

CMB anisotropy power spectrum using linear combinations of WMAP maps

Rajib Saha^{1,2,3}, Simon Prunet³, Pankaj Jain¹ & Tarun Souradeep²

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*

and

³*Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, F-75014 Paris, France.*

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 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 CMB angular power spectrum solely from multi-frequency observations has been proposed and implemented on the first year WMAP data by Saha et al. 2006. All previous estimates of power spectrum of CMB are based upon foreground templates using data sets from different experiments. However our methodology demonstrates that *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 1 year and 3 year data.

I. INTRODUCTION

Starting from the end of the last millenium remarkable progress in cosmology has been made by the precise measurements of the anisotropies in CMB from different ground based as well as satellite observations [1, 2, 3, 4]. The Wilkinson Microwave Anisotropy Probe (WMAP) measures the CMB anisotropy over the 5 frequency bands at 23 GHz (K), 33 GHz (Ka), 41 GHz (Q), 61 GHz (V) and 94 GHz (W). The observation system of the WMAP satellite consists of 10 differencing assemblies (DA), [5, 6, 7, 8], one each for K and Ka bands, two for Q band, two for V band and four for W band. They are labeled as K, Ka, Q1, Q2, V1, V2, W1, W2, W3, and W4 DA respectively. In the 1 year and 3 year data release the WMAP science team has provided the science community with large amount of high quality data sets measured by these 10 DAs. However, extracting the primordial signal from these large data set is a non trivial task. The anisotropies in 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 CMB signal from the multicomponent foreground contaminated data is thus complicated. There exists several methods in literature to remove foregrounds using foreground tracer templates [10, 11, 12] built from observations from other experiments. However, this requires a prior model of spatio-spectral dependence for all the foreground components. The effect of uncertainties in the foreground models to estimate CMB anisotropies have been discussed in Refs. [13, 14]. The foreground cleaning is applicable in the region away from the galactic plane. 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 the small angular scales. The detector noise, being a random quantity, is treated differently from foreground contamination which are treated here as fixed templates on the sky. However each of the DA's of WMAP has uncorrelated noise property [10, 12, 15, 16]. WMAP science team used this property to remove detector noise bias in a cross correlated power spectrum obtained from two different DA [10, 12] using DA maps with frequencies 41 GHz, 61 GHz and 94 GHz.

An interesting model independent method to remove foregrounds from the multi-frequency observations of CMB without any assumption about foreground components has been proposed in Ref. [17] and implemented in Ref. [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 same in all the bands in unit 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 Ref. [18] is not directly usable for primordial power spectrum estimation at smaller angular scale. This is because the detector noise bias dominates over the CMB power spectrum at smaller angular scale, beyond the beam width of WMAP detectors.

In earlier publications [21, 22, 23] we extended this model independent foreground removal method to remove

detector noise bias also. In this work we describe the basic formalism of our previous work in detail. We apply our method both on the WMAP 1 year and WMAP 3 year data to estimate CMB angular power spectrum. We form several cleaned maps using different cross combinations of the DA maps. Finally we form several cross power spectra where detector noise bias is removed. The results from this analysis are summarized in figure 9. In this figure we show the power spectra estimated from the WMAP's 1 year and 3 year data using the multi-frequency combination of CMB maps. The bottom part of this figure shows the residual unresolved point source contamination that were corrected for to obtain these two spectra. These power spectra are obtained without making any explicit model of the foregrounds or the detector noise. In most power spectrum extraction procedures, only the three highest frequency channels observed by WMAP have been used to extract CMB power spectrum. We present a more general procedure where we use observations from all the five frequency channels of WMAP. The primary merit of the foreground removal method is that it avoids any need to remove foregrounds based upon extrapolated flux measurements at frequencies far away from observational frequencies of WMAP [10, 24, 25].

Presence of a bias in the internally cleaned maps has been reported earlier in the Refs. [12, 26, 27]. In this work we also perform a detailed study of the nature of bias in the cleaned power spectrum. We show that the cleaned power spectrum is not an unbiased estimator of the underlying CMB spectrum. Naively one expects that there might be some residual foreground contamination causing a positive bias in the power spectrum. For a simplified approach, which cleans the entire sky simultaneously, without sub-dividing into regions of varying foreground contamination, we are able to analytically compute the cleaned power spectrum in terms of a CMB signal and the foreground plus detector noise covariance matrix. The existence of bias is easily identified from these results. We also report and quantify an interesting negative bias in the cleaned power spectrum. This negative bias is directly determined by the underlying CMB power spectrum and is strongest at the lowest multipoles. The analytical results for the bias estimations make the model independent foreground removal method more interesting for use in cosmology. The bias in the cleaned power spectrum can be removed following models of foreground and noise covariance matrices after a model independent foreground removal is performed. This makes the estimated power spectrum less prone to uncertainties of the foreground modeling compared to a method which tries to minimize foreground using external templates. In this work we debias the CMB anisotropy spectra obtained from WMAP data only at the large multipole regime, $l \geq 400$. It turns out that because of large noise level of WMAP the point source bias is the only issue in this multipole range. We do not attempt to employ a debiasing method at the low multipoles where the negative bias is expected to dominate, because of its complicated nature in the case of the iterative, multiregion cleaning method. A more detailed study on this issue will be reported in a future publication.

The plan of this paper is as follows. The basic formalism and methodology to obtain power spectrum is described in the section II. Here we also discuss the bias present in this method. The implementation of our methodology on the WMAP 1 year and 3 year data is discussed in section III. Here we also obtain an analytic expression for the residual unresolved point source power contamination. The results are described in section IV and finally we conclude in section V. Some of the notations used in this report are as follows. We denote matrices and vectors by **boldfaced** 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^c .

II. BASIC FORMALISM AND METHODOLOGY

In this section we outline the mathematical formalism to obtain the underlying power spectrum. We also quantify the bias present in the estimated power spectrum using our procedure.

A. Foreground Removal

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

$$\tilde{\Delta T}^i(\hat{n}) = \int (\Delta T^c(\hat{n}') + \Delta T^{fi}(\hat{n}')) B^i(\hat{n} \cdot \hat{n}') d\hat{n}' + \Delta T^{ni}(\hat{n}). \quad (1)$$

Here $\Delta T^c(\hat{n})$ and $\Delta T^{fi}(\hat{n})$ are respectively CMB and foreground component of the anisotropy in the channel, i . The detector noise in the channel, i , is ΔT^{ni} . The beam function $B^i(\hat{n} \cdot \hat{n}')$ represents the smoothing of the map due to finite resolution of the antenna of channel, i . The beam is assumed to be circularly symmetric as done in most analysis. We note that the detector noise is not affected by the beam function. An experiment such as WMAP provides multi-frequency maps $\tilde{\Delta T}^i(\hat{n})$, $i = 1, 2, \dots, n_c$ corresponding to observation of the CMB at n_c different frequency bands.

Equivalently, in the spherical harmonic representation

$$\tilde{a}_{lm}^i = (a_{lm}^c + a_{lm}^{fi}) B_l^i + a_{lm}^{ni}. \quad (2)$$

where a_{lm} are respective spherical harmonic contributions to the maps and B_l are Legendre transform coefficients of the beam $B^i(\hat{n} \cdot \hat{n}')$. The aim is to linearly combine the maps with appropriate weights to get an optimal estimator of the CMB anisotropy $\Delta T^c(\hat{n}')$ that minimizes the contribution from $\Delta T^f(\hat{n}')$. The linear combination of the multi-frequency maps available can be carried out in pixel space, or, in the equivalent representation in terms of the spherical harmonic coefficients. The former has been followed by the WMAP team to produce Internal Linear Combination (ILC) map and also in a related, but more elaborate approach in Ref. [27] to produce LILC map. The approach of carrying out a multi-frequency maps combination in the spherical harmonic space was proposed in Ref. [17]. For the first year of WMAP data this was implemented in Ref. [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,

$$a_{lm}^{\text{Clean}} = \sum_{i=1}^{i=n_c} \hat{W}_l^i \frac{a_{lm}^i}{B_l^i}. \quad (3)$$

Here \hat{W}_l^i is a weight factor which depends upon the multipole l and the frequency channel, i . 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

$$\hat{C}_l^{\text{Clean}} = \frac{1}{2l+1} \sum_{m=-l}^{m=l} a_{lm}^{\text{Clean}} a_{lm}^{\text{Clean}*}. \quad (4)$$

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

$$\hat{C}_l^{\text{Clean}} = \hat{\mathbf{W}}_l \hat{\mathbf{C}}_l \hat{\mathbf{W}}_l^T, \quad (5)$$

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

$$\hat{\mathbf{C}}_l = \begin{pmatrix} \frac{\hat{C}_l^{11}}{B_l^1 B_l^1} & \cdots & \cdots & \frac{\hat{C}_l^{1n_c}}{B_l^1 B_l^{n_c}} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\hat{C}_l^{n_c 1}}{B_l^{n_c} B_l^1} & \cdots & \cdots & \frac{\hat{C}_l^{n_c n_c}}{B_l^{n_c} B_l^{n_c}} \end{pmatrix}, \quad (6)$$

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

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

and $\hat{C}_l^{i,j}$ is the cross power spectrum between the i^{th} and j^{th} channel,

$$\hat{C}_l^{i,j} = \sum_{m=-l}^{m=l} \frac{a_{lm}^i a_{lm}^{j*}}{2l+1} = \frac{a_{l0}^i a_{l0}^j}{2l+1} + 2\Re \sum_{m=1}^{m=l} \frac{a_{lm}^i a_{lm}^{j*}}{2l+1}. \quad (8)$$

By construction, the $\hat{\mathbf{C}}_l$ matrix is symmetric. The spherical harmonic coefficients extracted from a map contain CMB signal and foregrounds, both smoothed by the beam function of the optical instrument used in the experiment, as well as detector noise. Using eq. (2) we obtain,

$$a_{lm} = B_l a_{lm}^S + \underbrace{B_l a_{lm}^F + a_{lm}^N}_{a_{lm}^{junk}}, \quad (9)$$

where a_{lm}^{junk} is used to denote the total non-CMB contamination in the map. Since the CMB contribution is independent of frequency, the expression for $\frac{\hat{C}_l^{ij}}{B_l^i B_l^j}$ simplifies to

$$\frac{\hat{C}_l^{ij}}{B_l^i B_l^j} = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \frac{(B_l^i a_{lm}^S + a_{lm}^{i(junk)})}{B_l^i} \times \frac{(B_l^j a_{lm}^{S*} + a_{lm}^{j(junk)*})}{B_l^j} = \hat{C}_l^S + \hat{C}_l^{ij(junk)},$$

for all values of i and j . Now using eq. (5) we obtain

$$\hat{C}_l^{Clean} = \hat{C}_l^S \hat{\mathbf{W}}_1 \mathbf{e}_0 \mathbf{e}_0^T \hat{\mathbf{W}}_1^T + \hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1^{(junk)} \hat{\mathbf{W}}_1^T, \quad (10)$$

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

$$\mathbf{e}_0 = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ 1 \end{pmatrix}. \quad (11)$$

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

$$\hat{\mathbf{W}}_1 \mathbf{e}_0 = \mathbf{e}_0^T \hat{\mathbf{W}}_1^T = 1, \quad (12)$$

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

$$\hat{C}_l^{Clean} = \hat{C}_l^S + \hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1^{junk} \hat{\mathbf{W}}_1^T. \quad (13)$$

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^{Clean} also minimizes the combined contamination coming from foreground and detector noise. It may be shown easily, using the Lagrange's multiplier method, that minimizing \hat{C}_l^{Clean} subject to the constraint eq. (12), gives the following expression for the weight factors [17, 26],

$$\hat{\mathbf{W}}_1 = \frac{\mathbf{e}_0^T \hat{\mathbf{C}}_1^{-1}}{\mathbf{e}_0^T \hat{\mathbf{C}}_1^{-1} \mathbf{e}_0}. \quad (14)$$

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

$$\hat{C}_l^{Clean} = \frac{1}{\mathbf{e}_0^T \hat{\mathbf{C}}_1^{-1} \mathbf{e}_0}. \quad (15)$$

It is important to note that in cases when $\hat{\mathbf{C}}_1$ is singular, it is possible to generalize the above expressions for weights and cleaned power spectrum in terms of the *Moore-Penrose Generalized inversion* (MPGI) of $\hat{\mathbf{C}}_1$. The generalized weights and cleaned power spectrum are then given by

$$\hat{\mathbf{W}}_1 = \frac{\mathbf{e}_0^T \hat{\mathbf{C}}_1^\dagger}{\mathbf{e}_0^T \hat{\mathbf{C}}_1^\dagger \mathbf{e}_0}, \quad (16)$$

$$\hat{C}_l^{Clean} = \frac{1}{\mathbf{e}_0^T \hat{\mathbf{C}}_1^\dagger \mathbf{e}_0}. \quad (17)$$

For further details of the analytic derivation of the weights and cleaned power spectrum we refer to appendix A.

B. Biases in the foreground cleaning method

Although the foreground cleaning is performed with the constraint that CMB power spectrum remains preserved, the method biases the final power spectrum. This section is devoted to a discussion of the full bias in the method.

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 increase in number of channels that are combined.

Let us consider n_c number of channels in the linear combinations for the cleaned map. The maps at each channel consists of the CMB signal and foreground contamination coming from the different foreground components (synchrotron, free-free, dust etc). Although, the number of distinct foreground components could be arbitrary, for brevity

and simplicity, here we club the contributions from all the components into a single foreground term. The foreground contribution can be described by a covariance matrix. Later, after simplification of the expressions, we can split up the total foreground covariance matrix in terms of the covariance matrices of the distinct constituent components. In the entire discussion that follows we do not treat foregrounds as a stochastic component. Instead we consider them as fixed templates in all the realizations. This is entirely justified. We are simply interested in computing a foreground free template rather than estimating information about the distribution of foreground from which they are drawn. We discuss the bias in few different cases depending upon the rank of the covariance matrix $\hat{\mathbf{C}}_1$.

1. *Case: Rank $\hat{\mathbf{C}}_1 \leq n_c - 1$*

First we consider an ideal case wherein detector noise remains absent and each foreground component follows a rigid frequency scaling on the entire sky. Let the number of foreground components be n_f . Then the rank of the foreground covariance matrix, \mathbf{C}_1^f , is also n_f .

The p^{th} foreground component for channel i is denoted by $F_0^p(\theta, \phi)f_p^i$, where the frequency dependence, f_p^i and the spatial (sky) dependence $F_0^p(\theta, \phi)$ are explicitly separable in the rigid scaling assumption. (Here, $F_0^p(\theta, \phi)$ is the p^{th} foreground template based on frequency ν_0 , so that $f_p^i = 1$, for frequency ν_0 .) We denote the CMB component by $C(\theta, \phi)$. Full signal map at frequency channel, i , is then given by

$$S^i(\theta, \phi) = C(\theta, \phi) + \sum_{p=1}^{n_f} F_0^p(\theta, \phi)f_p^i. \quad (18)$$

Alternatively, in the spherical harmonic space,

$$a_{lm}^i = a_{lm}^c + \sum_{p=1}^{n_f} f_p^i a_{lm}^{p0}. \quad (19)$$

The auto power spectrum of the i^{th} channel

$$\hat{C}_l^i = \hat{C}_l^c + 2 \sum_{p=1}^{n_f} f_p^i \hat{C}_l^{cf(p)0} + \sum_{p,p'}^{n_f} f_p^i f_{p'}^i C_l^{(pp')0}. \quad (20)$$

In the above equation, $C_l^{(pp')0}$ is the correlation between any two foreground components p, p' and $\hat{C}_l^{cf(p)0}$ denotes the chance correlation between CMB signal and p^{th} foreground component. The cross power spectrum between two channels i, j is given by

$$\hat{C}_l^{ij} = \hat{C}_l^c + \sum_{p=1}^{n_f} f_p^i \hat{C}_l^{cf(p)0} + \sum_{p=1}^{n_f} f_p^j \hat{C}_l^{cf(p)0} + \sum_{p,p'}^{n_f} f_p^i f_{p'}^j C_l^{(pp')0}. \quad (21)$$

It is convenient to define the vectors,

$$\hat{\mathbf{f}}_1^{p0} = \hat{C}_l^{cf(p)0} \begin{pmatrix} f_p^1 \\ f_p^2 \\ \vdots \\ f_p^{n_c} \end{pmatrix}, \quad (22)$$

and

$$\hat{\mathbf{f}}_1^0 = \sum_{p=1}^{n_f} \hat{\mathbf{f}}_1^{p0}. \quad (23)$$

A little algebraic manipulation allows us to recast eq. (21) as

$$\hat{\mathbf{C}}_1 = \hat{C}_l^c \mathbf{e}_0 \mathbf{e}_0^T + \hat{\mathbf{f}}_1^0 \mathbf{e}_0^T + \mathbf{e}_0 \hat{\mathbf{f}}_1^{0T} + \mathbf{C}_1^f. \quad (24)$$

The above equation will be useful to compute the bias in the cleaned power spectrum. On the ensemble average the cleaned power spectrum as given by eq. (17) could be simplified with the help of successive use of a set of theorems

reported in Refs. [28, 29]. An elaborate discussion of these theorems is given in appendix E. Assuming statistically isotropic CMB sky we obtain the following expression for the ensemble average of the cleaned power spectrum

$$\langle \hat{C}_l^{Clean} \rangle = \langle \hat{C}_l^c \rangle - n_f \frac{\langle \hat{C}_l^c \rangle}{2l+1}. \quad (25)$$

We easily note a few interesting aspects of the above equation. First of all there exists a negative bias. Secondly, 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 debias 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 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 $n_c \geq n_f + 1$ ensure that sufficient amount of spectral information are available to remove all the foregrounds. The negative bias arises as the weights are to be determined from the empirical covariance matrix to take into account information available from the observed data.

