CMB anisotropy power spectrum using linear combinations of WMAP maps

Rajib Saha, Simon Prunet, Pankaj Jain, Tarun Souradeep

CMB anisotropy power spectrum using linear combinations of WMAP maps

Rajib Saha,[1,2,3] Simon Prunet,[3] Pankaj Jain,[1] and Tarun Souradeep[2]
[1]*Department of Physics, Indian Institute of Technology, Kanpur, U.P, 208016, India*
[2]*Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411007, India*
[3]*Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, F-75014 Paris, France*
(Received 28 June 2007; published 3 July 2008)

In recent years the goal of estimating different cosmological parameters precisely has set new challenges in the effort to accurately measure the angular power spectrum of the CMB. This has required removal of foreground contamination as well as detector noise bias with reliability and precision. Recently, a novel, *model-independent* method for the estimation of the CMB angular power spectrum solely from multifrequency observations has been proposed and implemented on the first year WMAP data by Saha *et al.* 2006. All previous estimates of the power spectrum of the CMB are based upon foreground templates using data sets from different experiments. However, our methodology demonstrates that the *CMB angular spectrum can be reliably estimated with precision from a self-contained analysis of the WMAP data*. In this work we provide a detailed description of this method. We also study and identify the biases present in our power spectrum estimate. We apply our methodology to extract the power spectrum from the WMAP data.

DOI: 10.1103/PhysRevD.78.023003 PACS numbers: 98.70.Vc, 07.05.Kf, 98.80.Es

I. INTRODUCTION

Starting from the end of the last millennium, remarkable progress in cosmology has been made by the precise measurements of the anisotropies in the CMB from different ground based as well as satellite observations [1–4]. In the 1-year and 3-year data release, the WMAP science team has provided the science community with a large amount of high quality CMB anisotropy data sets measured by its 10 independent differencing assemblies (DA) [5–8]. However, extracting the primordial CMB signal from these large data sets is a nontrivial task. The anisotropies in the CMB are weak in comparison to those originating due to radiation in the local universe, which inevitably contaminate the observed signal. These dominant foreground emissions are from within the Milky Way as well as from extragalactic point sources [9]. In the low frequency microwave regime the strongest contamination comes from the synchrotron and free-free emission [5,10]. At higher frequencies, where synchrotron and free-free emissions are low, dust emission dominates. A reliable extraction of the CMB signal from the multicomponent foreground contaminated data is thus complicated. There exist several methods in the literature to remove foregrounds using foreground tracer templates [10–12] built from observations from other experiments. However, this requires a prior model of (joint) spatial and spectral dependence for all the foreground components. The effect of uncertainties in the foreground models to estimate the CMB anisotropies have been discussed in [13,14]. Detector noise is another important concern that has to be addressed in order to precisely measure the angular power spectrum. The angular power spectrum from detector noise is dominant over the CMB power spectrum at small angular scales. The detector noise, being a random quantity, is treated differently from foreground contamination which is treated here as fixed templates on the sky. However, the noise is uncorrelated between different DAs [10,12,15,16]. The WMAP science team used this property to remove detector noise bias in a cross-correlated power spectrum obtained from two different DAs [10,12] using DA maps at frequencies 41 GHz, 61 GHz, and 94 GHz.

An interesting model-independent method to remove foregrounds from the multifrequency observations of the CMB has been proposed in [17] and implemented in [18] in order to extract the CMB signal from the WMAP data. The foreground emissions were removed by exploiting the fact that their contributions in different spectral bands are considerably different, while the CMB anisotropy power spectrum is the same in all the bands in units of thermodynamic temperature due to the Planck blackbody energy spectrum of the CMB [19,20]. The main advantage of this foreground cleaning method is that it is totally free from any assumption about foreground modeling. Another advantage is that it is computationally fast. However, the auto-power spectrum obtained from a single cleaned map as reported in [18] is not directly usable for primordial power spectrum estimation at small angular scales. This is because the detector noise bias dominates over the CMB power spectrum at small angular scales, beyond the beam width of the WMAP detectors.

In earlier publications [21–23] we extended this model-independent foreground removal method to remove detector noise bias. Here we describe the basic formalism of our previous work in detail, while also performing a detailed study of the nature of bias in the cleaned power spectrum. We apply our method to the WMAP 3-year data to estimate the CMB angular power spectrum.

TABLE I.

Symbols	Definitions
C_l^s	Theoretical CMB angular power spectrum
$\hat{C}_l^s$	An estimator of C_l^s
$B(\hat{n}\cdot\hat{n}')$	Circularly symmetric beam function
B_l	Legendre transform of $B(\hat{n}\cdot\hat{n}')$
$\hat{w}_l^a$	Weight for channel a and multipole l
$\hat{\mathbf{C}}_l$	Empirical covariance matrix for the CMB, foreground, and detector noise
$\hat{\mathbf{C}}_l^{f+N}$	Empirical covariance matrix for foreground and detector noise
$\hat{\mathbf{C}}_l^N$	Empirical detector noise covariance matrix
$\mathbf{C}_l^N$	Theoretical detector noise covariance matrix
$\mathbf{C}_l^f$	Foreground covariance matrix
$\mathbf{C}_l^{\mathrm{ps}}$	Unresolved point source covariance matrix
$\hat{C}_l^{(sp)0}$	A chance correlation between the CMB and the pth foreground at some reference frequency ν_0
$\hat{\mathbf{f}}_l^{p0}$	Shape vector of the pth foreground component scaled by $\hat{C}_l^{sf(p)0}$
$\hat{\mathbf{f}}_l^0$	$\sum_{p=1}^{n_f}\hat{\mathbf{f}}_l^{p0}$
$\mathbf{A}^+$	Moore-Penrose generalized inverse of a matrix $\mathbf{A}$
n_c	Number of channels linearly combined
n_f	Number of foreground components
$\mathcal{C}(\mathbf{A})$	Column space (range) of matrix $\mathbf{A}$
Null $(\mathbf{A})$	Null space of matrix $\mathbf{A}$

The plan of this paper is as follows. We describe the foreground removal methodology in Sec. II. The cross-power spectrum estimation methodology is described in Sec. III. In Sec. III B we describe the method of estimating the residual, unresolved, point source power spectrum. Section IV deals with the estimation of error bars on the final power spectrum. In Sec. V we analytically derive the bias in the cleaned power spectrum. The results are described in Sec. VI, and finally we conclude in Sec. VII.

In this paper we use indices a, b, c, d to denote frequency channels, and the index n to denote a sky region or a mask. A foreground component is denoted by the index p. We use bold faced roman indices $\mathbf{i}$, $\mathbf{j}$ to denote cleaned maps. The multipole moments are represented by l or l'. The Greek letter α is used to denote the index for the binned power spectrum. We denote matrices and vectors by bold faced letters. Any variable (scalar, vector, or matrix) which depends on the stochastic component is denoted by a hat on the top. As an example, we note that the CMB power spectrum is a stochastic variable which is represented as $\hat{C}_l^s$. In Table I we define commonly used symbols in this article.

II. FOREGROUND REMOVAL

In this section we outline the mathematical formalism to obtain the foreground-cleaned map and describe its implementation on the WMAP data.

A. Formalism

The observed signal at frequency channel a in a differential telescope like WMAP can be modeled as

$$\Delta\tilde{T}^a(\hat{n}) = \int(\Delta T^{\mathrm{s}}(\hat{n}') + \Delta T^{\mathrm{f}a}(\hat{n}'))B^a(\hat{n}\cdot\hat{n}')d\hat{n}' + \Delta T^{\mathrm{n}a}(\hat{n}). \quad (1)$$

Here $\Delta T^{\mathrm{s}}(\hat{n})$ and $\Delta T^{\mathrm{f}a}(\hat{n})$ are, respectively, the CMB and foreground component of the anisotropy in the channel a. We note that the CMB component does not contain any index, as this component is independent of frequency. The detector noise in channel a is $\Delta T^{\mathrm{n}a}$. The beam function $B^a(\hat{n}\cdot\hat{n}')$ represents the smoothing of the map due to finite resolution of the antenna of channel, a. The beam is assumed to be circularly symmetric, as done in most analyses. We note that the detector noise is not affected by the beam function. An experiment such as WMAP provides multifrequency maps $\Delta\tilde{T}^a(\hat{n})$, $a = 1, 2, \ldots, n_c$, corresponding to observations of the CMB at n_c different frequency bands. Equivalently, in the spherical harmonic representation,

$$a_{lm}^a = (a_{lm}^{\mathrm{s}} + a_{lm}^{\mathrm{f}a})B_l^a + a_{lm}^{\mathrm{n}a}, \quad (2)$$

where a_{lm}^a, a_{lm}^{s}, and $a_{lm}^{\mathrm{f}a}$ are, respectively, the spherical harmonic transforms of $\Delta\tilde{T}^a(\hat{n})$, $\Delta T^{\mathrm{s}}(\hat{n})$, and $\Delta T^{\mathrm{f}a}(\hat{n})$, and B_l^a's are the Legendre transform coefficients of the beam $B^a(\hat{n}\cdot\hat{n}')$. Our aim is to linearly combine the maps with appropriate weights to get an optimal estimator of the CMB anisotropy $\Delta T^{\mathrm{s}}(\hat{n}')$ that minimizes the contribution from the foregrounds. The linear combination of the multifrequency maps can be carried out in pixel space, or equivalently, in terms of the spherical harmonic coefficients. The former has been followed by the WMAP team to produce its Internal Linear Combination (ILC) map, and also, in a related, but more elaborate approach,

in [24] to produce the Lagrange Internal Linear Combination map. The approach of carrying out a multi-frequency map combination in the spherical harmonic space was proposed in [17]. For the first year of WMAP data this was implemented in [18]. This method has the advantage that we can simultaneously take into account variation of foreground with sky positions and with different multipoles for a given sky position.

We define a cleaned map as a linearly weighted sum of the maps at different frequencies,

$$a_{lm}^{\rm Clean} = \sum_{a=1}^{n_c} \hat{w}_l^a \frac{a_{lm}^a}{B_l^a}. \tag{3}$$

Here $\hat{w}_l^a$ is a weight factor which depends upon the multipole l and the frequency channel a. Since each of the channels has a different beam resolution, the maps are deconvolved by the corresponding circularly symmetric beam transform functions B_l prior to the linear combination. The total power in the cleaned map at a given multipole l is then

$$\hat{C}_l^{\rm Clean} = \frac{1}{2l+1} \sum_{m=-l}^{m=l} a_{lm}^{\rm Clean} a_{lm}^{\rm Clean*}. \tag{4}$$

Substituting Eq. (3) in Eq. (4), we obtain

$$\hat{C}_l^{\rm Clean} = \hat{\mathbf{W}}_l \hat{\mathbf{C}}_l \hat{\mathbf{W}}_l^T, \tag{5}$$

where the matrix $\hat{\mathbf{C}}_l$ is given by

$$\hat{\mathbf{C}}_l = \begin{pmatrix} \frac{\hat{C}_l^{11}}{B_l^1 B_l^1} & .. & .. & \frac{\hat{C}_l^{1n_c}}{B_l^1 B_l^{n_c}} \\ .. & .. & .. & .. \\ .. & .. & .. & .. \\ \frac{\hat{C}_l^{n_c 1}}{B_l^{n_c} B_l^1} & .. & .. & \frac{\hat{C}_l^{n_c n_c}}{B_l^{n_c} B_l^{n_c}} \end{pmatrix}, \tag{6}$$

$\hat{\mathbf{W}}_l$ is a row vector describing the weights for different channels,

$$\hat{\mathbf{W}}_l = (\hat{w}_l^1, \hat{w}_l^2, , \ldots, \hat{w}_l^{n_c}), \tag{7}$$

and $\hat{C}_l^{ab}$ is the cross-power spectrum between the ath and bth channels,

$$\hat{C}_l^{ab} = \sum_{m=-l}^{m=l} \frac{a_{lm}^a a_{lm}^{b*}}{2l+1} = \frac{a_{l0}^a a_{l0}^b}{2l+1} + 2{\rm Re} \sum_{m=1}^{m=l} \frac{a_{lm}^a a_{lm}^{b*}}{2l+1}. \tag{8}$$

By construction, the matrix $\hat{\mathbf{C}}_l$ is symmetric. If we denote the total non-CMB contamination at channel a by $\bar{a}_{lm}^a = B_l^a a_{lm}^{fa} + a_{lm}^{na}$, then using Eq. (2) we can simplify the expression for $\frac{\hat{C}_l^{ab}}{B_l^a B_l^b}$ as

$$\frac{\hat{C}_l^{ab}}{B_l^a B_l^b} = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \frac{(B_l^a a_{lm}^s + \bar{a}_{lm}^a)}{B_l^a} \times \frac{(B_l^b a_{lm}^{s*} + \bar{a}_{lm}^{b*})}{B_l^b} = \hat{C}_l^s + \hat{C}_l^{ab({\rm junk})},$$

where

$$\hat{C}_l^{ab({\rm junk})} = \sum_{m=-l}^{m=l} \frac{\bar{a}_{lm}^a \bar{a}_{lm}^{b*}}{2l+1} + \sum_{m=-l}^{m=l} \frac{a_{lm}^s \bar{a}_{lm}^{b*}}{2l+1} + \sum_{m=-l}^{m=l} \frac{\bar{a}_{lm}^a a_{lm}^{s*}}{2l+1}, \tag{9}$$

for all values of a and b. Now using Eq. (5) we obtain

$$\hat{C}_l^{\rm Clean} = \hat{C}_l^s \hat{\mathbf{W}}_l \mathbf{e}_0 \mathbf{e}_0^T \hat{\mathbf{W}}_l^T + \hat{\mathbf{W}}_l \hat{\mathbf{C}}_l^{({\rm junk})} \hat{\mathbf{W}}_l^T, \tag{10}$$

where $\mathbf{e}_0$ is a column vector with unit elements

$$\mathbf{e}_0 = \begin{pmatrix} 1 \\ .. \\ .. \\ 1 \end{pmatrix}. \tag{11}$$

The CMB signal power in the cleaned map is kept unaltered by imposing the constraint

$$\hat{\mathbf{W}}_l \mathbf{e}_0 = \mathbf{e}_0^T \hat{\mathbf{W}}_l^T = 1, \tag{12}$$

on weights $\hat{w}_l^a$. Using Eqs. (10) and (12), the total power at a multipole l in the cleaned map is

$$\hat{C}_l^{\rm Clean} = \hat{C}_l^s + \hat{\mathbf{W}}_l \hat{\mathbf{C}}_l^{\rm junk} \hat{\mathbf{W}}_l^T. \tag{13}$$

Thus the CMB signal power is only an additive positive constant in the expression of the total power of the weighted map. Hence, choosing weights that minimize $\hat{C}_l^{\rm Clean}$ also minimizes the combined contamination coming from foreground, detector noise, and a chance correlation between the CMB and the non-CMB components. It may be shown, using the Lagrange multiplier method, that minimizing $\hat{C}_l^{\rm Clean}$ subject to the constraint Eq. (12), gives the following expression for the weight factors [17,25],

$$\hat{\mathbf{W}}_l = \frac{\mathbf{e}_0^T \hat{\mathbf{C}}_l^{-1}}{\mathbf{e}_0^T \hat{\mathbf{C}}_l^{-1} \mathbf{e}_0}. \tag{14}$$

Following Eqs. (5) and (14) we can express the power in the cleaned maps neatly as

$$\hat{C}_l^{\rm Clean} = \frac{1}{\mathbf{e}_0^T \hat{\mathbf{C}}_l^{-1} \mathbf{e}_0}. \tag{15}$$

Equation (15) is valid for the estimation of the autopower spectrum of a cleaned map obtained using a global method of foreground cleaning. However, as discussed later, we estimate the power spectrum using cross correlation. In this case Eq. (15) does not remain valid. The numerical implementation also involves an iterative method of foreground cleaning and application of a sky mask before power spectrum estimation. These effects are also not incorporated in Eq. (15).

We note that in certain cases $\hat{\mathbf{C}}_l$ may be singular. This may happen if the number of channels is more than the number of unknown components, assuming negligible detector noise. In this case it is useful to employ the *Moore-*

Penrose generalized inverse (MPGI) of $\hat{\mathbf{C}}_l$ in the above expressions for the weights and cleaned power spectrum. The MPGI is well defined even when the matrix $\hat{\mathbf{C}}_l$ is singular.

The generalized weights and cleaned power spectrum are then given by

$$\hat{\mathbf{W}}_l = \frac{\mathbf{e}_0^T \hat{\mathbf{C}}_l^+}{\mathbf{e}_0^T \hat{\mathbf{C}}_l^+ \mathbf{e}_0}, \tag{16}$$

$$\hat{C}_l^{\text{Clean}} = \frac{1}{\mathbf{e}_0^T \hat{\mathbf{C}}_l^+ \mathbf{e}_0}, \tag{17}$$

where $\hat{\mathbf{C}}_l^+$ is the MPGI of $\hat{\mathbf{C}}_l$. A detailed derivation of the generalized weights and cleaned power spectrum is given in Appendix A.

B. Implementation on WMAP data

We use the "raw" DA maps (i.e., that have not undergone any foreground cleaning process) both for the WMAP 1-year and WMAP 3-year data release from the LAMBDA website. These maps follow the HEALPIX[1] pixelization scheme at a resolution level $N_{\text{side}} = 512$ corresponding to approximately 3×10^6 sky pixels. During spherical harmonic transform of these maps, we restrict ourselves to a maximum multipole, $l_{\text{max}} = 1024$.

The spectra of foreground emission components as well as their amplitudes depend on the location in the sky. Therefore, foreground removal becomes more efficient by allowing the weights to depend on the sky parts rather than using a single set of weights for the entire sky [11,18]. We partition the sky into a total of $r = 9$ regions, depending upon the level of foreground contamination, following [18] with minor variations.[2] We describe our sky partitioning method, which depends upon the level of foreground contamination, in detail in Appendix B. An alternative approach may be to partition the sky based upon the information about the spectral index variation of the foreground components.

