Blind correction of the EB-leakage in the pixel domain

Hao Liu, James Creswell, Konstantina Dachlythra

The Niels Bohr Institute & Discovery Center, Blegdamsvej 17, DK-2100 Copenhagen, Denmark
Key Laboratory of Particle and Astrophysics, Institute of High Energy Physics, CAS, 19B YuQuan Road, Beijing, China

E-mail: liuhao@nbi.dk, james.creswell@nbi.ku.dk, kwn.dachlythra@gmail.com

Abstract.
We study the problem of EB-leakage that is associated with incomplete polarized CMB sky. In the blind case that assumes no additional information about the statistical properties and amplitudes of the signal from the missing sky region, we prove that the recycling method (Liu et al. 2018) gives the unique best estimate of the EB-leakage and reduces the error to a level of the tensor-to-scalar ratio $r = 10^{-4}$ to $10^{-3}$. Compared to the previous method, our method reduces the uncertainties in the BB power spectrum due to EB-leakage by more than one order of magnitude in the most interesting domain of multipoles, where $\ell$ is between 100 and 200. This work also provides a useful guideline for observational design of future CMB experiments.

Keywords: cosmic background radiation; cosmology: observations; gravitational waves
1 Introduction

The most promising way to detect primordial gravitational waves is by measuring the B-mode polarization of the Cosmic Microwave Background (CMB) radiation. This requires first separating the B-mode polarization from the dominant E-mode polarization. However, because the E/B decomposition is non-local, when only part of the sky is visible, the calculated B-mode is corrupted by power originating in the E-mode, which is called the E-to-B leakage \cite{1, 2}. Since neither current experiments nor those upcoming in the next few decades will be able to provide reliable full sky background data, the E-to-B leakage is a problem that must be solved before one can detect primordial gravitational waves. In this paper we show how to make the best blind estimate (BBE) of this E-to-B leakage using the data within the available sky region and the shape of this region.

Depending on the prior assumptions and the starting point, there are three types of estimations: the blind, prior, and posterior estimations. Blind estimation means there are no prior assumptions, while the other two are based on the context of data \(x\) described by a model with parameters \(\Theta\). Prior estimation means to make an estimate of the data from model parameters, and posterior estimation means to estimate the model parameters from the data. In practice, these two concepts are connected through Bayes’ theorem \(P(x|\Theta)P(\Theta) = P(\Theta|x)P(x)\), where \(P(x|\Theta)\) is the conditional probability of \(x\) with given \(\Theta\), and \(P(\Theta|x)\) is the conditional probability of \(\Theta\) with given \(x\). Although posterior estimation is only found in the context of parameter estimation because it targets model parameters, the idea of prior estimation can be extended to include a more general notion of estimation with prior constraints, and does not necessarily require a physical model.

Strictly speaking, posterior estimation is unsuitable for the EB-leakage problem, because the goal is the data \(x\), not model parameters \(\Theta\). Indirect usage of a posterior estimation might be possible, but several problems have to be solved in advance, which will be discussed in section 4. Besides the posterior estimation, we have two other options for the EB-leakage problem: either blind estimation or prior estimation. This paper focuses on the blind estimation, and the prior estimation will be studied in a future work.

Normally, the best estimation is defined to be the unbiased estimation with the smallest error. However, in the case of blind estimation of EB-leakage, where we allow no constraints on the unavailable sky region, the error of the estimation is completely undetermined and cannot be presented in the form of error bars. Therefore, the BBE is defined as follows:
Let \( S(p_1, \ldots, p_n, q_1, \ldots, q_m) \) be the real EB-leakage, where \( p_i \) are the available pixels, and \( q_i \) are the unavailable pixels. If \( S \) can be decomposed into

\[
S(p_1, \ldots, p_n, q_1, \ldots, q_m) = E(p_1, \ldots, p_n) + \Delta(q_1, \ldots, q_m) + \text{const},
\]

then \( E(p_1, \ldots, p_n) \) is the BBE, and \( \Delta(q_1, \ldots, q_m) \) is the error. With the Taylor series expansion of \( S \), one can easily prove that, if the decomposition in eq. (1.1) exists, then \( E \) is unique, thus the BBE is unique except for a constant offset. In this work, we not only give the mathematical form of \( E \), but also point out how to calculate it efficiently.

