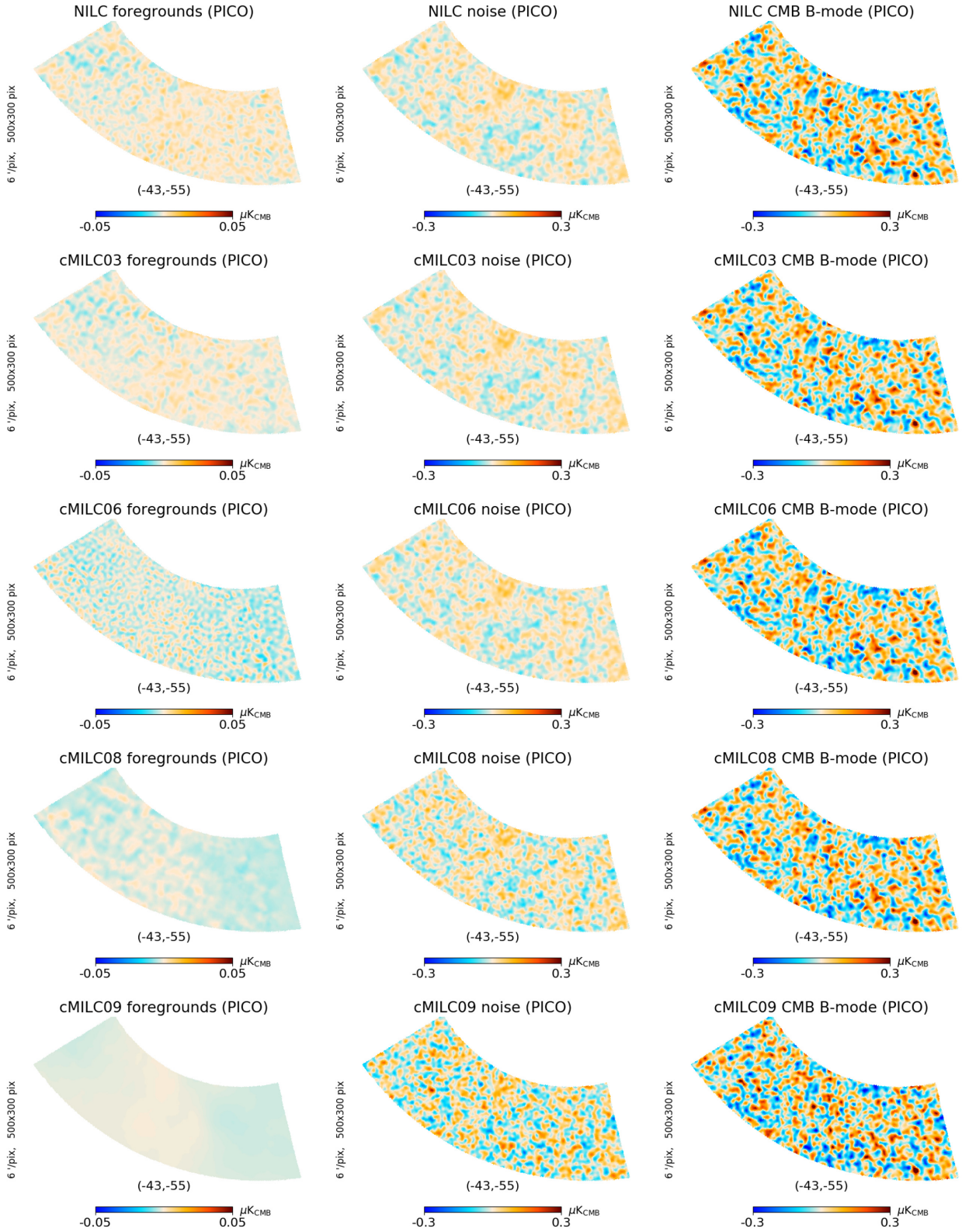
Figure 5. *Top to bottom rows*: cMILC results for *PICO* in the BICEP2 region when deprojecting more and more foreground moments. *Left column*: residual foregrounds. *Middle column*: residual noise. *Right column*: reconstructed CMB *B*-mode map. Deprojecting moments with cMILC significantly reduces the residual foreground contamination in the recovered CMB *B*-mode map, while the noise penalty is still reasonably low for *PICO*. CMILC08 is the optimal choice in terms of trade-off between residual foreground bias and noise penalty on the tensor-to-scalar ratio *r* (see Section 4.2.4).
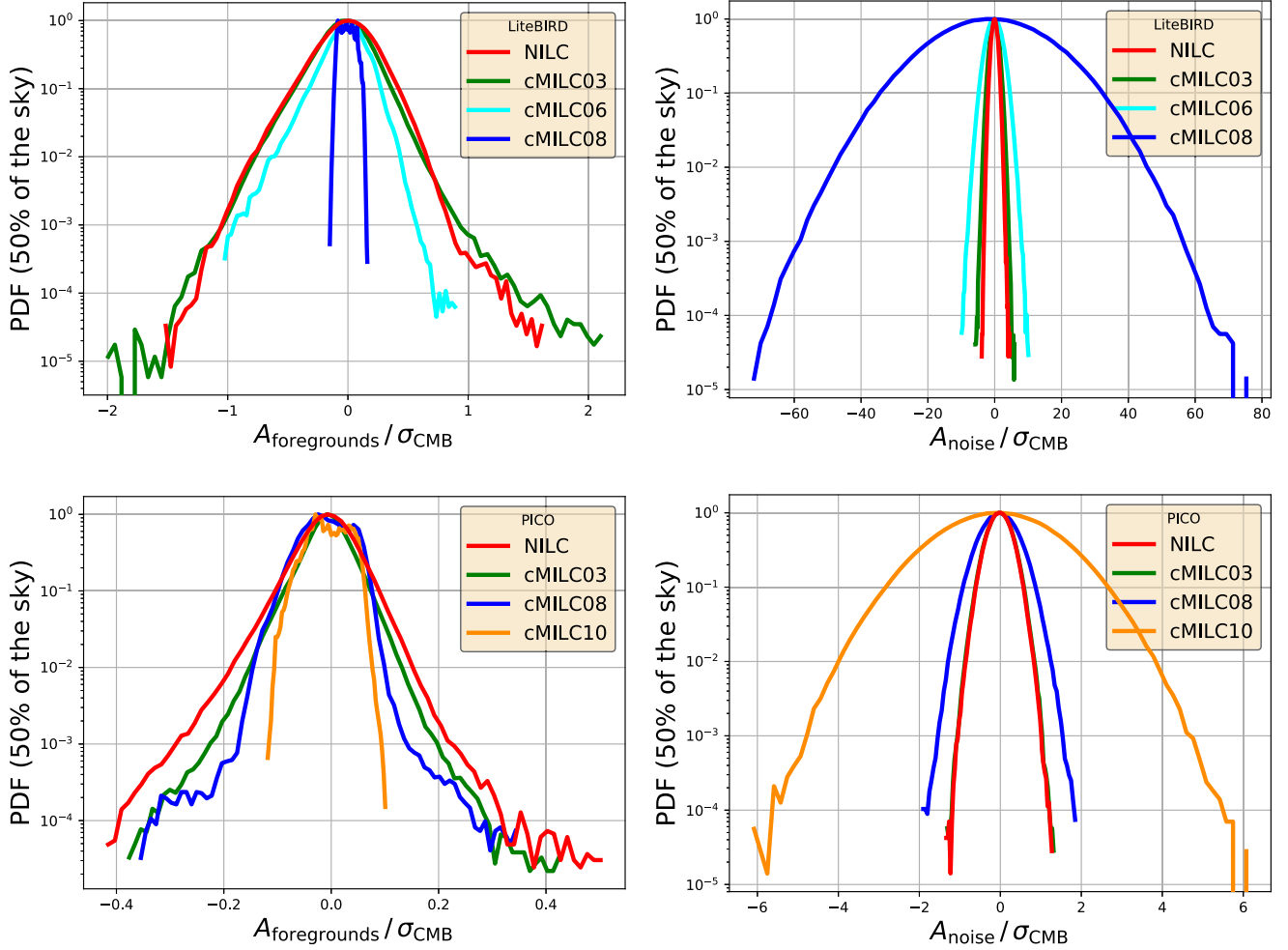
Figure 6. One-point statistics (probability distribution function – PDF) of foreground residuals (*left-hand panels*) and noise residuals (*right-hand panels*) on $f_{sky} = 50$ per cent of the sky for *LiteBIRD* (*upper panels*) and *PICO* (*lower panels*). Deprojecting more and more moments with cMILC significantly reduces the variance and skewness of non-Gaussian foreground residuals in the recovered CMB B-mode map. Besides B modes, this would be of great benefit also for the search for primordial non-Gaussianity and for CMB lensing reconstruction (see Section 5 for discussion).

primordial CMB B-mode power spectrum at $r = 10^{-3}$. In contrast, by adding constraints on more and more dust and synchrotron moments, cMILC helps reducing the level of residual foreground contamination in the recovered CMB B-mode power spectrum down to $r < 10^{-3}$ at recombination scales. In particular, deprojecting all zeroth- and first-order moments of the dust and synchrotron (cMILC08; blue line) enables significant reduction of the residual foreground contamination to CMB B modes for *LiteBIRD*, down to levels well below $r = 10^{-3}$ at all multipoles.