2. Case: Rank (\mathbf{C}_1^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 Ref. [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 of all the foreground components exceeds the number of maps available for linear combination then \mathbf{C}_1^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 $\frac{\langle \hat{C}_l^c \rangle}{2l+1}$. Keeping aside the detector noise for the moment, the ensemble average of the cleaned power spectrum is given by

$$\langle \hat{C}_l^{Clean} \rangle = \langle \hat{C}_l^c \rangle + \frac{\mathbf{1}}{\mathbf{e}_0^T (\mathbf{C}_1^f)^{-1} \mathbf{e}_0} + (1 - n_c) \frac{\langle \hat{C}_l^c \rangle}{2l+1}. \quad (26)$$

Detailed derivation of the above equation is similar to derivation of eq. (25). As \mathbf{C}_1^f is of full rank and a positive definite matrix, the second term on the right causes a positive foreground bias.

3. Noise and foreground case

The discussion in the last two sections 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 derivation of eqs. (25) and (26) we can show that, on the ensemble average the cleaned power spectrum in this case is given by

$$\langle \hat{C}_l^{Clean} \rangle = \langle \hat{C}_l^c \rangle + \left\langle \frac{\mathbf{1}}{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{f+N})^{-1} \mathbf{e}_0} \right\rangle + (1 - n_c) \frac{\langle \hat{C}_l^c \rangle}{2l+1}. \quad (27)$$

We have carried out Monte-Carlo simulations to verify the analytical result given by eq. (25). We perform simulations of foreground cleaning using $n_c = 3$ channels, corresponding to 41, 61 and 94 GHz frequencies. The foreground model consists of a synchrotron and free-free emission. Each of the foreground component is assumed to follow a rigid frequency scaling all over the sky. First we generate synchrotron and free-free templates using the Planck Sky Model ¹ at the 3 different frequencies. Since these templates do not follow rigid frequency scaling all over the sky, (

¹ 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.

eg., synchrotron spectral index varies with the position of the sky), we developed a method to regenerate rigid scaling synchrotron and free templates at these 3 different frequency channels²

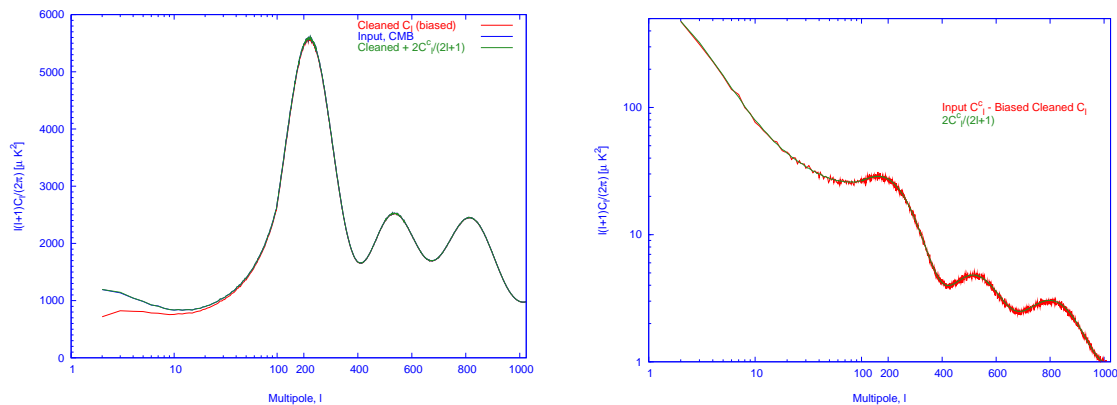


FIG. 1: The negative bias in the extracted power spectrum at low l is shown by the red line in the left panel. The bias corrected spectrum is plotted in blue line. However it lies entirely behind the green line and is not visible. The right panel shows how well the analytic results for bias match with those obtained from Monte Carlo simulations.

The left hand panel of the fig. 1 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. (25) contributes $-2\langle\hat{C}_l^c\rangle/2l+1$ as a negative bias. In the right hand panel of the fig. 1 we explicitly show that the magnitude of the negative bias is exactly compensated by $2\frac{\langle\hat{C}_l^c\rangle}{2l+1}$. The negative bias is important at the lowest multipoles and becomes negligible at high l , e.g. the bias is only $\approx 3\mu K^2$ at multipole $l \approx 800$. The bias corrected spectrum plotted in green color in the left panel of this fig. 1 is completely hidden by the blue curve corresponding to the input CMB power spectrum which is used to generate random realizations of CMB maps.

The bias in the cleaned power spectrum for different cases described so far is computed following the assumption that foreground cleaning is simultaneously carried out over the entire sky. However for the cleaning purpose of WMAP maps we follow a more sophisticated method. The main reason behind opting for a sophistication is that the spectrum of foreground emission as well the amplitude strongly depend on the location in the sky. The effectiveness of the foreground cleaning in varying foreground spectral index has been studied in the existing literature, [9, 14]. Following the discussion in section II B 2 a varying foreground spectra would cause a positive foreground bias in the cleaned power spectrum. To minimize this bias we followed Ref. [18] and divide the sky into 9 regions based on a simple estimation of the level of foreground contamination (see appendix B). We then carry out foreground cleaning iteratively taking one region at a time starting with dirtiest region. The advantage of such iterative method is that diffuse foreground contamination which is dominant at the low multipoles compared to the detector noise could be very effectively minimized without a precise model of residual bias. The main principle for iterative cleaning is to search for sky regions where foreground emission could be assumed approximately constant or at least nearly constant. Such a philosophy of foreground cleaning has been adopted by WMAP team to produce their Internal Linear Combination Map [11]. In a future publication we would generalize the bias results reported in this article to the most general method of iterative cleaning. Another advantage of iterative method is that one has the freedom to adjust weights on different sky positions depending upon the amplitude of the foreground components leading to a better cleaning.

² For this purpose, we first note that for each component the approximate normalization $\sqrt{C_l^i/C_l^{23}}$ remains roughly constant for all l . Here C_l^{23} is the synchrotron (or free-free) power spectrum for frequency 23 GHz and C_l^i is the synchrotron (or free-free) power spectrum at any of other 3 frequencies. We generate a synchrotron (or free-free) template at the i^{th} frequency channel following a scaling of the 23 GHz template by the number $\sqrt{C_l^i/C_l^{23}}$. This ensures that the emission for each foreground component at different frequencies follows rigid frequency scaling.

n_c -channel combination	Cleaned Map
$D_1^1 + D_1^2 + \dots + D_1^{n_c}$	C_1
$D_2^1 + D_2^2 + \dots + D_2^{n_c}$	C_2
$D_3^1 + D_3^2 + \dots + D_3^{n_c}$	C_3
.....
$D_d^1 + D_d^2 + \dots + D_d^{n_c}$	C_d

TABLE I: List of several possible cleaned maps with uncorrelated detector noise properties. Here we assume n_c number of frequency bands with each band having d number of independent detectors.

C. Power spectrum Estimation

The power spectrum is estimated by cross correlating multichannel foreground cleaned maps from independent Differencing Assemblies (DA). The alternative is to use auto correlation and use a detector noise model to debias the power spectrum. In this case the detector noise model uncertainty affects the mean of the cleaned power spectrum. We note that the WMAP team too used cross-correlation of independent DA to obtain their power spectrum. Let us consider an hypothetical experiment with n_c number of frequency channels each of which is denoted by i . Each frequency channel consists of d number of independent detectors denoted by D_j^i . Thus D_j^i is the map observed by j^{th} detector of i^{th} channel. We then propose to form several cleaned maps in such a way that each cleaned map consists of one map from each of the n_c number of channels. By choosing appropriate combinations of DA maps we identify two different cleaned maps having totally disjoint set of detectors. Then assuming that the noise properties are uncorrelated in two different detectors, the noise bias will not affect the cross power spectrum of two such cleaned maps at the 1st order. The final power spectrum could be a simple average of all possible such cross power spectra. In table I we show a set of possible cleaned maps $C_j, j = 1, 2, \dots, d$ with uncorrelated detector noise properties. Hence $\frac{d(d-1)}{2}$ number of cross power spectra can be obtained that do not have any noise bias. In general the number of possible cleaned maps is more than d . If we divide the combination of cleaned map 1 in p number of subsets and the combination of cleaned map 2 in q number of subsets then one only needs to make sure that these p and q subsets have uncorrelated noise properties.

III. IMPLEMENTATION ON WMAP DATA

In our method we linearly combine maps corresponding to a set of 4 DAs at different frequencies.³ We treat the K and Ka maps effectively as observations of the CMB sky in two different DA at the low frequencies. Therefore, we use either K or Ka maps in any combination. In case of the W band, 4 DA maps are available. We simply form pairwise averaged map taking two of them at a time and form 6 effective DA maps corresponding to W band. W_{ij} represents simply an averaged map obtained from the i^{th} and j^{th} DA of W band. In table II we list all the 48 possible linear combinations of the DA maps that lead to ‘cleaned’ maps, C_i and CA_i ’s, where $i = 1, 2, \dots, 24$. In an alternative approach, we form combinations to form cleaned maps excluding the most foreground contaminated K and Ka bands (referred to as the three-channel $n_c = 3$ case). This leads to a total of 24 cleaned maps. All such combinations are shown in the bottom panel of the table II. In this combination the cleaned maps are labeled as C_i ’s, where $i = 1, 2, \dots, 24$.

The entire method leading to the power spectrum estimation consists of three main steps. First we perform foreground cleaning using several multi-channel combinations of WMAP maps. At the second step we obtain cross power spectra from these foreground cleaned maps. The foreground cleaning is similar to Refs. [17, 18]. Finally we correct for the estimated residual unresolved point source contamination. We note in passing that each of these steps are logically modular and each of them could be modified or improved independent of the other.

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

4-channel combinations ($n_c = 4$)	
(K,KA)+Q1+V1+W12=(C1,CA1)	(K,KA)+Q1+V2+W12=(C13,CA13)
(K,KA)+Q1+V1+W13=(C2,CA2)	(K,KA)+Q1+V2+W13=(C14,CA14)
(K,KA)+Q1+V1+W14=(C3,CA3)	(K,KA)+Q1+V2+W14=(C15,CA15)
(K,KA)+Q1+V1+W23=(C4,CA4)	(K,KA)+Q1+V2+W23=(C16,CA16)
(K,KA)+Q1+V1+W24=(C5,CA5)	(K,KA)+Q1+V2+W24=(C17,CA17)
(K,KA)+Q1+V1+W34=(C6,CA6)	(K,KA)+Q1+V2+W34=(C18,CA18)
(K,KA)+Q2+V2+W12=(C7,CA7)	(K,KA)+Q2+V1+W12=(C19,CA19)
(K,KA)+Q2+V2+W13=(C8,CA8)	(K,KA)+Q2+V1+W13=(C20,CA20)
(K,KA)+Q2+V2+W14=(C9,CA9)	(K,KA)+Q2+V1+W14=(C21,CA21)
(K,KA)+Q2+V2+W23=(C10,CA10)	(K,KA)+Q2+V1+W23=(C22,CA22)
(K,KA)+Q2+V2+W24=(C11,CA11)	(K,KA)+Q2+V1+W24=(C23,CA23)
(K,KA)+Q2+V2+W34=(C12,CA12)	(K,KA)+Q2+V1+W34=(C24,CA24)
3-channel combinations ($n_c = 3$)	
Q1+V1+W12=C1	Q1+V2+W12=C13
Q1+V1+W13=C2	Q1+V2+W13=C14
Q1+V1+W14=C3	Q1+V2+W14=C15
Q1+V1+W23=C4	Q1+V2+W23=C16
Q1+V1+W24=C5	Q1+V2+W24=C17
Q1+V1+W34=C6	Q1+V2+W34=C18
Q2+V2+W12=C7	Q2+V1+W12=C19
Q2+V2+W13=C8	Q2+V1+W13=C20
Q2+V2+W14=C9	Q2+V1+W14=C21
Q2+V2+W23=C10	Q2+V1+W23=C22
Q2+V2+W24=C11	Q2+V1+W24=C23
Q2+V2+W34=C12	Q2+V1+W34=C24

TABLE II: The table on the top shows 48 different combinations of the DA maps used in our 4 channel cleaning method. The bottom table shows the 24 possible combinations in the 3 channel cleaning method.

A. Map cleaning

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 HEALPix⁴ pixelization scheme at a resolution level $N_{\text{side}} = 512$ corresponding to approximately 3 million sky pixels. All these maps are provided in the ‘nested’ pixelization scheme which is suitable for nearest neighbor searches. However for converting maps to spherical harmonic space and vice versa it is computationally advantageous to convert them to ring format that facilitates the use of Fast Fourier transformation method along equal latitudes. Therefore prior to the analysis all the maps corresponding to 10 DA’s were converted to ‘ring’ pixelization scheme. When converting a map to spherical harmonic space we restricted ourselves to a maximum multipole, $l_{\text{max}} = 1024$.

The spectrum of foreground emission has some dependency on the location on the sky. A better cleaning may be achieved if we partition the entire sky into certain number of sky parts depending upon the level of foreground contaminations [18]. Then cleaning is done for each sky-parts iteratively. For each regions we will have then different weights which are chosen to minimize foreground contamination from that particular region. Cleaning 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. Following Ref. [18] we partition the sky depending upon the level of foreground contamination.⁵

The entire sky is partitioned in a total of $r = 9$ regions depending upon their foreground contamination. All the r sky parts are shown in the fig. 2. Individual sky parts are color-coded according to the number r assigned to them. The dirtiest part is labeled with the maximum index. The procedure of forming these sky parts are similar to Ref. [18]

⁴ For comprehensive studies about HEALPix we refer to Ref. [30, 31, 32].

⁵ 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 rigid scaling approximation. In an ongoing study we have investigated and demonstrated improvement in foreground cleaning for different levels of partitioning [33].

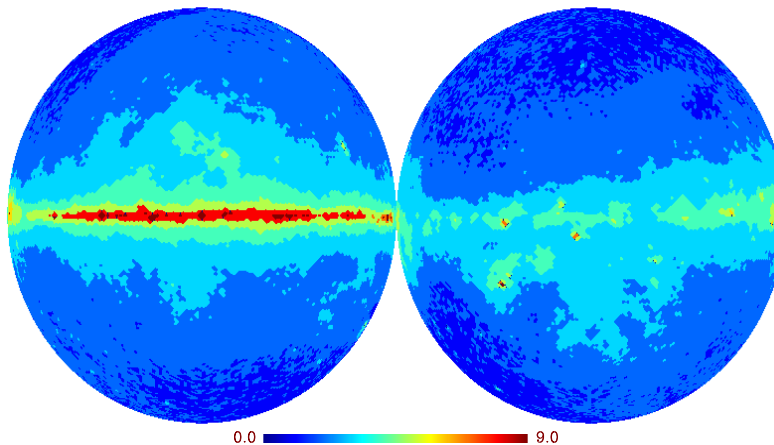


FIG. 2: The 9 different masks in orthographic projection that correspond to the partitions of the sky. Each region is color coded to distinguish them visually. The cleanest region is the region farthest away from the galactic plane. The dirtiest region (black spots) lie in the galactic plane.

with small variants. We describe our sky partitioning to identify these r regions in detail in appendix B.

For each combination, the cleaning procedure is performed in r iterations starting from the dirtiest region. As in Ref. [18] we call the initial foreground contaminated maps as the initial temporary maps. At the end of each of iteration we obtain a set of partially cleaned temporary maps which are used as input in the next iteration. In i^{th} iteration ($i = 1, 2, 3, \dots, r$) we perform the following three steps :

1. Using the i^{th} mask we select the i^{th} region of the sky from the temporary maps. Then we obtain power spectrum matrix from the i^{th} region only.⁶
2. Using the weight factors we obtain a full sky cleaned map using eq. (3).
3. Next we replace the i^{th} region of the temporary maps by the corresponding region of the cleaned map. Before replacement the cleaned map is smoothed to the resolution of the different frequency channels. After replacement the maps define new temporary maps for the next iteration.

The entire cleaning procedure is automated to carry the three steps iterated r times to obtain the final cleaned maps. The final cleaned maps have the resolution of the W band.

In eq. (7) the weights could have numerical errors if the C_l matrix becomes ill-conditioned at specific values of l . This can happen if by chance any mode has almost equal contribution combined from the CMB, foreground and detector noise. Hence, in practice, the numerical implementation obtains weights using the C_l matrix smoothed over a range of $\Delta l = 11$ prior to inversion.

As shown in table II, for the four channel combination ($n_c = 4$), a total of 48 cleaned maps can be obtained using all the possible combinations of the DA maps for each of the WMAP 1 year and WMAP 3 year data. The cleaned map C8 for WMAP 1 year data is shown in the left panel of the figure 3. There is some residual foreground left in the galactic plane as seen in this map. We apply 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. A similar map is obtained from the WMAP 3 year data also. The right panel of the figure 3 shows CA8 map from WMAP-3 data. Both the maps look similar.