The definition of the BBE means that the error is a function only of the missing sky pixels, and the BBE depends on the available sky pixels in exactly the same way as the true leakage does. Therefore, any further improvement of the BBE requires additional information about the missing sky region.

In ref. [3], we introduced the background of the EB-leakage problem, and used the “recycling method” to correct the E-to-B leakage. The method is fast, simple, and provides much better results than previous corrections. In this work, we prove that the recycling method gives exactly the BBE of EB-leakage. Meanwhile, it should also be noted that the detection of the primordial B-mode is very complicated – as mentioned in ref. [3], there are at least five main obstacles: foreground removal, delensing, noise, systematics, and the \( EB \) leakage. A successful solution of the EB-leakage is just one step of the whole effort.

This paper is organized as follows: in section 2, we introduce the basis and notations, and prove that the recycling method gives the BBE of EB-leakage in section 3. The possibility of further correction with additional information is discussed in section 4, and the conclusion is given in section 5.

2 Basis and notations

We briefly review the calculation of \( E \)- and \( B \)-family maps by pixel domain convolution. More details can be found in refs. [4–6].

Given a polarized sky map \( P(n) = (Q(n), U(n)) \), the true \( E \)- and \( B \)-family maps are:

\[
\begin{align*}
P_E(n) &= \begin{pmatrix} Q_E \\ U_E \end{pmatrix}(n) = \int G_E(n, n') P(n') dn', \\
P_B(n) &= \begin{pmatrix} Q_B \\ U_B \end{pmatrix}(n) = \int G_B(n, n') P(n') dn',
\end{align*}
\]

with \( P_E(n) + P_B(n) = P(n) \) and:

\[
\begin{align*}
G_E(n, n') &= \begin{pmatrix} G_1 + G_2 \\ +G_3 & G_4 \end{pmatrix}(n, n'), \\
G_B(n, n') &= \begin{pmatrix} G_4 -G_3 \\ -G_2 & G_1 \end{pmatrix}(n, n').
\end{align*}
\]

Note that the true EB-leakage \( S \) is not unique with missing data, as described by \( \Delta \). Only \( E \) is unique.
The $G_{1-4}$ functions are defined as:

$$G_1(n, n') = \sum_{\ell,m} F_{+ \ell m}(n) F_{+ \ell m}^*(n'),$$

$$G_2(n, n') = \sum_{\ell,m} F_{+ \ell m}(n) F_{- \ell m}^*(n'),$$

$$G_3(n, n') = \sum_{\ell,m} F_{- \ell m}(n) F_{+ \ell m}^*(n'),$$

$$G_4(n, n') = \sum_{\ell,m} F_{- \ell m}(n) F_{- \ell m}^*(n'),$$

(2.3)

and the $F_{+, -}$ functions are defined in terms of the spin-2 spherical harmonics as:

$$F_{+ \ell m}(n) = -\frac{1}{2} \left[ 2 Y_{\ell m}(n) + 2 Y_{\ell m}(n) \right],$$

$$F_{- \ell m}(n) = -\frac{1}{2i} \left[ 2 Y_{\ell m}(n) - 2 Y_{\ell m}(n) \right].$$

(2.4)

Note that $G_4$ are real, $G_2 = G_3$, and $G_1 + G_4 = \delta$.

As an augmentation, the $G_E$ and $G_B$ kernels can be written in terms of a common kernel $G$ and a delta function:

$$G_E(n, n') = \frac{1}{2} \delta(n - n') + G(n, n'),$$

$$G_B(n, n') = \frac{1}{2} \delta(n - n') - G(n, n').$$

(2.5)

In practice, when applied to pixelized sky maps, the sums in eq. (2.3) are not taken to $\ell = \infty$ but instead to a finite $\ell_{\text{max}}$. In this case, the identities $G_2 = G_3$ and $G_1 + G_4 = \delta$ are broken and the delta functions in eq. (2.5) are replaced by “bandpassed” delta functions, which behave similarly but maintain the orthogonality of the $E$ and $B$ modes for any finite $\ell_{\text{max}}$.

For convenience, the operation of extracting $P_E(n)$ or $P_B(n)$ from $P(n)$ using eq. (2.1) is written in form of operators as follows:

$$\Psi_E(P) \Rightarrow P_E(n),$$

$$\Psi_B(P) \Rightarrow P_B(n).$$

(2.6)

3 Proof of best correction

In this section, we give the mathematical form of the BBE of EB-leakage, and point out how to calculate it efficiently in practice.