Similarly, the lower left panel of Fig. 7 shows the power spectra of the residual foreground contamination for *PICO* after component separation with the NILC and cMILC methods. Due to its higher sensitivity and broader spectral coverage, *PICO* allows reducing foreground residuals well below the primordial CMB B-mode power spectrum at $r = 10^{-3}$ across the full range of multipoles covering both reionization and recombination scales, even with a blind NILC method. Still, cMILC enables further reduction of the residual foreground contamination to the CMB B-mode power spectrum by deprojecting more and more foreground moments. We find that first- and second-order dust temperature moments (cMILC08, cMILC10, and cMILC12) lead to the lowest biases due to residual foreground

contamination, reaching below the detection limit $r = 5 \times 10^{-4}$ of *PICO* (Hanany et al. 2019).

While deprojecting successively more moments with cMILC reduces the level of residual foreground contamination across a wide range of multipoles, for *PICO* the overall decrease of residuals actually breaks at the largest scales $\ell \lesssim 15$, suggesting that our first-guess pivots, based on zeroth-order fits in the literature (e.g. Planck Collaboration X 2016a), might not be appropriate for the largest angular scales, where averaging processes, and thus high-order moments, are the most significant. For sensitive experiments like *PICO*, high-order moments should thus be included in parametric SED fitting to revise zeroth-order pivots. In Section 5.1 and Fig. 10, we update the dust pivot temperature $\overline{T}_d$ to show that, with revised pivots, cMILC is able to reduce the residual foreground contamination consistently at all mutipoles. However, a clear computational procedure for optimizing the pivots still requires more work, as we also explain in Section 5.1.

To conclude, cMILC provides quite spectacular results on foregrounds removal across all multipoles, thus significantly reducing biases on the CMB B-mode power spectrum. However, there is a clear trade-off between residual foregrounds mitigation (left-hand
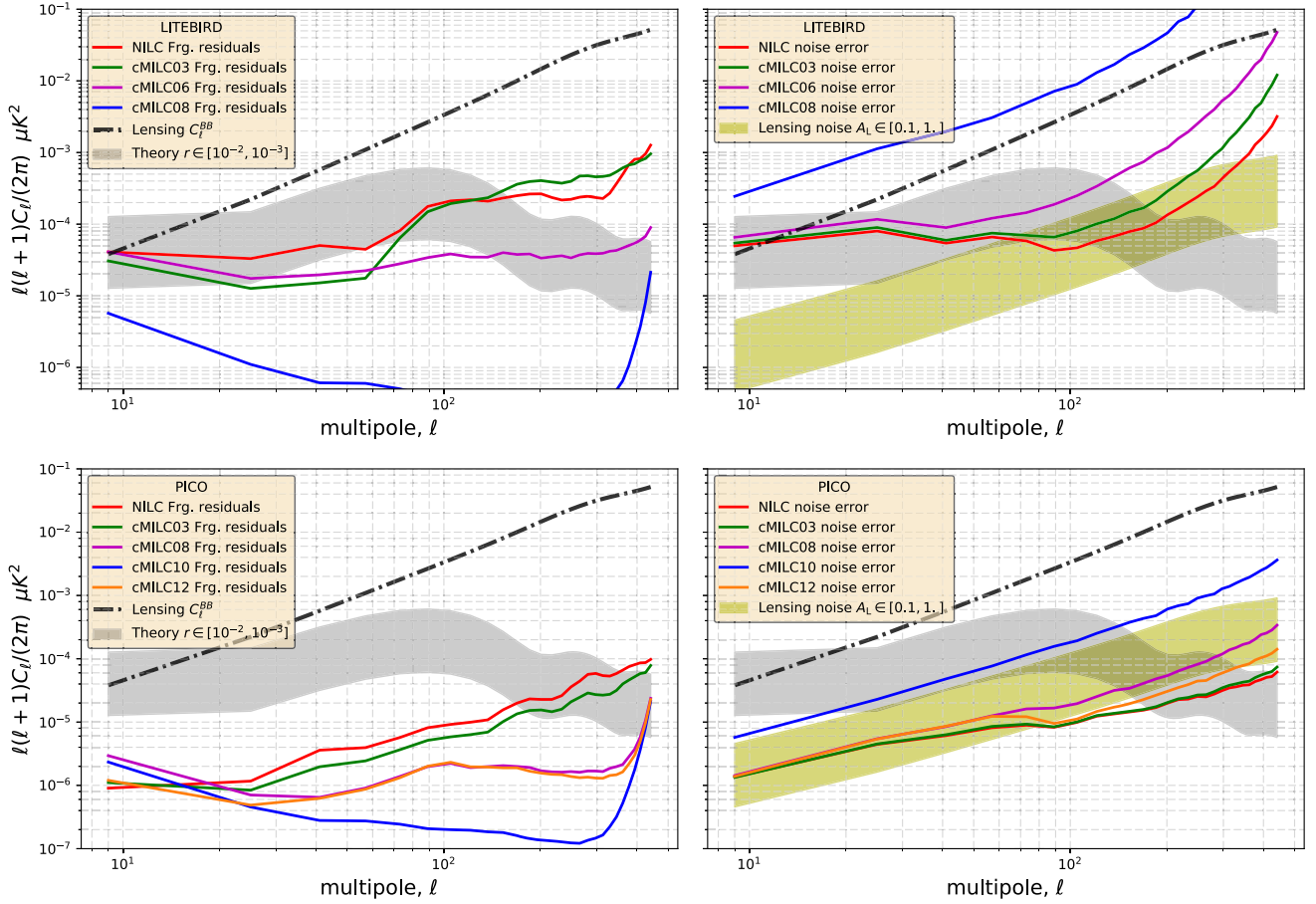
Figure 7. *Left-hand panels*: Power spectrum of the residual foreground contamination, C_ℓ^{fgds}, when deprojecting more and more foreground moments with cMILC, for *LiteBIRD* (upper panel) and *PICO* (lower panel) for the PySM d1s1 simulation. *Right-hand panels*: Standard deviation of the residual noise and lensing power spectra, $\sqrt{2/(2\ell+1)f_{\text{sky}}}\,C_\ell^{\text{noise/lensing}}$. The residual foreground contamination across multipoles is significantly reduced by deprojecting more and more moments with cMILC, as compared to NILC.

panels of Fig. 7) and noise degradation (right-hand panels of Fig. 7) that needs to be appreciated to find the best combination of moments in cMILC that will lead to less biased but still sensitive constraints on the observable of interest, which here is *r* (see Section 4.2.4).

4.2.4 Likelihood estimation of the tensor-to-scalar ratio

In this section, we perform the likelihood estimation of the recovered probability distribution of the tensor-to-scalar ratio *r* after foreground removal with NILC and cMILC. We emphasize two important aspects: the irreducible bias on *r*, or *systematic error* arising from residual foreground contamination (noting that noise and lensing biases on the power spectrum can usually be corrected for), and the *statistical uncertainty* $\sigma(r)$. The latter has contributions from cosmic variance of the primordial signal, residual foregrounds, residual lensing signals, and noise.

The power spectrum of the reconstructed CMB *B*-mode map from either NILC or cMILC is

$$\hat{C}_\ell^{\text{BB}} = C_\ell^{\text{lens}} + \hat{C}_\ell^{\text{fgds}} + \hat{C}_\ell^{\text{noise}}, \qquad (45)$$

where $C_{\ell_b}^{\text{lens}}$ is the power spectrum of the CMB lensing *B*-mode signal ($r = 0$, $A_{\text{L}} = 1$), $\hat{C}_\ell^{\text{fgds}}$ the power spectrum of residual foregrounds (left-hand panels of Fig. 7), and $\hat{C}_\ell^{\text{noise}}$ the noise power spectrum.

The reconstructed CMB *B*-mode map can be corrected for part of the cosmic variance of the lensing signal, either through internal delensing (Larsen et al. 2016; Carron, Lewis & Challinor 2017; Millea, Anderes & Wandelt 2019) or through external lensing tracers (Sherwin & Schmittfull 2015; Planck Collaboration VIII 2020b) such as cosmic infrared background maps (Planck Collaboration Int. XLVIII 2016).