The weights \tilde{W}_l are shown in the figures 4 for one of the cleaned map (CA1) for the cleanest and second cleanest region. For low l where the diffuse foregrounds are dominant, the weights take large positive and negative values to

⁶ These power spectrum matrix is obtained from the partial sky spherical harmonic coefficients. In practice while using 1 year WMAP data we have not combined K band map for $l > 603$ ($B_l < 0$ for K band for $l > 603$ and the negative values of beam function specified is unphysical.). KA and Q band beam functions are available till $l = 850$ and $l = 1000$. Therefore we do not combine these bands beyond these values. Similar limits determined by the condition $B_l > 0$ were used for 3 year data analysis also.

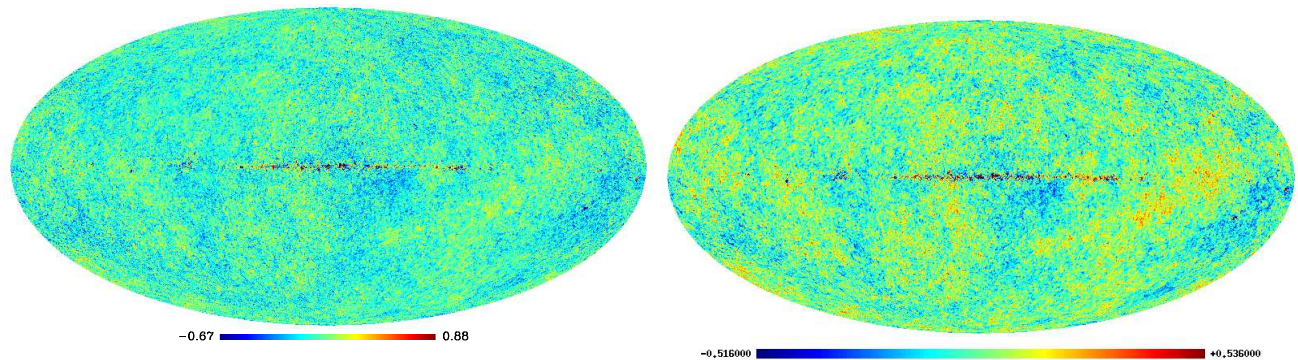


FIG. 3: The left panel shows the cleaned map C8 using the 1 year WMAP data. There is some residual foreground contamination near the galactic regions. The right panel shows the CA8 map using the 3 year WMAP data. The temperature scale of both the figures are chosen from the sky part outside the Kp2 mask.

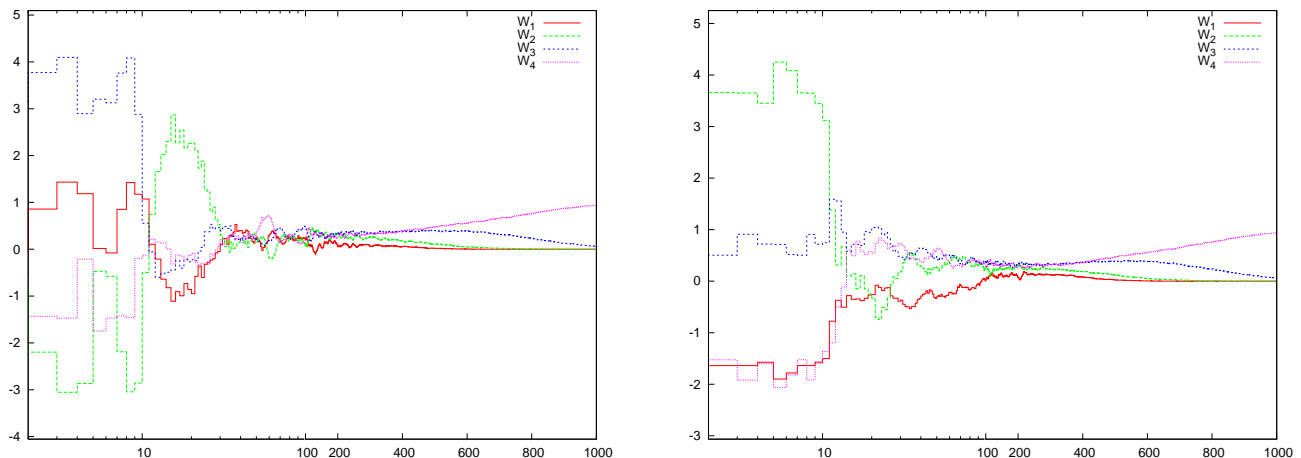


FIG. 4: The left panel of the figure shows the weights for the cleanest region for the combination KA, Q1, V1, W12. The right panel show the weights for the same combination but for the second cleanest region. At low multipole foregrounds dominate. Therefore, weights take large positive and negative values to subtract foregrounds.

subtract the foregrounds. At large l the maximum weight is given to the W band channel since it has the highest resolution.

B. Power spectrum estimation

Our final power spectrum is based upon the MASTER estimate introduced in Ref. [34]. Even after performing foreground cleaning the galactic disk region remains significantly contaminated. Hence one needs to exclude this region before estimating the cosmological power spectrum. However, flagging the contaminated sky region effectively introduces uneven weighting of the pixels. Those flagged have effectively zero weight and rest have unit weight. Weighting a map in the pixel space causes neighboring multipoles to get correlated. The correlation of the multipoles are described by the coupling matrix which is determined by the geometry of the sky coverage. Moreover, the power spectrum of such a map is biased low because of less sky coverage. The MASTER method debiases the partial sky power spectrum on the ensemble average by inverting the coupling matrix.

The MASTER method originally implemented in Ref. [34] deals with auto power spectrum only. This can readily be generalized to the case of cross power spectrum [35]. We first exclude the galactic region from all the 48 cleaned maps (Table II, $n_c = 4$) using Kp2 mask as supplied by the WMAP team. Then map pairs C_i & C_j to be cross correlated are chosen such that they do not have any DA common between them. This choice ensures that the detector noise bias does not affect any of the cross power spectrum. A total of 24 cross power spectra could be obtained which

are unaffected by the detector noise bias. Each of these cross spectra are then debiased from the partial sky effect following Ref. [35] using the coupling (bias) matrix corresponding to the Kp2 mask. The small scale systematic effects of beam and pixel smoothing were removed using appropriate circularized beam transform [34] and pixel window functions as supplied by the HEALPix package. The 24 cross power spectra are then combined with equal weights into a single ‘Uniform average’ power spectrum.⁷ The final power spectrum is binned in the same manner as the WMAP’s published result for ease of comparison.

Both one and three year power spectra were obtained following this method. We defer a more detailed discussion on the 3 year power spectrum results to the section IV B. We also estimate a residual contamination in the ‘Uniform average’ power spectrum for both WMAP 1 year and WMAP 3 year data from the unresolved point sources. A point source model used by WMAP team [10, 12] and estimated entirely from the WMAP data [11] is sufficient for our estimation. We recover a flat, (approximately $140\mu K^2$ for WMAP 1 year data and $100\mu K^2$ for WMAP 3 year data), residual point source contamination in the two ‘Uniform average’ power spectra for $l \gtrsim 400$. This residual is much less than actual point source contamination in Q, KA or K band and intermediate between V and W band point source contamination.

C. Estimation of residual unresolved point source power spectrum

There have been several studies regarding the point source contamination in the CMB maps [36, 37, 38]. 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]. The WMAP point source model consists of a point source covariance matrix in thermodynamic temperature unit following $C_l^{ps(ij)} = c_{(ij)} A (\nu_i/\nu_0)^{-2} (\nu_j/\nu_0)^{-2}$ for two different frequency channel i and j . Here A is the amplitude of unresolved point source spectrum in antenna temperature at a reference frequency ν_0 and c_{ij} is the conversion factor from antenna to thermodynamic temperature. We compute the residual unresolved point source contamination in each of the 24 cross power spectra assuming that the residual (extra galactic) point source contamination has a statistically isotropic distribution in the sky. As described in section III A, the foreground cleaning procedure consists of r number of iterative steps. The total point source contamination in any of the cross power spectra consists of contribution from each of the individual sky parts. The final residual unresolved point source contamination is the uniform average over all such 24 residual point source contamination. Below we describe in detail residual point source contamination in a cross combination of two maps (\mathbf{i}, \mathbf{j}) .

As stated in section III B we apply Kp2 mask on the cleaned maps prior to cross power spectrum estimation. Apart from removing the galactic region, Kp2 mask also removes a circular region of 0.6 degree radius around each of the resolved point sources. Therefore the resolved point sources cannot affect the cross power spectra. However in the detector noise dominated region, $l \geq 400$, the iterative foreground cleaning method cannot remove all power from the unresolved point sources. As resolved point sources are already masked it is important to estimate only residual unresolved point source contamination in each of the cross power spectrum.

Let us assume that $a_{lm}^{\mathbf{i}}$ is the spherical harmonic coefficients obtained from cleaned map $C_{\mathbf{i}}$, where $\mathbf{i} = 1, 2, \dots, 48$,

$$a_{lm}^{\mathbf{i}} = B_l^{\mathbf{i}} a_{lm}^c + B_l^{\mathbf{i}} a_{lm}^{p(\mathbf{i})}. \quad (28)$$

Here $a_{lm}^{p(\mathbf{i})}$ is the residual unresolved point source contamination. We note that we have not considered any detector noise contribution in this equation. However this does not imply any loss of generality in the discussion of cross power spectrum. The noise bias does not affect the cross power spectrum. Also we do not include any residual diffuse foreground in our study in this section. At the large multipole region where point source contribution becomes significant diffuse galactic contaminations are entirely subdominant and thus they are not a concern.

We assume a statistically isotropic model of the residual unresolved point source distribution over the sky. Point sources are uncorrelated with CMB. On the ensemble average a partial cross power spectrum obtained from cleaned maps \mathbf{i} and \mathbf{j} is related to the CMB and point source power spectra as follows

$$\langle \tilde{C}_l^{\mathbf{ij}} \rangle = M_{ll'} \left(\langle \hat{C}_l^c \rangle + \langle \hat{C}_l^{p(\mathbf{ij})} \rangle \right) B_l^{\mathbf{i}} B_{l'}^{\mathbf{j}} p_{l'}^2. \quad (29)$$

⁷ There exists the additional freedom to choose optimal weights for combining the 24 cross-power spectra which we do not discuss in this work.

Here, $M_{ll'}$ is the coupling matrix, $\hat{C}_{l'}^{p(\mathbf{ij})}$ is the estimate for the full sky residual unresolved point source power spectrum and $B_l^i, B_{l'}^j, p_l^2$ denote combined effect of beam and pixel smoothing. We can recast the above equation as,

$$\langle \hat{C}_l^c \rangle = \frac{M_{ll'}^{-1} \langle C_{l'}^{\mathbf{ij}} \rangle}{B_l^i B_{l'}^j p_l^2} - \langle \hat{C}_{l'}^{p(\mathbf{ij})} \rangle. \quad (30)$$

The next task is to obtain the estimates $\hat{C}_{l'}^{p(\mathbf{ij})}$ themselves in each cross spectra. To compute them, let us assume that $f^{\mathbf{i}}(\theta, \phi)$ is the residual point source function present in the \mathbf{i}^{th} cleaned map. Our cleaning method partitions the sky into 9 parts. The entire sky is then cleaned in a total of 9 iterations. If $g_k(\theta, \phi)$ is the point source residual present in the k^{th} sky part, we have

$$f^{\mathbf{i}}(\theta, \phi) = \sum_{k=1}^{k=9} g_k(\theta, \phi). \quad (31)$$

After expanding both sides in spherical harmonics we obtain

$$a_{lm}^{\mathbf{i}} = \sum_{k=1}^{k=9} \tilde{a}_{lm}^k. \quad (32)$$

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

$$\tilde{a}_{lm}^k = \sum_{l'm'} M_{lml'm'}^k a_{l'm'}^k. \quad (33)$$

We note that, $a_{l'm'}^k$ represents the residual point source contamination in the *entire cleaned map* obtained after the k^{th} iteration. This is different from the contamination, \tilde{a}_{lm}^k , that is actually present in the k^{th} sky part. The symbol $M_{lml'm'}^k$ denotes the mode mode coupling matrix for the spherical harmonic modes for the k^{th} sky part. We can rewrite $a_{l'm'}^k$ in terms of the spherical harmonic coefficients of the temporary maps obtained at k^{th} iteration as,

$$a_{l'm'}^k = \sum_{i=1}^{i=4} a_{l'm'}^{ki} \hat{w}_{l'}^{ki}, \quad (34)$$

where $\hat{w}_{l'}^{ki}$ is the weight for the i^{th} channel at the k^{th} iteration for multipole l . Following eqs. (32), (33) and (34) we obtain

$$a_{lm}^{\mathbf{i}} = \sum_{k=1}^{k=9} \sum_{l'm'} M_{lml'm'}^k \sum_{i=1}^{i=4} a_{l'm'}^{ki} \hat{w}_{l'}^{ki}. \quad (35)$$

The residual unresolved point source spectrum in one cross combination (\mathbf{i}, \mathbf{j}) is given by

$$\langle \hat{C}_l^{ps(\mathbf{ij})} \rangle = \sum_{m=-l}^{m=l} \frac{\langle a_{lm}^{\mathbf{i}} a_{lm}^{*\mathbf{j}} \rangle}{2l+1}. \quad (36)$$

Substituting $a_{lm}^{\mathbf{i}}$ and $a_{lm}^{*\mathbf{j}}$ we find,

$$\langle \hat{C}_l^{ps(\mathbf{ij})} \rangle = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{k,k'=1}^{k,k'=9} \sum_{l',m'} \sum_{l'',m''} M_{lml'm'}^k M_{lml''m''}^{*k'} \sum_{i,i'=1}^{i,i'=4} \langle a_{l'm'}^{ki} a_{l''m''}^{*k'i'} \hat{w}_{l'}^{ki} \hat{w}_{l''}^{k'i'} \rangle. \quad (37)$$

We note that the two set of weights corresponding to the two cleaned maps which are cross correlated are not strictly identical. Therefore, in the above equation, we have used a prime to distinguish between the weights for two cleaned maps.

We note that the primary foreground contamination at large multipole region comes from (extra-galactic) point sources. Diffuse galactic foregrounds are subdominant at large l . The beam deconvolution process leads to dominance of the effective noise of the DA maps over the unresolved point source contamination. As a result weights become entirely determined by the noise level of the maps and asymptotically ($l \rightarrow \infty$) become constant for all realizations.

Moreover, small differences in weights from realization to realization due to fluctuations in noise from the mean level is not a concern. These fluctuations could be further suppressed by computing binned estimate of the power spectrum. Using simulations of the cleaning procedure with realistic model of point sources and detector noise we have verified that the weights remain effectively unchanged whether we use point sources or not. Since the pairs of cleaned maps that are cross-correlated have uncorrelated noise, the corresponding weights \hat{w}^{ki} and $\hat{w}'^{k'i'}$ could be treated as uncorrelated with one another. Also, weights are effectively uncorrelated with unresolved point sources as they are determined by the detector noise only. In this case the residual unresolved point source contamination becomes

$$\langle C_l^{psij} \rangle = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{k,k'=1}^{k,k'=9} \sum_{l',m'} \sum_{l'',m''} M_{lm'l'm'}^k M_{lm'l''m''}^{*k'} \sum_{i,i'=1}^{i,i'=4} \langle a_{l'm}^{ki} a_{l''m''}^{*k'i'} \rangle \langle \hat{w}_{l'}^{ki} \hat{w}_{l''}^{k'i'} \rangle. \quad (38)$$

Next we assume that the distribution of unresolved point sources is statistically isotropic over the full-sky, i.e.

$$\langle a_{l'm}^{ki} a_{l''m''}^{*k'i'} \rangle = C_l^{ps(ii')} \delta_{l'l''} \delta_{m'm''}. \quad (39)$$

Here $C_l^{ps(ii')}$ is the point source model as supplied by the WMAP team. This simplifies the expression for the residual unresolved point source contribution,

$$\langle C_l^{psij} \rangle = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{k,k'=1}^{k,k'=9} \sum_{l',m'} M_{lm'l'm'}^k M_{lm'l'm'}^{*k'} \sum_{i,i'=1}^{i,i'=4} C_l^{ps(ii')} \langle \hat{w}_{l'}^{ki} \hat{w}_{l'}^{k'i'} \rangle. \quad (40)$$

The residual unresolved point source contamination then could be written as,

$$\langle \hat{C}_l^{ps(ij)} \rangle = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{k,k'=1}^{k,k'=9} \sum_{l',m'} M_{lm'l'm'}^k M_{lm'l'm'}^{*k'} \langle \hat{\mathbf{W}}_l^k \mathbf{C}_l^{ps} \hat{\mathbf{W}}_l^{k'} \rangle. \quad (41)$$

Here, $\hat{\mathbf{W}}_l^k$ is the row vector of weights for sky part k as in eq. (7). The elements of the point source power spectrum matrix are given by $C_l^{ps(ii')}$. We note in passing that this matrix is not symmetric because we are interested in residual point source power in the cross combination of two maps obtained from two cleaned maps which are linear combinations of (K, Q, V, W) and (KA,Q,V, W) respectively. Explicitly we use the following form of the matrix

$$C_l^{ps(ii')} = \begin{pmatrix} C_l^{KKA} & C_l^{KQ} & C_l^{KV} & C_l^{KW} \\ C_l^{KAQ} & C_l^{QQ} & C_l^{QV} & C_l^{QW} \\ C_l^{KAV} & C_l^{QV} & C_l^{VV} & C_l^{VW} \\ C_l^{KAW} & C_l^{WQ} & C_l^{WV} & C_l^{WW} \end{pmatrix}. \quad (42)$$

We note that, in general, point sources may not be perfectly correlated in all the frequencies. The point source power spectrum assumed by the WMAP team and used in this work assume the existence of a single point source template which may not be perfectly true. However, our point source correction method could easily incorporate the extra information of point source decoherence from frequency to frequency in the matrix in eq. (42).