All the sky regions obtained by us are similar to those obtained in [18]. However, instead of forming a single cleaned map by combining averaged DA maps for each of the frequency channels as in [18], we form 48 foreground-cleaned maps taking different possible combinations of DA maps. These maps are labeled as C**i** and CA**i** where **i** $= 1, 2, \ldots, 24$. The detailed nomenclature of these 48 maps is given in Table 1 of [21]. For all bands, except the W band, we use data from individual DAs, since at most two different DAs are available per band. For the case of the W band we have four different DAs and we use averaged maps of two DAs. This reduces the contamination in any particular cleaned map. Let Q1 (V1) and Q2 (V2) denote the maps corresponding to the two Q (V) band DAs and Wnm denote the averaged map of the DA n and m for the W channel. The cleaned map C1, for example, is obtained by a linear combination of the K, Q1, V1, W12 maps. Here W12 denotes the averaged map for the first and second W band DAs. Similarly, the cleaned map CA1 is obtained by a linear combination of KA, Q1, V1, W12 maps. We defer a discussion about the philosophy of forming several cleaned maps till Sec. III. For each combination we perform foreground cleaning in an iterative fashion following the method described in [18]. The entire cleaning procedure of all the 48 maps is automated.

We note that the weights given by Eq. (14) are such that they tend to weigh the maps in an inverse-noise variance manner at large l. One can easily see this using Eq. (14), considering for simplicity only 2 channels. Assuming that the noise is uncorrelated between two detectors and effective (beam-deconvolved) detector noise dominates over other components at large l, the weights for the 2 channels are given by

$$\hat{w}_l^1 = \frac{(1/\hat{N}_l^1)}{(1/\hat{N}_l^1) + (1/\hat{N}_l^2)}, \qquad \hat{w}_l^2 = \frac{(1/\hat{N}_l^2)}{(1/\hat{N}_l^1) + (1/\hat{N}_l^2)}, \tag{18}$$

where $\hat{N}_l^1$ and $\hat{N}_l^2$ are the beam-deconvolved noise power spectra of the two maps. At large l, due to inverse-noise weighting, the weights for all channels are negligible compared to the highest resolution channel. We explicitly confirm this with our numerical simulations. For WMAP, the W band has the highest resolution. Hence, while all DA maps contribute at low l, only the W band survives at high

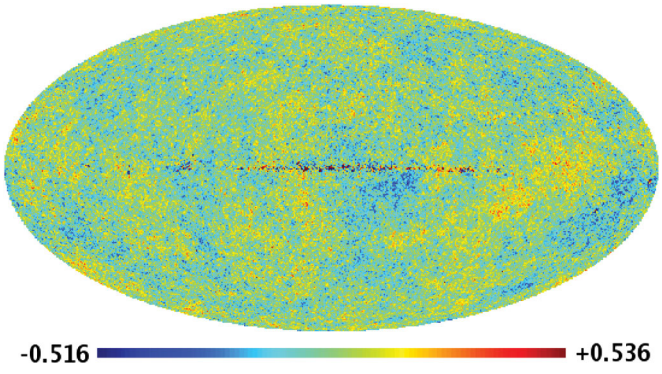

FIG. 1 (color online). The figure shows cleaned map CA8 using WMAP 3-year data. There is some residual foreground contamination near the galactic regions.

[1]For a comprehensive description of HEALPIX, we refer to [26–28].

[2]Although this scheme is found to be effective, it is possible to envisage other schemes such as those that make foreground contamination in each part closer to the rigid scaling approximation. The emission pattern, $g(\theta, \phi, \nu)$, of a foreground component satisfying the rigid scaling approximation follows $g(\theta, \phi, \nu) = g(\theta, \phi, \nu_0) f_\nu$, with $f_\nu = 1$ for $\nu = \nu_0$, where ν_0 is a reference frequency. In an ongoing study we have investigated and demonstrated improvement in foreground cleaning for different levels of partitioning [29].

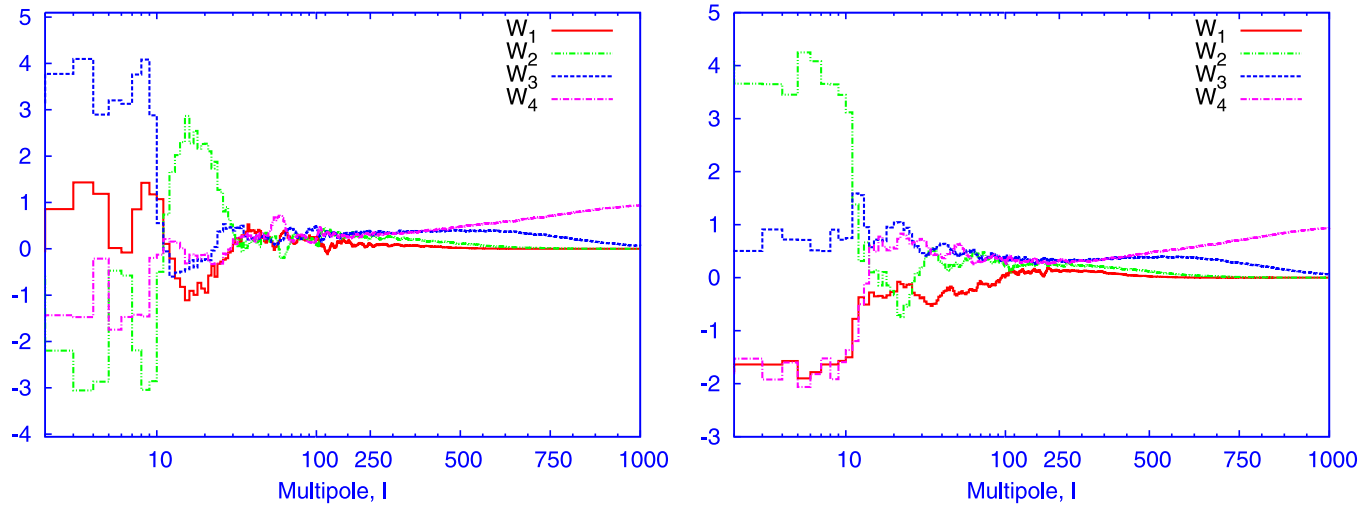

FIG. 2 (color online). The left panel of the figure shows the weights for the cleanest region for the combination KA, Q1, V1, W12. The weights W_1, W_2, W_3, W_4 correspond to the channels KA, Q1, V1, W12, respectively. The right panel shows the weights for the same combination but for the second cleanest region. At low multipoles foregrounds dominate. Therefore, weights take large positive and negative values to subtract foregrounds.

l. This results in cleaned maps which have the same resolution as the W band.

In Eq. (7) the weights could have numerical errors if the matrix $\hat{\mathbf{C}}_l$ becomes ill conditioned at specific values of l. Hence, in practice, the numerical implementation obtains weights using the $\hat{\mathbf{C}}_l$ matrix smoothed over a range of $\Delta l = 11$ prior to inversion.

The cleaned map CA8 for WMAP 3-year data is shown in Fig. 1. There is some residual foreground left in the galactic plane as seen in this map. We apply the Kp2 mask, supplied by the WMAP team, to flag the contaminated pixels near the galactic plane. Therefore, these residuals do not affect our final estimated power spectrum.

The weights $\hat{w}_l$ are shown in Fig. 2 for one of the cleaned maps (CA1) for the cleanest and second cleanest regions. For low l where the diffuse foregrounds are dominant, the weights take large positive and negative values to subtract the foregrounds. At large l the maximum weight is given to the W band channel since it has the highest resolution.

III. POWER SPECTRUM ESTIMATION

In this section we describe our procedure for estimating the power spectrum of the WMAP CMB signal. The power spectrum is obtained by cross correlating the 48 foreground-cleaned maps, C1, C2, ..., C24 and CA1, CA2, ..., CA24. We also describe our method for estimation of the residual unresolved point source contamination.

A. Cross-power spectrum estimation using foreground-cleaned maps

We estimate the CMB power spectrum by cross correlating pairs of foreground-cleaned maps. In order to remove detector noise bias we cross correlate two cleaned maps which do not have a common detector map. For example, cross correlating the two cleaned maps C1 and CA12, obtained from the individual maps (K, Q1, V1, W12) and (KA, Q2, V2, W34), respectively, yields a power spectrum with substantially reduced detector noise bias.[3] The WMAP-supplied Kp2 mask is applied to each of the cleaned maps prior to cross correlation in order to remove residual contamination from the galactic region and resolved point sources. Using all 48 cleaned maps we obtain a set of 24 cross-power spectra [21], shown in Table II. We note that the WMAP team also used cross correlation of independent DA maps to remove detector noise bias from their power spectrum [10,12]. The alternative to cross correlation is to use autocorrelation and use a detector noise model to remove bias from the power spectrum. In the autocorrelation case the detector noise model uncertainty affects the mean of the cleaned power spectrum, whereas the cross-correlation method is advantageous since the power spectrum is not affected by the uncertainty in the noise model.

We estimate the 24 full-sky power spectra using partial-sky spectra following the MASTER method [10,30,31]. The basic idea of the MASTER method is that removing a sky region effectively introduces uneven weighting of the pixels. Those removed have effectively zero weight and the

[3]In [18] a single foreground-cleaned map is obtained by linearly combining 5 maps corresponding to each of the different WMAP frequency channels. For each of the Q, V, and W frequency channels, where more than one DA map is available, an averaged map is formed using all the DA maps for that channel. However, averaging over *all* the DA maps in a given frequency channel precludes any possibility of removing detector noise bias using cross correlation, as one obtains only one cleaned map at the end.

TABLE II. List of all the 24 cross-power spectra using the 48 cleaned maps, C1, C2, ..., C24 and CA1, CA2, ..., CA24.

C1 ⊗ CA12	C2 ⊗ CA11	C3 ⊗ CA10	C4 ⊗ CA9	C5 ⊗ CA8	C6 ⊗ CA7
C7 ⊗ CA6	C8 ⊗ CA5	C9 ⊗ CA4	C10 ⊗ CA3	C11 ⊗ CA2	C12 ⊗ CA1
C13 ⊗ CA24	C14 ⊗ CA23	C15 ⊗ CA22	C16 ⊗ CA21	C17 ⊗ CA20	C18 ⊗ CA19
C19 ⊗ CA18	C20 ⊗ CA17	C21 ⊗ CA16	C22 ⊗ CA15	C23 ⊗ CA14	C24 ⊗ CA13

rest have unit weight. Weighting a map in pixel space induces correlations between neighboring multipoles. The correlation of the multipoles is described by the coupling matrix which is determined by the geometry of the sky coverage. Moreover, the partial-sky power spectrum of such a map has a lower bias because of the reduced coverage. The MASTER method removes bias from the partial-sky power spectrum on the ensemble average by inverting the coupling matrix.

We convert each of the 24 partial-sky cross-power spectra to full-sky estimates using the coupling (bias) matrix corresponding to the Kp2 mask. The small scale systematic effects of beam and pixel smoothing are removed using appropriate circularized beam transform [30] and pixel window functions. The resulting 24 full-sky cross-power spectra are then combined with equal weights into a single "uniform average" power spectrum.[4] There exists some residual unresolved point source contamination in the uniform average power spectrum. We next describe the method we used for removing this contamination.

B. Estimation of the residual, unresolved, point source power spectrum

There have been several studies regarding the point source contamination in the CMB maps [32–34]. We estimate residual contamination due to unresolved point sources in the uniform average power spectrum following the point source model constructed by the WMAP team [10,12]. After masking resolved point sources using the Kp2 mask, the WMAP science team directly fits the amplitude A of the unresolved point sources following a likelihood analysis using the cross-power spectra of the template cleaned maps and a model CMB power spectrum. The unresolved point source covariance matrix modeled by WMAP is given by $C_l^{\mathrm{ps}(ab)} = c_{ab}A(\nu_a/\nu_0)^{-2}(\nu_b/\nu_0)^{-2}$ in thermodynamic temperature units and for two different frequency channels, a and b. Here A is the amplitude of the unresolved point source spectrum in antenna temperature at a reference frequency ν_0 and c_{ab} is the conversion factor from antenna to thermodynamic temperature. Using the best fit value of A, the WMAP science team subtracts out the unresolved point source contamination.

The residual unresolved point source contamination is much smaller in our case since much of the contamination is removed by our foreground removal method. Therefore, we need to estimate unresolved point source contamination in our cross spectra in a method different from WMAP. We model unresolved point source contamination in the uncleaned maps by WMAP's point source model with their fitted values. Then we compute the residual unresolved point source contamination in each of the 24 cross-power spectra assuming that the residual unresolved point source contamination has a statistically isotropic distribution in the sky. Unresolved point source contamination in each cross-power spectra is effectively computed by summing up contributions from individual sky parts formed during cleaning. As described in detail in Appendix C, our estimate of the residual unresolved point source contamination in a cross-power spectrum of the cleaned maps C**i** and CA**j** reads

$$\hat{C}_l^{\mathrm{ps}(\mathbf{ij})} = \sum_{n=1}^{n=9}\sum_{l'} M_{ll'}^{n}\hat{\mathbf{W}}_{l'}^{n}\mathbf{C}_{l'}^{\mathrm{ps}}\hat{\mathbf{W}}_{l'}^{ln^T} + \sum_{n,n'n\neq n'}\sum_{l'} M_{ll'}^{nn'}\hat{\mathbf{W}}_{l'}^{n}\mathbf{C}_{l'}^{\mathrm{ps}}\hat{\mathbf{W}}_{l'}^{lk^T}, \quad (19)$$

where $M_{ll'}^{nn'}$ is the cross-coupling matrix of two masks n, n'. Here $\hat{\mathbf{W}}_l^n$ and $\hat{\mathbf{W}}_l^{ln}$ are, respectively, the weight vectors of C**i** and CA**j** for the sky region n. The final residual unresolved point source contamination in the uniform aver-

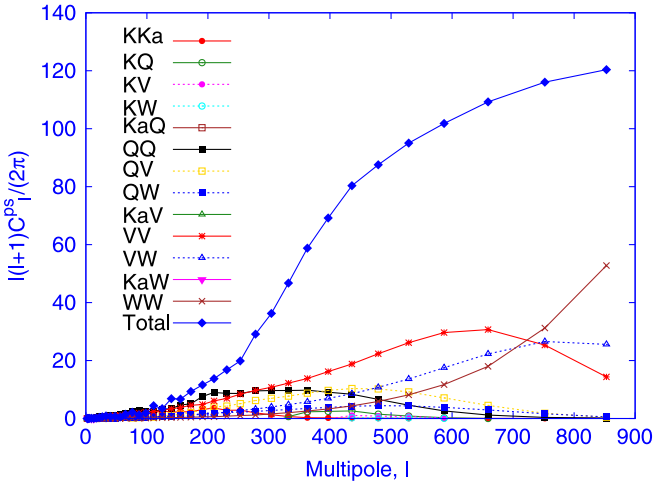

FIG. 3 (color online). The effective contribution from each of the elements (a, b) of the matrix C_l^{ps} (see text) in the total residual point source estimation. The total residual unresolved point source spectrum is also shown (blue filled diamonds).

[4]There exists the additional freedom to choose optimal weights for combining the 24 cross-power spectra which we do not discuss in this work.

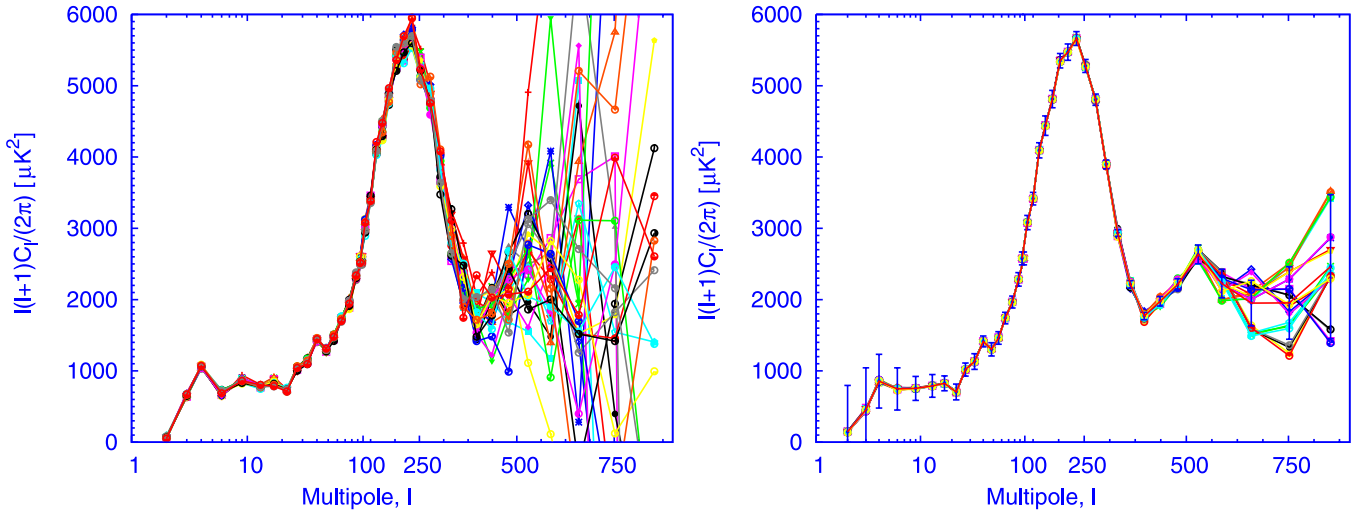

FIG. 4 (color online). The left panel shows 28 cross-power spectra obtained by the WMAP team's 1-year analysis without removing the unresolved point source contamination. The right panel shows our 24 cross-power spectra based on 1-year data.

age power spectrum is obtained by simply averaging over all such 24 residual unresolved point source spectra.[5]

Figure 3 shows the residual unresolved point source contamination $\sum_{n=1}^{n=9}\sum_{l'} M^{n}_{ll'}\hat{w}^{na}_{l'}C^{\mathrm{ps}(ab)}_{l'}\hat{w}^{nb}_{l'}$ for different values of a, b after averaging over all the 24 cross-power spectra, i.e., on (**ij**). The total residual unresolved point source spectrum is also shown. The dominant contributors to the total unresolved point source spectrum at large l are the WW, VW, and VV combinations. This is expected since the V and W bands share most of the weights at large l because of their higher angular resolutions. Weights are negligible for K, Ka, Q bands in the large l limit, and hence point source contamination from these bands is heavily suppressed.

The WMAP team's 1-year power spectrum is derived from the 28 cross-power spectra which are available from the LAMBDA website in the unbinned form. These 28 cross-power spectra are not corrected for the residual unresolved point sources. The point source correction is, of course, applied before using these for power spectrum estimation. Following the WMAP team's 1-year bins we compare these 28 cross-power spectra in Fig. 4 with 24 cross-power spectra (Table II) obtained from our own 1-year results. These 24 cross-power spectra also are not corrected for the residual unresolved point sources and show very little dispersion compared to WMAP's point source contaminated cross-power spectra. The uniform average power spectrum plotted in starred (green) points in Fig. 5 has less excess power near the second acoustic peak compared to a uniform average of WMAP's 28 cross-power spectra (filled circles in red). This merely shows that we have removed some amount of point source contamination at this range of l during our cleaning.