For our likelihood analysis, we will assume three levels of delensing: no delensing ($A_{\text{L}} = 1$), 60 per cent delensing ($A_{\text{L}} = 0.4$), and full delensing ($A_{\text{L}} = 0$), such that after component separation and delensing the measured CMB *B*-mode power spectrum is

$$\tilde{C}_\ell^{\text{BB}} \equiv \hat{C}_\ell^{\text{BB}} - (1 - A_L) C_\ell^{\text{lens}} = A_L C_\ell^{\text{lens}} + \hat{C}_\ell^{\text{fgds}} + \hat{C}_\ell^{\text{noise}}. \qquad (46)$$

The measured CMB *B*-mode power spectrum, $\tilde{C}_\ell^{\text{BB}}$, can in principle be corrected for the noise bias $\hat{C}_\ell^{\text{noise}}$ and for the residual lensing bias $A_L C_\ell^{\text{lens}}$, so that after these corrections, any irreducible bias on the measured CMB *B*-mode power spectrum arises from the residual foreground contamination:

$$\hat{C}_\ell^{\text{fgds}} = \tilde{C}_\ell^{\text{BB}} - A_L C_\ell^{\text{lens}} - \hat{C}_\ell^{\text{noise}}. \qquad (47)$$

The binned likelihood (Hamimeche & Lewis [2008]) on the tensor-to-scalar ratio r is thus built as

$$-2\ln\mathcal{L}(r) = \sum_{\ell_b, \ell_b'} \left(\hat{C}_{\ell_b}^{\text{fgds}} - r\, C_{\ell_b}^{\text{prim}}\right) M_{\ell_b \ell_b'}^{-1} \left(\hat{C}_{\ell_b'}^{\text{fgds}} - r\, C_{\ell_b'}^{\text{prim}}\right),$$

(48)

where $\hat{C}_{\ell}^{\text{fgds}}$ is the power spectrum of residual foregrounds, $C_{\ell_b}^{\text{prim}}$ is the primordial CMB B-mode power spectrum model for a tensor-to-scalar ratio $r = 1$, and $M_{\ell\ell'}$ is the covariance matrix for a fiducial cosmological model with $r = 0$. Given that the input sky simulations do not contain any primordial signal, i.e. $r = 0$, the likelihood equation (48) thus computes the equivalent r bias due to residual foreground contamination, and its significance with respect to the overall uncertainty due to cosmic variance of residual lensing, noise, and residual foregrounds, which are all included in the covariance matrix.

The binning, $\ell_b \pm \Delta\ell/2$ with $\Delta\ell = 16$, of the power spectra mitigates correlations between different ℓ modes, so that off-diagonal terms of the covariance matrix can be neglected, thus leaving only the diagonal elements:

$$M_{\ell_b \ell_b} = \frac{2}{(2\ell_b + 1)\, f_{\text{sky}} \Delta\ell} \left(\hat{C}_{\ell_b}^{\text{BB}} - (1 - A_L)\, C_{\ell_b}^{\text{lens}}\right)^2,$$

(49)

which, according to equation (46), accounts for the cosmic variance of the residual lensing signal $2(A_L C_\ell^{\text{lens}})^2/(2\ell + 1)$, the sample variance of the residual foreground power spectrum $2(\hat{C}_\ell^{\text{fgds}})^2/(2\ell + 1)$, the sample variance of the noise power spectrum $2(\hat{C}_\ell^{\text{noise}})^2/(2\ell + 1)$, and their cross-terms. The likelihood results on r are obtained by summing modes between $\ell_{\min} = 2$ and $\ell_{\max} = 450$, thus accounting for both the reionization and recombination bumps.

Fig. [8] summarizes our results on the recovered tensor-to-scalar ratio r after component separation with NILC or cMILC for both *LiteBIRD* and *PICO*, and for different levels of delensing. Biased detections (SNR ≥ 2) due to significant foreground residuals are shown in red with both 1σ (thick red) and 2σ (thin red) error bars, while for unbiased (SNR < 2) measurements, the 95 per cent upper bounds on r are depicted by the thick green lines. As already observed on map reconstruction (Sections 4.2.1 and 4.2.2) and power-spectrum analysis (Section 4.2.3), adding constraints on an increasing number of moments with cMILC generally increases the uncertainty on r due to noise degradation, but allows for eliminating biases on r due to residual foreground contamination. This thus can help to robustly turn high significance but *false* detections with NILC into unbiased upper limits with the optimal choice of moment number in cMILC. However, the details depend on the experimental configuration and the overall level of delensing as we explain now.

The goal for full success of the *LiteBIRD* mission (Hazumi et al. [2019]) is to achieve $\delta r \lesssim 10^{-3}$, including both statistical and systematic uncertainties. For additional comparison, we thus define the simple metric

$$\delta r \equiv \sqrt{r^2 + \sigma^2(r = 0)},$$

(50)

which accounts for both the *statistical error* $\sigma(r = 0)$ and the *systematic error*, i.e. the residual bias r with respect to $r = 0$. For *LiteBIRD* (upper panels of Fig. [8]), NILC clearly leads to a significant bias on r in our simulations, with $r = (2.5 \pm 1.0) \times 10^{-3}$ in the absence of any delensing and $r = (3.2 \pm 0.4) \times 10^{-3}$ in case of full delensing, which corresponds to a 3σ to 8σ bias on $r = 0$ and thus a cumulative systematic and statistical error of $\delta r \gtrsim 3 \times 10^{-3}$. Therefore, the systematic foreground residuals of NILC would not allow *LiteBIRD* to robustly reach its goals, even if the statistical uncertainties are sufficient. In contrast, for cMILC

the bias on r progressively decreases by deprojecting more and more foreground moments, with cMILC06, cMILC07, and cMILC08 all leading to unbiased measurements of $r = 0$, i.e. 2σ upper limits. Among these three unbiased versions, we see that for *LiteBIRD* cMILC06 (f_{sync}, f_{dust}, $\partial_\beta f_{\text{dust}}$) provides the best trade-off between bias mitigation and noise degradation on r, with $r = (0.7 \pm 1.1) \times 10^{-3}$ consistent with unbiased measurement of $r = 0$, while accumulating a systematic and statistical errors as low as $\delta r \simeq \sigma(r = 0) \simeq 10^{-3}$. This would allow a constraint on r that is in line with the goals for *LiteBIRD*. Deprojecting dust temperature moments $\partial_T f_{\text{dust}}$ (cMILC08) would further reduce residual foreground biases down to $r \lesssim 0.3 \times 10^{-3}$, but the limited spectral coverage of *LiteBIRD* at high frequency does not seem to provide enough sensitivity to constrain dust temperature moments and the noise degradation for cMILC08 starts blowing up. Finally, increasing the level of delensing helps reducing statistical uncertainties on r for NILC and low-order cMILC. In contrast, the gain is less significant for the most-constrained versions of cMILC like cMILC06, because noise dominates over the lensing error for *LiteBIRD* (see upper right panel of Fig. [7]).

For *PICO* (lower panels of Fig. [8]), we can generally say that the overall degradation of the statistical error is not as large when including progressively more moments until about cMILC10. This is due to the broad frequency coverage and high channel sensitivity of *PICO*, which allows us to extract a lot of extra information even beyond r. However, the overall performance of the method for recovery of r depends significantly on the level of delensing. Assuming full delensing (left-hand panel), NILC would yield $r = (2.2 \pm 0.4) \times 10^{-4}$, which corresponds to a 5σ bias on $r = 0$ due to residual foreground contamination, while cMILC allows to reduce the bias on r by progressively deprojecting more moments of the foregrounds. For now omitting the optimized hybrid case cMILC12 (which we will return to in Section 5.2), for *PICO*, cMILC08 (f_{sync}, f_{dust}, $\partial_\beta f_{\text{sync}}$, $\partial_\beta f_{\text{dust}}$, $\partial_T f_{\text{dust}}$) provides the best results, giving $r = (0.7 \pm 0.9) \times 10^{-4}$ consistent with $r = 0$, and hence a cumulative systematic and statistical error as low as $\delta r \simeq 10^{-4}$. Due to high-frequency coverage above 500 GHz, *PICO* allows to constrain dust temperature moments $\partial_T f_{\text{dust}}$ (cMILC08) with enough sensitivity to control residual foreground biases, while still mitigating the noise degradation. Nevertheless, for $A_{\text{lens}} = 0$, we can see that *all* considered methods, including NILC, provide tight constraints on r, with biases typically smaller than $r = 5 \times 10^{-4}$, which is the target detection of *PICO* (Hanany et al. [2019]). For partial delensing, all the considered methods provide quite consistent results on r, compatible with $r = 0$, since the residual lensing error dominates over the noise for *PICO* (see lower right panel of Fig. [7]). However, as we will show below, pivot optimization can affect this behaviour.