From a detailed study of a single iteration cleaning method we verified that a residual point source bias in the auto or cross power spectrum of the cleaned maps could be very well approximated by $\hat{\mathbf{W}}_l^k \mathbf{C}_l^{ps} \hat{\mathbf{W}}_l^{k'T}$, without a need to compute $\langle \hat{\mathbf{W}}_l^k \mathbf{C}_l^{ps} \hat{\mathbf{W}}_l^{k'T} \rangle$. A detailed discussion of this analysis and the corresponding simulations are given in appendix D. Thus we propose $\langle \hat{C}_l^{ps(ij)} \rangle \approx \hat{C}_l^{ps(ij)}$. This simplifies the expression for the residual unresolved point source contribution

$$\hat{C}_l^{ps(ij)} = \frac{1}{2l+1} \sum_{m=-l}^{m=l} \sum_{k,k'=1}^{k,k'=9} \sum_{l',m'} M_{lm'l'm'}^k M_{lm'l'm'}^{*k'} \hat{\mathbf{W}}_l^k \mathbf{C}_l^{ps} \hat{\mathbf{W}}_l^{k'T}. \quad (43)$$

Eq. 43 can further be written as

$$\hat{C}_l^{ps(ij)} = \sum_{k=1}^{k=9} \sum_{l'} M_{ll'}^k \hat{\mathbf{W}}_l^k \mathbf{C}_l^{ps} \hat{\mathbf{W}}_l^{k'T} + \frac{1}{2l+1} \left(\sum_{m=-l}^{m=l} \sum_{k,k' (k \neq k')} \sum_{l',m'} M_{lm'l'm'}^k M_{lm'l'm'}^{*k'} \hat{\mathbf{W}}_l^k \mathbf{C}_l^{ps} \hat{\mathbf{W}}_l^{k'T} \right). \quad (44)$$

Here $M_{ll'}$ is the mode mode coupling matrix. The first term of the bracket is obtained in analogous method as shown in Ref. [34]. The last term can also be further simplified. For this purpose we note that

$$M_{lml'm'}^k = \sum_{l''m''} w_{l''m''}^k (-1)^{m'} \left[\frac{(2l+1)(2l'+1)(2l''+1)}{4\pi} \right]^{1/2} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} l & l' & l'' \\ m & -m' & m'' \end{pmatrix}. \quad (45)$$

where w_{lm}^k are the spherical harmonic coefficients from the k^{th} mask. Hence

$$\begin{aligned} \sum_{mm'} M_{lml'm'}^k M_{lml'm'}^{*k'} &= \frac{(2l+1)(2l'+1)}{4\pi} \sum_{l''m''} \sum_{l'''m'''} w_{l''m''}^k w_{l'''m'''}^{k'} ((2l''+1)(2l'''+1))^{1/2} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \\ &\times \begin{pmatrix} l & l' & l''' \\ 0 & 0 & 0 \end{pmatrix} \sum_{mm'} \begin{pmatrix} l & l' & l'' \\ m & -m' & m'' \end{pmatrix} \begin{pmatrix} l & l' & l''' \\ m & -m' & m''' \end{pmatrix}. \end{aligned} \quad (46)$$

Using the property of the Wigner $3jm$ symbol we have

$$\begin{aligned} \sum_{mm'} M_{lml'm'}^k M_{lml'm'}^{*k'} &= \frac{(2l+1)(2l'+1)}{4\pi} \sum_{l''m''} \sum_{l'''m'''} w_{l''m''}^k w_{l'''m'''}^{k'} ((2l''+1)(2l'''+1))^{1/2} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix} \\ &\times \begin{pmatrix} l & l' & l''' \\ 0 & 0 & 0 \end{pmatrix} \delta_{l''l'''} \delta_{m''m'''} \frac{1}{2l''+1}. \end{aligned} \quad (47)$$

Hence

$$\frac{1}{2l+1} \sum_{mm'} M_{lml'm'}^k M_{lml'm'}^{*k'} = \frac{(2l'+1)}{4\pi} \sum_{l''m''} \sum_{l'''m'''} \hat{w}_{l''m''}^k \hat{w}_{l'''m'''}^{k'} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}^2. \quad (48)$$

This is easily written in terms of the cross power spectra $w_{l''}^{kk'}$ of two masks k, k'

$$\frac{1}{2l+1} \sum_{mm'} M_{lml'm'}^k M_{lml'm'}^{*k'} = \frac{(2l'+1)}{4\pi} \sum_{l''} (2l''+1) w_{l''}^{kk'} \begin{pmatrix} l & l' & l'' \\ 0 & 0 & 0 \end{pmatrix}^2 = M_{ll'}^{kk'}. \quad (49)$$

We have defined $M_{l''}^{kk'}$ as the cross coupling matrix of two masks k, k' . Finally we can write eq. (37) as

$$\hat{C}_l^{ps(ij)} = \sum_{k=1}^{k=9} \sum_{l'} M_{ll'}^k \hat{\mathbf{W}}_l^k \mathbf{C}_l^{ps} \hat{\mathbf{W}}_l^{kT} + \sum_{k,k' \neq k'} \sum_{l'} M_{ll'}^{kk'} \hat{\mathbf{W}}_l^k \mathbf{C}_l^{ps} \hat{\mathbf{W}}_l^{k'T}, \quad (50)$$

which estimates unresolved residual point source contamination in individual cross power spectrum. After correcting each of the cross power spectra we form a simple average to obtain final point source corrected spectrum. We note that the second term on the right hand side of the above equation gives entirely 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 limit.* In Ref. [22], we have shown that the residual point source contamination is significantly smaller than the contamination arising from K, Ka, Q or V band. In fact point source residual is intermediate between V and W band point source power. Figure 5 shows, residual unresolved point source contamination $\sum_{k=1}^{k=9} \sum_{l'} M_{ll'}^k \hat{W}_l^{ki} C_l^{ps(ij)} \hat{W}_l^{kj}$ for different values of i, j . Here i, j are the index representing the 4 DA. All possible combinations (i, j) are explicitly shown in eq. (42). The total unresolved point source spectrum is shown as the pink line with filled circular points. 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 band 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 basis of the WMAP team's 1 year power spectrum are 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. Following the WMAP team's 1 year bins we compare these 28 cross power spectra in Fig. 6 with 24 cross spectra obtained from our own 1 year results. These 24 cross spectra also are not corrected for the residual unresolved point sources and show very little dispersion compared to the WMAP results. The 'uniform average' power spectrum plotted in green line in fig. 7 has less excess power near the second acoustic peak compared to an 'uniform average' of the WMAP's 28 cross spectra (red line). This merely shows that we have removed some amount of point sources at this range of l during our cleaning.

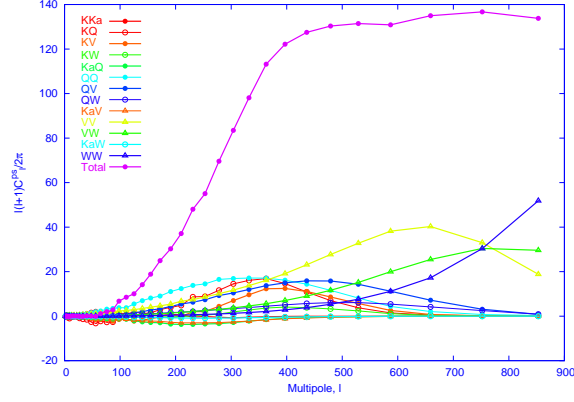


FIG. 5: The effective contribution from each of the elements of the matrix $C_U^{ps(ij)}$ in the total residual point source estimation, where i, j are the index representing the 4 DA we use for combinations.

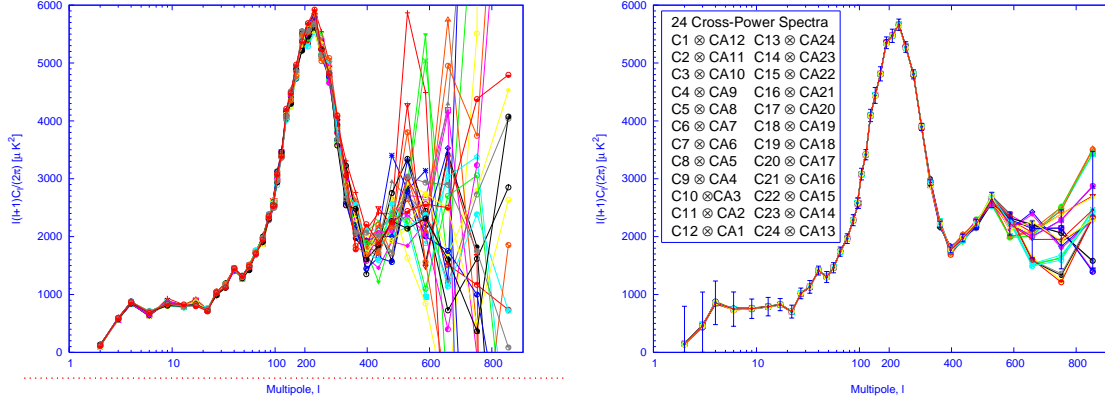


FIG. 6: The left panel shows 28 cross power spectra that would have been obtained by WMAP team's 1 year analysis without any point source subtraction. The right panel shows our 24 cross spectra based on 1 year data.

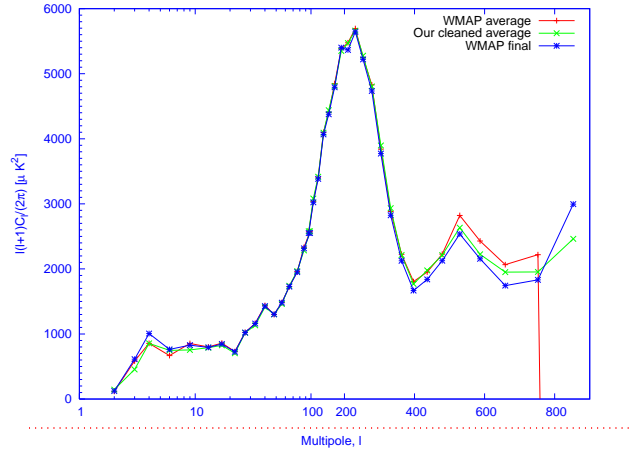


FIG. 7: Comparison of the WMAP's average power spectrum (red) with our binned power spectrum without any point source subtraction (green). Clearly starting from the first acoustic peak we have less point source contamination. WMAP's final binned power spectrum is also shown in blue for comparison. Interestingly the notch at $l=4$ appears to be reduced in the WMAP's average power spectrum (red).

D. Computing error bars

We rely upon Monte-Carlo simulation to compute the error bars on our power spectrum. We generate synchrotron, free-free and thermal dust maps corresponding to different frequencies in a given combination using the publicly available Planck Sky Model. Each of the random realizations of 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 the number $\sigma_0/\sqrt{N_p}$ 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 .

Due to 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^c becomes stronger. The covariance matrix $\langle \Delta \hat{C}_l^c \Delta \hat{C}_{l'}^c \rangle = \langle (\hat{C}_l^c - \langle \hat{C}_l^c \rangle)(\hat{C}_{l'}^c - \langle \hat{C}_{l'}^c \rangle) \rangle$ obtained from simulations is therefore expected to have non-diagonal 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}_b denotes the binned power spectrum. Then the covariance matrix of the binned spectrum is obtained as $\langle \Delta \hat{C}_b^c \Delta \hat{C}_{b'}^c \rangle = \langle (\hat{C}_b^c - \langle \hat{C}_b^c \rangle)(\hat{C}_{b'}^c - \langle \hat{C}_{b'}^c \rangle) \rangle$. The standard deviation obtained from the diagonal elements of the binned covariance matrix was used as the error-bars on the binned final spectrum extracted from the WMAP data. We also define a normalized covariance matrix of the binned power spectrum following,

$$C_{bb'} = \frac{\langle \Delta \hat{C}_b^c \Delta \hat{C}_{b'}^c \rangle}{\sqrt{\langle (\Delta \hat{C}_b^c)^2 \rangle \langle (\Delta \hat{C}_{b'}^c)^2 \rangle}}, \quad (51)$$

wherein all the elements of this matrix are bound to lie between $[-1, 1]$. This correlation matrix represents the actual bin to bin correlation matrix following the cleaning method. In the left panel of the figure 8 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 sufficiently diagonal dominated.

IV. RESULTS

Figure 9 shows the main result of the power spectrum for CMB anisotropy estimated using our analysis of WMAP 1 year and WMAP 3 year data. The blue line shows the WMAP 3 year power spectrum and the red line shows the 1 year spectrum. All 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 1 year and 3 year respectively. In what follows we describe the power spectrum obtained by us for WMAP 1 year data and WMAP 3 year data respectively.

A. WMAP 1 year data

Using the 48 foreground cleaned maps obtained from the WMAP 1 year maps we obtain a ‘Uniform average’ power spectrum following the method mentioned in section III B. The residual unresolved point source contamination is removed following III C. The estimated power spectrum with error bars is plotted in red in figure 9.

We find a suppression of power in the quadrupole and octupole moments consistent with the results published by the WMAP team. However, our quadrupole moment ($146\mu K^2$) is little larger than the quadrupole moment estimated by WMAP team ($123\mu K^2$) and Octupole ($455\mu K^2$) is less than the WMAP team result ($611\mu K^2$). The ‘Uniform average’ power spectrum does not show the ‘bite’ like feature present at the first acoustic peak in the power spectrum reported by WMAP [10]. 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 [39]. For the residual point source corrected power spectrum we obtain the first acoustic peak at $l = 219.8 \pm 0.8$ with the peak amplitude $\Delta T_l = 74.1 \pm 0.3\mu K$, the second acoustic peak at $l = 544 \pm 17$ with the peak amplitude $\Delta T_l = 48.3 \pm 1.2\mu K$ and the first trough at $l = 419.2 \pm 5.6\mu K$ with peak amplitude $\Delta T_l = 41.7 \pm 1\mu K$. The left panel of figure 10 shows the three different ranges of multiples used to find out peak and trough positions and their corresponding amplitudes ΔT_l . (A similar plot for the WMAP 3 year data is shown in the right panel of this figure. The results for WMAP 3 year analysis are summarized in section IV B.)

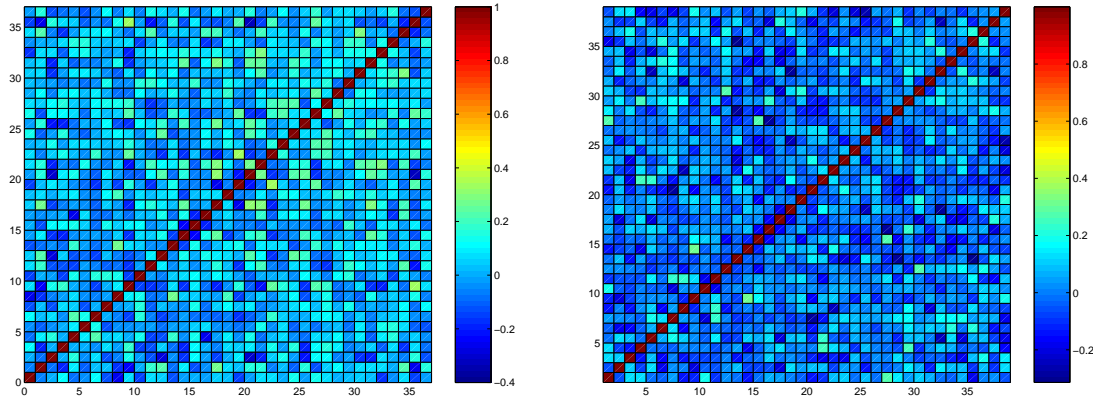


FIG. 8: Correlation matrix $\frac{\langle \Delta C_b \Delta C_{b'} \rangle}{\sqrt{\langle (\Delta C_b)^2 \rangle \langle (\Delta C_{b'})^2 \rangle}}$ from our simulation plotted with respect to the bin index. As the figure shows the matrix is mostly dominated by the diagonal elements.

As a cross check of the method, we have carried out the analysis with other possible combinations of the DA maps.

1. The WMAP team also provide foreground cleaned maps corresponding to Q1 to W4 DA (LAMBDA). The Galactic foreground signal, consisting of synchrotron, free-free, and dust emission, was removed using the 3-band, 5-parameter template fitting method [11]. We also include K and Ka band maps which are not foreground cleaned. The resulting power spectrum from our analysis matches closely to the ‘Uniform average’ power spectrum.
2. Excluding the K and Ka band from our analysis we get a power spectrum close to the ‘Uniform average’ results. Notably, we find a more prominent notch at $l = 4$ similar to WMAP’s published results.