IV. COMPUTING ERROR BARS

We rely upon Monte Carlo simulation to compute the error bars in the power spectrum. We generate synchrotron, free-free, and thermal dust maps corresponding to different frequencies in a given combination using the publicly

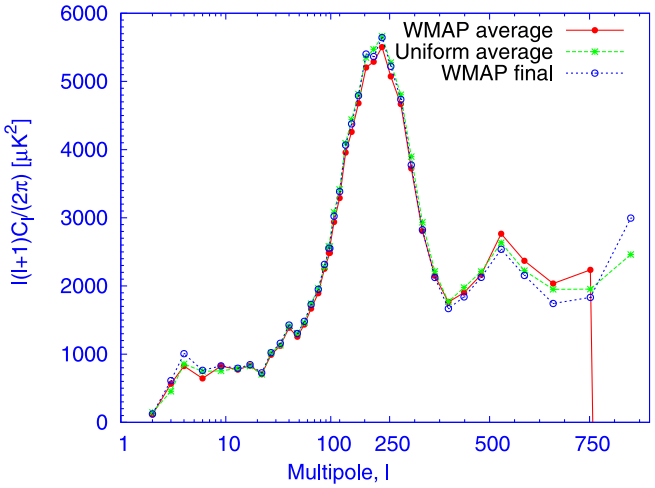

FIG. 5 (color online). Comparison of WMAP's 1-year average power spectrum (filled circles in red) with our binned power spectrum without any point source subtraction (starred green points). WMAP's final binned power spectrum is shown by unfilled blue circular points. Interestingly, the notch at $l = 4$ appears to be reduced in WMAP's average power spectrum.

[5]We note that the second term on the right-hand side of Eq. (19) gives a negligible contribution to the total estimate of the residual point source correction. This is a result of the fact that weights are effectively determined by the detector noise in the large l limit. Two different sky parts have uncorrelated noises. So weights from two different sky regions are uncorrelated at large l, making the second term much smaller compared to the first term.

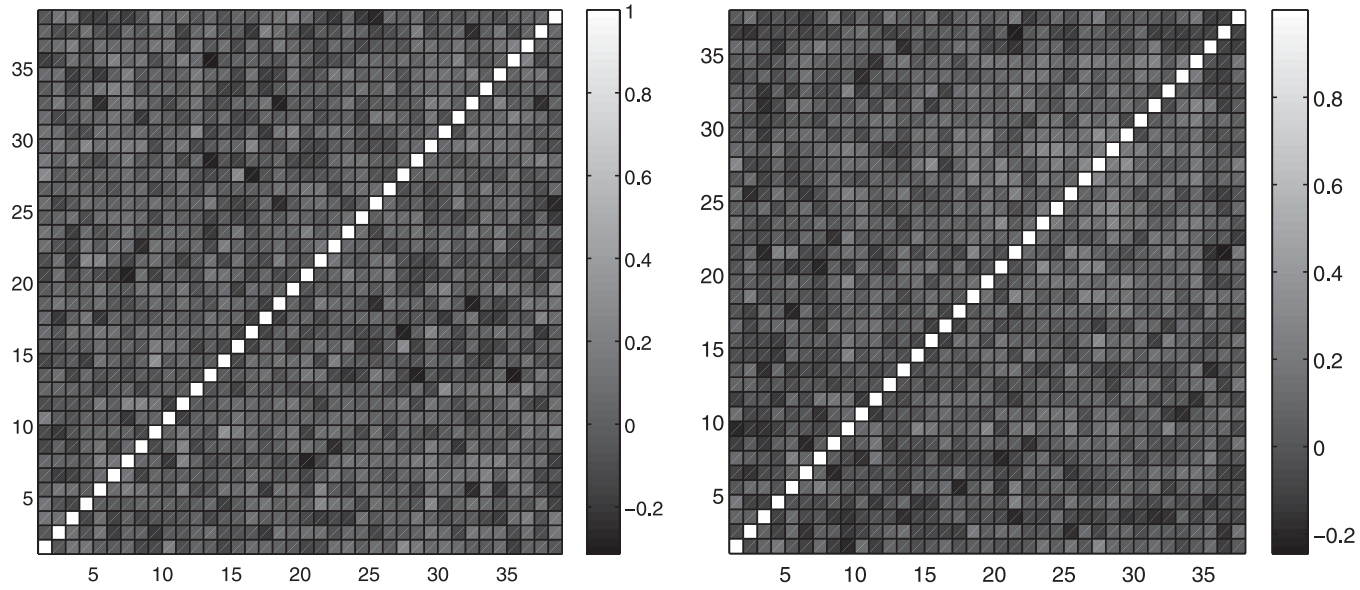

FIG. 6. Correlation matrix $\frac{\langle \Delta C_\alpha \Delta C_{\alpha'} \rangle}{\sqrt{\langle (\Delta C_\alpha)^2 \rangle \langle (\Delta C_{\alpha'})^2 \rangle}}$ from our simulation plotted with respect to the bin index for the WMAP 1-year (left panel) and 3-year (right panel) analyses. The matrix is found to be dominated mostly by the diagonal elements.

available Planck Sky Model[6] (PSM). Each of the random realizations of the CMB and foreground maps are convolved by the appropriate beam function for each detector. Random noise maps corresponding to each detector are generated by first sampling a Gaussian distribution with unit variance. In the final step we multiply each Gaussian variable by $\sigma_0/\sqrt{N_p}$ to form realistic detector noise maps. Here σ_0 is the noise per observation of the detector under consideration and N_p the effective number of observations at each pixel. These realistic maps with detector noise, foreground, and CMB signal are then passed through the cleaning pipeline. The error bars for our power spectrum correspond to the standard deviation of the power spectrum obtained from Monte Carlo simulations.

Because of the (Kp2) mask applied to remove potential foreground contaminated regions, the neighboring multipoles become coupled. In the presence of detector noise, correlation between neighboring $\hat{C}_l^s$ becomes stronger. The covariance matrix $\langle \Delta \hat{C}_l^s \Delta \hat{C}_{l'}^s \rangle = \langle (\hat{C}_l^s - \langle \hat{C}_l^s \rangle)(\hat{C}_{l'}^s - \langle \hat{C}_{l'}^s \rangle) \rangle$ obtained from simulations is therefore expected to have nondiagonal elements. It is convenient to bin the power spectrum in order to minimize the correlations and errors. We have considered a binning identical to that used by the WMAP team in their analysis. Let $\hat{C}_\alpha$ denote the binned power spectrum. Then the covariance matrix of the binned spectrum is obtained as $\langle \Delta \hat{C}_\alpha^s \Delta \hat{C}_{\alpha'}^s \rangle = \langle (\hat{C}_\alpha^s - \langle \hat{C}_\alpha^s \rangle) \times (\hat{C}_{\alpha'}^s - \langle \hat{C}_{\alpha'}^s \rangle) \rangle$. The standard deviation obtained from the diagonal elements of the binned covariance matrix was used as the error bars on the binned final spectrum. We also define a correlation matrix of the binned power spectrum using

$$C_{\alpha\alpha'} = \frac{\langle \Delta \hat{C}_\alpha \Delta \hat{C}_{\alpha'} \rangle}{\sqrt{\langle (\Delta \hat{C}_\alpha)^2 \rangle \langle (\Delta \hat{C}_{\alpha'})^2 \rangle}}, \tag{20}$$

wherein all the elements of this matrix are bound to lie between $[-1, 1]$. In the left panel of Fig. 6 we show the correlation matrix for WMAP 1-year simulations. The right-hand panel of this figure is the corresponding plot for the WMAP 3-year analysis. Both these matrices are seen to be dominated by diagonal elements.

The results of Monte Carlo simulations of the foreground removal and power spectrum estimation method reveal a negative bias in the low l moments. The left and right panels of Fig. 7 show the existence of such bias from 110 simulations using PSM generated foreground maps with WMAP 1-year and 3-year noise models, respectively. The origin of this negative bias is explained in Sec. V. For $l = 2$ and $l = 3$ the bias is -27.4% and -13.8%, respectively. However, this bias becomes negligible at higher l; e.g., at $l = 22$, it is only -0.8%.

V. BIAS ANALYSIS

Although foreground cleaning is performed with the constraint that the CMB power spectrum is preserved, the method biases the final power spectrum. This section is devoted to a discussion of the full bias in the method. In the entire discussion that follows we do not treat foregrounds

[6]We acknowledge the use of version 1.1 of the Planck reference sky model, prepared by the members of Working Group 2 and available at www.planck.fr/heading79.html.

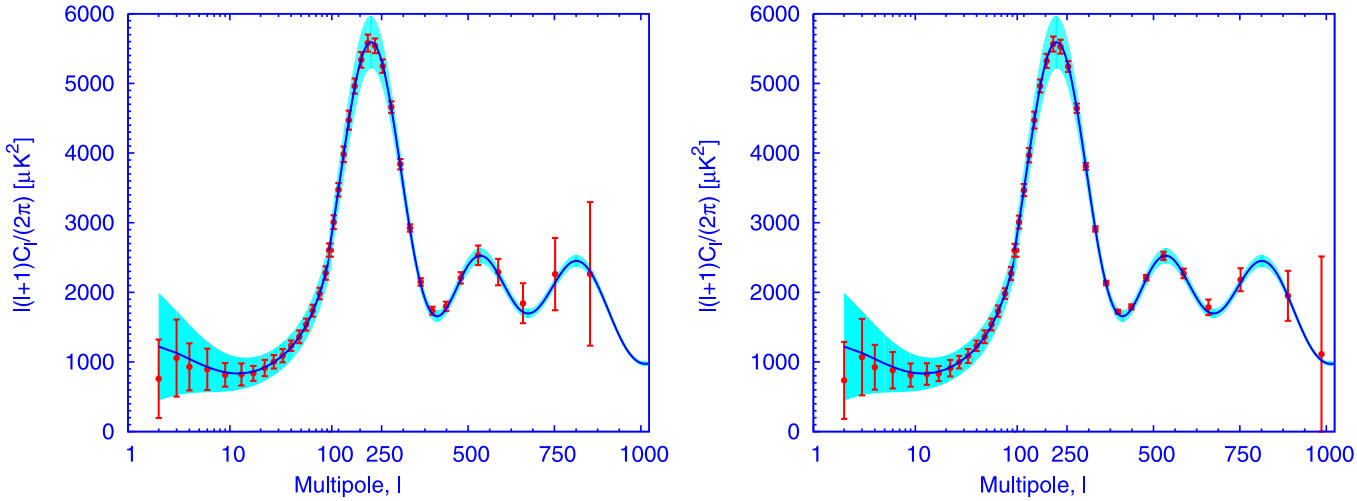

FIG. 7 (color online). The left panel (filled red circular points) shows the ensemble-averaged power spectrum from 110 Monte Carlo simulations of our power spectrum estimation method. The simulations were carried out using the 1-year WMAP detector noise maps available from the LAMBDA website. We use publicly available PSM to generate the diffuse foreground models. The recovered spectrum is binned in the same manner as the WMAP 1-year power spectrum. The input theoretical spectrum is shown by the solid (blue) line along with cosmic variance. The right panel is the same as the left but with 3-year noise maps. The 3-year noise maps were generated following the method described in the text. The spectrum is binned using the binning scheme of the WMAP team's analysis of 3-year data.

as a stochastic component. Instead we consider them as fixed templates in all the realizations. This is entirely justified since we are interested in computing a foreground-free CMB map rather than extracting information about the distribution of the foreground from which they are drawn. We discuss the bias in a few different cases depending upon the rank of the covariance matrix $\hat{\mathbf{C}}_l$.

The existence of some bias is not difficult to anticipate and understand. The method is intended to perform a minimization of the foreground power spectrum which is a positive-definite quantity. Unless the foreground cleaning is fully effective at all multipoles, the minimization would leave some residual foreground which naively would give rise to a positive bias in the cleaned power spectrum. However, it is very interesting that there exists an additional negative bias in the method. This negative bias is strongest at the lowest multipoles and increases in magnitude with an increase in the number of channels that are combined.

We carry out analytical calculations for bias assuming that the entire sky is cleaned simultaneously. Our analysis elucidates the generic features of the power spectrum estimated using the foreground removal algorithm. A generalization of the analysis to the case of the iterative method of foreground removal is straightforward, however, with a complicated algebra. The analytic results in this section are being presented mainly to get an understanding of the origin of the bias and to verify simulation results. The actual value of the bias is estimated by direct numerical simulations of the complete iterative procedure used for power spectrum estimation, as described in Sec. IV.

A. Case: Rank $(\mathbf{C}_l^f) \leq n_c - 1$

In the signal-dominated regime one can ignore detector noise from analysis. If detector noise is subdominant, the rank of the empirical full covariance matrix is determined by the number of independent components. Therefore, for n_f independent foreground components (all of which follow rigid frequency scaling) and a CMB component, the rank of the full covariance matrix is given by $n_f + 1$. If the number of bands n_c to be linearly combined becomes equal to the rank, $n_f + 1$, the full covariance matrix is of full rank and hence invertible. In this section we are interested in computing bias in the cleaned power spectrum when $\text{rank}(\mathbf{C}_l^f) \leq n_c - 1$. Such a case arises when $n_c > n_f + 1$.

Let the pth foreground component for channel a be denoted by $F_0^p(\theta, \phi) f_p^a$. Here, $F_0^p(\theta, \phi)$ is the pth foreground template based on frequency ν_0, so that $f_p^a = 1$, for frequency ν_0. The full signal map at the ath frequency channel is given by

$$S^a(\theta, \phi) = C(\theta, \phi) + \sum_{p=1}^{n_f} F_0^p(\theta, \phi) f_p^a. \tag{21}$$

Alternatively, in spherical harmonic space,

$$a_{lm}^a = a_{lm}^s + \sum_{p=1}^{n_f} f_p^a a_{lm}^{p0}. \tag{22}$$

The autopower spectrum of the ath channel

$$\hat{C}_l^a = \hat{C}_l^s + 2\sum_{p=1}^{n_f} f_p^a \hat{C}_l^{(sp)0} + \sum_{p,p'}^{n_f} f_p^a f_{p'}^a C_l^{(pp')0}, \quad (23)$$

where $C_l^{(pp')0}$ is the correlation between any two foreground components p, p', and $\hat{C}_l^{(sp)0}$ is the chance correlation between the CMB and the pth foreground component. The cross-power spectrum between two channels a, b is given by

$$\hat{C}_l^{ab} = \hat{C}_l^s + \sum_{p=1}^{n_f} f_p^a \hat{C}_l^{(sp)0} + \sum_{p=1}^{n_f} f_p^b \hat{C}_l^{(sp)0} + F_l^{ab}, \quad (24)$$

where the foreground covariance matrix

$$F_l^{ab} = \sum_{p,p'}^{n_f} f_p^a f_{p'}^b C_l^{(pp')0}. \quad (25)$$

Introducing an explicit matrix notation, we write

$$\begin{aligned}\hat{\mathbf{C}}_l = \hat{C}_l^s &\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ . & . & \dots & . \\ 1 & . & \dots & 1 \end{pmatrix}_{(n_c\times n_c)} \\ &+ \sum_{p=1}^{n_f} \hat{C}_l^{(sp)0} \begin{pmatrix} f_p^1 & f_p^1 & \dots & f_p^1 \\ f_p^2 & f_p^2 & \dots & f_p^2 \\ . & . & \dots & . \\ f_p^{n_c} & f_p^{n_c} & \dots & f_p^{n_c} \end{pmatrix}_{(n_c\times n_c)} \\ &+ \sum_{p=1}^{n_f} \hat{C}_l^{(sp)0} \begin{pmatrix} f_p^1 & f_p^2 & \dots & f_p^{n_c} \\ f_p^1 & f_p^2 & \dots & f_p^{n_c} \\ . & . & \dots & . \\ f_p^1 & f_p^2 & \dots & f_p^{n_c} \end{pmatrix}_{(n_c\times n_c)} \\ &+ \begin{pmatrix} F_l^{11} & F_l^{12} & \dots & F_l^{1n_c} \\ F_l^{21} & F_l^{22} & \dots & F_l^{2n_c} \\ . & . & \dots & . \\ F_l^{n_c 1} & F_l^{n_c 2} & \dots & F_l^{n_c n_c} \end{pmatrix}_{(n_c\times n_c)}. \end{aligned} \quad (26)$$

We define

$$\hat{\mathbf{f}}_l^{p0} = \hat{C}_l^{(sp)0} \begin{pmatrix} f_p^1 \\ f_p^2 \\ . \\ f_p^{n_c} \end{pmatrix}. \quad (27)$$

Clearly, $\hat{\mathbf{f}}_l^{p0}, \mathbf{e}_0 \in \mathbb{R}_{n_c,1}$, where $\mathbb{R}_{n_f,n_c}$ denotes the set of real $n_f \times n_c$ matrices. The full covariance matrix can then be written as

$$\hat{\mathbf{C}}_l = \hat{C}_l^s \mathbf{e}_0 \mathbf{e}_0^T + \hat{\mathbf{A}}_1, \quad (28)$$

where $\hat{\mathbf{A}}_1 = \hat{\mathbf{f}}_l^0 \mathbf{e}_0^T + \hat{\mathbf{A}}_2$, $\hat{\mathbf{A}}_2 = \mathbf{e}_0 \hat{\mathbf{f}}_l^{0T} + \mathbf{A}_3$, $\mathbf{A}_3 \equiv \mathbf{F}_l$ is the foreground covariance matrix, and

$$\hat{\mathbf{f}}_l^0 = \sum_{p=1}^{n_f} \hat{\mathbf{f}}_l^{p0}, \quad (29)$$

with $\hat{\mathbf{f}}_l^0 \in \mathbb{R}_{n_c,1}$.