Overall, our results indicate NILC alone does not lead to unbiased estimates of r in the case of *LiteBIRD*, but has a good performance for *PICO*. By applying cMILC, these limitations can be overcome and even *PICO*'s constraints on r could be further improved. It also seems that with the dls1 simulations inclusion of first-order moments suffices for constraints on r, even if *PICO* does show sensitivity to the second-order moment of T_d.

4.2.5 More complex foreground models: MHD simulation with line-of-sight contributions

As stressed by Tassis & Pavlidou ([2015]), the effective SED of the Galactic thermal dust foreground emission must be less trivial than a simple modified blackbody because of the average of multiple cloud
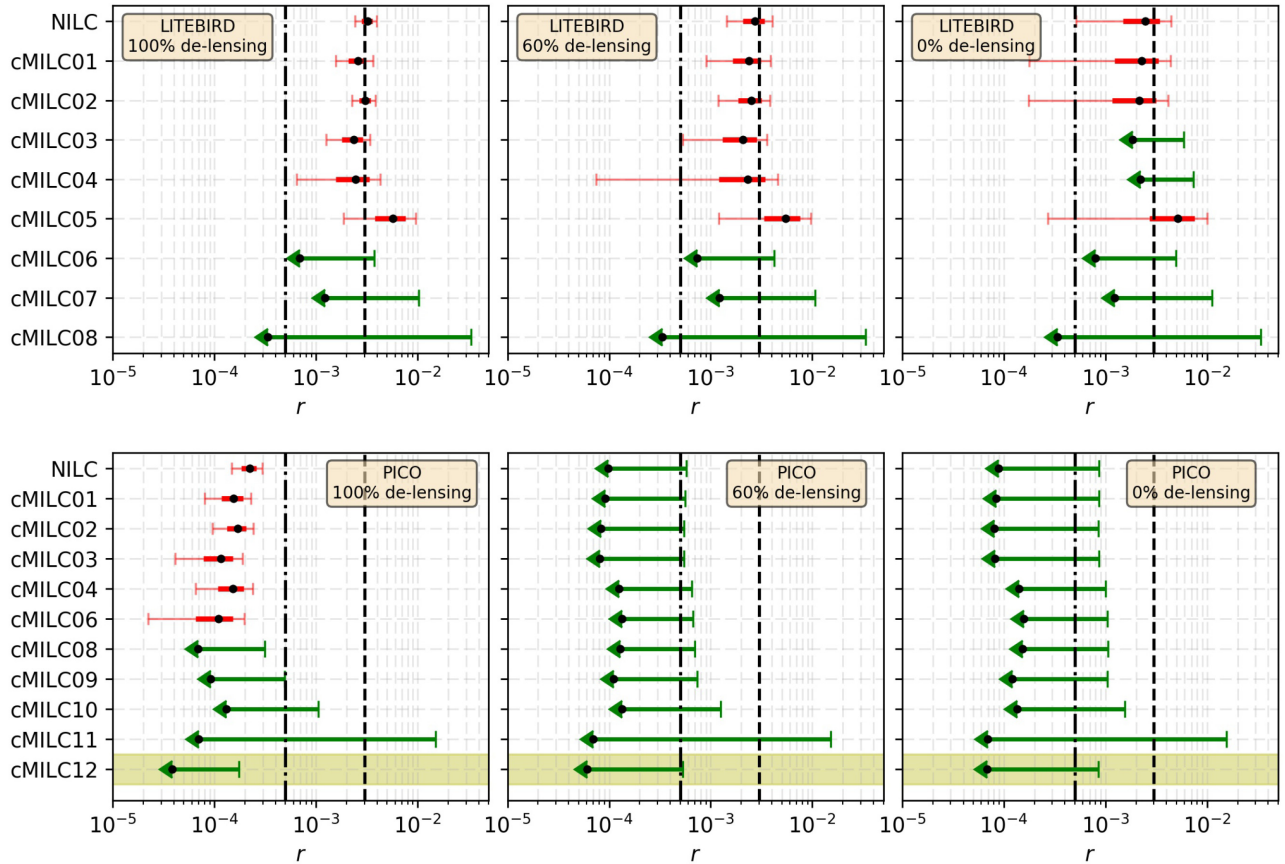
Figure 8. *Left-hand panels*: Bias and uncertainty on the recovered tensor-to-scalar ratio *r* for *LiteBIRD* (*upper panels*) and *PICO* (*lower panels*) after component separation on the `PySM` `d1s1` simulation with the `NILC` and `cMILC` methods, for different levels of residual lensing contamination. Biased detections at more than 2σ due to residual foregrounds are shown in red with 1σ (bold) and 2σ (thin) error bars, while unbiased detections consistent with $r = 0$ are shown in green with 2σ upper limits. The dashed vertical line marks $r = 3 \times 10^{-3}$ and the dash–dotted line $r = 5 \times 10^{-4}$. CMILC06 provides the best result on *r* for *LiteBIRD* (upper panel), by showing minimal bias due to foreground residuals, while not paying much noise penalty with respect to `NILC` in terms of 2σ upper limit. For *PICO* (lower panel), cMILC08 and cMILC12 provide the lowest biases on *r*, without paying much noise penalty with respect to `NILC`.

contributions of various emissivities, temperatures, and magnetic fields orientations along the line of sight, thus leading to some decorrelation of the polarized dust emission across frequencies. The moments of the effective foreground emission resulting from line-of-sight averaging effects can in principle be deprojected with `cMILC`. In this section, we thus investigate the performance of `cMILC` on sky simulations having non-trivial foreground complexity, by using publicly available MHD-model simulations[7] that were produced for the *PICO* data challenge.

In this sky simulation, the dust and synchrotron emissions are consistently derived from MHD simulations of the magnetic field turbulence in the interstellar medium (Kritsuk et al. 2018; Kim, Choi & Flauger 2019), with the dust model described in Hensley (2015). The MHD-model simulation is particularly interesting because of integrating multiple modified blackbodies of varying temperatures and spectral indices along the line of sight, so that the resulting dust SED in each pixel will not be a perfect modified blackbody, which makes this foreground model both more realistic and more challenging for parametric fitting methods aiming at *B*-mode component separation. The CMB and noise realizations in the MHD simulation are the same as those in the `d1s1` simulation.

[7]https://zzz.physics.umn.edu/ipsig/20180424_dc_maps (model 96).

Fig. 9 summarizes the `cMILC` results on the residual foregrounds and noise contamination to the CMB *B*-mode power spectrum for the MHD simulation (upper panels), together with the results on the recovered tensor-to-scalar ratio (lower panels). By deprojecting foregrounds moments with the semiblind method `cMILC`, we are able to tackle the foreground complexity arising from line-of-sight effects in the MHD simulation, and again reduce the residual foreground foreground contamination further than `NILC`. The best results on *r* for the MHD simulation are obtained with cMILC03 and cMILC06. For 60 per cent delensing, cMILC06 provides $r = (0.3 \pm 0.8) \times 10^{-4}$ consistent with $r = 0$, and a cumulation of systematic and statistical errors of $\delta r = \sqrt{r^2 + \sigma^2(r)} \lesssim 2.1 \times 10^{-4}$, which reduces down to $\delta r \lesssim 8.5 \times 10^{-5}$ in the case of full delensing.