In case of ‘Uniform average’ a maximum difference of $92 \mu K^2$ is observed only for octupole. For the large multipole range the difference is small and for $l = 752$ it is approximately $48 \mu K^2$. This is well within the 1σ error bar ($510 \mu K^2$) obtained from the simulation. This shows that our foreground cleaning is comparable in efficiency to that obtained by employing template fitting methods that rely on a model of foreground emission to estimate the contamination at the CMB dominated frequencies.

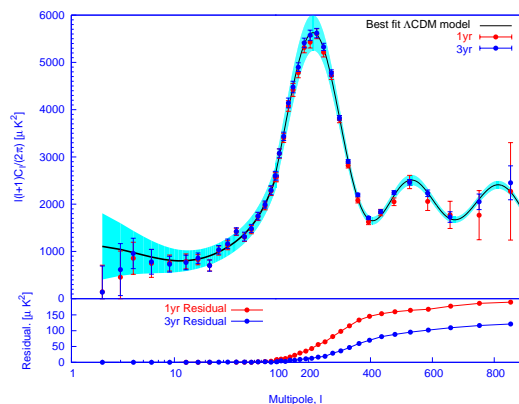


FIG. 9: Comparison of 1 year 4 channel with that of 3 year 4 channel power spectrum. The best fit WMAP’s power spectrum is shown in black line along with cosmic variance band. The bottom panel of this figure shows the residual unresolved point source contamination for both the power spectra.

The Monte Carlo simulations of our cleaning method also reveals the negative bias in the low l moments. The origin of this negative bias is explained in section II B. For $l = 2$ and $l = 3$ the bias is -27.4% and -13.8% respectively. However this bias become negligible at higher l , e.g. at $l = 22$, it is only -0.8% . This bias can be explained in terms of an anti-correlation of the CMB with the residual foregrounds in the cleaned map. For further details of the bias we refer to appendix E. The standard deviation obtained from the diagonal elements of the covariance matrix is used

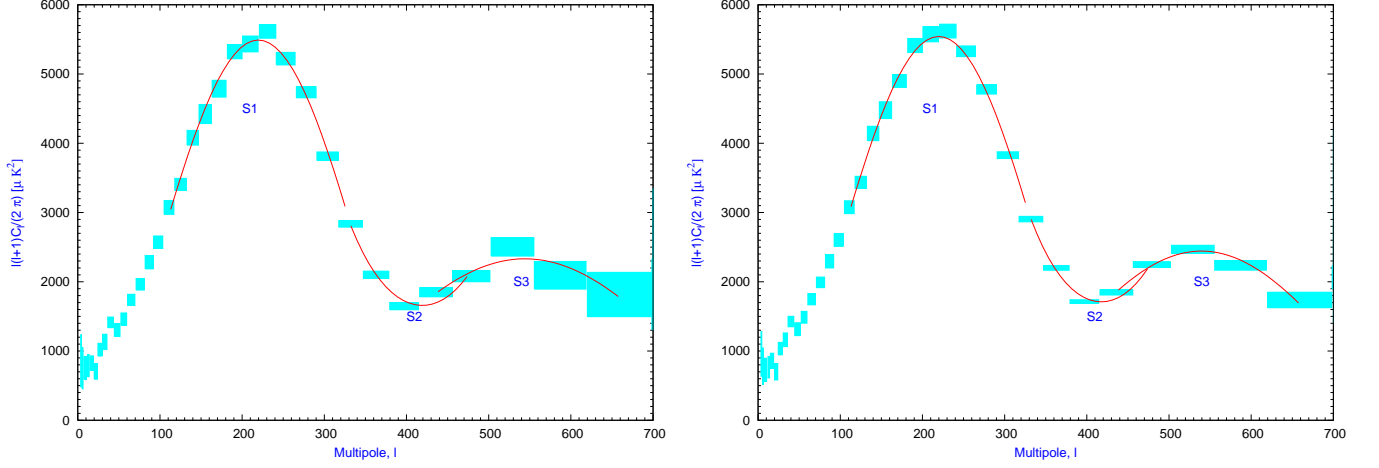


FIG. 10: The left panel shows the 3 different multipole ranges used to obtain positions of the first peak, the first trough and the second peak from our point source subtracted power spectrum using 1 year WMAP data. Before fitting the 1 year power spectrum was binned in the same manner as the WMAP’s 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.

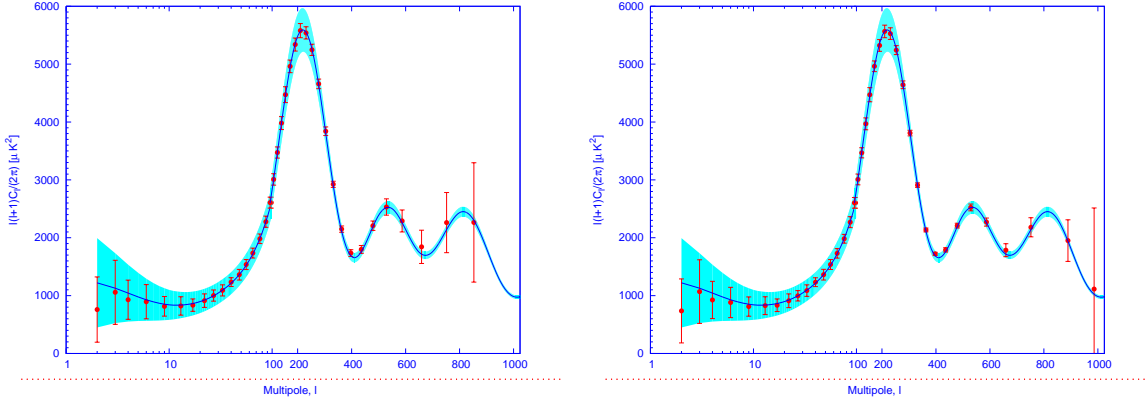


FIG. 11: The left panel shows (in red points) ensemble averaged power spectrum from 110 Monte Carlo simulations of our power spectrum estimation method. The simulations were carried out using the 1 year WMAP detector noise maps available from the LAMBDA website. We use publicly available Planck Sky Model to generate the diffuse foreground models. The recovered spectrum is binned in the same manner as WMAP 1 year power spectrum. The input theoretical spectrum is shown in blue line with cosmic variance. The right panel is same figure but with 3 year noise maps. The 3 year noise maps were generated following the method described in the text. The spectrum is binned following the binning scheme of the WMAP team’s analysis of 3 year data.

as the error bars on the C_l ’s obtained from the data. The ensemble average of 110 cleaned power spectrum is shown in the left panel of the figure 11. The presence of bias at the low multipole moments is visible in this figure.

B. WMAP 3 year data

1. 4 channel combinations

We analyse the 3 year WMAP data by using a procedure identical to that used for the 1 year data. The details are given in section II. In figure 12 we show each of the 24 cross power. The ‘Uniform average’ of these 24 cross power spectra is also shown in this figure in red line with blue error-bar. This figure is similar to the 24 individual cross power spectra obtained for WMAP 1 year data [21, 22]. 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 shows small dispersion only till the second acoustic peak [21]. This may be explained due to the effectively lower detector noise in the 3 year data compared to

the 1 year data.

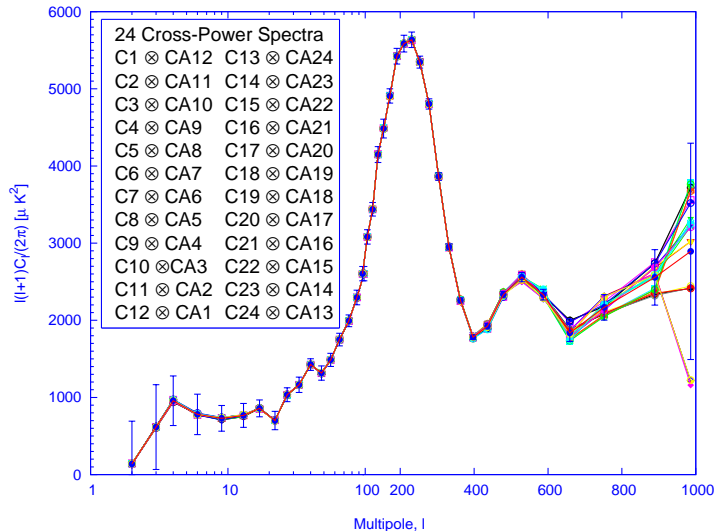


FIG. 12: The 24 cross power spectra for the 3 year WMAP data are shown with the detector noise bias removed. The red line with blue error-bars is the ‘Uniform average’ power spectrum. In the inset of this figure we show all the possible cross correlations of the cleaned maps that give rise to the 24 cross spectra.

We perform a peak fitting to the peaks and troughs of the 3 year power spectrum as well. A parabolic function of the form $\Delta T_l = \Delta T_{l_0} + \alpha(l - l_0)^2$ is fitted to the peaks and troughs of the power spectrum amplitude. We use the 3 year binned data for the fitting purpose. For the point source corrected power spectra, the first acoustic peak has an amplitude $\Delta T_{l_0} = 74.4 \pm 0.3 \mu K$ at the multipole position $l = 219.9 \pm 0.8$. The first trough is located at $l = 417.7 \pm 3.2$ with an amplitude $\Delta T_{l_0} = 41.4 \pm 0.6 \mu K$. The amplitude and position of the second acoustic peak are given by $\Delta T_{l_0} = 49.4 \pm 0.4 \mu K, l = 539.5 \pm 3.7$. In the right panel of figure 10 we show the different multipole ranges used to obtain the position of the peaks and troughs and their amplitudes.

2. 3 channel combinations

We also follow an alternative approach in which we use only the Q,V and W band DA maps. This is similar to the 4 channel cleaning, however now we only get 24 cleaned maps. We form a total of 12 cross power spectra from these cleaned maps after applying 3 year Kp2 mask. After debiasing all the power spectra by the coupling matrix and removing beam and pixel effect we obtain an ‘3 channel uniform average’ power spectrum. We find that the 3 channel spectrum matches well with the 4 channel spectrum.

C. Comparison of 1 year and 3 year power spectra

Fig. 9 compares the power spectra obtained from the 1 year and 3 WMAP data binned identically using the WMAP team’s 1 year binning method. They match closely with one another and also with the WMAP’s best fit power spectrum available from the LAMBDA website. For the residual point source corrected power spectrum we obtain for WMAP 1 year (WMAP 3 year) data the first acoustic peak at $l = 219.8 \pm 0.8$ (219.9 ± 0.8) with the peak amplitude $\Delta T_l = 74.1 \pm 0.3 \mu K$ ($74.4 \pm 0.3 \mu K$), the second acoustic peak at $l = 544 \pm 17$ (539.5 ± 3.7) with the peak amplitude $\Delta T_l = 48.3 \pm 1.2 \mu K$ ($49.4 \pm 0.4 \mu K$) and the first trough at $l = 419.2 \pm 5.6 \mu K$ (417.7 ± 3.2) with peak amplitude $\Delta T_l = 41.7 \pm 1 \mu K$ ($41.4 \pm 0.6 \mu K$).

We note that our cleaning method significantly removes unresolved point source contamination. The original l^2 dependence of the unresolved point source power spectrum present in the foreground contaminated maps (as well present in template cleaned maps) is significantly reduced and becomes independent of l at large l . We also note that the unresolved residual point source contamination is less by about $\approx 50 \mu K^2$ in 3 year power spectrum than the 1 year power spectrum. This is expected. The WMAP supplied 3 year Kp2 mask removes more point sources than the 1 year Kp2 mask.

V. CONCLUSION

The rapid improvement in the sensitivity and resolution of the CMB experiments has posed increasingly stringent requirements on the level of separation and removal of the foreground contaminants. We carry out an estimation of the CMB power spectrum from the WMAP data that is independent of foreground model. The method does not rely upon any foreground template and employs the lack of noise correlation between independent channels. This paper is a detailed description of the first estimate of the CMB angular power spectrum solely based upon the WMAP data. In this paper, we present an indepth study of the biases that arise in the foreground cleaning. In particular, we provide an understanding and correction for the negative bias at low multipoles reported in our earlier work [21].

Usual approaches to foreground removal, usually incorporate the extra information about the foregrounds available at other frequencies, the spatial structure and distribution in constructing a foreground template at the frequencies of the CMB measurements. These approaches may be susceptible to uncertainties and inadequacies of modeling involved in extrapolating from the frequency of observation to CMB observations that a blind approach, such as presented here, evades. The understanding of polarized foreground for CMB polarization maps is rather scarce. Hence modeling uncertainties could dominate the systematics error budget of conventional foreground cleaning. The blind approach extended to estimating polarization spectra after cleaning CMB polarization maps could prove to be particularly advantageous.

Acknowledgments

The analysis pipeline as well as the entire simulation pipeline is based on primitives from the HEALPix package.⁸ 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. RS acknowledges the Indo-French Sandwich Fellowship granted by the French Embassy in India and EGIDE in Paris, France. RS thanks IAP for hosting his visit. RS 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 Garry Hinshaw on the unresolved point source model. We thank Amir Hajian, Subharthi Ray and Sanjit Mitra in 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 minimizing the total power \hat{C}_l^{Clean} of the cleaned map. However, we ensure that the CMB angular power spectrum is conserved during cleaning by imposing the constraint $\hat{\mathbf{W}}_1 \mathbf{e}_0 = \mathbf{e}_0^T \hat{\mathbf{W}}_1^T = 1$ on the weights. The solution for the weights that satisfy these conditions is the point in weight space where normals to the functions $f(\hat{\mathbf{W}}_1) = \hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T$ and $g(\hat{\mathbf{W}}_1) = \hat{\mathbf{W}}_1 \mathbf{e}_0$ are parallel to one another. Following Lagrange's multiplier method, this is cast to an equivalent problem of minimizing

$$\hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T - \lambda \hat{\mathbf{W}}_1 \mathbf{e}_0. \quad (\text{A1})$$

Here λ is the unknown Lagrange multiplier parameter which could be determined from variational principle. At the extrema, the expression in eq. (A1) is unchanged under small variations in $\hat{\mathbf{W}}_1$ leading to

$$\Delta \hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T + \hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1 \Delta \hat{\mathbf{W}}_1^T - \lambda \Delta \hat{\mathbf{W}}_1 \mathbf{e}_0 = 0. \quad (\text{A2})$$

Since the power spectrum matrix, $\hat{\mathbf{C}}_1$ is a symmetric matrix, the first two terms of the left hand side are equal to one another. Hence we obtain,

$$\Delta \hat{\mathbf{W}}_1 \left[2 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T - \lambda \mathbf{e}_0 \right] = 0. \quad (\text{A3})$$

⁸ The HEALPix distribution is publicly available from the website <http://healpix.jpl.nasa.gov>.

Since this relation is true for any arbitrary variation $\Delta \hat{\mathbf{W}}_1$, we obtain

$$\left[2\hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T - \lambda \mathbf{e}_0 \right] = 0. \quad (\text{A4})$$

We introduce a (non zero) square matrix $\hat{\mathbf{G}}_1$. Later we will identify $\hat{\mathbf{G}}_1$ as the *Moore Penrose Generalized Inverse* (MPGI) of the covariance matrix $\hat{\mathbf{C}}_1$. After multiplication from left by this matrix we can rewrite the above equation as

$$2\hat{\mathbf{G}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T - \lambda \hat{\mathbf{G}}_1 \mathbf{e}_0 = 0. \quad (\text{A5})$$

Hence we obtain

$$\lambda = \frac{2\mathbf{e}_0^T \hat{\mathbf{G}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T}{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \mathbf{e}_0}. \quad (\text{A6})$$

Now we use the constraint $\hat{\mathbf{W}}_1 \mathbf{e}_0 = 1$. Assuming $\hat{\mathbf{W}}_1 \neq \mathbf{0}$ we multiply eq. (A4) from left by $\hat{\mathbf{W}}_1$ and to obtain

$$2\hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T = \lambda \hat{\mathbf{W}}_1 \mathbf{e}_0 = \lambda. \quad (\text{A7})$$

Using eq. (A6) and eq. (A7) we obtain

$$2\hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T = 2 \frac{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T}{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \mathbf{e}_0}, \quad (\text{A8})$$

or,

$$\left(\hat{\mathbf{W}}_1 - \frac{\mathbf{e}_0^T \hat{\mathbf{G}}_1}{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \mathbf{e}_0} \right) \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T = 0. \quad (\text{A9})$$

If we neglect solutions which belongs to null space of $\hat{\mathbf{C}}_1$,⁹ i.e., assuming $\hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T \neq \mathbf{0}$, we obtain

$$\hat{\mathbf{W}}_1 = \frac{\mathbf{e}_0^T \hat{\mathbf{G}}_1}{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \mathbf{e}_0}. \quad (\text{A10})$$

We impose the restriction on $\hat{\mathbf{G}}_1$ that it is symmetric (since the MPGI of a symmetric matrix is symmetric) to obtain

$$\hat{\mathbf{W}}_1^T = \frac{\hat{\mathbf{G}}_1 \mathbf{e}_0}{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \mathbf{e}_0}. \quad (\text{A11})$$

The corresponding power spectrum of the cleaned map is given by

$$\hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T = \frac{\mathbf{e}_0^T \hat{\mathbf{G}}_1}{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \mathbf{e}_0} \hat{\mathbf{C}}_1 \frac{\hat{\mathbf{G}}_1 \mathbf{e}_0}{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \mathbf{e}_0} = \frac{\mathbf{1}}{\mathbf{e}_0^T \hat{\mathbf{G}}_1 \mathbf{e}_0}, \quad (\text{A12})$$

where we have imposed the condition $\hat{\mathbf{G}}_1 \hat{\mathbf{C}}_1 = \hat{\mathbf{C}}_1 \hat{\mathbf{G}}_1$. It is easy to note that the choice of symmetric $\hat{\mathbf{G}}_1$ helps to obtain a simplified expression for the cleaned power spectrum. Also one can verify that, if $\hat{\mathbf{G}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{G}}_1 = \hat{\mathbf{G}}_1$ is satisfied and $\hat{\mathbf{G}}_1$ is symmetric then $\hat{\mathbf{G}}_1$ satisfies all the defining conditions of MPGI of $\hat{\mathbf{C}}_1$. Hence, we can identify $\hat{\mathbf{G}}_1$ as the Moore-Penrose Generalized Generalized Inverse (MPGI) of $\hat{\mathbf{C}}_1$ (and vice versa).