Given a sky mask $M(n)$, which takes values of 1 or 0 depending on whether $n$ is available or
not, and use the notation in eq. (2.6), the true EB-leakage is

\[
L(n)_{\text{true}} = \Psi_B(M \Psi_E(P)) = \int G_B(n, n') M(n') P_E(n') \, dn'
\]

\[
= \int G_B(n, n') M(n') \left[ \int G_E(n', n'') P(n'') \, dn'' \right] \, dn'
\]

\[
= \int P(n'') \, dn'' \int G_B(n, n') G_E(n', n'') M(n') \, dn'
\]

\[
= \int G_E(n, n'') P(n'') \, dn'',
\]

where

\[
G_E(n, n'') \equiv \int G_B(n, n') G_E(n', n'') M(n') \, dn'
\]

is the EB-leakage convolution kernel. It is fully determined by \( M(n') \). If \( M(n') = 1 \) (no mask), then \( G_E(n, n'') = 0 \).

Given a mask, the EB-leakage is affected by outside-to-inside propagation and vice versa, which makes the estimation complicated. Eq. (3.1) gives the leakage as a convolution of the kernel \( G_E(n, n'') \) and the input sky map \( P(n'') \). The kernel describes the full propagation effect without involving the particular sky map. This clear separation makes the following study much easier: when a mask is present and given no additional information of the missing sky region, all available information is fully described by \( M(n'') P(n'') \), hence the BBE of EB-leakage is

\[
L(n)_{\text{best}} = \int G_E(n, n'') M(n'') P(n'') \, dn''.
\]

Eq. (3.3) can also be understood through the error of \( L(n)_{\text{best}} \):

\[
L(n)_{\text{error}} = L(n)_{\text{true}} - L(n)_{\text{best}} = \int G_E(n, n'') [1 - M(n'')] P(n'') \, dn'',
\]

which is 100% determined by the missing sky region. Therefore, without additional information about the missing sky region, it is impossible to reduce \( L(n)_{\text{error}} \). This fully satisfies the definition of the BBE in section 1.

Eqs. (3.1) and (3.3) are difficult to calculate directly. However, using the recycling method from ref. [3] it is possible to calculate \( L(n)_{\text{best}} \) without the computationally expensive kernel \( G_E(n, n'') \):

\[
L(n)_{\text{template}} = \int G_B(n, n') M(n') P_E(n') \, dn'
\]

\[
= \int G_B(n, n') M(n') \left[ \int G_E(n', n'') P(n'') \, dn'' \right] \, dn'
\]

\[
= \int M(n'') P(n'') \, dn'' \int G_B(n, n') G_E(n', n'') M(n') \, dn'
\]

\[
= \int G_E(n, n'') P(n'') \, dn''
\]

\[
= L(n)_{\text{best}}.
\]
Because eq. (3.3) and eq. (3.5) are identical, the template from the recycling method is exactly the BBE of EB-leakage. As shown by eq. (3.5), \( L(n)_{\text{best}} \) can be easily calculated by two steps: 1) Apply the mask to \( P(n) \) and get \( P'_E \). 2) Apply the mask to \( P'_E \) and get \( P''_B \) as the leakage estimate. Using the notations in eq. (2.6), the final form of the BBE is the following:

\[
L(n)_{\text{best}} = \Psi_B(M\Psi_E(MP)).
\] (3.6)

4  Possibility of further improvement with additional information

All the above analysis is blind and makes no assumptions about the missing sky signal, not even Gaussianity or isotropy. In this case, the recycling method gives the BBE of EB-leakage.

Given additional information, it might be possible to partly reconstruct the missing sky region, e.g., using lossless Fisher estimators [7], as was done in ref. [8] for the temperature case. If this can be done properly for polarized maps, then the EB-leakage estimation can certainly be improved. However, there is an important constraint that was repeatedly mentioned in refs. [7, 8] and other works, that both Gaussianity and isotropy have to be assumed for current Fisher estimators, because only then can the statistical properties of the covariance matrix be fully determined by the angular power spectrum.