We use Eq. (28) to compute the bias in the cleaned power spectrum. On the ensemble average the cleaned power spectrum, as given by Eq. (17), can be simplified with the help of successive use of a set of theorems reported in [35,36]. An extensive discussion of these theorems is given in Appendix D. Assuming a statistically isotropic CMB sky we obtain the following expression for the ensemble average of the cleaned power spectrum:

$$\langle \hat{C}_l^{\text{Clean}} \rangle = \langle \hat{C}_l^s \rangle - n_f \frac{\langle \hat{C}_l^s \rangle}{2l+1}. \quad (30)$$

We easily note a few interesting aspects of the above equation. First of all, there exists a negative bias. Second, because of $\sim 1/(2l+1)$ decay, this bias is important at the lowest multipoles. Another important point to note is that the bias depends on the underlying CMB power spectrum. Therefore it is possible to remove bias in any given statistically isotropic CMB model in this approach by constructing an appropriately scaled estimator. Lastly, we find that there is no foreground bias. One would naively expect that the foregrounds would contribute a positive bias in the power spectrum. However, this does not happen in this case as the rigid scaling assumption, along with the condition $n_c \geq n_f + 1$, ensures that a sufficient amount of spectral information is available to remove all the foregrounds. The negative bias arises since the weights are determined from the empirical covariance matrix in order to take into account only the information available from the observed data.

B. Case: Rank $(\mathbf{C}_l^f) = n_c$

The rigid scaling assumption for the foreground contaminants considered in the previous section is at best a reasonable approximation and is known not to be valid in general. As mentioned in [9] a foreground component with varying spectral index over the sky could be approximated in terms of two templates, provided the variation is small compared to the mean spectral index over the sky. A stronger variation will need more than two templates for reasonable modeling. In such a situation, if the number of templates required for modeling all the foreground components exceeds the number of maps available for a linear combination, then $\mathbf{C}_l^f$ is of full rank. In this case a positive foreground bias appears along with a negative bias. The negative bias is similar to the previous case in that it remains proportional to $\frac{\langle \hat{C}_l^s \rangle}{2l+1}$. Keeping aside the detector noise for the moment, the ensemble average of the cleaned power spectrum is given by

$$\langle \hat{C}_l^{\text{Clean}} \rangle = \langle \hat{C}_l^s \rangle + \frac{1}{\mathbf{e}_0^T (\mathbf{C}_l^f)^{-1} \mathbf{e}_0} + (1 - n_c) \frac{\langle \hat{C}_l^s \rangle}{2l+1}. \quad (31)$$

A detailed derivation of the above equation is similar to derivation of Eq. (30). Since $\mathbf{C}_l^f$ is of full rank and a

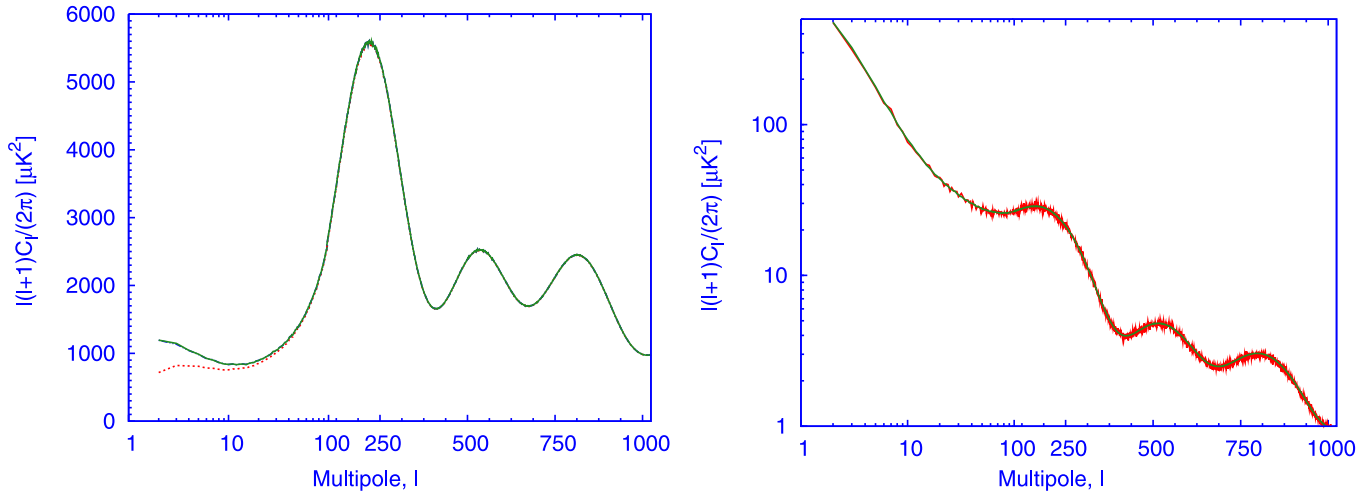

FIG. 8 (color online). The negative bias in the extracted power spectrum at low l is shown by the dashed (red) line in the left panel. The bias-corrected spectrum is plotted in the solid (green) line. The input CMB power spectrum (dashed-dotted blue line) lies entirely behind the solid line and is not visible. The right panel shows how well the analytic results for bias match with those obtained from Monte Carlo simulations. The wiggly (red) line of this figure shows the magnitude of negative bias obtained from simulations. The solid (green) line shows $(l(l+1)/2\pi)2C_l^s/(2l+1)$.

positive-definite matrix, the second term on the right causes a positive foreground bias.

C. Noise and foreground case

The discussion in the last two subsections does not consider any detector noise. We now consider the most general case where we have both foreground and detector noise. Following a method similar to that used in the derivation of Eqs. (30) and (31), we find that on the ensemble average the cleaned power spectrum in this case is given by

$$\langle \hat{C}_l^{\rm Clean} \rangle = \langle \hat{C}_l^s \rangle + \left\langle \frac{1}{\mathbf{e}_0^T (\hat{\mathbf{C}}_l^{f+N})^{-1} \mathbf{e}_0} \right\rangle + (1 - n_c) \frac{\langle \hat{C}_l^s \rangle}{2l+1}. \tag{32}$$

Here the foreground covariance matrix $\mathbf{C}_l^f$ need not be of full rank. However, the foreground plus noise covariance matrix $\hat{\mathbf{C}}_l^{f+N}$ is assumed to have full rank.

We have carried out Monte Carlo simulations to verify the analytical result given by Eq. (30). We perform simulations of foreground cleaning using $n_c = 3$ channels, corresponding to 41, 61, and 94 GHz frequencies. The foreground model consists of synchrotron and free-free emission. Each of the foreground components is assumed to follow a rigid frequency scaling all over the sky. First we generate synchrotron and free-free templates using the PSM at the three different frequencies. Since these templates do not follow rigid frequency scaling all over the sky (e.g., the synchrotron spectral index varies with the position of the sky), we developed a method to regenerate rigid scaling synchrotron and free-free templates at these three different frequency channels.[7]

The left-hand panel of Fig. 8, based on 1000 Monte Carlo simulations, shows that there exists a negative bias in the cleaning method. For the assumed model of foreground components with rigid frequency scaling, the second term on the right in Eq. (30) contributes $-2\langle \hat{C}_l^s \rangle/(2l+1)$ as a negative bias. In the right-hand panel of Fig. 8 we explicitly show that the magnitude of the negative bias is exactly compensated by $2\langle \hat{C}_l^s \rangle/(2l+1)$. The negative bias is important at the lowest multipoles and becomes negligible at high l; e.g., the bias is only $\approx 3\mu\mathrm{K}^2$ at multipole $l \approx 800$. The bias-corrected spectrum plotted by the solid (green) line in the left panel of Fig. 8 completely hides the dashed-dotted (blue) curve corresponding to the input CMB power spectrum which is used to generate random realizations of the CMB maps.

VI. RESULTS

All the 24 cross-power spectra estimated using WMAP 3-year data prior to point source correction are shown in Fig. 9. The uniform average of these 24 cross-power spec-

[7]For this purpose, we first note that for each component the approximate normalization $\sqrt{C_l^a/C_l^{23}}$ remains roughly constant for all l. Here C_l^{23} is the synchrotron (or free-free) power spectrum for frequency of 23 GHz and C_l^a is the synchrotron (or free-free) power spectrum at any of the other three frequencies. We generate a synchrotron (or free-free) template at the ath frequency channel by scaling the 23 GHz template by the number $\sqrt{C_l^a/C_l^{23}}$. This ensures that the emission for each foreground component at different frequencies follows rigid frequency scaling.

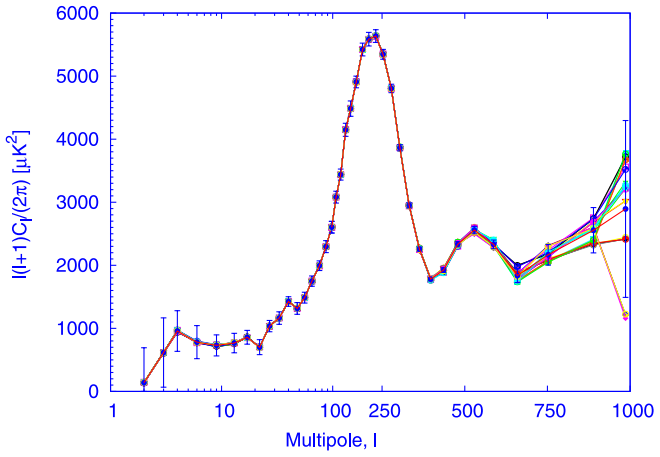

FIG. 9 (color online). The 24 cross-power spectra for the 3-year WMAP data are shown with the detector noise bias removed. The solid (red) line with (blue) error bars is the uniform average power spectrum.

tra is also shown in this figure. A similar figure for WMAP 1-year data is shown in the right panel of Fig. 4. Both the figures are similar. However, for the 3-year data all the 24 cross-power spectra show very little dispersion till the second trough. In comparison, the 1-year data show small dispersion only till the second acoustic peak. This may be explained due to the fact that the detector noise in the 3-year data is effectively lower compared to the 1-year data.

Subtracting the residual unresolved point source contamination estimated in Sec. III B from the uniform average power spectrum, we obtain the final CMB power spectrum as shown in Fig. 10. The unfilled (blue) points

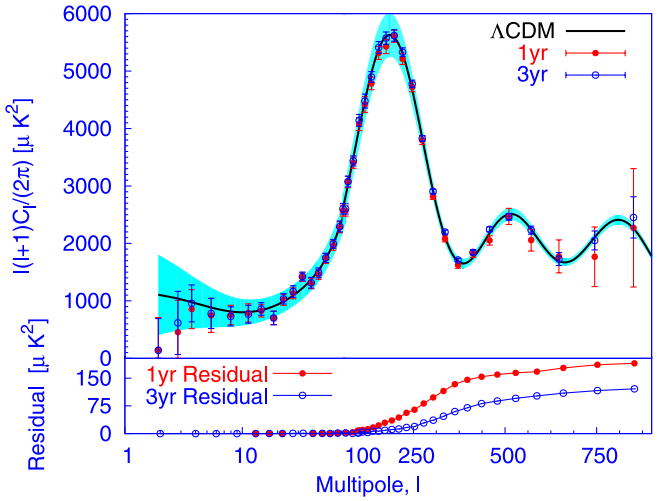

FIG. 10 (color online). Comparison of the 1-year 4-channel with that of the 3-year 4-channel power spectrum. The best fit WMAP power spectrum is shown by the solid (black) line along with the cosmic variance band. The bottom panel of this figure shows the residual unresolved point source contamination for both the power spectra.

show the WMAP 3-year power spectrum and the filled (red) points show the 1-year spectrum. Both of these spectra are corrected for residual unresolved point source contamination. In the lower panel of this figure we show the residual unresolved point source contamination for the 1-year and 3-year data, respectively.

As shown in the lower panel of Fig. 10, residual unresolved point source contamination is less by about 50 μK^2 in the 3-year power spectrum in comparison to the 1-year power spectrum. This is because the WMAP-supplied 3-year Kp2 mask removes more point sources than the 1-year Kp2 mask. Also, the original l^2 dependence of the unresolved point source power spectrum present in the foreground contaminated maps (as well as present in template cleaned maps) is significantly reduced and becomes almost independent of l at large l indicating that a significant amount of point source removal is achieved during cleaning. In [22], we have shown that the residual point source contamination is smaller than the contamination arising from K, Ka, Q, or V bands and intermediate between V and W bands.

We perform a quadratic fit of the form $\Delta T_l = \Delta T_{l_0} + \alpha(l - l_0)^2$ to the peaks and troughs of the binned spectrum similar to WMAP analysis [37] using our point source corrected 1-year and 3-year power spectra. For the 1-year (3-year) power spectrum we obtain the first acoustic peak at $l = 219.8 \pm 0.8$ (219.9 ± 0.8) with the peak amplitude $\Delta T_l = 74.1 \pm 0.3$ μK (74.4 ± 0.3 μK), the second acoustic peak at $l = 544 \pm 17$ (539.5 ± 3.7) with the peak amplitude $\Delta T_l = 48.3 \pm 1.2$ μK (49.4 ± 0.4 μK), and the first trough at $l = 419.2 \pm 5.6$ μK (417.7 ± 3.2) with peak amplitude $\Delta T_l = 41.7 \pm 1$ μK (41.4 ± 0.6 μK). The left panel (right panel) of Fig. 11 shows the three different ranges of multipoles used to find out the peak and the trough positions and their corresponding amplitudes ΔT_l using the 1-year (3-year) power spectrum.

VII. CONCLUSION

We have developed a procedure to estimate the CMB power spectrum directly from the raw CMB data without the need of foreground templates. The procedure relies on the spectral dependence of the foregrounds and the fact that the CMB signal is independent of frequency. We construct foreground-cleaned maps by combining subsets of four DAs out of the ten available from WMAP. These foreground-cleaned maps are then cross correlated to remove the detector noise bias. We have successfully applied this procedure to the WMAP data. In our analysis we find that a relatively small contamination remains at large l due to unresolved point sources. We have developed a procedure to remove this contamination. The resulting power spectrum agrees well with that obtained by the WMAP science team.

A very promising feature of our foreground removal method is that one can model the power spectrum of

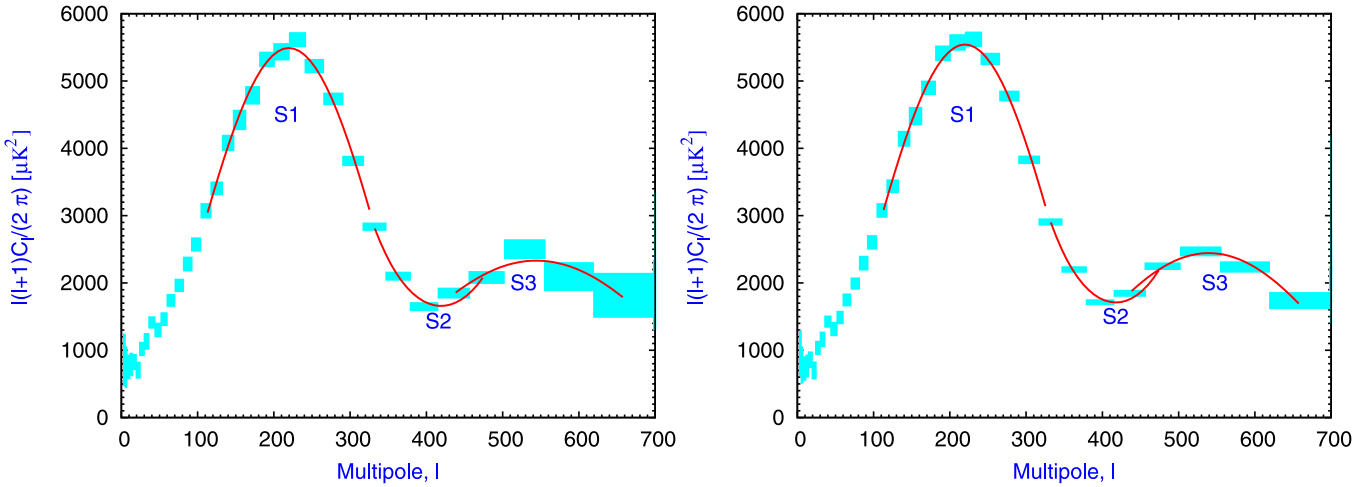

FIG. 11 (color online). The left panel shows the three different multipole ranges used to obtain positions of the first peak, the first trough, and the second peak from our point source subtracted power spectrum using 1-year WMAP data. Before fitting, the 1-year power spectrum was binned in the same manner as the WMAP's binning. The box error bars are used to indicate x and y error bars. The right panel shows the same figure but using the WMAP 3-year data. A correction due to residual unresolved point sources was performed prior to fitting.

residual contamination analytically. We have presented an in-depth study of such residuals for single iteration cleaning using the autopower spectrum of the cleaned maps. These residuals can be modeled by the empirical foreground and noise covariance matrix. We have also shown that there exists a negative bias in the power spectrum which is strongest at the lowest multipoles. This negative bias is directly determined by the underlying CMB power spectrum. Hence, as long as the negative bias is not very large, we may estimate it by using the extracted CMB power spectrum.

ACKNOWLEDGMENTS

The analysis pipeline as well as the entire simulation pipeline is based on primitives from the HEALPIX package. (The HEALPIX distribution is publicly available from the website http://healpix.jpl.nasa.gov.) We acknowledge the use of version 1.1 of the Planck reference sky model, prepared by the members of Working Group 2 and available at www.planck.fr/heading79.html. The entire analysis procedure was carried out on the IUCAA HPC facility as well as on the computing facilities at IAP. R. S. acknowledges support from the French Embassy in India and EGIDE in Paris, France. R. S. thanks IAP for hosting his visit. R. S. thanks Francois Bouchet, Christophe Pichon, Karim Benabed, Pawel Bielewicz, and Planck group members at IAP for useful and illuminating discussions. We are grateful to Lyman Page, Olivier Dore, Charles Lawrence, Kris Gorski, Hans Kristian Eriksen, and Max Tegmark for thoughtful comments and suggestions. We acknowledge a private communication with Gary Hinshaw on the unresolved point source model. We thank Amir Hajian, Subharthi Ray, and Sanjit Mitra at IUCAA for helpful discussions. We thank the WMAP team for producing excellent quality CMB maps and making them publicly available.