Interestingly, deprojecting the first- and second-order moments of synchrotron in cMILC08, cMILC09, cMILC10, and cMILC11 tends to degrade the results for the MHD simulation. We interpret this behaviour as being caused by the way the MHD simulation has been built: while the dust emission has spectral variations along the line of sight and across the sky, the synchrotron index is uniform in the MHD simulation (Brandon S. Hensley, private communication). This implies that higher order moments of the synchrotron spectral index do not contribute in this simulation, and any constraints that nulls synchrotron components in `cMILC` will only increase the variance of the unconstrained foregrounds and noise. This also highlights
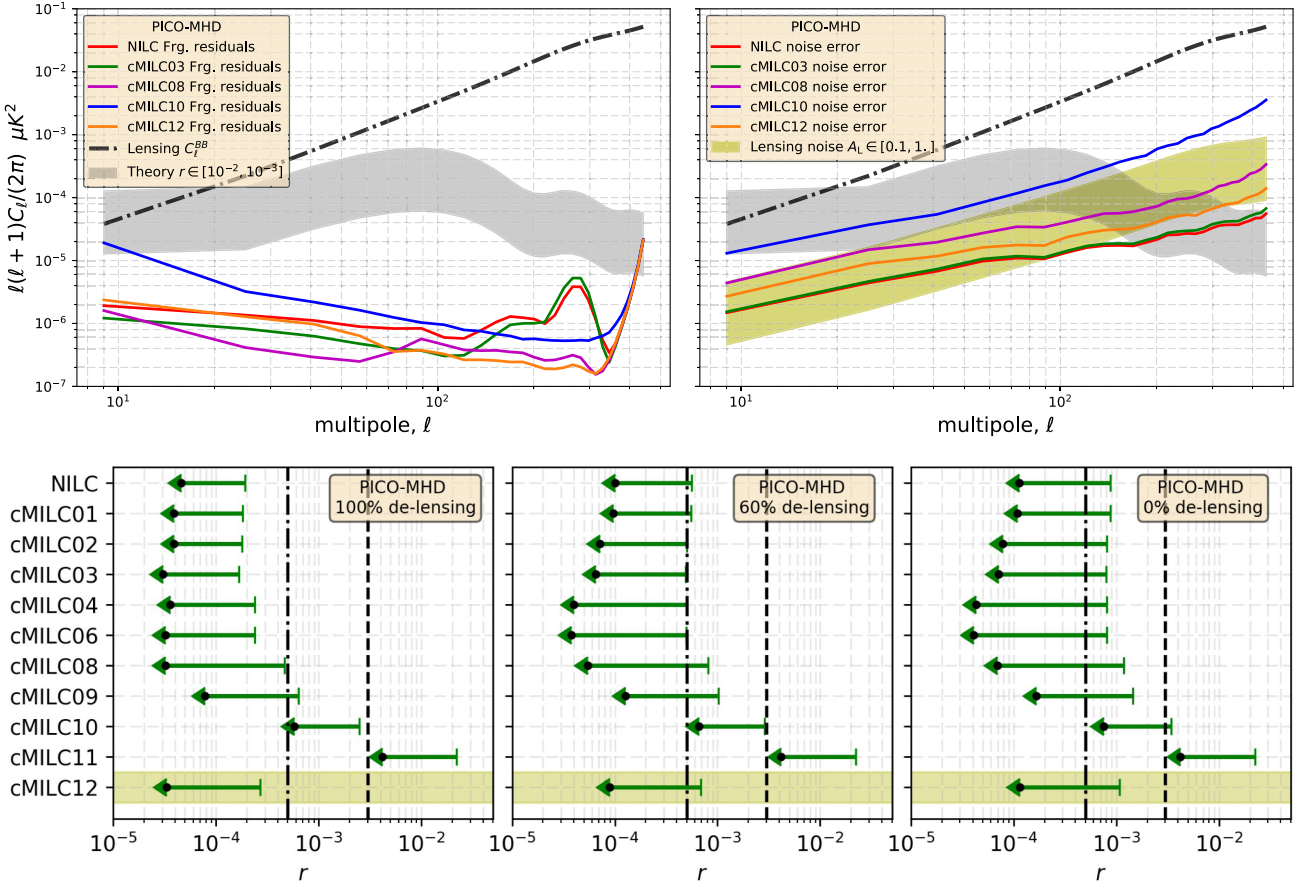
Figure 9. Idem Figs 7 and 8, but for the MHD foreground simulation with line-of-sight effects. By deprojecting moments arising from line-of-sight contributions, cMILC helps reducing biases on r without paying much noise penalty with respect to NILC in terms of 2σ upper limit. cMILC03, cMILC04, and cMILC06 provide the best results on r in terms of trade-off between residual bias and noise penalty.

that cMILC can provide a solid diagnosis of the level of spectral complexity of each foreground.

Overall, our analysis shows that cMILC is a quite robust component separation method, which without any extra modifications also allows to handle foregrounds complexity arising from line of sight and other averaging effects. In addition, the effect of unmodelled spectral complexity is successfully reduced by means of the blind variance minimization within cMILC.

5 DISCUSSION

While the previous sections demonstrated some general aspects of cMILC, reaching the full potential of the method and comparison to other methods requires more work. Here, we first discuss some possible optimizations of cMILC (Sections 5.1 and 5.2), and then briefly mention the relevance to ground-based CMB experiments (Section 5.3), and other observables and figures of merit (Section 5.4).

5.1 On the importance of optimal pivot parameters

Throughout our analysis, we have been using first-guess pivots $\{\overline{\beta}_s, \overline{\beta}_d, \overline{T}_d\}$ in cMILC (equation 30), based on the mean spectral index and temperature of dust and synchrotron issued from the current knowledge on intensity foregrounds (e.g. Planck Collaboration

X 2016a). In the language of moment expansion, these first-guess pivots from the literature typically result from fitting a zeroth-order model to the effective SED of the foreground emission:

$$I(\nu) = A_1 \, f_{\text{sync}}(\nu, \beta_s) + A_2 \, f_{\text{dust}}(\nu, \beta_d, T_d), \quad (51)$$

where the pivots, i.e. β_s, β_d, and T_d, are fit in each pixel and then averaged. As such, the zeroth-order model equation (51) might not provide the best fit at the largest angular scales, where averaging effects and higher order moment corrections become relevant to the SED. As pointed out above, in Fig. 7 for *PICO* d1s1 one can see that the performance of cMILC in reducing foreground contamination is spectacular on a large range of multipoles but non-optimal at the lowest multipoles $\ell \lesssim 15$, which are most relevant for constraining r. This suggests that first-guess pivots derived from traditional zeroth-order moment fits (equation 51) might not be optimal for cleaning foregrounds at the largest angular scales, where averaging effects, and thus higher order moments, become relevant. Another complication is that the optimal pivots will generally depend on the type of method that is employed, given that analysis choices vary but directly introduce averaging effects.

To investigate this point further, we thus modified the pivot values in the first needlet band to evaluate the impact on the performance of cMILC at low ℓ. In particular, we found that lowering the pivot value for the dust temperature down to $\overline{T}_d = 16\,\text{K}$ actually improves the performance of cMILC in cleaning foregrounds at low ℓ, and shows a
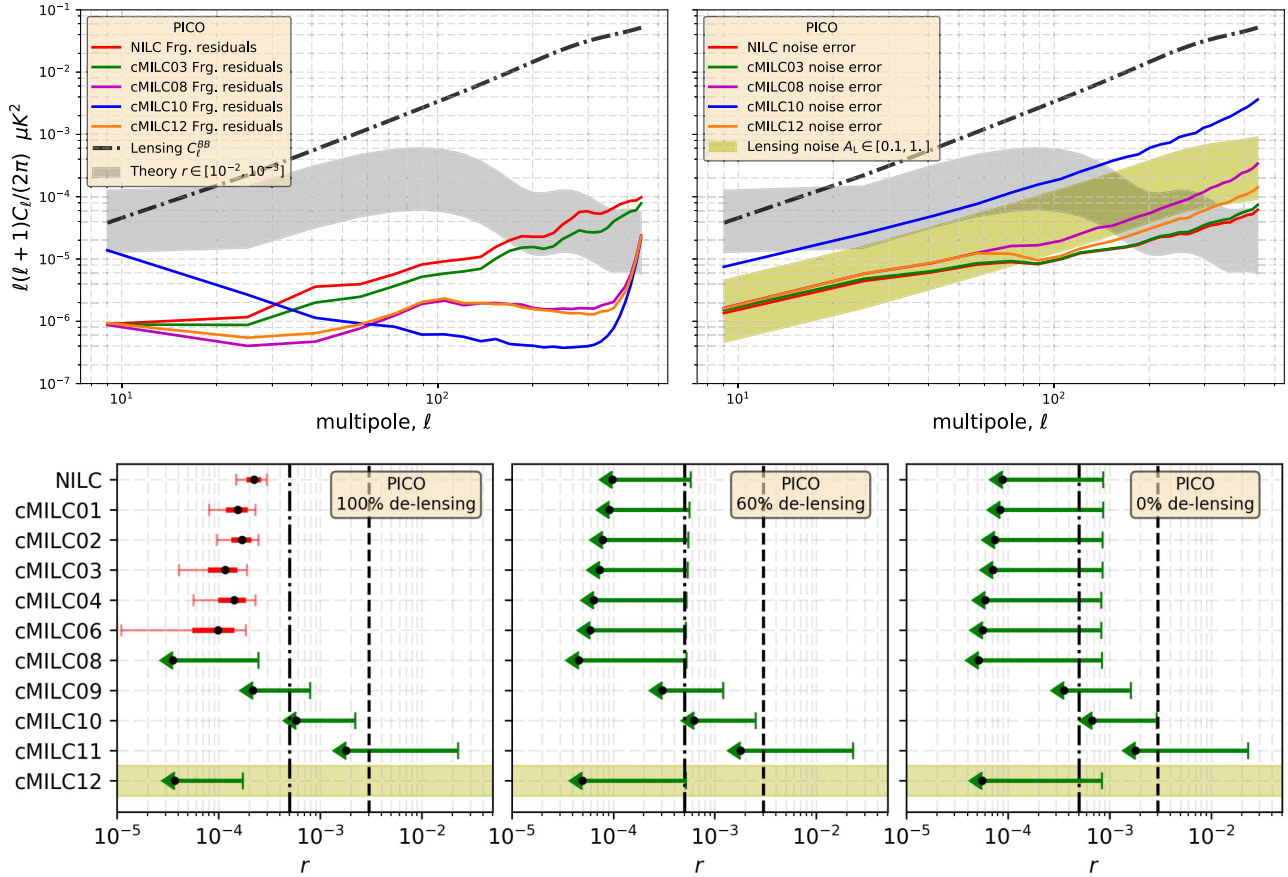
Figure 10. Update of Figs 7 and 8 for *PICO* d1s1 by adopting a different pivot temperature, $\overline{T}_d = 16\,$K, for cMILC at low ℓ. By updating the pivot at low ℓ, cMILC allows to reduce the residual foreground contamination across the whole range of multipoles, and thus progressively reduces the bias on r by deprojecting more and more moments, without paying much noise penalty.