Unfortunately, the EB-leakage is highly non-isotropic, and therefore cannot be estimated using current Fisher estimators. It could be possible to redesign the Fisher estimator and remove the requirement for isotropy. However, several difficulties must be solved: for example, the covariance matrix is non-analytic and is always singular (due to the missing region), and the map of a single component (like the EB-leakage alone) is not available at the beginning of estimation. Alternatively, instead of the posterior Fisher estimator, one could use a prior estimator that incorporates given prior information. This approach will be investigated in future work. However, some simple assumptions can be made that give minor, though immediate, improvements, such as assuming that \( L(n)_{\text{best}} \) is uncorrelated with \( P_B(n) \), in which case the correction can be slightly improved by removing the template using linear fitting. This was adopted in ref. [3].

5  Conclusion and discussion

In this work, by proving that the recycling method gives the BBE of EB-leakage, the problem of EB-leakage is completely solved in the blind case. To illustrate the correction method, we run a test that is similar to the one in figure 5 of ref. [3], differing only in that we skip the linear fitting procedure (see section 4) to make the estimation completely blind. In the test, we calculate the EB-leakage correction using either our recycling method or the PURE-method [2, 9–14]. The PURE method is also blind and was previously the best one. In both cases, the full sky B-mode spectrum is reconstructed using the MASTER method [15] and the NMASTER code [16, 17]. The results are shown in figure 1, which is almost the same as figure 5 of ref. [3]: the correction is 1–2 orders of magnitude better than the PURE method, and in the most important multiple range for detecting primordial gravitational waves, i.e. \( 80 \leq \ell \leq 200 \), our result is good enough to detect \( r \approx 10^{-4} \), which is sufficient for the next few decades. Comparison with figure 5 of ref. [3] shows that the linear fitting used in ref. [3] helps to improve the result by about 40%. This nicely illustrates the final option mentioned in section 4: the improvement is not big, but still good because it costs almost nothing.

As mentioned in ref. [3], the five main obstacles in detection of CMB B-modes are foreground removal, delensing, noise, systematics, and the EB-leakage. In the blind case, this work reduces the list to four, which is a solid step towards real detection of the primordial gravitational waves.
Figure 1. Comparison of the errors of EB leakage correction: red for MASTER+PURE and blue for MASTER+our method. The primordial $B$-mode spectra for $r = 10^{-2} \sim 10^{-4}$ (black solid) and the lensing $B$-mode spectrum (black dashed) are also added for comparison. Everything is done under the same conditions (resolution, simulated maps, sky region, apodization, etc.). This is similar to figure 5 of ref. [3], with only one difference that the linear fitting procedure is skipped to make the estimation completely blind.

This work is also relevant for the design of future CMB experiments: if there is no additional information about the missing region, then the ability to detect $r$ in a certain observation region is limited by the BBE of EB-leakage for that region. Therefore it is unnecessary to increase the detector sensitivity to an extent that significantly exceeds this limit. In order to be sensitive to even lower $r$, it would be necessary to enlarge the region of observation, which will probably also require longer observing time and more accurate foreground removal.

Acknowledgments

We sincerely thank Sebastian von Hausegger and Pavel Naselsky for very helpful discussions. This research has made use of the HEALPix [18] package and the NAMASTER/PYMASTER package [16, 17], and was partially funded by the Danish National Research Foundation (DNRF) through establishment of the Discovery Center and the Villum Fonden through the Deep Space project. Hao Liu is also supported by the Youth Innovation Promotion Association, CAS.
Figure 2. Examples of the EB-convolution kernel $G_{EB}(n_0, n)$. Cross for $n_0$ (fixed in each example) and circle for the edge of the mask. Upper: $U = 0$. Lower: $Q = 0$. $N_{side} = 32$ and $\ell_{max} = 16$.

A Examples of the EB convolution kernel

The convolution kernel $G_{EB}(n, n'')$ can be calculated from eq. (3.2). Compared to direct calculation (which is quite difficult), a more convenient way is to do it from eq. (3.1) using a Dirac delta function:

$$
G_{EB}(n, n_0) = L(n)_{true, n_0} = 
\int G_B(n, n') M(n') \left[ \int G_E(n', n'') P(n'') \delta(n'' - n_0) \, dn'' \right] \, dn',
$$

(A.1)

In practice, eq. (A.1) means to obtain $G_{EB}(n, n_0)$ as follows:

1. Start from a zero map and set $Q(n_0)$ or $U(n_0)$ to 1.
2. Calculate $P_E$ without mask.
3. Calculate $P_B$ from the output of step 2 with a mask.