Eq. (A12) has an interesting property. It remains valid even when the full covariance matrix $\hat{\mathbf{C}}_1$ is singular. A singular full covariance matrix is encountered in noiseless case (or numerically very low noise case) when the foreground components follow a rigid frequency scaling and total number of components (all foregrounds and CMB) become less than the number of available channels. If $\hat{\mathbf{C}}_1$ is non singular we can replace $\hat{\mathbf{G}}_1$ everywhere by $\hat{\mathbf{C}}_1^{-1}$. This is because MPGI of a non-singular matrix is its inverse.

⁹ There is a physical justification behind neglecting this solution. The weights satisfying $\hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T = \mathbf{0}$ do not preserve CMB power while minimizing total power in the cleaned map. These solution for the weights merely sets the total power in the cleaned map $\hat{\mathbf{W}}_1 \hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T = \mathbf{0}$. Thus we are not interested in the space of solutions to $\hat{\mathbf{C}}_1 \hat{\mathbf{W}}_1^T = \mathbf{0}$.

APPENDIX B: PARTITIONING THE SKY

An important advantage of the foreground cleaning analyzed here 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 partition could be done directly using the knowledge of expected spectral index dependence on the sky.) We followed the procedure of Ref. [18] to partition the sky. Each of these partitions are identified with a sky masks. The mask takes non zero value at all pixels contained within the sky partition represented and is zero outside. In this section we describe the procedure of constructing these masks.

There are 2 Difference Assemblies for Q band, 2 for V band and 4 for W band. To make the masks we first averaged all the DA maps for a given frequency band. Correspondingly we averaged the beam functions for each frequency band. For the K and KA bands there are only one difference assembly in each case. Therefore for these bands no averaging was done. We smoothed all the 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 5 smoothed maps. The smoothing was performed first obtaining the a_{lm} coefficients. As each averaged map was effectively smoothed by the averaged beam function (corresponding to each channel) during the observations, we decided to first remove the beam effect. Then we smoothed each map by the common resolution of the lowest frequency band. Mathematically, the difference maps were obtained as follows:

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

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

$$a_{lm}^{Q-K} = a_{lm}^Q \frac{B_l^K}{B_l^Q} - a_{lm}^K, \quad (B3)$$

$$a_{lm}^{K-KA} = a_{lm}^K - a_{lm}^{KA} \frac{B_l^K}{B_l^{KA}}. \quad (B4)$$

The a_{lm} were converted to difference maps using 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 ugrade to a resolution of $N_{\text{side}} = 64$. We identify 7 different sky mask partitions from this low resolution junk map after applying cutoff corresponding to the following temperature thresholds (in μK) $T > 30000$, $30000 \geq T > 10000$, $10000 \geq T > 3000$, $3000 \geq T > 1000$, $1000 \geq T > 300$, $300 \geq T > 100$ and $T < 100$. The partition $T > 30000$ is maximally contaminated by foreground emission. The second dirtiest partition is disjoint on the sky and we use 3 separate masks corresponding this partition. The resulting 9 masks are then converted back to the HEALPix resolution $N_{\text{side}} = 512$ using ugrade routine of HEALPix. Next we smooth each mask using a Gaussian beam of FWHM $30'$ and redefine smoother mask boundaries at the threshold of 0.5.¹⁰ In this case we found that almost the entire sky is covered by the 9 masks (except for a few pixels in the sky).

To convert the second dirtiest region into 3 mask files we upgraded the $N_{\text{side}} = 64$ resolution map to the $N_{\text{side}} = 512$. Then using IDL task mollcursor, we found out the extension in galactic θ, ϕ coordinates of the 3 different parts of the

¹⁰ We found that if we smooth them by a Gaussian function of FWHM 2 degree (as mentioned in Ref. [18]) and then define the boundaries at the threshold of 0.5 the resulting 9 masks cannot cover the entire sky. In that case there are some regions near the galactic plane which do not belong to any of our 9 masks. Obviously if such regions are not covered by any of the masks and remain present in our final cleaned map then final power spectrum will be contaminated by the foreground. Therefore we chose to use a Gaussian function of lower FWHM of $30'$ for smoothing. The reason why in our case, masks smoothed by a 2 degree Gaussian function cannot 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 found that the masks near the galactic regions contain a few small isolated regions. Therefore smoothing by a Gaussian with FWHM as large as 2 degree, gives this isolated small regions maximum value far less than unity, in fact maximum value becomes quite near 0.5, the cutoff value, after smoothing. Consequently applying a cut of 0.5 removes most of the part of these isolated regions, leading to some part of the sky, uncovered by the masks.

second dirtiest region. We then implemented a method which determines the pixel index in ‘ring’ format for these 3 regions. Finally we converted them to 3 different masks at $N_{\text{side}} = 512$. The masks were then smoothed by Gaussian function of FWHM = 30'. We applied a cutoff of 0.5 to each of them. The 8th, 7th and 6th masks were numbered according to the descending order of maximum pixel value in the junk map at resolution $N_{\text{side}} = 512$.

Figure 2 shows our 9 different masks that partition the sky based on estimated level of foreground contamination. These regions are similar to what is shown in Ref. [18]. As the figure shows, one side of the band near the galactic plane is more severely foreground contaminated.

APPENDIX C: COMBINING CROSS POWER SPECTRA

The basis of our final power spectrum are a set of 24 cross power spectra where the detector noise bias has been removed. A uniform weighting of cross spectra is used to obtain the final power spectrum. In this section we describe the procedure to obtain the 24 cross spectra and their combination to form the final spectrum.

Let us assume that $\Delta T^i(\theta, \phi)$ represents one of the 48 final cleaned map, where ($i = 1, 2, 3, \dots, 48$). To remove the residual foreground contamination near the galactic plane (figure 3) we apply the Kp2 mask supplied by the WMAP team on each of these 48 maps. The masked map can be represented as

$$\Delta T'^i(\hat{n}) = W(\hat{n})\Delta T^i(\hat{n}). \quad (\text{C1})$$

Here $W(\theta, \phi)$ represents the Kp2 mask. In the next step, we obtain a cross power spectrum between pairs of foreground cleaned maps that have uncorrelated noise (recall that noise in different DAs are uncorrelated). If \tilde{a}_{lm}^i and \tilde{a}_{lm}^j are the spherical harmonic coefficients obtained from two such maps

$$\tilde{a}_{lm}^{i,j} = \int W(\hat{n})\Delta T^{i,j}(\hat{n})d\Omega. \quad (\text{C2})$$

The cross power spectrum is obtained using,

$$\tilde{C}_l^{ij} = \sum_{m=-l}^{m=l} \frac{\tilde{a}_{lm}^i \tilde{a}_{lm}^{j*}}{2l+1}. \quad (\text{C3})$$

Here the superscript ij represents the cross power spectrum obtained from the i^{th} and j^{th} foreground cleaned maps. Following Ref. [34], the ensemble average of the C_l estimated from the partial sky is related to the C_l from the full sky as

$$\langle \tilde{C}_l \rangle = \sum_{l'} M_{ll'} \langle \hat{C}_{l'} \rangle, \quad (\text{C4})$$

where $M_{ll'}$ is a coupling matrix. This matrix represents the fact that when we multiply our map by the weight function (Kp2 mask) in the pixel space we are effectively performing a smoothing operation and neighboring spherical harmonic coefficients get coupled. An analytic expression for this coupling matrix is given in Ref. [34],

$$M_{l_1 l_2} = \frac{2l_2 + 1}{4\pi} \sum_{l_3=|l_1-l_2|}^{l_3=l_1+l_2} (2l_3 + 1) W_{l_3} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2, \quad (\text{C5})$$

where the last term is the Wigner-3j symbol and W_l is the power spectrum of the mask under consideration. Although in our case we are interested in the partial sky power spectrum which has been obtained by cross correlating two different maps, it is easy to show that in this case also eq. (C4) remains valid. The rank of the coupling matrix is limited to $l_{\text{max}} = 1024$. We use numerical routines from the NETLIB package [40] to compute the Wigner -3j symbol.

Eq. (C4) is true only for an ideal observation with infinite angular resolution. In practice all instruments have finite angular resolution given by the beam function $B(\theta, \phi)$ of the instrument. Mathematically, we may write the observed temperature anisotropy,

$$\Delta T(\hat{n}) = \int \Delta T'(\hat{n}') B(\hat{n}, \hat{n}') d\Omega_{\hat{n}'}, \quad (\text{C6})$$

where $\Delta T'(\hat{n}')$ is now the full sky map in the absence of the beam. In most of the CMB experiments the beam function is circularly symmetric to a good approximation, i.e., depends only on the angle $\theta'' = \cos^{-1}(\hat{n} \cdot \hat{n}')$ between two directions \hat{n} and \hat{n}' ,

$$B(\hat{n}, \hat{n}') \equiv B(\hat{n} \cdot \hat{n}'). \quad (\text{C7})$$

We may expand this function in terms of Legendre polynomials,

$$B(\hat{n} \cdot \hat{n}') = \sum_{l=0}^{l=\infty} \frac{2l+1}{4\pi} B_l P_l(\hat{n} \cdot \hat{n}'). \quad (\text{C8})$$

Here, B_l 's are Legendre transform of the beam function. Substituting this in eq. (C6) and expanding ΔT and $\Delta T'$ in spherical harmonics, we obtain

$$\sum_{l=0}^{l=\infty} \sum_{m=-l}^{m=l} a_{lm} Y_{lm}(\hat{n}) = \sum_{l'=0}^{l'=\infty} \sum_{m'=-l'}^{m'=l'} a'_{l'm'} \int Y_{l'm'}(\hat{n}') \sum_{l''=0}^{l''=\infty} \frac{2l''+1}{4\pi} B_{l''} P_{l''}(\hat{n} \cdot \hat{n}') d\Omega_{\hat{n}'}. \quad (\text{C9})$$

Using the addition formula

$$P_{l''}(\hat{n} \cdot \hat{n}') = \frac{4\pi}{2l''+1} \sum_{m''=-l''}^{m''=l''} Y_{l''m''}(\hat{n}) Y_{l''m''}^*(\hat{n}') \quad (\text{C10})$$

and the orthonormality property of spherical harmonics

$$\int Y_{l'm'}(\hat{n}') Y_{l''m''}^*(\hat{n}') = \delta_{l'l''} \delta_{m'm''}, \quad (\text{C11})$$

we obtain,

$$\sum_{l=0}^{l=\infty} \sum_{m=-l}^{m=l} a_{lm} Y_{lm}(\hat{n}) = \sum_{l'=0}^{l'=\infty} \sum_{m'=-l'}^{m'=l'} a'_{l'm'} Y_{l'm'}(\hat{n}) B_{l'}. \quad (\text{C12})$$

We again use orthonormality of spherical harmonics to obtain

$$a_{lm} = a'_{lm} B_l. \quad (\text{C13})$$

This relation shows that due to the finite resolution of the instrument spherical harmonic coefficients get multiplied by the Legendre transform of the beam function. Hence it is easier to account for the effect of a circular beam in the spherical harmonic space than deconvolving the map by the beam function in the pixel space.

The effect of finite pixel size of the map has similar effect on the recovered spherical harmonic coefficients. The pixel window functions p_l have been supplied with the HEALPix distribution for different resolutions. Taking into account finite pixel size we have

$$a_{lm} = a'_{lm} B_l p_l. \quad (\text{C14})$$

Hence the recovered power spectrum is related to the actual CMB sky power spectrum as

$$\hat{C}_l = \hat{C}'_l B_l^2 p_l^2. \quad (\text{C15})$$

In presence of both beam and finite pixel effects, eq. (C4) obtained for the ideal case ¹¹ is now modified to

$$\langle \tilde{C}_l \rangle = \sum_{l'} M_{ll'} \langle \hat{C}'_{l'} \rangle B_{l'}^2 p_{l'}^2. \quad (\text{C16})$$

A generalization of this expression in case of cross power spectrum has been also reported in Ref. [35]. The partial sky cross power spectrum is related to the full sky power spectrum in the following manner

$$\langle \tilde{C}_l^{ij} \rangle = \sum_{l'} M_{ll'} \langle \hat{C}'_{l'} \rangle B_{l'}^i B_{l'}^j p_{l'}^2. \quad (\text{C17})$$

¹¹ In Ref. [34] the authors followed the convention that B_l represents combined smoothing effect due to pixel as well as beam. Our notation is different from them.

Here we note that the noise terms drop out in the cross-correlation procedure.

The final estimate of the full-sky spectrum is obtained by simply inverting the coupling matrix. We have used a singular value decomposition technique to invert our matrix. We checked that this matrix gives unbiased estimates of the full sky power spectrum using Monte Carlo simulations of CMB maps and using Kp2 cut.

The final full sky estimate of the power spectrum is obtained from

$$C_l^{\mathbf{i}} = \sum_{l'} (M^{-1})_{ll'} \tilde{C}_{l'}^{ij} / B_l^i B_l^j p_l^2. \quad (\text{C18})$$

with $\mathbf{i} \equiv (i, j)$, $\mathbf{i} = (1, 2, 3, 4, \dots, 24)$ as there are 24 cross-correlations possible that do not have detector noise bias. Using the 24 cross-power spectra we form an averaged power spectrum following

$$\bar{C}_l = \sum_{i=1}^{i=24} C_l^i N_{cross} \quad (\text{C19})$$

where $N_{cross} = 1/24$. Now we bin them in the same manner as the WMAP. The binned power spectrum is defined as

$$C_b = \frac{1}{\Delta l} \sum_{l=l_{min}}^{l=l_{max}} \frac{l(l+1)}{2\pi} \bar{C}_l, \quad (\text{C20})$$

with $\Delta l = l_{max} + 1 - l_{min}$. Eqn. (C20) defines our final power spectrum.

APPENDIX D: ESTIMATION OF RESIDUAL UNRESOLVED POINT SOURCE CONTAMINATION

In this appendix we study the point source bias in the auto power spectrum of a cleaned map and in the cross power of two cleaned maps. From a detailed study of a single iteration cleaning method we verify that a residual point source bias in the auto or cross power spectrum of the cleaned maps could be very well approximated by $\hat{\mathbf{W}} \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^{\text{T}}$, without a need to compute $\langle \hat{\mathbf{W}} \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^{\text{T}} \rangle$. For the auto power spectrum we obtain an analytic point source bias in terms of the noise covariance matrix and the point source covariance matrix. We note that, in practice when we estimate cosmological power spectrum from the WMAP data we do not use the auto power spectrum of a cleaned map. Nevertheless the auto power spectrum is important to study along with the cross power spectrum for a deeper understanding of the point source bias. We note that in all the Monte Carlo simulations in this section we treat point sources as fixed templates. The sources of randomness come from CMB and detector noises.