more consistent picture of monotonic reduction of the bias on r when deprojecting more and more moments with cMILC. The results for *PICO* d1s1 simulation with the updated pivot temperature are shown in Fig. 10. A progressive reduction of the bias on r, without much noise penalty, is clear from cMILC01 to cMILC08, while it actually breaks for higher order constraints (cMILC09, cMILC10, cMILC11), suggesting that a proper optimization of *all* the pivots would probably be required instead of this simple ad hoc variation of the pivot temperature. Still, these results highlight how the performance of cMILC is sensitive to the choice of pivots at low ℓ, and the importance of choosing the most appropriate pivot parameters for the moment expansion. While a full optimization of the pivots using sophisticated parametric pivot fitting algorithms deserves an in-depth work that is beyond the scope of this paper, hereafter we highlight a few important aspects.

We attempted optimizing the pivots by extracting mean polarization SEDs from our simulation. However, for polarization the definition of an average SED is not as straightforward, given that no polarization 'monopole' exists. This makes it hard to define a concrete metric for the pivot optimization. This metric is also expected to directly depend on other analysis choices such as real-space versus harmonic-space analysis.

Besides, further optimization of cMILC would probably be possible by using local pivots across different regions of the sky. While here we have been using uniform pivot parameters across the sky throughout our analysis, the all-sky average temperature and spectral

index of the dust and synchrotron differ from the local mean index and temperature in a certain region of the sky, e.g. in the BICEP2 region. Therefore, a possible improvement of the cMILC method would be to adopt different pivot parameters in different sky regions and perform a local moment expansion. One open question is then how to optimally combine the results from the various regions to constraint on r.

Finally, further optimization of cMILC could potentially be achieved for the localization of filtering in both pixel space and harmonic space, through the investigation of different needlet windows in terms of number, shape, and width in order to determine whether the performance of the foreground cleaning would benefit from more localization of the filter in pixel space or more localization in harmonic space. All this outlines a high-dimensional optimization problem that will be carried out in the future.

5.2 Hybrid moments across multipole ranges

The relative contribution of the different moments of dust and synchrotron, e.g. $A_{\nu_s}(\beta_s - \overline{\beta}_s)\partial_{\beta_s} f_{\text{sync}}$ versus $A_{\nu_d}(\beta_d - \overline{\beta}_d)\partial_{\beta_d} f_{\text{dust}}$, to the overall foreground power spectrum is likely to vary across multipoles. In particular, we experienced that the first-order synchrotron index moment had more constraining power at low multipoles $\ell \leq 50$, while the first-order dust index moment was helping more at $\ell > 50$. Given that cMILC is performed on a needlet frame (see Fig. 3), it allows us to set different combinations of moment constraints

for different multipole windows in order to optimize the trade-off between residual foreground biases and noise degradation across multipoles.

To explore this idea further, we thus implemented a hybrid version cMILC12 (see Table 1), which deprojects the moments f_{sync}, f_{dust}, $\partial_\beta f_{\text{sync}}$, $\partial_T f_{\text{dust}}$ in the first two needlet bands and the moments f_{sync}, f_{dust}, $\partial_\beta f_{\text{dust}}$, $\partial_T f_{\text{dust}}$ at higher multipoles $\ell > 50$. As shown in Figs 8 and 10, the hybrid version cMILC12 provides the best trade-off and result of the analysis, with both the lowest bias on r due to residual foregrounds and the smallest statistical uncertainty due to low noise degradation: $r = (0.4 \pm 0.5) \times 10^{-4}$ consistent with an unbiased recovery of $r = 0$, and a cumulative systematic and statistical errors of $\delta r = \sqrt{r^2 + \sigma(r)^2} < 6 \times 10^{-5}$. cMILC is thus quite flexible by allowing for different combinations of moments across multipoles.

With future foreground observations, it will be possible to use cMILC for diagnostics of the most relevant moments of the foreground emission for each ranges of multipoles. Furthermore, given that the various moments of the non-Gaussian foreground contamination are spatially correlated with each other, it would in principle be possible to compress the number of moments into a small set of independent components ranked by their relevance to optimize foreground cleaning with cMILC. These ideas also have to be worked out more cleanly in the future.

5.3 Ground-based surveys

For future ground-based CMB surveys, such as SO (Ade et al. 2019) and CMB-S4 (Abazajian et al. 2016), it is more difficult to directly probe the reionization peak of the primordial CMB *B*-mode at the lowest multipoles $\ell \lesssim 15$ because only a fraction of the sky from the ground is accessible. Hence, future ground-based CMB experiments will have to rely mainly on information from the recombination peak (i.e. $30 \leq \ell \leq 200$) to constrain r. In addition, the number of available frequency bands for ground-based experiments is limited by atmospheric windows, which will limit the number of foreground moments that can be deprojected by cMILC.

However, as evident from the left-hand panels of Fig. 7, the deprojection of only two or three foreground moments with cMILC is already very successful around the recombination peak, with the reduction of the residual foreground contamination of cMILC in comparison to NILC being most significant at intermediate multipoles $30 \leq \ell \leq 200$. Therefore, cMILC would be particularly helpful also for future ground-based CMB surveys in removing residual foreground biases on r. With upcoming ground-based experiments like CCAT-prime (Aravena et al. 2019), which will provide wider frequency coverage owing to improved atmospheric transparency, we may anticipate significant benefits of using cMILC. However, dedicated forecasts are required taking differences in the frequency coverage and angular resolution into account.

5.4 Other observables and figures of merit

While in this work we focused on component separation for *B* modes, with r defining the figure of merit, cMILC is also directly applicable to CMB temperature and *E*-mode analyses relevant to other observables. As shown in Fig. 7, the potential of cMILC in removing foregrounds through moments significantly improves towards small angular scales. The details may change with further optimizations of cMILC, but generally we expect that cMILC could also help improve constraints on cosmological parameters such as N_{eff} and $\sum m_\nu$, which depend on damping tail physics

(e.g. Abazajian et al. 2015). Additional benefits may become apparent for studies of the Sunyaev–Zeldovich (SZ) power spectrum (e.g. Komatsu & Seljak 2002; Hill & Pajer 2013; Bolliet et al. 2020), which again rely on information gleaned from small-scale fluctuations.

Our method could also provide gains in future searches for primordial non-Gaussinity and CMB lensing/delensing analyses. As shown for the one-point statistics of foreground residuals in Fig. 6, cMILC allows significantly reducing *non-Gaussian* foreground contamination in the CMB map by deprojecting moments. Since CMB lensing (e.g. Hu & Okamoto 2002) and bispectrum estimators (e.g. Bucher, van Tent & Carvalho 2010) rely on the extraction of non-Gaussian features (no matter if primordial or generated by lensing), they are prone to non-Gaussian foreground residuals in the CMB map (e.g. van Engelen et al. 2014; Hill 2018), and thus to be potentially biased depending on the sensitivity of future experiments. This aspect deserves further investigation but is beyond the present scope.

6 CONCLUSIONS

For robust measurements of the tensor-to-scalar ratio r as a signature of primordial gravitational waves from inflation, it is essential to minimize both statistical and systematic errors due to residual foregrounds in component separation analyses. Statistical uncertainties $\sigma(r)$ arise from the cosmic/sample variance of the residual lensing, foreground, and noise contaminations after component separation and delensing, while systematic errors, leading to biases on r, arise from the residual foreground contamination to the CMB *B*-mode power spectrum after component separation. Hence, reliable foreground cleaning is achieved by minimizing cumulative systematic and statistical errors $\delta r = \sqrt{(r - r^{\text{true}})^2 + \sigma^2(r)}$.