APPENDIX A: ANALYTIC DERIVATION OF WEIGHTS AND CLEANED POWER SPECTRUM

The main idea behind the blind foreground cleaning method used here is entirely based upon the minimization of total power in the cleaned map in multipole space [18]. Weights for different channels are obtained by minimizing the total power $\hat{C}_l^{\rm Clean}$ of the cleaned map. However, we ensure that the CMB angular power spectrum is conserved during cleaning by imposing the constraint $\hat{\mathbf{W}}_l \mathbf{e}_0 \mathbf{e}_0^T \hat{\mathbf{W}}_l^T = 1$ on the weights. The solution for the weights that satisfy these conditions is the point in weight space where normals to the functions $f(\hat{\mathbf{W}}_l) = \hat{\mathbf{W}}_l \hat{\mathbf{C}}_l \hat{\mathbf{W}}_l^T$ and $g(\hat{\mathbf{W}}_l) = \hat{\mathbf{W}}_l \mathbf{e}_0$ are parallel to one another. Following Lagrange's multiplier method, this is equivalent to minimizing

$$\hat{\mathbf{W}}_l \hat{\mathbf{C}}_l \hat{\mathbf{W}}_l^T - \lambda \hat{\mathbf{W}}_l \mathbf{e}_0. \tag{A1}$$

Here λ is the unknown Lagrange multiplier parameter which can be determined by the variational principle. At the extrema, the expression in Eq. (A1) is unchanged under small variations in $\hat{\mathbf{W}}_l$ leading to

$$\Delta\hat{\mathbf{W}}_l \hat{\mathbf{C}}_l \hat{\mathbf{W}}_l^T + \hat{\mathbf{W}}_l \hat{\mathbf{C}}_l \Delta\hat{\mathbf{W}}_l^T - \lambda \Delta\hat{\mathbf{W}}_l \mathbf{e}_0 = 0. \tag{A2}$$

Since the power spectrum matrix $\hat{\mathbf{C}}_l$ is a symmetric matrix, the first two terms of the left-hand side are equal to one another. Hence we obtain

$$\Delta\hat{\mathbf{W}}_l [2\hat{\mathbf{C}}_l \hat{\mathbf{W}}_l^T - \lambda \mathbf{e}_0] = 0. \tag{A3}$$

Since this relation is true for any arbitrary variation $\Delta\hat{\mathbf{W}}_l$,

we obtain

$$2\hat{\mathbf{C}}_l\hat{\mathbf{W}}_l^T - \lambda\mathbf{e}_0 = 0. \tag{A4}$$

We now multiply both sides of the above equation from the left by $\hat{\mathbf{C}}_l\hat{\mathbf{G}}_l$, where $\hat{\mathbf{G}}_l$ is the MPGI of $\hat{\mathbf{C}}_l$. Using the condition of the MPGI, $\hat{\mathbf{C}}_l = \hat{\mathbf{C}}_l\hat{\mathbf{G}}_l\hat{\mathbf{C}}_l$, we obtain

$$2\hat{\mathbf{C}}_l\hat{\mathbf{W}}_l^T - \lambda\hat{\mathbf{C}}_l\hat{\mathbf{G}}_l\mathbf{e}_0 = 0. \tag{A5}$$

Using Eqs. (A4) and (A5) we find that $\mathbf{e}_0 = \hat{\mathbf{C}}_l\hat{\mathbf{G}}_l\mathbf{e}_0$, where $\hat{\mathbf{C}}_l\hat{\mathbf{G}}_l$ is the orthogonal projector on $\mathcal{C}(\hat{\mathbf{C}}_l)$. This implies that $\mathbf{e}_0 \in \mathcal{C}(\hat{\mathbf{C}}_l)$. If we multiply Eq. (A4) by $\hat{\mathbf{G}}_l$ from the left, we obtain

$$2\hat{\mathbf{G}}_l\hat{\mathbf{C}}_l\hat{\mathbf{W}}_l^T - \lambda\hat{\mathbf{G}}_l\mathbf{e}_0 = 0. \tag{A6}$$

Since the matrix $\hat{\mathbf{C}}_l$ is symmetric, using the conditions of the MPGI, one can verify that $\hat{\mathbf{G}}_l$ is also a symmetric matrix. Using one of the defining properties of the MPGI, namely, that $\hat{\mathbf{C}}_l\hat{\mathbf{G}}_l$ is Hermitian, we obtain

$$\hat{\mathbf{C}}_l\hat{\mathbf{G}}_l = \hat{\mathbf{G}}_l\hat{\mathbf{C}}_l, \tag{A7}$$

i.e., $\hat{\mathbf{G}}_l$ and $\hat{\mathbf{C}}_l$ commute. Using this condition in Eq. (A6), we obtain

$$2\hat{\mathbf{C}}_l\hat{\mathbf{G}}_l\hat{\mathbf{W}}_l^T - \lambda\hat{\mathbf{G}}_l\mathbf{e}_0 = 0. \tag{A8}$$

We note that if $\hat{\mathbf{W}}_l^T \in \text{Null}\,(\hat{\mathbf{C}}_l)$, i.e., if $\hat{\mathbf{C}}_l\hat{\mathbf{W}}_l^T = 0$, then the total power in the cleaned map, $\hat{\mathbf{W}}_l\hat{\mathbf{C}}_l\hat{\mathbf{W}}_l^T$, vanishes. Therefore we require that the weights satisfy $\hat{\mathbf{W}}_l^T \notin \text{Null}\,(\hat{\mathbf{C}}_l)$. If we decompose $\hat{\mathbf{W}}_l^T$ in terms of two components $\hat{\mathbf{W}}_{1l}^T \in \text{Null}\,(\hat{\mathbf{C}}_l)$ and $\hat{\mathbf{W}}_{2l}^T \in \mathcal{C}(\hat{\mathbf{C}}_l)$, then it is easy to see that the total power in the cleaned map is determined by the $\hat{\mathbf{W}}_{2l}^T$ component, i.e., $\hat{\mathbf{W}}_l\hat{\mathbf{C}}_l\hat{\mathbf{W}}_l^T = \hat{\mathbf{W}}_{2l}\hat{\mathbf{C}}_l\hat{\mathbf{W}}_{2l}^T$. In other words the null space component, $\hat{\mathbf{W}}_{1l}^T$, does not have any effect on the cleaned map power spectrum. Using the condition $\mathbf{e}_0 \in \mathcal{C}(\hat{\mathbf{C}}_l)$, we note that $\mathbf{e}_0^T\hat{\mathbf{W}}_l^T = \mathbf{e}_0^T\hat{\mathbf{W}}_{1l}^T + \mathbf{e}_0^T\hat{\mathbf{W}}_{2l}^T = \mathbf{e}_0^T\hat{\mathbf{W}}_{2l}^T = 1$ since $\mathbf{e}_0^T\hat{\mathbf{W}}_{1l}^T = 0$. Thus $\hat{\mathbf{W}}_{2l}^T$ satisfies the condition required to preserve the CMB component in the cleaned map. For this reason we consider only the component of weight that lies on $\mathcal{C}(\hat{\mathbf{C}}_l)$ and ignore any component that belongs to Null $(\hat{\mathbf{C}}_l)$ in the following discussion. Effectively this implies that $\hat{\mathbf{W}}_l^T \in \mathcal{C}(\hat{\mathbf{C}}_l)$. Since $\hat{\mathbf{W}}_l^T \in \mathcal{C}(\hat{\mathbf{C}}_l)$ one can show that

$$\hat{\mathbf{C}}_l\hat{\mathbf{G}}_l\hat{\mathbf{W}}_l^T = \hat{\mathbf{W}}_l^T. \tag{A9}$$

Using this relation in Eq. (A8), we obtain

$$\hat{\mathbf{W}}_l^T = \frac{\lambda}{2}\hat{\mathbf{G}}_l\mathbf{e}_0. \tag{A10}$$

Multiplying both sides of the above equation from the left by $\mathbf{e}_0^T$ and using $\mathbf{e}_0^T\hat{\mathbf{W}}_l^T = 1$, we obtain

$$\lambda = \frac{2}{\mathbf{e}_0^T\hat{\mathbf{G}}_l\mathbf{e}_0}. \tag{A11}$$

Using Eqs. (A10) and (A11), we obtain

$$\hat{\mathbf{W}}_l^T = \frac{\hat{\mathbf{G}}_l\mathbf{e}_0}{\mathbf{e}_0^T\hat{\mathbf{G}}_l\mathbf{e}_0}. \tag{A12}$$

The corresponding power spectrum of the cleaned map is given by

$$\hat{\mathbf{W}}_l\hat{\mathbf{C}}_l\hat{\mathbf{W}}_l^T = \frac{1}{\mathbf{e}_0^T\hat{\mathbf{G}}_l\mathbf{e}_0}, \tag{A13}$$

where we have used the first condition of the MPGI, $\hat{\mathbf{G}}_l\hat{\mathbf{C}}_l\hat{\mathbf{G}}_l = \hat{\mathbf{G}}_l$. Equation (A13) has an interesting property. It remains valid even when the full covariance matrix $\hat{\mathbf{C}}_l$ is singular. A singular full covariance matrix is encountered in the noiseless case (or numerically, the very low noise case) when the effective total number of foreground and CMB components becomes less than the number of bands available for the linear combination. If $\hat{\mathbf{C}}_l$ is not singular we can replace $\hat{\mathbf{G}}_l$ everywhere by $\hat{\mathbf{C}}_l^{-1}$. This is because the MPGI of a nonsingular square matrix is its inverse.

APPENDIX B: PARTITIONING THE SKY

An important advantage of our foreground cleaning procedure is that we can allow the weights to vary with sky positions as well as with the multipole moment. To allow the weights to vary with sky positions, we can partition the sky into several regions depending upon the level of foreground contamination. (Alternatively, the partitioning could be done directly using the knowledge of the expected spectral index dependence on the sky.) We followed the procedure of [18] to partition the sky. Each of these partitions is identified with a sky mask. The mask values are nonzero within the sky partition and zero outside. In this section we describe the procedure of constructing these masks.

There are 2 DAs for the Q band, 2 for the V band, and 4 for the W band. To make the masks we first average all the DA maps for a given frequency band. Correspondingly, we average the beam functions for each frequency band. For the K and KA bands there is only one DA in each case. Therefore, for these bands no averaging was done. We smoothed all five maps (e.g., K, Ka, Q, V, W) by the resolution function of the K band, which has the lowest resolution. We obtained four difference maps, W-V, V-Q, Q-K, K-Ka, out of these five smoothed maps. The smoothing was performed by first obtaining the a_{lm} coefficients. Since each averaged map was effectively smoothed by the averaged beam function (corresponding to each channel) during the observations, we first remove the beam effect. Then we smooth each map by the common resolution of the lowest frequency band. Mathematically, the difference maps are obtained as follows:

$$a_{lm}^{W-V} = a_{lm}^W\frac{B_l^K}{B_l^W} - a_{lm}^V\frac{B_l^K}{B_l^V}, \tag{B1}$$

$$a_{lm}^{V-Q} = a_{lm}^{V}\frac{B_l^K}{B_l^V} - a_{lm}^{Q}\frac{B_l^K}{B_l^Q}, \tag{B2}$$

$$a_{lm}^{Q-K} = a_{lm}^{Q}\frac{B_l^K}{B_l^Q} - a_{lm}^{K}, \tag{B3}$$

$$a_{lm}^{K-KA} = a_{lm}^{K} - a_{lm}^{KA}\frac{B_l^K}{B_l^{KA}}. \tag{B4}$$

The a_{lm} are converted to difference maps using the HEALPIX supplied subroutine `alm2map`.

Next we construct a junk map out of these four difference maps assigning at each pixel the absolute maximum value among the four difference maps. We down-sample the junk map using the HEALPIX supplied program `udgrade` to a resolution of $N_{\rm side} = 64$. We identify seven different sky mask partitions from this low resolution junk map after applying cutoffs corresponding to the following temperature thresholds (in μK): $T > 30\,000$, $30\,000 \geq T > 10\,000$, $10\,000 \geq T > 3000$, $3000 \geq T > 1000$, $1000 \geq T > 300$, $300 \geq T > 100$, and $T < 100$. The partition $T > 30\,000$ is maximally contaminated by foreground emission. The second dirtiest partition is disjoint on the sky, and we use three separate masks corresponding to this partition. The resulting nine masks are then converted back to the HEALPIX resolution $N_{\rm side} = 512$ using the `udgrade` routine of HEALPIX. Next we smooth each mask using a Gaussian beam of FWHM $30'$ and redefine smoother mask boundaries at the threshold of 0.5.[8] In this case we find that almost the entire sky is covered by the nine masks (except for a few pixels in the sky).

To convert the second dirtiest region into three mask files, we upgrade the $N_{\rm side} = 64$ resolution map to the $N_{\rm side} = 512$. Then using the `mollcursor` facility, we find out the extension in galactic θ, ϕ coordinates of the three different parts of the second dirtiest region. We then implement a method which determines the pixel index in "ring" format for these three regions. Finally, we converted them to three different masks at $N_{\rm side} = 512$. The masks are then smoothed by the Gaussian function of FWHM $= 30'$. We applied a cutoff of 0.5 to each of them. The 8th, 7th, and 6th masks are numbered according to the descending order of maximum pixel value in the junk map at resolution $N_{\rm side} = 512$.

[8]We find that if we smooth them by a Gaussian function of 2 degree FWHM (as mentioned in [18]) and then define the boundaries at the threshold of 0.5, the resulting nine masks do not cover the entire sky. In that case there are some regions near the galactic plane which do not belong to any of our nine masks. Obviously, if such regions are not covered by any of the masks and remain present in our final cleaned map, then the final power spectrum will be contaminated by the foreground. Therefore, we choose to use a Gaussian function of lower FWHM of $30'$ for smoothing. The reason why in our case masks smoothed by a 2° Gaussian function do not cover the entire sky is clear. This is actually dependent on the common beam function by which we are smoothing each map before forming the difference maps. We find that the masks near the galactic regions contain a few small isolated regions. Therefore smoothing by a Gaussian with FWHM as large as 2° gives this isolated small regions a maximum value far less than unity; in fact, the maximum value becomes quite near 0.5, the cutoff value, after smoothing. Consequently, applying a cut of 0.5 removes a considerable portion of these isolated regions, leading to some parts of the sky not being covered by the masks.

APPENDIX C: ESTIMATION OF RESIDUAL UNRESOLVED POINT SOURCE CONTAMINATION

In this appendix we first present a detailed analytical method for computing the unresolved point source contamination in the power spectrum from WMAP data. The analytic approach assumes that at large l the weights do not show significant fluctuations. Hence they are approximately the same for all random realizations. We demonstrate the validity of this assumption by detailed analytical calculations and Monte Carlo simulations.

Let us assume that $a_{lm}^{\mathbf{i}}$ are the spherical harmonic coefficients obtained from cleaned map C$\mathbf{i}$ or CA$\mathbf{i}$, where $\mathbf{i} = 1, 2, \ldots, 24$,

$$a_{lm}^{\mathbf{i}} = B_l^{\mathbf{i}} a_{lm}^{s} + B_l^{\mathbf{i}} a_{lm}^{\rm ps(i)}. \tag{C1}$$

Here $a_{lm}^{\rm ps(i)}$ is the contribution due to the residual unresolved point source contamination. We do not consider any detector noise contribution in this equation. However, this does not imply any loss of generality, as we wish to compute point source bias in the cross-power spectrum where detector noise bias drops out. Also, we do not include any residual diffuse foreground. In the high l regime where point source contribution becomes significant, the diffuse galactic contamination is entirely subdominant and thus is not of concern.

We assume a statistically isotropic model of the residual unresolved point source distribution over the sky. The point sources are uncorrelated with the CMB. On the ensemble average a partial cross-power spectrum, $\tilde{C}_l^{\mathbf{ij}}$, obtained by cross correlating cleaned maps C$\mathbf{i}$ and CA$\mathbf{j}$ is related to the full-sky CMB and point source power spectra as follows:

$$\langle \tilde{C}_l^{\mathbf{ij}} \rangle = M_{ll'}(\langle \hat{C}_{l'}^{s} \rangle + \langle \hat{C}_{l'}^{\rm ps(ij)} \rangle) B_{l'}^{\mathbf{i}} B_{l'}^{\mathbf{j}} p_{l'}^2. \tag{C2}$$

Here, $M_{ll'}$ is the coupling matrix corresponding to the Kp2 mask, $\hat{C}_l^{\rm ps(ij)}$ is the estimate of the full-sky residual unresolved point source power spectrum, and $B_{l'}^{\mathbf{i}} B_{l'}^{\mathbf{j}} p_{l'}^2$ denotes the combined effect of beam and pixel smoothing. One may rewrite the above equation as

$$\langle \hat{C}_l^{s} \rangle = \frac{M_{ll'}^{-1} \langle \tilde{C}_{l'}^{\mathbf{ij}} \rangle}{B_l^{\mathbf{i}} B_l^{\mathbf{j}} p_l^2} - \langle \hat{C}_l^{\rm ps(ij)} \rangle. \tag{C3}$$

The next task is to obtain the estimates $\hat{C}_l^{\rm ps(ij)}$ themselves in each cross-power spectra. To compute them, let us assume

that $f^{\mathbf{i}}(\theta, \phi)$ is the residual point source function present in the **i**th cleaned map. If $g_n(\theta, \phi)$ is the point source residual present in the nth sky part, we have

$$f^{\mathbf{i}}(\theta, \phi) = \sum_{n=1}^{9} g_n(\theta, \phi). \tag{C4}$$

After expanding both sides in spherical harmonics, we obtain

$$a^{\mathbf{i}}_{lm} = \sum_{n=1}^{9} \tilde{a}^n_{lm}. \tag{C5}$$

The partial-sky unresolved residual point source modes could be written in terms of the full-sky modes using

$$\tilde{a}^{\,n}_{lm} = \sum_{l'm'} M^n_{lml'm'} a^n_{l'm'}. \tag{C6}$$

We note that $a^n_{l'm'}$ represents the residual point source contamination in the *entire cleaned map* obtained after the nth iteration which is obviously different from the contamination, $\tilde{a}^n_{lm}$, that is actually present in the nth sky part. The symbol $M^n_{lml'm'}$ denotes the mode coupling matrix for the spherical harmonic modes for the nth sky part [30],

$$M^n_{lml'm'} = \sum_{l''m''} w^n_{l''m''}(-1)^{m'}\left[\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi}\right]^{1/2} \times \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} l & l' & l'' \\ m & -m' & m'' \end{pmatrix}, \tag{C7}$$

where w^n_{lm} are the spherical harmonic coefficients of the nth mask. We can rewrite $a^n_{l'm'}$ in terms of the spherical harmonic coefficients of the temporary maps obtained at nth iteration as

$$a^n_{l'm'} = \sum_{a=1}^{4} a^{na}_{l'm'} \hat{w}^{na}_{l'}, \tag{C8}$$

where $\hat{w}^{na}_l$ is the weight for the ath channel at the nth iteration for multipole l. Using Eqs. (C5), (C6), and (C8) we obtain

$$a^{\mathbf{i}}_{lm} = \sum_{n=1}^{9}\sum_{l'm'} M^n_{lml'm'} \sum_{a=1}^{4} a^{na}_{l'm'} \hat{w}^{na}_{l'}. \tag{C9}$$