1. Auto Power Spectrum

Following eq. (27) the cleaned auto power spectrum obeys,

$$\langle \hat{C}_l^{\text{Clean}} \rangle = \langle \hat{C}_l^c \rangle + \left\langle \frac{1}{\mathbf{e}_0^{\text{T}} (\hat{\mathbf{C}}_1^{\text{f+N}})^{-1} \mathbf{e}_0} \right\rangle + (1 - n_c) \frac{\langle \hat{C}_l^c \rangle}{2l+1}. \quad (\text{D1})$$

Clearly the positive bias is given by $\langle 1/\mathbf{e}_0^{\text{T}} (\hat{\mathbf{C}}_1^{\text{f+N}})^{-1} \mathbf{e}_0 \rangle$ 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 and since diffuse foreground is not a concern at large l , we do not include any diffuse foreground component in this study. Considering the second term on the right hand side of eq. (D1) 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}_1^{N} , not the empirical noise covariance matrix, $\hat{\mathbf{C}}_1^{\text{N}}$, where $\hat{\mathbf{C}}_1^{\text{N}} = \mathbf{C}_1^{\text{N}} + \delta \hat{\mathbf{C}}_1^{\text{N}}$ and $\delta \hat{\mathbf{C}}_1^{\text{N}}$ denotes noise fluctuation on the true noise level. Therefore we express $\langle 1/\mathbf{e}_0^{\text{T}} (\hat{\mathbf{C}}_1^{\text{f+N}})^{-1} \mathbf{e}_0 \rangle$ 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 WMAP detector noise level is much larger than the point source spectrum,

we simplify $\langle 1/\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{f}+\mathbf{N}})^{-1}\mathbf{e}_0 \rangle$ analytically by making expanding to first order ¹², in terms of $\mathbf{C}_1^{\text{ps}}(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}$,

$$(\hat{\mathbf{C}}_1^{\mathbf{f}+\mathbf{N}})^{-1} = (\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1} - (\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{C}_1^{\text{ps}}(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}, \quad (\text{D2})$$

Due to the low point source contamination relative to the WMAP noise level it is reasonable to assume that $(\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{C}_1^{\text{ps}}(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{e}_0)/(\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{e}_0) \ll 1$. Expanding to first order we obtain,

$$\left\langle \frac{1}{\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{f}+\mathbf{N}})^{-1}\mathbf{e}_0} \right\rangle = \left\langle \frac{1}{\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{e}_0} \right\rangle + \left\langle \frac{\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{C}_1^{\text{ps}}(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{e}_0} \right\rangle. \quad (\text{D3})$$

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 upto second order in $\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}$,

$$\left\langle \frac{1}{\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}\mathbf{e}_0} \right\rangle = \left\langle \frac{1}{\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0} \left(1 + \frac{\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0} - \frac{\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0} + \frac{(\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0)^2}{(\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0)^2} \right) \right\rangle. \quad (\text{D4})$$

The noise fluctuations are assumed to have zero mean, $\langle \delta\hat{\mathbf{C}}_1^{\mathbf{N}} \rangle = 0$. Hence the second term in the bracket vanishes on the ensemble average. Only the numerators of the third and fourth terms in the bracket on the right hand side of eq. (D4) are stochastic variables. On the ensemble average the numerator of the third term becomes,

$$\left\langle \mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0 \right\rangle = \left\langle \sum_{i,j,k} C_l^{Ni}\delta\hat{C}_l^{N(ij)}C_l^{Nj}\delta\hat{C}_l^{N(jk)}C_l^{Nk} \right\rangle = \sum_{i,j,k} C_l^{Ni}C_l^{Nj}C_l^{Nk} \left\langle \delta\hat{C}_l^{N(ij)}\delta\hat{C}_l^{N(jk)} \right\rangle,$$

where we have assumed that the noise covariance matrix of WMAP is diagonal. We simplify the ensemble averaged quantities using the relation eq. (28) of Ref. [35]. If we consider a general term of the form $\langle \delta\hat{C}_l^{N(ij)}\delta\hat{C}_l^{N(kp)} \rangle$ then for uncorrelated detector noises, the ensemble averaged quantities will survive only for the same pair, i.e. when $(i,j) = (k,p)$. We further note that,

$$\left\langle \delta\hat{C}_l^{N(ij)^2} \right\rangle = \begin{cases} \frac{1}{2l+1}C_l^{Ni}C_l^{Nj}, & i \neq j \\ \frac{2}{2l+1}C_l^{Ni}C_l^{Ni}, & i = j \end{cases}. \quad (\text{D5})$$

Using above relations the third and fourth term in the bracket of eq. (D4) become

$$\left\langle \frac{\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0} \right\rangle = \frac{(1+n_c)}{2l+1}, \quad (\text{D6})$$

Following a similar method we also simplify the fourth term in eq. (D4). The final result is

$$\left\langle \frac{(\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\delta\hat{\mathbf{C}}_1^{\mathbf{N}}(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0)^2}{(\mathbf{e}_0^T(\mathbf{C}_1^{\mathbf{N}})^{-1}\mathbf{e}_0)^2} \right\rangle = \frac{2}{2l+1}. \quad (\text{D7})$$

¹² We note that we cannot use Sherman-Morrison formula to decouple the foreground and detector noise bias from the term $\left\langle \frac{1}{\mathbf{e}_0^T(\hat{\mathbf{C}}_1^{\mathbf{f}+\mathbf{N}})^{-1}\mathbf{e}_0} \right\rangle$ in an useful form.

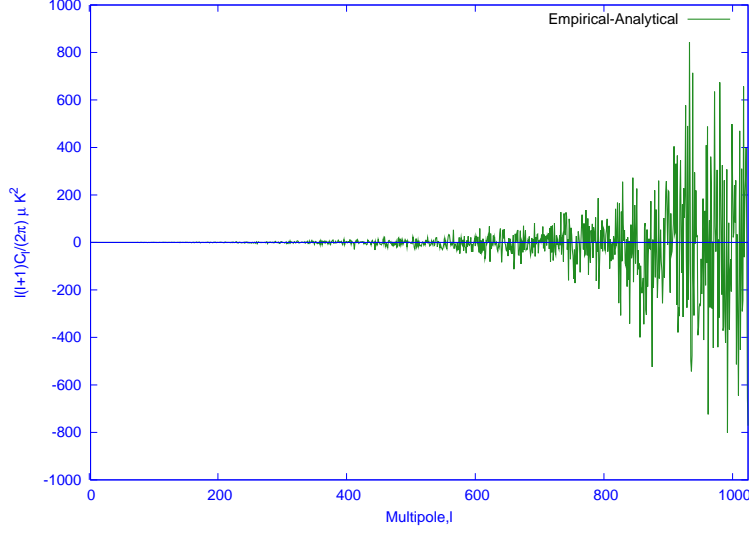


FIG. 13: The difference between the empirical noise bias and the theoretical estimate of noise bias given in eq. D8 for V and W bands.

Using eqs. (D4), (D6), (D7) we obtain,

$$\left\langle \frac{1}{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1} \mathbf{e}_0} \right\rangle = \left\langle \frac{1}{\mathbf{e}_0^T (\mathbf{C}_1^{\mathbf{N}})^{-1} \mathbf{e}_0} \left(1 - \frac{1+n}{2l+1} + \frac{2}{2l+1} \right) \right\rangle = \frac{1}{\mathbf{e}_0^T (\mathbf{C}_1^{\mathbf{N}})^{-1} \mathbf{e}_0} \frac{2l+2-n_c}{2l+1}. \quad (\text{D8})$$

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 from which these empirical noise covariance matrices were computed were formed following a method similar to that mentioned in section III D. In Fig. 13 we show the difference between the empirical noise bias $\langle 1/\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1} \mathbf{e}_0 \rangle$ and the second order analytical expression for noise bias, given by the right hand side of eq. (D8). The difference is consistent with zero. The large fluctuations at large l is 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 in a single iteration cleaning with CMB, point sources and detector noise having WMAP's noise level. We compute ensemble averaged cleaned power spectrum $\langle \hat{C}_l^{\text{Clean}} \rangle$ from 1000 such simulations. The validity of eq. (D8) 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 $\langle \hat{C}_l^{\text{Clean}} \rangle - \left(1/\mathbf{e}_0^T (\mathbf{C}_1^{\mathbf{N}})^{-1} \mathbf{e}_0 \right) ((2l+2-n_c)/(2l+1))$ still has a residual positive bias at multipole range $l > 400$. This excess is a clear demonstration of the presence of point source bias. The brown line in Fig. 14 is the power spectrum, corrected for the noise and the negative CMB bias of the form $-C_l^c/(2l+1)$. The excess is caused by the point source bias. We analytically compute the point source bias in terms of theoretical noise covariance matrix similar to eq (D8). Following a first order expansion in $\mathbf{C}_1^{\text{ps}} (\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1}$ and a second order expansion in terms of $\delta \hat{\mathbf{C}}_1^{\mathbf{N}} (\mathbf{C}_1^{\mathbf{N}})^{-1}$ we obtain,

$$\left\langle \frac{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1} \mathbf{C}_1^{\text{ps}} (\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1} \mathbf{e}_0}{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\mathbf{N}})^{-1} \mathbf{e}_0} \right\rangle = \frac{\mathbf{e}_0^T (\mathbf{C}_1^{\mathbf{N}})^{-1} \mathbf{C}_1^{\text{ps}} (\mathbf{C}_1^{\mathbf{N}})^{-1} \mathbf{e}_0}{\mathbf{e}_0^T (\mathbf{C}_1^{\mathbf{N}})^{-1} \mathbf{e}_0} \frac{2l}{2l+1} + \frac{\text{tr}(\mathbf{C}_1^{\text{ps}} (\mathbf{C}_1^{\mathbf{N}})^{-1})}{(2l+1)(\mathbf{e}_0^T (\mathbf{C}_1^{\mathbf{N}})^{-1} \mathbf{e}_0)}. \quad (\text{D9})$$

Again we verify the validity of the above expression following Monte Carlo simulations. We compute the point source bias from 1000 cleaning simulations following $\langle \hat{C}_l^{\text{Clean}} \rangle - C_l^c - C_l^c/(2l+1) - \left(1/\mathbf{e}_0^T (\mathbf{C}_1^{\mathbf{N}})^{-1} \mathbf{e}_0 \right) ((2l+2-n_c)/(2l+1))$

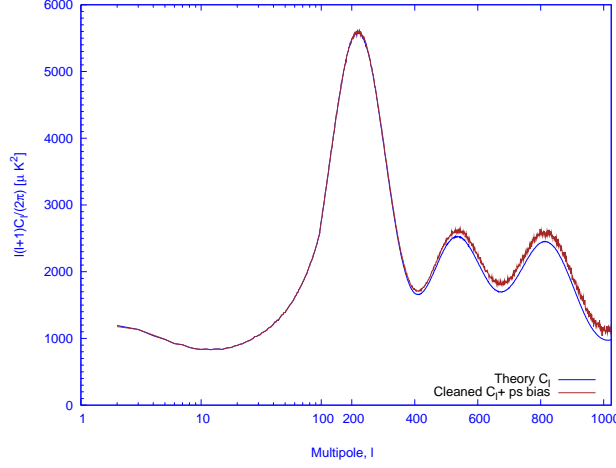


FIG. 14: The auto power spectrum without correcting for the residual unresolved point source bias is shown in brown line. The blue line shows the theoretical CMB power spectrum.

where C_l^c denotes the theoretical CMB power spectrum. In Fig. 15 the green line shows 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 black line with filled points shows the analytical estimate of the point source bias computed from the right hand side of the eq. (D9). 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.

Given the large noise level of WMAP maps, one might expect that $\langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle$ would be a good approximation of the point source bias that is given by the second term of eq. (D3). In the limit $\hat{\mathbf{C}}_1 \rightarrow \hat{\mathbf{C}}_1^{\text{N}}$, 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 $\mathbf{C}_1^{\text{ps}} \hat{\mathbf{C}}_1^{\text{N}-1}$,

$$\begin{aligned} \langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle &= \left\langle \frac{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1}}{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{e}_0} \mathbf{C}_1^{\text{ps}} \frac{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1}}{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{e}_0} \right\rangle - \frac{C_l^c}{2l+1} \left\langle \frac{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{C}_1^{\text{ps}} (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{e}_0}{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{e}_0} \right\rangle \\ &\quad + \frac{C_l^c}{2l+1} \text{tr}(\mathbf{C}_1^{\text{ps}} \langle (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \rangle). \end{aligned} \quad (\text{D10})$$

Here, C_l^c is the theoretical CMB power spectrum. However, 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 $0.1 \mu K^2$ 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, $\mathbf{C}_1^{\text{ps}} (\hat{\mathbf{C}}_1^{\text{N}})^{-1}$, the first term on the right hand side go as $\hat{\mathbf{C}}_1^{\text{N}}$, whereas the last two terms goes as $C_l^c/(2l+1)$. Indeed, the noise becomes much larger than CMB power spectrum at large l due to beam deconvolution, making the two last two terms insignificant compared to the first term. In figure 16 we show magnitude of the individual contributions arising from the second and third term following 1000 Monte Carlo simulations of the cleaning method. We use V and W bands in the simulations using realistic WMAP noise levels and residual unresolved point source model. The red line shows the difference between the two terms which is less than $0.01 \mu K^2$ or even smaller at $l > 400$. Hence it is justified to assume,

$$\langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle = \left\langle \frac{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1}}{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{e}_0} \mathbf{C}_1^{\text{ps}} \frac{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1}}{\mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{e}_0} \right\rangle. \quad (\text{D11})$$

In Fig. 14 we show the point source bias computed following $\langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle$ in blue line. This matches well with the black line with filled circle, justifying eq. (D11).

The unresolved point source bias is then $\langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle$. Now we propose that, the point source bias could also be taken into account following $\hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \sim \langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle$. 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

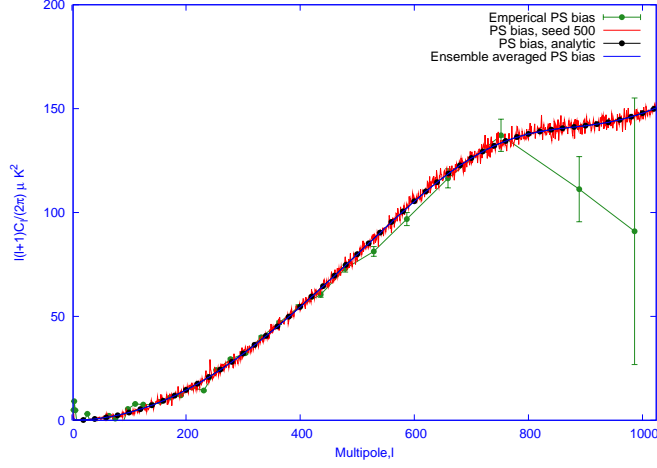


FIG. 15: The point source bias, $\langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle$ (blue line), and the approximation $\hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T$ (red line) for the case of auto power spectrum of the cleaned maps. The empirical point source bias computed by using $\langle \hat{C}_l^{\text{Clean}} \rangle - C_l^c - C_l^c / (2l + 1) - \left(1 / \mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{e}_0 \right) ((2l + 2 - n_c) / (2l + 1))$ is shown in green line. The error bars are computed only for the term $\langle 1 / \mathbf{e}_0^T (\hat{\mathbf{C}}_1^{\text{N}})^{-1} \mathbf{e}_0 \rangle$ without including the cosmic variance. All the data points are binned following the WMAP's 3 year binning method. The analytical point source bias computed following eq. (D9) is also shown in black line with black dots.

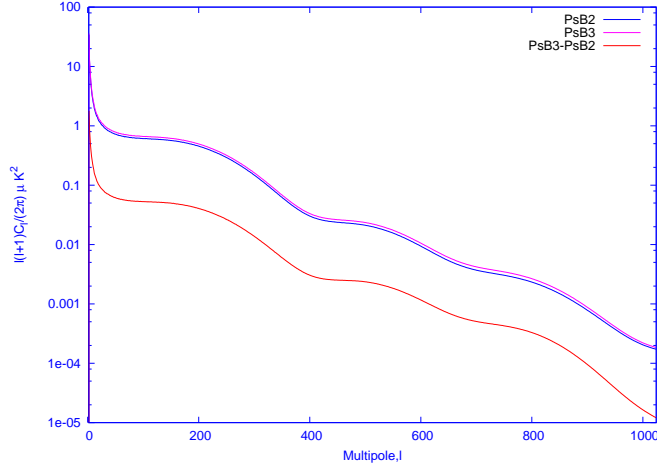


FIG. 16: The contributions arising from the third (blue line) and fourth (pinkline) terms of eq. (D10) and their difference.

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 $\langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle$. 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 the figure 15 we show how well $\hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T$ (red line) matches with $\langle \hat{\mathbf{W}}_1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^T \rangle$. This result is obtained from 1000 Monte Carlo simulations of our cleaning method using V and W bands.

2. Cross Power Spectrum

We now consider point source bias in case of cross power spectrum of two cleaned maps. In this case the cleaned power spectrum is given by,

$$\langle \hat{C}_l^{\text{Clean}} \rangle = C_l^c + 2(1 - n_c) \frac{C_l^c}{2l + 1} + \langle \hat{\mathbf{W}}_1^1 \mathbf{C}_1^{\text{ps}} \hat{\mathbf{W}}_1^{2T} \rangle + \langle \hat{\mathbf{W}}_1^1 \hat{\mathbf{C}}_1^{\text{N}(12)} \hat{\mathbf{W}}_1^{2T} \rangle. \quad (\text{D12})$$

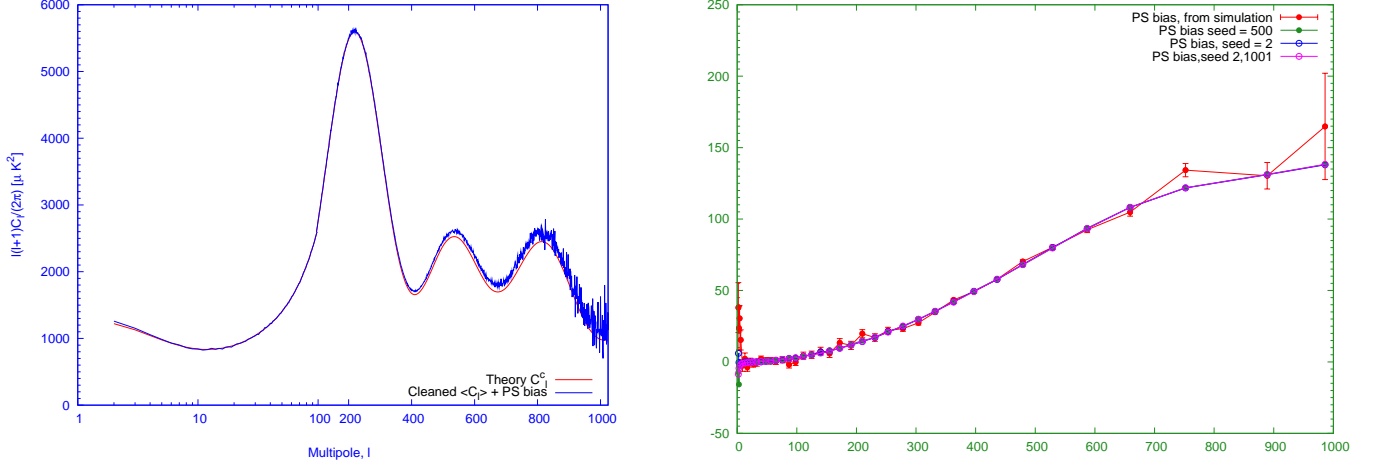


FIG. 17: *Left*: The point source bias at large multipole range obtained from 1000 Monte Carlo simulations of the cleaning method involving V and W bands is shown in blue line. The CMB bias $\frac{1-n_c}{2l+1}C_l^c$ has been removed before plotting the blue line to show the bias coming from point source only. The red line shows the theoretical power spectrum from which random realization of CMB were drawn. *Right*: Point source bias computed from $\langle \hat{C}_l^{Clean} \rangle - C_l^c - 2(1-n_c)\frac{C_l^c}{2l+1}$ is shown in red line. The magenta line shows point source bias computed from $\langle \hat{\mathbf{W}}_1^1 \mathbf{C}_1^{ps} \hat{\mathbf{W}}_1^{2T} \rangle$ which agrees excellently with $\hat{\mathbf{W}}_1^1 \mathbf{C}_1^{ps} \hat{\mathbf{W}}_1^{2T}$ computed from two arbitrary realizations shown in green and blue lines.