Steps 1–3 give $G_{EB}(n, n_0)$. However, $G_{EB}(n_0, n)$ is easier to understand, because the EB-leakage at $n_0$ (point of interest) is simply

$$
L(n_0)_{true} = \int G_{EB}(n_0, n) P(n) \, dn.
$$

(A.2)

This can be done by repeating steps 1–3 for all possible $n_0$, and deriving $G_{EB}(n_0, n)$ from the results. The calculation is time consuming, and here we present the results only for low resolution.

In figure 2, we show examples of $G_{EB}(n_0, n)$ for $N_{side} = 32$ and $\ell_{max} = 16$. The mask is a $r = 20^\circ$ disk mask located in the center of the map. One can see that when $n_0$ (cross) is in the middle of the mask, the kernel is relatively much weaker than when $n_0$ is at the edge, which is consistent to the well-known fact that the center region contains much less EB-leakage than the edge region.
References

[1] A. Lewis, A. Challinor and N. Turok, *Analysis of CMB polarization on an incomplete sky*, Phys. Rev. D 65 (Jan., 2002) 023505, [astro-ph/0106536].

[2] E. F. Bunn, M. Zaldarriaga, M. Tegmark and A. de Oliveira-Costa, *E/B decomposition of finite pixelized CMB maps*, Phys. Rev. D67 (2003) 023501, [astro-ph/0207338].

[3] H. Liu, J. Creswell, S. von Hausegger and P. Naselsky, *Methods for pixel domain correction of EB leakage*, arXiv e-prints (Nov, 2018) arXiv:1811.04691, [1811.04691].

[4] H. Liu, J. Creswell and P. Naselsky, *E and B families of the Stokes parameters in the polarized synchrotron and thermal dust foregrounds*, JCAP 5 (May, 2018) 059, [1804.10382].

[5] A. Rotti and K. Huffenberger, *Real-space computation of E/B-mode maps. Part I. Formalism, compact kernels, and polarized filaments*, JCAP 1901 (2019) 045, [1807.11940].

[6] H. Liu, *Fingerprint of Galactic Loop I on polarized microwave foregrounds*, Astr. Astrophys. 617 (Sept., 2018) A90, [1806.06532].

[7] M. Tegmark, *How to measure CMB power spectra without losing information*, Phys. Rev. D 55 (May, 1997) 5895–5907, [astro-ph/9611174].

[8] G. Efstathiou, Y.-Z. Ma and D. Hanson, *Large-angle correlations in the cosmic microwave background*, Mon. Not. R. Astr. Soc. 407 (Oct., 2010) 2530–2542, [0911.5399].

[9] A. Lewis, *Harmonic E/B decomposition for CMB polarization maps*, Phys. Rev. D 68 (Oct., 2003) 083509, [astro-ph/0305545].

[10] E. F. Bunn, *Separating E from B*, New Astronomy Reviews 47 (Dec., 2003) 987–994, [astro-ph/0306003].

[11] M. Zaldarriaga, *The Polarization of the Cosmic Microwave Background, Measuring and Modeling the Universe* (2004) 309, [astro-ph/0305272].

[12] W. Zhao and D. Baskaran, *Separating e and b types of polarization on an incomplete sky*, Phys. Rev. D 82 (Jul, 2010) 023001.

[13] K. M. Smith, *Pseudo-C estimators which do not mix E and B modes*, Phys. Rev. D 74 (Oct., 2006) 083002, [astro-ph/0511629].

[14] K. M. Smith and M. Zaldarriaga, *General solution to the E-B mixing problem*, Phys. Rev. D 76 (Aug., 2007) 043001, [astro-ph/0610059].

[15] E. Hivon, K. M. Górski, C. B. Netterfield, B. P. Crill, S. Prunet and F. Hansen, *MASTER of the Cosmic Microwave Background Anisotropy Power Spectrum: A Fast Method for Statistical Analysis of Large and Complex Cosmic Microwave Background Data Sets*, Astrophys. J. 567 (Mar., 2002) 2–17, [astro-ph/0105302].

[16] LSST DARK ENERGY SCIENCE collaboration, D. Alonso, J. Sanchez and A. Slosar, *A unified pseudo-Cℓ framework*, 1809.09603.

[17] D. Alonso, J. Sanchez and A. Slosar, “Namaster: A unified pseudo-cl framework.” https://github.com/LSSTDESC/NaMaster.

[18] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke et al., *HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere*, Astrophys. J. 622 (Apr., 2005) 759–771, [astro-ph/0409513].