Now using $\langle \hat{C}^{\mathrm{ps}(\mathbf{ij})}_l \rangle = \sum_{m=-l}^{m=l} \langle a^{\mathbf{i}}_{lm} a^{\mathbf{j}*}_{lm} \rangle / (2l+1)$ we obtain

$$\langle \hat{C}^{\mathrm{ps}(\mathbf{ij})}_l \rangle = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{n,n'=1}^{9} \sum_{l',m'} \sum_{l'',m''} M^n_{lml'm'} M^{n'*}_{lml''m''} \times \sum_{a,a'=1}^{4} \langle a^{na}_{l'm'} a^{n'a'*}_{l''m''} \hat{w}^{na}_{l'} \hat{w}^{\prime n'a'}_{l''} \rangle. \tag{C10}$$

Because of beam deconvolution the effective noise of the DA maps dominates over the unresolved point source contamination at the large l regime. Hence weights become entirely determined by the noise level of the maps and asymptotically ($l \to \infty$) become constant for all realizations.[9] Using simulations of the cleaning procedure with a realistic model of point sources and detector noise, we have verified that the weights remain effectively unchanged whether we use point sources or not. Also, weights are uncorrelated with unresolved point sources, as they are determined by the detector noise only. In this case the residual unresolved point source contamination becomes

$$\langle C^{\mathrm{ps}(\mathbf{ij})}_l \rangle = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{n,n'=1}^{9} \sum_{l',m'} \sum_{l'',m''} M^n_{lml'm'} M^{n'*}_{lml''m''} \times \sum_{a,a'=1}^{4} \langle a^{na}_{l'm'} a^{n'a'*}_{l''m''} \rangle \langle \hat{w}^{na}_{l'} \hat{w}^{\prime n'a'}_{l''} \rangle. \tag{C11}$$

Now using isotropic distribution of point sources over the sky, i.e., $\langle a^{na}_{l'm'} a^{n'a'*}_{l''m''} \rangle = C^{\mathrm{ps}(aa')}_{l'} \delta_{l'l''} \delta_{m'm''}$, where $C^{\mathrm{ps}(ab)}_{l'}$ is the point source model as supplied by the WMAP team, we simplify Eq. (C11) as

$$\langle C^{\mathrm{ps}(\mathbf{ij})}_l \rangle = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{n,n'=1}^{9} \sum_{l',m'} M^n_{lml'm'} M^{n'*}_{lml'm'} \times \sum_{a,a'=1}^{4} C^{\mathrm{ps}(aa')}_{l'} \langle \hat{w}^{na}_{l'} \hat{w}^{\prime n'a'}_{l'} \rangle, \tag{C12}$$

which can also be written as

$$\langle \hat{C}^{\mathrm{ps}(\mathbf{ij})}_l \rangle = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{n,n'=1}^{9} \sum_{l',m'} M^n_{lml'm'} M^{n'*}_{lml'm'} \times \langle \hat{\mathbf{W}}^n_{l'} \mathbf{C}^{\mathrm{ps}}_{l'} \hat{\mathbf{W}}^{\prime n'}_{l'} \rangle. \tag{C13}$$

Here, $\hat{\mathbf{W}}^n_l$ is the row vector of weights for sky part n as in Eq. (7). The elements of the point source power spectrum matrix, $\mathbf{C}^{\mathrm{ps}}_l$, are given by $C^{\mathrm{ps}(ab)}_l$. We note that, in general, point sources may not be perfectly correlated at all the frequencies. The point source power spectrum assumed by the WMAP team and used in this work assumes the existence of a single point source template which may not be perfectly true. However, our point source correction method can easily incorporate more detailed models of the point source power spectrum.

The weights are asymptotically constant from realization to realization. Hence one can approximate $\langle \hat{\mathbf{W}}_l \mathbf{C}^{\mathrm{ps}}_l \hat{\mathbf{W}}^T_l \rangle$ as $\hat{\mathbf{W}}_l \mathbf{C}^{\mathrm{ps}}_l \hat{\mathbf{W}}^T_l$; i.e. one does not need to compute the ensemble average. Thus we have $\langle \hat{C}^{\mathrm{ps}(\mathbf{ij})}_l \rangle \approx \hat{C}^{\mathrm{ps}(\mathbf{ij})}_l$, where

[9] Small differences in weights from realization to realization due to fluctuations in noise from the mean level are not a concern. These fluctuations could be further suppressed by computing the binned power spectrum.

$$\hat{C}_l^{\mathrm{ps(ij)}} = \frac{1}{2l+1} \times \sum_{m=-l}^{m=l} \sum_{n,n'=1}^{9} \sum_{l',m'} M^n_{lml'm'} M^{n'*}_{lml'm'} \hat{\mathbf{W}}^n_{l'} \mathbf{C}^{\mathrm{ps}}_{l'} \hat{\mathbf{W}}^{l'n'T}_{l'}. \tag{C14}$$

Equation (C14) can further be written as

$$\hat{C}_l^{\mathrm{ps(ij)}} = \sum_{n=1}^{9} \sum_{l'} M^n_{ll'} \hat{\mathbf{W}}^n_{l'} \mathbf{C}^{\mathrm{ps}}_{l'} \hat{\mathbf{W}}^{l'n'T}_{l'} + \frac{1}{2l+1} \times \left(\sum_{m=-l}^{m=l} \sum_{n,n'(n\neq n')} \sum_{l',m'} M^n_{lml'm'} M^{n'*}_{lml'm'} \hat{\mathbf{W}}^n_{l'} \mathbf{C}^{\mathrm{ps}}_{l'} \hat{\mathbf{W}}^{l'n'T}_{l'} \right). \tag{C15}$$

Here $M^n_{ll'}$ is the coupling matrix for the sky part n. The first term on the right-hand side is computed by using the formula for $M^n_{ll'}$ given in [30]. The last term can also be further simplified. Using Eq. (C7) and properties of the Wigner $3jm$ symbol, we find

$$\frac{1}{2l+1} \sum_{mm'} M^n_{lml'm'} M^{n'*}_{lml'm'} = \frac{(2l'+1)}{4\pi} \sum_{l''m''} \sum_{l'''m'''} w^n_{l''m''} w^{l'n'}_{l'''m'''} \times \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}^2. \tag{C16}$$

This is easily written in terms of the cross-power spectra $w^{nn'}_{l''}$ of two masks n, n',

$$\frac{1}{2l+1} \sum_{mm'} M^n_{lml'm'} M^{n'*}_{lml'm'} = \frac{(2l'+1)}{4\pi} \sum_{l''} (2l''+1) w^{nn'}_{l''} \times \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}^2 = M^{nn'}_{ll'}. \tag{C17}$$

Here we have defined $M^{nn'}_{ll'}$ as the cross-coupling matrix of two masks n, n'. Using the above relation in Eq. (C15), we obtain Eq. (19) which estimates the unresolved residual point source contamination in the individual cross-power spectrum.

1. Autopower spectrum

In this section we show, using Monte Carlo simulations, that the residual point source bias in the autopower spectrum of the cleaned map can be modeled as $\hat{\mathbf{W}}_l \mathbf{C}^{\mathrm{ps}}_l \hat{\mathbf{W}}^T_l$. To estimate the point source bias we first obtain an analytical model of the detector noise bias. This noise bias is subtracted from the cleaned power spectrum. We also give an analytical model of the point source bias.

The cleaned autopower spectrum obeys Eq. (32). Clearly the positive bias is given by $\langle 1/\mathbf{e}_0^T(\hat{\mathbf{C}}_l^{f+N})^{-1}\mathbf{e}_0\rangle$ which is caused by the detector noise and foreground covariance matrix of the map. In this appendix we are basically interested in studying point source bias. Since diffuse foreground is not a concern at large l, we do not include it in this study. Considering the second term on the right-hand side of Eq. (32), we note that it is not possible to identify a bias which solely comes from point sources. Another point to note is that we know only the mean noise covariance matrix $\mathbf{C}_l^N$, not the empirical noise covariance matrix $\hat{\mathbf{C}}_l^N$, where $\hat{\mathbf{C}}_l^N = \mathbf{C}_l^N + \delta\hat{\mathbf{C}}_l^N$ and $\delta\hat{\mathbf{C}}_l^N$ denotes noise fluctuations. Therefore, we express $\langle 1/\mathbf{e}_0^T(\hat{\mathbf{C}}_l^{f+N})^{-1}\mathbf{e}_0\rangle$ in terms of a mean noise covariance matrix and the foreground covariance matrix. As we will see, such a simplification helps us to obtain an analytical expression for the point source bias. Keeping in mind that the WMAP detector noise level is much larger than the point source spectrum, we simplify $\langle 1/\mathbf{e}_0^T(\hat{\mathbf{C}}_l^{f+N})^{-1}\mathbf{e}_0\rangle$ analytically by expanding to first order,[10] in terms of $\mathbf{C}_l^{\mathrm{ps}}(\hat{\mathbf{C}}_l^N)^{-1}$,

$$(\hat{\mathbf{C}}_l^{f+N})^{-1} = (\hat{\mathbf{C}}_l^N)^{-1} - (\hat{\mathbf{C}}_l^N)^{-1}\mathbf{C}_l^{\mathrm{ps}}(\hat{\mathbf{C}}_l^N)^{-1}. \tag{C18}$$

Because of the low point source contamination relative to the WMAP noise level, it is reasonable to assume that $(\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{C}_l^{\mathrm{ps}}(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0)/(\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0) \ll 1$. Expanding to first order, we obtain

$$\left\langle \frac{1}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^{f+N})^{-1}\mathbf{e}_0} \right\rangle = \left\langle \frac{1}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0} \right\rangle + \left\langle \frac{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0} \mathbf{C}_l^{\mathrm{ps}} \frac{(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0} \right\rangle. \tag{C19}$$

We interpret the first term on the right-hand side as the detector noise bias. The second term is the point source bias. In the following subsections we recast these results in terms of the mean noise covariance matrix. We also verify our analytic expressions using Monte Carlo simulations.

a. Noise induced bias

We first turn our attention to the detector noise bias. We estimate this term up to second order in $\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}$,

$$\left\langle \frac{1}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0} \right\rangle = \left\langle \frac{1}{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0} \times \left(1 + \frac{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0} - \frac{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0} + \frac{(\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\mathbf{e}_0)^2}{(\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0)^2} \right) \right\rangle. \tag{C20}$$

[10] We note that we cannot use the Sherman-Morrison formula, discussed in Appendix D, to decouple the foreground and detector noise bias from the term $\langle 1/\mathbf{e}_0^T(\hat{\mathbf{C}}_l^{f+N})^{-1}\mathbf{e}_0\rangle$ in a useful form.

The noise fluctuations are assumed to have zero mean, $\langle \delta\hat{\mathbf{C}}_l^N \rangle = \mathbf{0}$. Hence the second term in the brackets vanishes on the ensemble average. Only the numerators of the third and fourth terms in the brackets on the right-hand side of Eq. (C20) are stochastic variables. On the ensemble average the numerator of the third term becomes

$$\langle \mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\mathbf{e}_0\rangle = \sum_{a,b,c} C_l^{Na}C_l^{Nb}C_l^{Nc}\langle \delta\hat{C}_l^{N(ab)}\delta\hat{C}_l^{N(bc)}\rangle,$$

where we have assumed that the noise covariance matrix of WMAP is diagonal. We simplify the ensemble-averaged quantities using Eq. (28) of [31]. If we consider a term of the form $\langle \delta\hat{C}_l^{N(ab)}\delta\hat{C}_l^{N(cd)}\rangle$, then for uncorrelated detector noises, the ensemble-averaged quantities will survive only for the same pair, i.e., when $(a, b) = (c, d)$. We further note that

$$\langle \delta\hat{C}_l^{N(ab)^2}\rangle = \begin{cases} \frac{1}{2l+1}C_l^{Na}C_l^{Nb} & a \neq b, \\ \frac{2}{2l+1}C_l^{Na}C_l^{Na} & a = b. \end{cases} \tag{C21}$$

Using the above relations the third and fourth terms in the brackets of Eq. (C20) become

$$\left\langle \frac{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0}\right\rangle = \frac{(1+n_c)}{2l+1}, \tag{C22}$$

$$\left\langle \frac{(\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}\mathbf{e}_0)^2}{(\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0)^2}\right\rangle = \frac{2}{2l+1}, \tag{C23}$$

respectively. Using Eqs. (C20), (C22), and (C23) we obtain

$$\left\langle \frac{1}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}\right\rangle = \frac{1}{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0}\frac{2l+2-n_c}{2l+1}. \tag{C24}$$

This enables us to compute the noise bias in terms of the theoretical noise models. We verified this equation by Monte Carlo simulations of 1000 noise covariance matrices corresponding to WMAP V and W bands. The noise maps used for computing these empirical noise covariance matrices are obtained using a method similar to that discussed in Sec. IV. In Fig. 12 we show the difference between the empirical noise bias $\langle 1/\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0\rangle$ and the second-order analytical expression for noise bias, given by the right-hand side of Eq. (C24). The difference is consistent with zero. The large fluctuations at large l are caused by the beam deconvolution effect in the effective noise covariance matrix.

b. Point source bias

Here we study the point source bias. We perform Monte Carlo simulations using single iteration cleaning. Besides CMB, the simulated maps contain contributions from point sources and detector noise, generated using

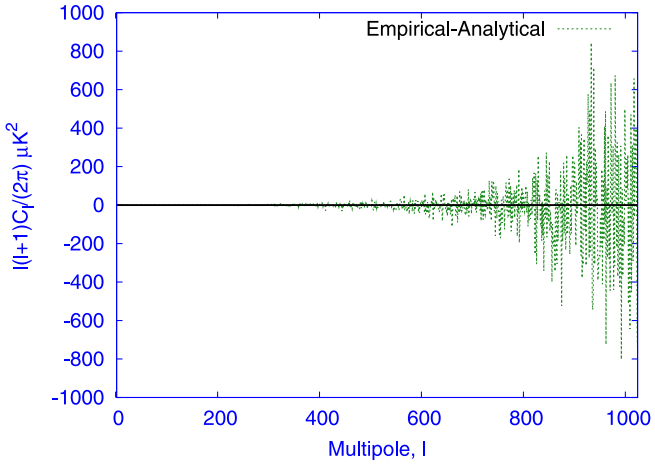

FIG. 12 (color online). The difference between the empirical noise bias and the theoretical estimate of noise bias given in Eq. (C24) for V and W bands.

WMAP's detector noise model. We compute the ensemble-averaged cleaned power spectrum $\langle \hat{C}_l^{\text{Clean}}\rangle$ from 1000 such simulations. The validity of Eq. (C24) allows us to compute noise bias in terms of the theoretical noise covariance matrix and subtract this analytical result from $\langle \hat{C}_l^{\text{Clean}}\rangle$. After correcting for noise bias we find that $\langle \hat{C}_l^{\text{Clean}}\rangle - (1/\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0)((2l+2-n_c)/(2l+1))$ still has a residual positive bias at the multipole range $l >$ 400. This excess is a clear demonstration of the presence of point source bias. The solid (brown) line in Fig. 13 is the power spectrum, corrected for the noise and the negative CMB bias of the form $-C_l^s/(2l+1)$. The excess is caused by the point source bias. We analytically compute the point source bias in terms of the theoretical noise covariance matrix similar to Eq. (C24). Following a first-order expan-

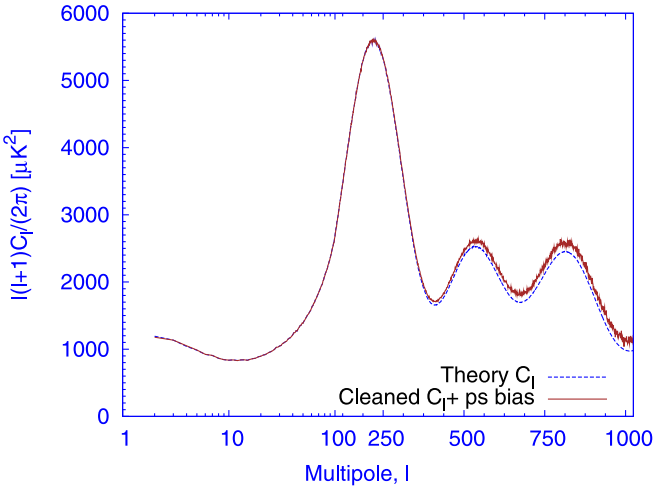

FIG. 13 (color online). The autopower spectrum without correcting for the residual unresolved point source bias is shown by the solid (brown) line. The dashed (blue) line shows the theoretical CMB power spectrum.

sion in $\mathbf{C}_l^{\rm ps}(\hat{\mathbf{C}}_l^N)^{-1}$ and a second-order expansion in terms of $\delta\hat{\mathbf{C}}_l^N(\mathbf{C}_l^N)^{-1}$, we obtain

$$\left\langle \frac{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}\mathbf{C}_l^{\rm ps}\frac{(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}\right\rangle = \frac{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}}{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0}\mathbf{C}_l^{\rm ps}\frac{(\mathbf{C}_l^N)^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0}\frac{2l}{2l+1} + \frac{\mathrm{tr}(\mathbf{C}_l^{\rm ps}(\mathbf{C}_l^N)^{-1})}{(2l+1)(\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0)}. \quad \text{(C25)}$$

We again verify the validity of the above expression using Monte Carlo simulations. We compute the point source bias from 1000 cleaning simulations using $\langle\hat{C}_l^{\rm Clean}\rangle - C_l^s - C_l^s/(2l+1) - (1/\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0)((2l+2-n_c)/(2l+1))$, where C_l^s denotes the theoretical CMB power spectrum. In Fig. 14 the circular (green) points show the binned estimate of the point source bias from simulations. The error bars plotted in this figure are valid for the mean of 1000 simulations of random samples. The starred (blue) points show the analytical estimate of the point source bias computed from the right-hand side of Eq. (C25). These two results match closely except at the last two points. However, one can easily conclude from the error bars plotted in this figure that differences between the two lines at these two points are not significant. We interpret these deviations as noise fluctuations.