In this work, we developed the new semiblind component separation method cMILC, which deprojects the main statistical moments of the foregrounds without altering the CMB *B*-mode signal, thus allowing to reduce systematic errors (or biases) on r due to residual foreground contamination, while still mitigating statistical uncertainties (see Figs 8–10). By applying cMILC to sky simulations with varying foreground complexity for experimental settings similar to those of *LiteBIRD* and *PICO*, we have identified specific combinations of foreground moments that optimize the trade-off between residual foreground biases and noise degradation for constraints on r. In particular, the cMILC method allows us to reach the sensitivity goals on r of both *LiteBIRD* and *PICO* mission concepts, by removing biases on r without excessive noise penalties, thus overcoming limitations of NILC.

We show that for LiteBIRD a deprojection of the first-order moment of the dust spectral index (Table 1: cMILC06) is needed for unbiased detection of $r = 0$ (see Fig. 8). In this case, the cumulative systematic and statistical errors (i.e. bias + variance) is $\delta r = \sqrt{r^2 + \sigma^2(r)} \lesssim 10^{-3}$, where the blind NILC method would exceed $\delta r \gtrsim 3 \times 10^{-3}$ due to more than 3σ residual foreground bias on $r = 0$. For *PICO*, deprojecting the first-order moments of the dust spectral index and dust temperature (Table 1: cMILC08) provides the best unbiased results on r without much noise penalty (see Figs 8 and 10). This yields $r = (0.7 \pm 0.9) \times 10^{-4}$, and hence a cumulative systematic and statistical error as low as $\delta r \simeq 10^{-4}$.

While our analysis shows that for constraints on r in principle only first moments are needed, *PICO* is further sensitive to the second moment of the dust temperature. This conclusion is expected to be a strong function of the complexity of the foreground simulation, and thus may not hold for more complicated skies. In the presence

of multiple dust modified blackbodies along the line of sight (Fig. 9 in Section 4.2.5), cMILC still performs well and helps reducing residual foreground contaminations and the bias on r without paying much noise penalty. Our analysis further shows that the optimization of pivots plays an important role (Section 5.1), requiring more investigation. These findings highlight the potential of cMILC as a robust foreground cleaning method but also diagnostic tool, enabling to test the sky complexity in future analysis.

The ability of cMILC to peel off foregrounds through the deprojection of their spectral moments is not limited to the search for primordial B modes and constraints on r, but could as well be exploited for CMB temperature and E-mode analyses to improve constraints on other cosmological parameters like $\sum m_\nu$ and N_{eff}. By getting rid of non-Gaussian residuals (Fig. 6 and Section 4.2.2), cMILC is also of great interest to searches for primordial non-Gaussianity and in the CMB lensing reconstruction. cMILC would also be helpful for the extraction of other faint cosmological signals such as SZ effects and anisotropic primordial spectral distortions. Forecasts for ground-based experiments should consider these additional observables in the analysis.

We also stress that moments of the foreground emission result not only from integrating multiple contributions along the line of sight, but also arise from several averaging processes across the sky like beam convolutions, spherical harmonic transforms, Q, U to E, B transforms, and other filtering processes (e.g. needlets). In fact, future CMB surveys will deliver sky maps of varying native beam resolutions at different frequencies. At low frequencies, the beams are typically larger than at high frequencies, and hence averaging processes are intrinsically larger at the map level. If not properly taken into account when combining maps, this will lead to inevitable decorrelation across frequencies. In addition, while most of the averaging effects can be propagated analytically, line-of-sight averaging is unavoidable, thus requiring a more general SED treatment even at the pixel level. All the above aspects pose a challenge for foreground modelling and the search for primordial B modes that cannot be ignored. Robustly deprojecting the moments of dust and synchrotron emissions that arise from averaging/decorrelation effects is thus crucial to avoid systematic errors from residual foreground contamination that will bias the faint primordial B-mode signal, hence the tensor-to-scalar ratio. The cMILC method provides an avenue forward this direction.

While here we specifically focused on a comparison to NILC, not only will it be important to extend this comparison to other methods but one should also think about augmenting other methods using moments. For instance, we anticipate that moment expansion approaches can help parametric component separation methods in the search for primordial B modes, providing a systematic way of extending the list of expected foreground parameters. The optimal method will also strongly depend on the observable under consideration, which further motivates more extensive comparisons of various cleaning methods in the future, ultimately preparing us for the analysis challenges in the years to come.

ACKNOWLEDGEMENTS

using the HEALPIX package (Górski et al. 2005) and the PSM package (Delabrouille et al. 2013).

DATA AVAILABILITY

REFERENCES

Abazajian K. N. et al., 2015, Astropart. Phys., 63, 66
Abazajian K. N. et al., 2016, preprint (arXiv:1610.02743)
Ade P. et al., 2019, J. Cosmol. Astropart. Phys., 2, 056
Alonso D., Dunkley J., Thorne B., Næss S., 2017, Phys. Rev. D, 95, 043504
Aravena M., et al., 2019, BAAS, 51, 213
Basak S., Delabrouille J., 2012, MNRAS, 419, 1163
Basak S., Delabrouille J., 2013, MNRAS, 435, 18
Basu K. et al., 2019, preprint (arXiv:1909.01592)
Baumann D. et al., 2009, in Dodelson S., Baumann D., Cooray A., Dunkley J., Fraisse A., Jackson M. G., Kogut A., Krauss L., Zaldarriaga M., Smith K., eds, AIP Conf. Proc. Vol. 1141, CMB Polarization Workshop: Theory and Foregrounds: CMBPol Mission Concept Study. Am. Inst. Phys., New York. p. 10
Bennett C. L. et al., 2003, ApJS, 148, 97
BICEP2 Collaboration, 2014, Phys. Rev. Lett., 112, 241101
BICEP2/Keck Collaboration, 2015, Phys. Rev. Lett., 114, 101301
Bolliet B., Brinckmann T., Chluba J., Lesgourgues J., 2020, MNRAS, 497, 1332
Bucher M., van Tent B., Carvalho C. S., 2010, MNRAS, 407, 2193
Cardoso J.-F., Le Jeune M., Delabrouille J., Betoule M., Patanchon G., 2008, IEEE J. Sel. Top. Signal Process., 2, 735
Carron J., Lewis A., Challinor A., 2017, J. Cosmol. Astropart. Phys., 2017, 035
Chluba J., Switzer E., Nelson K., Nagai D., 2013, MNRAS, 430, 3054
Chluba J., Hill J. C., Abitbol M. H., 2017, MNRAS, 472, 1195
Chluba J. et al., 2019, preprint (arXiv:1909.01593)
Delabrouille J., Cardoso J. F., Patanchon G., 2003, MNRAS, 346, 1089
Delabrouille J., Cardoso J.-F., Le Jeune M., Betoule M., Fay G., Guilloux F., 2009, A&A, 493, 835
Delabrouille J. et al., 2013, A&A, 553, A96
Delabrouille J. et al., 2019, preprint (arXiv:1909.01591)
Eriksen H. K., Banday A. J., Górski K. M., Lilje P. B., 2004, ApJ, 612, 633
Eriksen H. K., Jewell J. B., Dickinson C., Band ay A. J., Górski K. M., Lawrence C. R., 2008, ApJ, 676, 10
Errard J., Feeney S. M., Peiris H. V., Jaffe A. H., 2016, J. Cosmol. Astropart. Phys., 3, 052
Fernández-Cobos R., Vielva P., Barreiro R. B., Martínez-González E., 2012, MNRAS, 420, 2162
Flauger R., Hill J. C., Spergel D. N., 2014, J. Cosmol. Astropart. Phys., 2014, 039
Ghigna T., Matsumura T., Patanchon G., Ishino H., Hazumi M., 2020, J. Cosmol. Astropart. Phys., 2020, 030
Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Guilloux F., Faÿ G., Cardoso J.-F., 2009, Appl. Comput. Harmon. Anal., 26, 143
Hamimeche S., Lewis A., 2008, Phys. Rev. D, 77, 103013
Hanany S. et al., 2019, preprint (arXiv:1902.10541)
Hazumi M. et al., 2019, J. Low Temp. Phys., 194, 443
Hensley B. S., 2015, PhD thesis, Princeton University
Hensley B. S., Bull P., 2018, ApJ, 853, 127
Hill J. C., 2018, Phys. Rev. D, 98, 083542
Hill J. C., Pajer E., 2013, Phys. Rev. D, 88, 063526
Hivon E., Górski K. M., Netterfield C. B., Crill B. P., Prunet S., Hansen F., 2002, ApJ, 567, 2