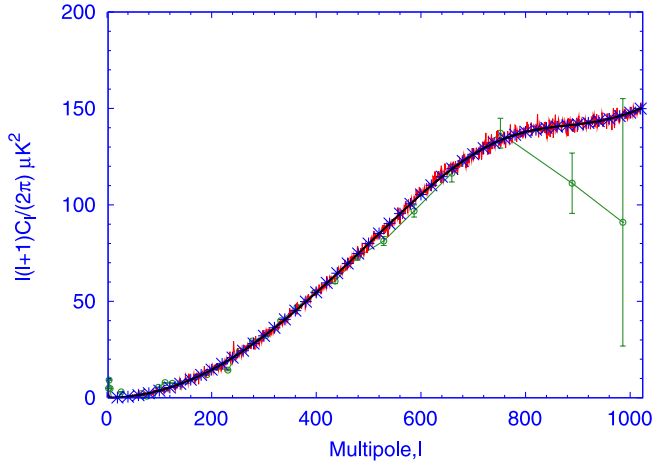

FIG. 14 (color online). The point source bias $\langle\hat{\mathbf{W}}_l\mathbf{C}_l^{\rm ps}\hat{\mathbf{W}}_l^T\rangle$, solid (black) line, and the approximation $\hat{\mathbf{W}}_l\mathbf{C}_l^{\rm ps}\hat{\mathbf{W}}_l^T$, wiggly (red) line, for the case of the autopower spectrum of the cleaned maps. The empirical point source bias computed by using $\langle\hat{C}_l^{\rm Clean}\rangle - C_l^s - C_l^s/(2l+1) - (1/\mathbf{e}_0^T(\mathbf{C}_l^N)^{-1}\mathbf{e}_0)((2l+2-n_c)/(2l+1))$ is shown by the unfilled circular points (green). The error bars are computed only for the term $\langle 1/\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0\rangle$ without including the cosmic variance. All the data points are binned following WMAP's 3-year binning method. The analytical point source bias computed using Eq. (C25) is also shown by the starred (blue) points.

Given the large noise level of WMAP maps, one might expect that $\langle\hat{\mathbf{W}}_l\mathbf{C}_l^{\rm ps}\hat{\mathbf{W}}_l^T\rangle$ would be a good approximation of the point source bias that is given by the second term of Eq. (C19). In the limit $\hat{\mathbf{C}}_l \to \hat{\mathbf{C}}_l^N$, applicable when noise dominates at large l and point source bias is also significant, these two terms are identical. Using the analytical form of weights as in Eq. (14), we find that up to the first order in $\mathbf{C}_l^{\rm ps}\hat{\mathbf{C}}_l^{N-1}$,

$$\langle\hat{\mathbf{W}}_l\mathbf{C}_l^{\rm ps}\hat{\mathbf{W}}_l^T\rangle = \left\langle \frac{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}\mathbf{C}_l^{\rm ps}\frac{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}\right\rangle - \frac{C_l^s}{2l+1}\left\langle\frac{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{C}_l^{\rm ps}(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\hat{\mathbf{C}}_l^N)^{-1}\mathbf{e}_0}\right\rangle + \frac{C_l^s}{2l+1}\mathrm{tr}(\mathbf{C}_l^{\rm ps}\langle(\hat{\mathbf{C}}_l^N)^{-1}\rangle). \quad \text{(C26)}$$

The last two terms in the above equations are negligibly small compared to the first term on the right-hand side. From simulations we find that these two terms contribute less than 0.1 μK^2 to the point source residual when multiplied by the prefactor $l(l+1)/(2\pi)$. The reason why these terms become negligible is that, apart from a multiplicative first-order term, $\mathbf{C}_l^{\rm ps}(\hat{\mathbf{C}}_l^N)^{-1}$, the first term on the right-hand side goes as $\mathbf{C}_l^N$, whereas the last two terms go as $C_l^s/(2l+1)$. Indeed, the noise becomes much larger than the CMB power spectrum at large l due to beam deconvolution, making the last two terms insignificant compared to the first term. In Fig. 15 we show the magnitude of the individual contributions arising from the second and third terms using 1000 Monte Carlo simulations of the cleaning method. We use V and W bands in the simulations using realistic WMAP noise levels and the residual unresolved point source model. The dashed-dotted (red) line shows the difference between the two terms which is less than 0.01 μK^2 or even smaller at $l > 400$. Hence it is justified

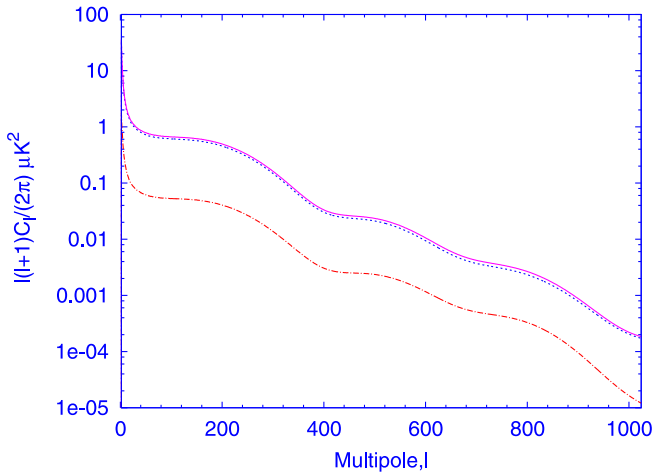

FIG. 15 (color online). The contributions arising from the third (dashed blue line) and fourth (solid pink line) terms of Eq. (C26) and their difference (dashed-dotted red line).

to assume

$$\langle \hat{\mathbf{W}}_l \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^T \rangle = \left\langle \frac{\mathbf{e}_0^T (\hat{\mathbf{C}}_l^N)^{-1}}{\mathbf{e}_0^T (\hat{\mathbf{C}}_l^N)^{-1} \mathbf{e}_0} \mathbf{C}_l^{\rm ps} \frac{\mathbf{e}_0^T (\hat{\mathbf{C}}_l^N)^{-1}}{\mathbf{e}_0^T (\hat{\mathbf{C}}_l^N)^{-1} \mathbf{e}_0} \right\rangle. \quad \text{(C27)}$$

In Fig. 14 we show the point source bias computed using $\langle \hat{\mathbf{W}}_l \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^T \rangle$ by the solid (black) line. This matches well with the starred (blue) points, justifying Eq. (C27).

The unresolved point source bias is then $\langle \hat{\mathbf{W}}_l \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^T \rangle$. Now we propose that the point source bias could also be approximated as $\hat{\mathbf{W}}_l \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^T \sim \langle \hat{\mathbf{W}}_l \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^T \rangle$. The reason is that, at large l where detector noise dominates, the weights are entirely determined by the mean noise level. Therefore weights become approximately constant from realization to realization. Small fluctuations in weights due to noise fluctuations are not important compared to the magnitude of total point source bias $\langle \hat{\mathbf{W}}_l \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^T \rangle$. The advantage of this method is that we can estimate point source bias in terms of the weights which are the results of a cleaning. In Fig. 14 we show how well $\hat{\mathbf{W}}_l \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^T$ (dashed-dotted red line) matches with $\langle \hat{\mathbf{W}}_l \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^T \rangle$. This result is obtained from 1000 Monte Carlo simulations of our cleaning method using V and W bands.

2. Cross-power spectrum

In this section we show, using Monte Carlo simulations, that a residual point source bias in the cross-power spectrum of the cleaned maps can be modeled as $\hat{\mathbf{W}}_l^1 \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^{2^T}$. Unlike the autopower case, we do not present any analytical estimate of the point source bias in this section. A detailed analytical result for the cross-power spectrum will be presented in a future publication.

The ensemble average of the cleaned cross-power spectrum is given by

$$\langle \hat{C}_l^{\rm Clean} \rangle = C_l^s + 2(1 - n_c) \frac{C_l^s}{2l+1} + \langle \hat{\mathbf{W}}_l^1 \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^{2^T} \rangle + \langle \hat{\mathbf{W}}_l^1 \hat{\mathbf{C}}_l^{N(12)} \hat{\mathbf{W}}_l^{2^T} \rangle. \quad \text{(C28)}$$

Here $\hat{\mathbf{W}}_l^1$ and $\hat{\mathbf{W}}_l^2$ are the weight vectors for the two cleaned maps and $\hat{\mathbf{C}}_l^{N(12)}$ denotes a chance noise correlation in the two sets of maps used to obtain the two cleaned maps. If significant, the last term on the right-hand side may cause a residual noise bias even in the cross-power spectrum. However, using Monte Carlo simulations with WMAP's noise level, we do not find any noise bias in cross-power spectra. Hence we ignore this quantity.

We study point source bias in the cross-power spectrum using 1000 Monte Carlo simulations with realistic point source and detector noise contamination. For simulation purpose we use V and W channels with the following cross-combinations, $v1w12 \otimes v2w34$, $v1w13 \otimes v2w24$, $v1w14 \otimes v2w23$, $v1w24 \otimes v2w13$, $v1w34 \otimes v2w12$. The ensemble-averaged cleaned power spectrum shows the existence of a point source bias at the large multipole range. The bias is clearly visible from the left panel of Fig. 16.

The quantity $\langle \hat{C}_l^{\rm Clean} \rangle - C_l^s - 2(1 - n_c) \frac{C_l^s}{2l+1}$ estimates point source bias in the cross-power spectrum. In the right panel of Fig. 16 we show this bias by the solid (red) line. We also compute $\langle \hat{\mathbf{W}}_l^1 \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^{2^T} \rangle$ using the weights from our

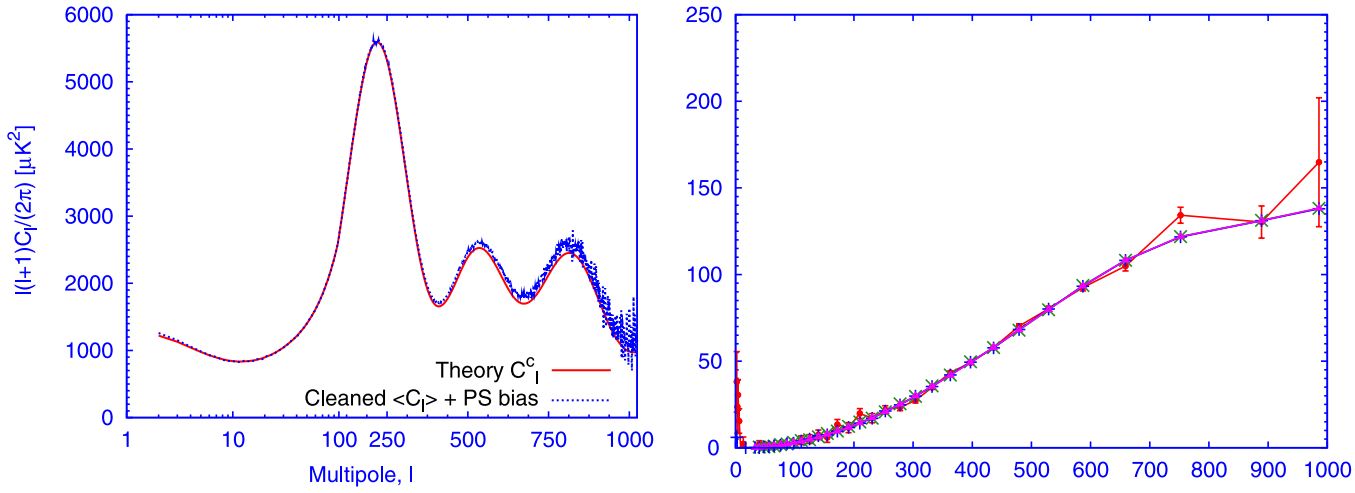

FIG. 16 (color online). Left panel: The point source bias at a large multipole range obtained from 1000 Monte Carlo simulations of the cleaning method involving V and W bands is shown in the dashed (blue) line. The negative bias $\frac{1-n_c}{2l+1} C_l^s$ has been removed before plotting the dashed (blue) line to show the bias coming from the point source only. The solid (red) line shows the theoretical power spectrum from which random realizations of CMB were drawn. Right panel: The point source bias computed from $\langle \hat{C}_l^{\rm Clean} \rangle - C_l^s - 2(1 - n_c) \frac{C_l^s}{2l+1}$ is shown by the filled circles with error bars (red). The triangular points (magenta) show point source bias computed using $\langle \hat{\mathbf{W}}_l^1 \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^{2^T} \rangle$. This shows excellent agreement with $\hat{\mathbf{W}}_l^1 \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^{2^T}$ computed from two arbitrary realizations shown by the plus (green) and cross (blue) symbols.

simulations. As this figure shows, the point source bias estimates, $\langle \hat{\mathbf{W}}_l^1 \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^{2^T} \rangle$ and $\hat{\mathbf{W}}_l^1 \mathbf{C}_l^{\rm ps} \hat{\mathbf{W}}_l^{2^T}$, from any two randomly chosen realizations match well with one another. Also, these quantities do not deviate significantly from the bias estimate $\langle \hat{C}_l^{\rm Clean} \rangle - C_l^s - 2(1-n_c)\frac{C_l^s}{2l+1}$. This justifies that the residual noise bias, $\langle \hat{\mathbf{W}}_l^1 \hat{\mathbf{C}}_l^{N(12)} \hat{\mathbf{W}}_l^{2^T} \rangle$, is ignorable in the cross-power spectra.

APPENDIX D: BIAS AT LOW l

In this appendix we derive Eq. (30). This derivation requires frequent use of generalized Sherman-Morrison formulas. In what follows we first state the generalized Sherman-Morrison formulas and then use them to derive Eq. (30).

1. Generalized Sherman-Morrison (GSM) formula for rank-one updates

The results of this section are mainly based upon the two papers [35,36]. In [35] analytic expressions for the MPGI of rank-one modified matrices of the form $\mathbf{M} = \mathbf{A} + \mathbf{bc}^\dagger$ (notations changed) are reported. Here $\mathbf{M}$, $\mathbf{A}$ are any $m \times n$ matrices, i.e., $\mathbf{M}, \mathbf{A} \in \mathbb{C}_{m,n}$, $\mathbf{b} \in \mathbb{C}_{m,1}$, and $\mathbf{c} \in \mathbb{C}_{n,1}$. $\mathbb{C}_{m,n}$ is the set of all $m \times n$ complex matrices, and $(\dagger)$ denotes conjugate transpose. The motivation of the work in [35] was to generalize the Sherman-Morrison formula

$$\mathbf{M}^{-1} = \mathbf{A}^{-1} - \frac{1}{\lambda}\mathbf{A}^{-1}\mathbf{bc}^\dagger\mathbf{A}^{-1} \tag{D1}$$

where $\lambda = 1 + \mathbf{c}^\dagger\mathbf{A}^{-1}\mathbf{b}$ in case of any arbitrary $m \times n$ matrix $\mathbf{A}$. The Sherman-Morrison formula given by the above equation is valid only for square and nonsingular matrices.

The main result of [35] is a set of formulas for the MPGI depending upon the different set of conditions, namely,

(i) $\mathbf{b} \notin \mathcal{C}(\mathbf{A})$ and $\mathbf{c} \notin \mathcal{C}(\mathbf{A}^\dagger)$,
(ii) $\mathbf{b} \in \mathcal{C}(\mathbf{A})$ and $\mathbf{c} \notin \mathcal{C}(\mathbf{A}^\dagger)$ and $\lambda = 0$,
(iii) $\mathbf{b} \in \mathcal{C}(\mathbf{A})$ and $\mathbf{c}$ arbitrary and $\lambda \neq 0$,
(iv) $\mathbf{b} \notin \mathcal{C}(\mathbf{A})$ and $\mathbf{c} \in \mathcal{C}(\mathbf{A}^\dagger)$ and $\lambda = 0$,
(v) $\mathbf{b}$ arbitrary and $\mathbf{c} \in \mathcal{C}(\mathbf{A}^\dagger)$ and $\lambda \neq 0$,
(vi) $\mathbf{b} \in \mathcal{C}(\mathbf{A})$ and $\mathbf{c} \in \mathcal{C}(\mathbf{A}^\dagger)$ and $\lambda = 0$,

where $\mathcal{C}(\mathbf{A})$ represents the column space of a matrix $\mathbf{A}$.

Later it was shown in [36] that it is sufficient to consider five independent cases only. All these five different cases are listed below along with a useful theorem regarding the rank modification that occurs during the rank-one update of the matrix $\mathbf{A}$.

Theorem.—For given $\mathbf{A} \in \mathbb{C}_{m,n}$ and nonzero $\mathbf{b} \in \mathbb{C}_{m,1}$ and $\mathbf{c} \in \mathbb{C}_{n,1}$, let $\mathbf{M}$ be the modifications of $\mathbf{A}$ of the form $\mathbf{M} = \mathbf{A} + \mathbf{bc}^\dagger$ and let $\lambda = 1 + \mathbf{c}^\dagger\mathbf{A}^+\mathbf{b}$. Then

$$r(\mathbf{M}) = r(\mathbf{A}) - 1 \Leftrightarrow \mathbf{b} \in \mathcal{C}(\mathbf{A}), \quad \mathbf{c} \in \mathcal{C}(\mathbf{A}^\dagger), \quad \lambda = 0,$$

$$r(\mathbf{M}) = r(\mathbf{A}) \Leftrightarrow \begin{cases} \mathbf{b} \in \mathcal{C}(\mathbf{A}), \mathbf{c} \in \mathcal{C}(\mathbf{A}^\dagger), \lambda \neq 0, \\ \mathbf{b} \in \mathcal{C}(\mathbf{A}), \mathbf{c} \notin \mathcal{C}(\mathbf{A}^\dagger), \\ \mathbf{b} \notin \mathcal{C}(\mathbf{A}), \mathbf{c} \in \mathcal{C}(\mathbf{A}^\dagger), \end{cases}$$

$$r(\mathbf{M}) = r(\mathbf{A}) + 1 \Leftrightarrow \mathbf{b} \notin \mathcal{C}(\mathbf{A}), \quad \mathbf{c} \notin \mathcal{C}(\mathbf{A}^\dagger), \quad \lambda = 0,$$

where $r(\mathbf{M})$ denotes the rank of the matrix $\mathbf{M}$. Analytical expressions of the MPGI $\mathbf{M}^+$ are given in [35,36] depending upon the five or six conditions that they find sufficient. The expressions are of the form

$$\mathbf{M}^+ = \mathbf{A}^+ + \mathbf{G}, \tag{D2}$$

where $\mathbf{G}$ is a matrix obtained from only sums and products of $\mathbf{A}$, $\mathbf{A}^+$, $\mathbf{b}$, $\mathbf{c}$, and their conjugate transposes. We do not mention the explicit form of all the cases here. Instead we give expressions for $\mathbf{M}^+$ for those cases which will be useful to us.

a. Case 1

If $\mathbf{b} \in \mathcal{C}(\mathbf{A})$, $\mathbf{c} \in \mathcal{C}(\mathbf{A}^\dagger)$, $\lambda \neq 0$, then

$$\mathbf{M}^+ = \mathbf{A}^+ - \lambda^{-1}\mathbf{de}^\dagger, \tag{D3}$$

where

$$\mathbf{d} = \mathbf{A}^+\mathbf{b}, \qquad \mathbf{e} = (\mathbf{A}^+)^\dagger\mathbf{c}. \tag{D4}$$

b. Case 2

If $\mathbf{b} \notin \mathcal{C}(\mathbf{A})$, $\mathbf{c} \notin \mathcal{C}(\mathbf{A}^\dagger)$, then

$$\mathbf{M}^+ = \mathbf{A}^+ - \mathbf{ku}^+ - \mathbf{v}^+\mathbf{h} + \lambda\mathbf{v}^+\mathbf{u}^+, \tag{D5}$$

where $\mathbf{k} = \mathbf{A}^+\mathbf{b}$, $\mathbf{u} = (\mathbf{I} - \mathbf{AA}^+)\mathbf{b}$, $\mathbf{v} = \mathbf{c}^\dagger(\mathbf{I} - \mathbf{A}^+\mathbf{A})$, $\mathbf{h} = \mathbf{c}^\dagger\mathbf{A}^+$.

c. Case 3

In [36], a relation between MM^+ and AA^+, the unique projectors on the column spaces of M and A, respectively, is reported. If $\mathbf{b} \notin \mathcal{C}(\mathbf{A})$, $\mathbf{c} \in \mathcal{C}(\mathbf{A}^\dagger)$, then

$$\mathbf{MM}^+ = \mathbf{AA}^+ - \eta^{-1}\mathbf{ee}^\mathbf{T} + \eta^{-1}\nu^{-1}\mathbf{qq}^\mathbf{T}, \tag{D6}$$

where $\mathbf{e} = (\mathbf{A}^+)^\dagger\mathbf{c}$, $\eta = \mathbf{e}^\dagger\mathbf{e}$, $\nu = \lambda\lambda^* + \eta\phi$, $\phi = \mathbf{f}^\mathbf{T}\mathbf{f}$, $\mathbf{f} = (\mathbf{I} - \mathbf{AA}^+)\mathbf{b}$.

2. Analytic computation of the bias

The analytic computation of bias employs the generalized Sherman-Morrison formula and the relation between orthogonal projectors mentioned in the above three cases. Consider the previously stated expression for the covariance matrix $\hat{\mathbf{C}}_l$, Eq. (28). First note that $\mathbf{e}_0 \notin \mathcal{C}(\mathbf{A}_3)$, $\hat{\mathbf{f}}_l^0 \in \mathcal{C}(\mathbf{A}_3^\dagger)$, so that this is identical to the conditions for case 3. Furthermore $\mathrm{rank}(\mathbf{A}_3) = \mathrm{rank}(\hat{\mathbf{A}}_2) = n_f$, where n_f is the number of foreground components. However, when we

consider matrix $\hat{\mathbf{A}}_2$ we see that $\hat{\mathbf{f}}_l^0 \notin \mathcal{C}(\hat{\mathbf{A}}_2)$, $\mathbf{e}_0 \notin \mathcal{C}(\hat{\mathbf{A}}_2^\dagger)$. So here the conditions of case 2 apply and rank $(\hat{\mathbf{A}}_1) =$ rank $(\hat{\mathbf{A}}_2) + 1 = n_f + 1$. On the other hand, $\mathbf{e}_0 \in \mathcal{C}(\hat{\mathbf{A}}_1)$. Hence rank $(\hat{\mathbf{C}}_1) =$ rank $(\hat{\mathbf{A}}_1) = n_f + 1$. We now carry out the analytic simplification of the expression for $\hat{C}_l^{\rm Clean}$.

We first note that $\mathbf{e}_0 \in \mathcal{C}(\hat{\mathbf{A}}_1)$. Therefore we are in a situation where conditions of case 1 are applicable. Our aim is to express $\hat{C}_l^{\rm Clean} = 1/\mathbf{e}_0^T \hat{\mathbf{C}}_l^+ \mathbf{e}_0$ in terms of $\hat{\mathbf{A}}_1$. For this purpose we simply need to express $\hat{\mathbf{C}}_l^+$ in terms of $\hat{\mathbf{A}}_1$ using the GSM formula appropriate for this case. A straightforward calculation gives

$$\hat{C}_l^{\rm Clean} = \hat{C}_l^s + \frac{1}{\mathbf{e}_0^T \hat{\mathbf{A}}_1^+ \mathbf{e}_0}. \tag{D7}$$

We next simplify $1/(\mathbf{e}_0^T \hat{\mathbf{A}}_1^+ \mathbf{e}_0)$ in terms of $\hat{\mathbf{A}}_2^+$. For this purpose we express $\hat{\mathbf{A}}_1^+$ in terms of $\hat{\mathbf{A}}_2^+$ using the GSM. We see that $\hat{\mathbf{f}}_l^0 \notin \mathcal{C}(\hat{\mathbf{A}}_2)$ and $\mathbf{e}_0 \notin \mathcal{C}(\hat{\mathbf{A}}_2^T)$. Therefore the GSM formula corresponding to case 2 is applicable and yields

$$\hat{\mathbf{A}}_1^+ = \hat{\mathbf{A}}_2^+ - \mathbf{k}\mathbf{u}^+ - \mathbf{v}^+\mathbf{h} + \lambda \mathbf{v}^+\mathbf{u}^+. \tag{D8}$$

The vectors $\mathbf{k}$ and $\mathbf{h}$ are given by $\mathbf{k} = \hat{\mathbf{A}}_2^+ \hat{\mathbf{f}}_l^0$ and $\mathbf{h} = \mathbf{e}_0^T \hat{\mathbf{A}}_2^+$, respectively. The Moore-Penrose inverse of the other two vectors $\mathbf{u}$ and $\mathbf{v}$ can be computed following the definition $\mathbf{u}^+ = \mathbf{u}^\dagger/\|\mathbf{u}\|$, where $\|\mathbf{u}\|$ denotes the vector norm. Thus we have

$$\mathbf{u}^+ = \frac{\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+)}{\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+)\hat{\mathbf{f}}_l^0}, \qquad \mathbf{v}^+ = \frac{(\mathbf{I} - \hat{\mathbf{A}}_2^+\hat{\mathbf{A}}_2)\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{I} - \hat{\mathbf{A}}_2^+\hat{\mathbf{A}}_2)\mathbf{e}_0}. \tag{D9}$$

After some algebra we find

$$\frac{1}{\mathbf{e}_0^T \hat{\mathbf{A}}_1^+ \mathbf{e}_0} = \frac{\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+)\hat{\mathbf{f}}_l^0}{\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+)\mathbf{e}_0}. \tag{D10}$$

We next notice that both the numerator and denominator of Eq. (D10) contain the orthogonal projector $\hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+$ on the column space of $\hat{\mathbf{A}}_2$. At this stage we only need to reexpress this term in terms of $\mathbf{A}_3\mathbf{A}_3^+$, which is also an orthogonal projector on the column space of $\mathbf{A}_3$. We recall that $\hat{\mathbf{A}}_2 = \mathbf{A}_3 + \mathbf{e}_0\hat{\mathbf{f}}_0^T$. The foreground shape vector $\hat{\mathbf{f}}_l^0$ lies on the column space of the foreground covariance matrix, $\mathbf{A}_3$. However, the CMB shape vector is linearly independent on the foreground templates. Therefore the shape vectors follow, $\mathbf{e}_0 \notin \mathcal{C}(\mathbf{A}_3)$, $\hat{\mathbf{f}}_l^0 \in \mathcal{C}(\mathbf{A}_3^\dagger)$. Thus the conditions of case 3 are applicable. Following the notations of [36], we obtain

$$\hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+ = \mathbf{A}_3\mathbf{A}_3^+ - \eta^{-1}\mathbf{e}\mathbf{e}^T + \eta^{-1}\nu^{-1}\mathbf{q}\mathbf{q}^T. \tag{D11}$$

To proceed further we need to express $\mathbf{e}\mathbf{e}^T$ and $\mathbf{q}\mathbf{q}^T$ in terms of $\mathbf{f}_l^0$ and $\mathbf{e}_0$. We note that $\mathbf{e} = \mathbf{A}_3^+\mathbf{f}_l^0$ and $\mathbf{q} = \lambda\mathbf{e} + \eta\mathbf{f}$. Following the notations of [36], we see that $\mathbf{f}$ is the component of $\mathbf{e}_0$ on a plane orthogonal to the column space of $\mathbf{A}_3$, i.e., $\mathbf{f} = (\mathbf{I} - \mathbf{A}_3\mathbf{A}_3^+)\mathbf{e}_0$. We obtain, after some algebra,

$$\begin{aligned}\hat{\mathbf{f}}_l^{0T}\hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+\hat{\mathbf{f}}_l^0 = {}& \hat{\mathbf{f}}_l^{0T}\mathbf{A}_3\mathbf{A}_3^+\hat{\mathbf{f}}_l^0 - \eta^{-1}(\hat{\mathbf{f}}_l^{0T}\mathbf{A}_3^+\hat{\mathbf{f}}_l^0)^2 \\ & + \eta^{-1}\nu^{-1}(\lambda^2(\hat{\mathbf{f}}_l^{0T}\mathbf{A}_3^+\hat{\mathbf{f}}_l^0)^2 \\ & + \eta^2(\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \mathbf{A}_3\mathbf{A}_3^+)\mathbf{e}_0)^2 \\ & + 2\lambda\eta(\hat{\mathbf{f}}_l^{0T}\mathbf{A}_3^+\mathbf{f}_l^0\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \mathbf{A}_3\mathbf{A}_3^+)\mathbf{e}_0)).\end{aligned} \tag{D12}$$

Although there are several terms on the right-hand side of the above equation, we find that some of them drop out. We note that $\hat{\mathbf{f}}_l^0 \in \mathcal{C}(\mathbf{A}_3)$, so that $(\mathbf{A}_3\mathbf{A}_3^+)\hat{\mathbf{f}}_l^0 = \hat{\mathbf{f}}_l^0$. Hence $\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \mathbf{A}_3\mathbf{A}_3^+)\mathbf{e}_0 = \mathbf{e}_0^T(\mathbf{I} - \mathbf{A}_3\mathbf{A}_3^+)\hat{\mathbf{f}}_l^0 = 0$. We define $Y = \hat{\mathbf{f}}_l^{0T}\mathbf{A}_3^+\hat{\mathbf{f}}_l^0$. Using Eq. (D12) and $\nu = \lambda^2 + \eta\phi$, we obtain the numerator of Eq. (D10) as

$$\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+)\hat{\mathbf{f}}_l^0 = \frac{Y^2\phi}{\nu}. \tag{D13}$$

The denominator of Eq. (D10) can be computed in a similar fashion. The result is

$$\hat{\mathbf{f}}_l^{0T}(\mathbf{I} - \hat{\mathbf{A}}_2\hat{\mathbf{A}}_2^+)\mathbf{e}_0 = -\frac{1}{\nu}Y\phi. \tag{D14}$$

Thus using Eqs. (D10), (D13), and (D14) we obtain

$$\frac{1}{\mathbf{e}_0^T\hat{\mathbf{A}}_1^+\mathbf{e}_0} = -Y = -\hat{\mathbf{f}}_l^{0T}\mathbf{A}_3^+\hat{\mathbf{f}}_l^0. \tag{D15}$$

The quantity $\langle\frac{1}{\mathbf{e}_0^T\mathbf{A}_1^+\mathbf{e}_0}\rangle = -\langle\hat{\mathbf{f}}_l^{0T}\mathbf{A}_3^+\hat{\mathbf{f}}_l^0\rangle$ constitutes the negative bias in the cleaned power spectrum. Below, we further simplify this term in terms of the rank of the foreground covariance matrix $\mathbf{A}_3$.

Simplification of the bias expression

We note that the elements of the foreground covariance matrix are given by

$$A_3^{ab} = \sum_{pp'}^{n_f} f_p^a f_{p'}^b C_l^{(pp')0}. \tag{D16}$$

Also, the elements of the vector $\hat{\mathbf{f}}_l^0$ denoting the chance correlation between the CMB and all the foreground components are given by

$$\hat{\mathbf{f}}_l^{0a} = \hat{f}_l^{0a} = \sum_p^{n_f} \hat{C}_l^{(sp)0} f_p^a. \tag{D17}$$

We rewrite the magnitude of the negative bias in terms of the above matrix elements and components of the foreground shape vector. After a little algebra we get

$$\langle\hat{\mathbf{f}}_l^{0T}\mathbf{A}_3^+\hat{\mathbf{f}}_l^0\rangle = \sum_{ab} A_3^{+ab}\left\langle\sum_{pp'}\hat{C}_l^{(sp)0}\hat{C}_l^{(sp')0}\right\rangle f_p^a f_{p'}^b. \tag{D18}$$

Using $\hat{C}_l^{(sp)0} = \sum_m (a_{lm}^s a_{lm}^{p0*})/(2l+1)$ and the fact that the

CMB anisotropies are statistically isotropic, i.e., $\langle a_{lm} a_{l'm'} \rangle = C_l^s \delta_{ll'} \delta_{mm'}$, we obtain

$$\left\langle \sum_{pp'} \hat{C}_l^{(sp)0} \hat{C}_l^{(sp')0} \right\rangle = \frac{C_l^s}{2l+1} \sum_{pp'} C_l^{(pp')0}. \tag{D19}$$

Using Eq. (D19) and $A_3^{ab} = \sum_{pp'}^{n_f} f_p^a f_{p'}^b C_l^{(pp')0}$, we can rewrite Eq. (D18) in terms of the following expression consisting of the CMB and the foreground covariance matrices,

$$\langle \mathbf{f}_l^{0T} \mathbf{A}_3^+ \mathbf{f}_l^0 \rangle = \frac{C_l^s}{2l+1} \sum_{ab} A_3^{+ab} A_3^{ab}. \tag{D20}$$

However, we observe that $\sum_{ab} A_3^{+ab} A_3^{ab} = \mathrm{tr}(\mathbf{A}_3^+ \mathbf{A}_3) = \mathrm{rank}\,(\mathbf{A}_3) = n_f$. Thus we have

$$\langle \hat{C}_l^{\mathrm{Clean}} \rangle = \langle \hat{C}_l^s \rangle - n_f \frac{\langle \hat{C}_l^s \rangle}{2l+1}, \tag{D21}$$

which is Eq. (30).

[1] A. Benoit *et al.*, Astropart. Phys. **17**, 101 (2002).
[2] G. F. Smoot *et al.*, Astrophys. J. Lett. **396**, L1 (1992).
[3] W. C. Jones *et al.*, Astrophys. J. **647**, 823 (2006).
[4] S. Padin *et al.*, Astrophys. J. **549**, L1 (2001).
[5] C. L. Bennett *et al.*, Astrophys. J. Suppl. Ser. **148**, 1 (2003).
[6] C. L. Bennett *et al.*, Astrophys. J. **583**, 1 (2003).
[7] M. Limon *et al.*, Wilkinson Microwave Anisotropy Probe (WMAP): Explanatory Supplement, Version 1.0, at the LAMBDA website http://lambda.gsfc.nasa.gov/product/map/dr1/pub_papers/firstyear/supplement/WMAP_supplement.pdf.
[8] G. Hinshaw *et al.*, Astrophys. J. Suppl. Ser. **148**, 63 (2003).
[9] F. R. Bouchet and R. Gispert, New Astron. Rev. **4**, 443 (1999).
[10] G. Hinshaw *et al.*, Astrophys. J. Suppl. Ser. **148**, 135 (2003).
[11] C. L. Bennett *et al.*, Astrophys. J. Suppl. Ser. **148**, 97 (2003).
[12] G. Hinshaw *et al.*, Astrophys. J. Suppl. Ser. **170**, 288 (2007).
[13] S. Dodelson, Astrophys. J. **482**, 577 (1997).
[14] M. Tegmark, Astrophys. J. **502**, 1 (1998).
[15] N. Jarosik *et al.*, Astrophys. J. Suppl. Ser. **145**, 413 (2003).
[16] N. Jarosik *et al.*, Astrophys. J. Suppl. Ser. **170**, 263 (2007).
[17] M. Tegmark and G. Efstathiou, Mon. Not. R. Astron. Soc. **281**, 1297 (1996).
[18] M. Tegmark, A. de Oliveira-Costa, and A. Hamilton, Phys. Rev. D **68**, 123523 (2003).
[19] J. Mather *et al.*, Astrophys. J. **420**, 439 (1994); **512**, 511 (1999).
[20] D. J. Fixsen *et al.*, Astrophys. J. **473**, 576 (1996).
[21] R. Saha, P. Jain, and T. Souradeep, Astrophys. J. Lett. **645**, L89 (2006).
[22] T. Souradeep, R. Saha, and P. Jain, New Astron. Rev. **50**, 854 (2006).
[23] H. K. Eriksen *et al.*, Astrophys. J. **656**, 641 (2007).
[24] H. K. Eriksen *et al.*, Astrophys. J. **612**, 633 (2004).
[25] M. Tegmark, D. J. Eisenstein, W. Hu, and A. Oliveira-Costa, Astrophys. J. **530**, 133 (2000).
[26] K. M. Gorski *et al.*, arXiv:astro-ph/9905275.
[27] K. M. Gorski *et al.*, arXiv:astro-ph/9812350.
[28] M. R. Calabretta, arXiv:astro-ph/0412607.
[29] R. Saha, A. Hajian, P. Jain, and T. Souradeep (unpublished).
[30] E. Hivon *et al.*, Astrophys. J. **567**, 2 (2002).
[31] M. Tristram, J. F. Macs-Pez, C. Renault, and D. Santos, Mon. Not. R. Astron. Soc. **358**, 833 (2005).
[32] M. Tegmark and A. de Oliveira-Costa, Astrophys. J. Lett. **500**, L83 (1998).
[33] E. Komatsu *et al.*, Astrophys. J. Suppl. Ser. **148**, 119 (2003).
[34] K. M. Huffenberger *et al.*, Astrophys. J. Lett. **651**, L81 (2006).
[35] C. D. Meyer, Jr., SIAM J. Appl. Math. **24**, 315 (1973).
[36] J. K. Baksalary *et al.*, Linear Algebra Appl. **372**, 207 (2003).
[37] L. Page *et al.*, Astrophys. J. Suppl. Ser. **148**, 233 (2003).