Hu W., Okamoto T., 2002, ApJ, 574, 566

Hurier G., Macías-Pérez J. F., Hildebrandt S., 2013, A&A, 558, A118

Ichiki K., Kanai H., Katayama N., Komatsu E., 2019, Prog. Theor. Exp. Phys., 2019, 033E01

Kim C.-G., Choi S. K., Flauger R., 2019, ApJ, 880, 106

Knox L., Song Y.-S., 2002, Phys. Rev. Lett., 89, 011303

Kogut A. et al., 2007, ApJ, 665, 355

Komatsu E., Seljak U., 2002, MNRAS, 336, 1256

Krachmalnicoff N. et al., 2018, A&A, 618, A166

Kritsuk A. G., Flauger R., Ustyugov S. D., 2018, Phys. Rev. Lett., 121, 021104

Larsen P., Challinor A., Sherwin B. D., Mak D., 2016, Phys. Rev. Lett., 117, 151102

Mangilli A., Aumont J., Rotti A., Boulanger F., Chluba J., Ghosh T., Montier L., 2021, A&A, 647, A52

Martin J., Ringeval C., Vennin V., 2014, Phys. Dark Universe, 5, 75

Millea M., Anderes E., Wandelt B. D., 2019, Phys. Rev. D, 100, 023509

Miville-Deschênes M.-A., Ysard N., Lavabre A., Ponthieu N., Macías-Pérez J. F., Aumont J., Bernard J. P., 2008, A&A, 490, 1093

Narcowich F., Petrushev P., Ward J., 2006, SIAM J. Math. Anal., 38, 574

Planck Collaboration XI, 2014, A&A, 571, A11

Planck Collaboration X, 2016a, A&A, 594, A10

Planck Collaboration XXII, 2016b, A&A, 594, A22

Planck Collaboration Int. XLVIII, 2016c, A&A, 596, A109

Planck Collaboration IV, 2020a, A&A, 647, A4

Planck Collaboration VIII, 2020b, A&A, 647, A8

Remazeilles M., Chluba J., 2020, MNRAS, 494, 5734

Remazeilles M., Delabrouille J., Cardoso J.-F., 2011a, MNRAS, 410, 2481

Remazeilles M., Delabrouille J., Cardoso J.-F., 2011b, MNRAS, 418, 467

Remazeilles M., Dickinson C., Eriksen H. K. K., Wehus I. K., 2016, MNRAS, 458, 2032

Remazeilles M. et al., 2018, J. Cosmol. Astropart. Phys., 4, 023

Rotti A., Chluba J., 2021, MNRAS, 500, 976

Sherwin B. D., Schmittfull M., 2015, Phys. Rev. D, 92, 043005

Starobinsky A. A., 1983, Sov. Astron. Lett., 9, 302

Stolyarov V., Hobson M. P., Lasenby A. N., Barreiro R. B., 2005, MNRAS, 357, 145

Stompor R., Errard J., Poletti D., 2016, Phys. Rev. D, 94, 083526

Sugai H. et al., 2020, J. Low Temp. Phys., 199, 1107

Tassis K., Pavlidou V., 2015, MNRAS, 451, L90

Tegmark M., de Oliveira-Costa A., Hamilton A. J., 2003, Phys. Rev. D, 68, 123523

Thorne B., Dunkley J., Alonso D., Næss S., 2017, MNRAS, 469, 2821

van Engelen A., Bhattacharya S., Sehgal N., Holder G. P., Zahn O., Nagai D., 2014, ApJ, 786, 13

APPENDIX A: FLEXIBILITY OF CMILC CONCERNING PIVOT PARAMETER VALUES

By construction, cMILC allows for some flexibility in the assumed pivot values of the foreground parameters. For a set of assumed pivot values, the nulling constraints of cMILC (equation 30) enable to null any part of the foreground emission that projects on to the moments centred on these pivots, without altering the CMB signal due to the conservation constraint. Should there be any small departures in the data from the assumed pivot values then a part of the foreground contamination cannot be fully deprojected, but such unconstrained foreground contamination is anyway handled by blind variance minimization, like NILC would proceed. To illustrate the margin on the choice of pivot values, in Fig. A1 we show the performance of cMILC in terms of residual foreground contamination for different pivot values, as compared to NILC.
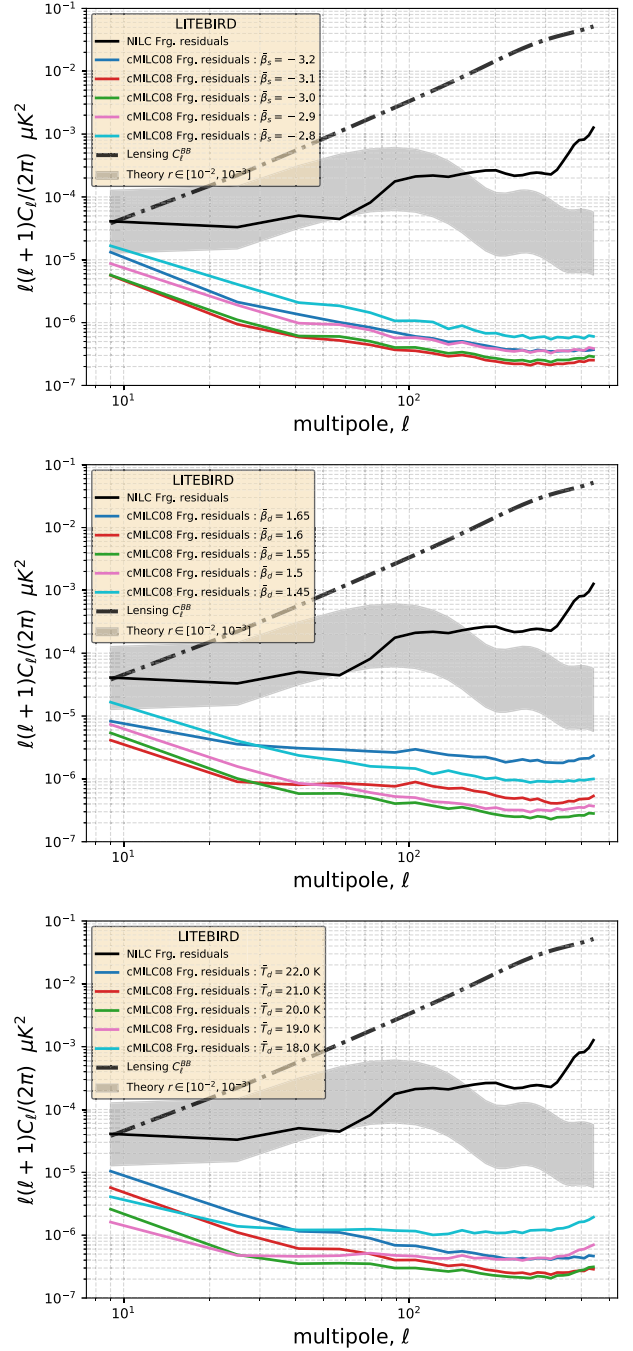

Figure A1. Performance of cMILC versus NILC in terms of residual foreground contamination when assuming different pivot values for $\overline{\beta}_s$, $\overline{\beta}_d$, and $\overline{T}_d$.

APPENDIX B: EXPERIMENTAL CONFIGURATIONS

Summary of experimental parameters for *LiteBIRD* (Table B1) and *PICO* (Table B2).

Table B1. Instrumental specifications of *LiteBIRD* (Hazumi et al. 2019) for the updated design (Sugai et al. 2020) with sensitivities taken from Ghigna et al. (2020).

Frequency [GHz]	Beam FWHM [arcmin]	Sensitivity [μK.arcmin]
40	69	39.76
50	56	25.76
60	48	20.69
68	43	12.72
78	39	10.39
89	35	8.95
100	29	6.43
119	25	4.30
140	23	4.43
166	21	4.86
195	20	5.44
235	19	9.72
280	24	12.91
337	20	19.07
402	17	43.53

Table B2. Instrumental specifications of *PICO* (Hanany et al. 2019) for the baseline design.

Frequency [GHz]	Beam FWHM [arcmin]	Sensitivity [μK.arcmin]
21	38.4	23.9
25	32.0	18.4
30	28.3	12.4
36	23.6	7.9
43	22.2	7.9
52	18.4	5.7
62	12.8	5.4
75	10.7	4.2
90	9.5	2.8
108	7.9	2.3
129	7.4	2.1
155	6.2	1.8
186	4.3	4.0
223	3.6	4.5
268	3.2	3.1
321	2.6	4.2
385	2.5	4.5
462	2.1	9.1
555	1.5	45.8
666	1.3	177
799	1.1	1050

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