Here $\hat{\mathbf{W}}_1^1$ and $\hat{\mathbf{W}}_1^2$ are the weight vectors for the two cleaned maps and $\hat{\mathbf{C}}_1^{N(12)}$ denotes a chance noise correlation in the two sets of maps used to obtain two cleaned maps. From Monte-Carlo simulations involving cross spectrum, $v1w12 \otimes v2w34$, $v1w13 \otimes v2w24$, $v1w14 \otimes v2w23$, $v1w24 \otimes v2w13$, $v1w34 \otimes v2w12$ we recover a point source bias at the large multipole range. In the left panel of the figure 17 we show the point source bias from 1000 such simulations. Like the auto power case, here too the point source bias could be obtained as $\hat{\mathbf{W}}_1^1 \mathbf{C}_1^{ps} \hat{\mathbf{W}}_1^{1T}$. We show this in the right panel of the figure 17. The red line shows bias computed from simulations, i.e. $\langle \hat{C}_l^{Clean} \rangle - C_l^c - 2(1-n_c)\frac{C_l^c}{2l+1}$. We compute $\hat{\mathbf{W}}_1^{1T} \mathbf{C}_1^{ps} \hat{\mathbf{W}}_1^2$ for any two arbitrary realizations. These are shown in blue and green colors respectively. The magenta colored line shows $\langle \hat{\mathbf{W}}_1^1 \mathbf{C}_1^{ps} \hat{\mathbf{W}}_1^{2T} \rangle$ which is in excellent agreement with blue and green lines. This justifies the approximation $\langle \hat{\mathbf{W}}_1^1 \mathbf{C}_1^{ps} \hat{\mathbf{W}}_1^{2T} \rangle \sim \hat{\mathbf{W}}_1^1 \mathbf{C}_1^{ps} \hat{\mathbf{W}}_1^{2T}$, so that point source bias could as well be computed from the weights from a single realization without any need of ensemble average. We have followed this approach of point source removal in the multi-region iterative cleaning method described in section III C.

APPENDIX E: BIAS AT LOW- l

Assume that there are n_f number of foreground components. Each of these follow rigid frequency scaling. The number of channels is n_c , where $n_c \geq n_f + 1$. The rank of full covariance matrix is determined by the number of independent components. Therefore, for n_f number of independent foreground components and an additional CMB component the rank of the full covariance matrix is $n_f + 1$. In many cases we choose total number of channels n_c to be equal to the rank of the matrix, $n_f + 1$. This ensures that the full covariance matrix is of full rank, (order=rank) so that it is invertible.

However it is also interesting to explore the freedom of using a singular covariance matrix $\hat{\mathbf{C}}_1$. This may arise, for example, if we consider n_f number of foreground components (each with rigid frequency scaling) but with n_c number of channels with $n_c > n_f + 1$. The motivation of this work lies in the fact we have shown that the cleaned power spectrum could still be defined in terms of Moore Penrose Inversion (in eq. (17)) in case we encounter a singular covariance matrix. So it is worth investigating whether there is a further simplified form of eq. (17).

The calculations proceed as follows:

- Obtain an analytic expression for the full covariance matrix. We find that the full covariance matrix can be obtained as three successive rank-one updates of three different matrices.
- Identify the appropriate cases of the generalized versions of Sherman Morrison formula which applies to each of

three rank-one updates that have to be carried out.

- Apply Sherman Morrison formula successively for the three rank one updates to obtain analytical expression for the bias.

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

$$S^i(\theta, \phi) = C(\theta, \phi) + \sum_{p=1}^{n_f} F_0^p(\theta, \phi) f_p^i. \quad (\text{E1})$$

Alternatively, in the spherical harmonic space,

$$a_{lm}^i = a_{lm}^c + \sum_{p=1}^{n_f} f_p^i a_{lm}^{p0}. \quad (\text{E2})$$

The auto power spectrum of the i^{th} channel

$$\hat{C}_l^i = \hat{C}_l^c + 2 \sum_{p=1}^{n_f} f_p^i \hat{C}_l^{cf(p)0} + \sum_{p,p'}^{n_f} f_p^i f_{p'}^i C_l^{(pp')0}, \quad (\text{E3})$$

where $C_l^{(pp')0}$ is the correlation between any two foreground components p, p' and $\hat{C}_l^{cf(p)0}$ is the chance correlation between CMB and p^{th} foreground component. The cross power spectrum between two channels i, j are given by

$$\hat{C}_l^{ij} = \hat{C}_l^c + \sum_{p=1}^{n_f} f_p^i \hat{C}_l^{cf(p)0} + \sum_{p=1}^{n_f} f_p^j \hat{C}_l^{cf(p)0} + \sum_{p,p'}^{n_f} f_p^i f_{p'}^j C_l^{(pp')0}, \quad (\text{E4})$$

or

$$\hat{C}_l^{ij} = \hat{C}_l^c + \sum_{p=1}^{n_f} f_p^i \hat{C}_l^{cf(p)0} + \sum_{p=1}^{n_f} f_p^j \hat{C}_l^{cf(p)0} + F_l^{ij}. \quad (\text{E5})$$

Introducing explicit matrix notations for the equation, we write

$$\begin{aligned} \hat{\mathbf{C}}_1 = \hat{C}_l^c & \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \cdot & \cdot & \dots & \cdot \\ 1 & \cdot & \dots & 1 \end{pmatrix}_{(n_c \times n_c)} + \sum_{p=1}^{n_f} \hat{C}_l^{cf(p)0} \begin{pmatrix} f_p^1 & f_p^1 & \dots & f_p^1 \\ f_p^2 & f_p^2 & \dots & f_p^2 \\ \cdot & \cdot & \dots & \cdot \\ f_p^{n_c} & f_p^{n_c} & \dots & f_p^{n_c} \end{pmatrix}_{(n_c \times n_c)} \\ + \sum_{p=1}^{n_f} \hat{C}_l^{cf(p)0} & \begin{pmatrix} f_p^1 & f_p^2 & \dots & f_p^{n_c} \\ f_p^1 & f_p^2 & \dots & f_p^{n_c} \\ \cdot & \cdot & \dots & \cdot \\ f_p^1 & f_p^2 & \dots & f_p^{n_c} \end{pmatrix}_{(n_c \times n_c)} + \begin{pmatrix} F^{11} & F^{12} & \dots & F^{1n_c} \\ F^{21} & F^{22} & \dots & F^{2n_c} \\ \cdot & \cdot & \dots & \cdot \\ F^{n_c 1} & F^{n_c 2} & \dots & F^{n_c n_c} \end{pmatrix}_{(n_c \times n_c)}. \end{aligned} \quad (\text{E6})$$

We define

$$\hat{\mathbf{f}}_1^{p0} = \hat{C}_l^{cf(p)0} \begin{pmatrix} f_p^1 \\ f_p^2 \\ \cdot \\ f_p^{n_c} \end{pmatrix}, \quad (\text{E7})$$

and

$$\mathbf{e}_0 = \begin{pmatrix} 1 \\ 1 \\ \cdot \\ 1 \end{pmatrix}. \quad (\text{E8})$$

Clearly, $\hat{\mathbf{f}}_1^{\text{p0}}, \mathbf{e}_0 \in \mathbb{R}_{n_c,1}$, where \mathbb{R}_{n_f, n_c} denotes set of real $n_f \times n_c$ matrices. Full covariance matrix can then be written as,

$$\hat{\mathbf{C}}_1 = \hat{C}_l^c \mathbf{e}_0 \mathbf{e}_0^T + \left(\sum_{p=1}^{n_f} \hat{\mathbf{f}}_1^{\text{p0}} \right) \mathbf{e}_0^T + \mathbf{e}_0 \left(\sum_{p=1}^{n_f} \hat{\mathbf{f}}_1^{\text{p0}} \right)^T + \mathbf{A}_3. \quad (\text{E9})$$

Define

$$\hat{\mathbf{f}}_1^0 = \sum_{p=1}^{n_f} \hat{\mathbf{f}}_1^{\text{p0}}, \quad (\text{E10})$$

then, it is possible to rewrite

$$\hat{\mathbf{C}}_1 = \underbrace{\hat{C}_l^c \mathbf{e}_0 \mathbf{e}_0^T + \hat{\mathbf{f}}_1^0 \mathbf{e}_0^T}_{\mathbf{A}_1} + \underbrace{\mathbf{e}_0 \hat{\mathbf{f}}_1^{0T} + \mathbf{A}_3}_{\mathbf{A}_2}, \quad (\text{E11})$$

with $\hat{\mathbf{f}}_1^0 \in \mathbb{R}_{n_c,1}$.

1. Generalized Sherman Morrison Formula for rank one updates

This section discusses some mathematical theorems that will be useful to us. The results are mainly based upon the two papers [28, 29]. In Ref. [28] analytic expressions for MPGI of rank one modified matrices of the form $\mathbf{M} = \mathbf{A} + \mathbf{b}\mathbf{c}^*$, (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, (*) denotes conjugate transpose. The motivation of the work in Ref. [28] was to generalize the Sherman Morrison formula

$$\mathbf{M}^{-1} = \mathbf{A}^{-1} - \frac{1}{\lambda} \mathbf{A}^{-1} \mathbf{b} \mathbf{c}^* \mathbf{A}^{-1} \quad (\text{E12})$$

where $\lambda = 1 + \mathbf{c}^* \mathbf{A}^{-1} \mathbf{b}$ 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 matrix.

The main results of the Ref. [28] are a set of formulas for the MPGI depending upon the different set of conditions, namely

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

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

Ref. [29] shows that it is sufficient to consider 5 independent cases only. All these 5 different cases are listed below along with an useful theorem proved regarding the rank modification that occurs during the rank one update of the matrix \mathbf{A} .

2. 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 $\mathbf{M} = \mathbf{A} + \mathbf{b}\mathbf{c}^$ and let $\lambda = 1 + \mathbf{c}^* \mathbf{A}^{-1} \mathbf{b}$. Then*

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

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

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

Analytical expressions of the MPGI \mathbf{M}^\dagger is given in Ref. [28] and in Ref. [29] depending upon the 6 or 5 conditions that they find sufficient. The expressions are of the form

$$\mathbf{M}^\dagger = \mathbf{A}^\dagger + \mathbf{G}, \quad (\text{E13})$$

where \mathbf{G} is a matrix obtained from only sums and products of \mathbf{A} , \mathbf{A}^\dagger , \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}^\dagger for those cases which will be useful to us.

a. Case 1

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

$$\mathbf{M}^\dagger = \mathbf{A}^\dagger - \lambda^{-1} \mathbf{d} \mathbf{e}^*, \quad (\text{E14})$$

where

$$\mathbf{d} = \mathbf{A}^\dagger \mathbf{b}, \mathbf{e} = (\mathbf{A}^\dagger)^* \mathbf{c}. \quad (\text{E15})$$

b. Case 2

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

$$\mathbf{M}^\dagger = \mathbf{A}^\dagger - \mathbf{k} \mathbf{u}^\dagger - \mathbf{v}^\dagger \mathbf{h} + \lambda \mathbf{v}^\dagger \mathbf{u}^\dagger, \quad (\text{E16})$$

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

c. Case 3

In Ref. [29], a relation between the unique projectors between on the column spaces of \mathbf{M}, \mathbf{A} is reported. If $\mathbf{b} \notin \mathcal{C}(\mathbf{A}), \mathbf{c} \in \mathcal{C}(\mathbf{A}^*)$ then

$$\mathbf{M} \mathbf{M}^\dagger = \mathbf{A} \mathbf{A}^\dagger - \eta^{-1} \mathbf{e} \mathbf{e}^T + \eta^{-1} \nu^{-1} \mathbf{q} \mathbf{q}^T, \quad (\text{E17})$$

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

3. 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,

$$\hat{\mathbf{C}}_1 = \hat{C}_l^c \mathbf{e}_0 \mathbf{e}_0^T + \underbrace{\hat{\mathbf{f}}_1^0 \mathbf{e}_0^T + \underbrace{\mathbf{e}_0 \hat{\mathbf{f}}_1^{0T}}_{\mathbf{A}_2}}_{\mathbf{A}_1} + \mathbf{A}_3. \quad (\text{E18})$$

First note that $\mathbf{e}_0 \notin \mathcal{C}(\mathbf{A}_3), \hat{\mathbf{f}}_1^0 \in \mathcal{C}(\mathbf{A}_3^*)$, so that this is identical to the conditions for case 3 and $rank(\mathbf{A}_3) = rank(\mathbf{A}_2) = n_f$, the number of foreground components. But when we consider matrix \mathbf{A}_2 we see that $\hat{\mathbf{f}}_1^0 \notin \mathcal{C}(\mathbf{A}_2), \mathbf{e}_0 \notin \mathcal{C}(\mathbf{A}_2^*)$. So here conditions of case 2 apply. Here $rank(\mathbf{A}_1) = rank(\mathbf{A}_2) + 1 = n_f + 1$. On the other hand, $\mathbf{e}_0 \in \mathcal{C}(\mathbf{A}_1), \mathbf{e}_0 \in \mathcal{C}(\mathbf{A}_1^*)$. Hence $rank(\hat{\mathbf{C}}_1) = rank(\mathbf{A}_1) = n_f + 1$. Now one can carry out the analytic simplification in three steps.

a. *Step 1 : First application of GSM*

We note that $\mathbf{e}_0 \in \mathcal{C}(\mathbf{A}_1)$. Therefore we are in a situation where conditions of case 1 are applicable. Our aim is to express $\hat{C}_l^{Clean} = 1/\mathbf{e}_0^T \mathbf{C}_1^\dagger \mathbf{e}_0$ in terms of \mathbf{A}_1 . For this purpose we simply need to express \mathbf{C}_1^\dagger in terms of \mathbf{A}_1 following GSM formula appropriate for this case. At the next step we compute $\mathbf{e}_0^T \mathbf{C}_1^\dagger \mathbf{e}_0$. The analysis is relatively simple and we do not show the detailed calculation here. Instead we give a similar calculation for a more complicated expression at the next subsection. The result of this section is,

$$\hat{C}_l^{Clean} = \hat{C}_l^c + \frac{1}{\mathbf{e}_0^T \mathbf{A}_1^\dagger \mathbf{e}_0}. \quad (\text{E19})$$

b. *Step 2 : Second application of GSM*

At this step we shall simplify $1/(\mathbf{e}_0^T \mathbf{A}_1^\dagger \mathbf{e}_0)$ further in terms of some function of \mathbf{A}_2^\dagger where $\mathbf{A}_1 = \mathbf{f}_1^0 \mathbf{e}_0^T + \mathbf{A}_2$. As mentioned in the previous subsection, this simplification method proceeds in two steps. First, we express \mathbf{A}_1^\dagger in terms of \mathbf{A}_2^\dagger following GSM. At the next stage we would compute $1/(\mathbf{e}_0^T \mathbf{A}_1^\dagger \mathbf{e}_0)$. We easily see that, $\mathbf{f}_1^0 \notin \mathcal{C}(\mathbf{A}_2)$ and $\mathbf{e}_0 \notin \mathcal{C}(\mathbf{A}_2^T)$. Therefore GSM formula corresponding to case 2 would be useful for us,

$$\mathbf{A}_1^\dagger = \mathbf{A}_2^\dagger - \mathbf{k} \mathbf{u}^\dagger - \mathbf{v}^\dagger \mathbf{h} + \lambda \mathbf{v}^\dagger \mathbf{u}^\dagger. \quad (\text{E20})$$

Each of the individual quantities on the right hand side are explained in the section E2. All the terms starting from the second on the right of above equation could be written in terms of the foreground and CMB shape vectors, \mathbf{f}_1^0 and \mathbf{e}_0 . The vectors \mathbf{k} and \mathbf{h} are given by $\mathbf{k} = \mathbf{A}_2^\dagger \mathbf{f}_1^0$ and $\mathbf{h} = \mathbf{e}_0^T \mathbf{A}_2^\dagger$ respectively. Moore Penrose Inverse of the other two vectors \mathbf{u} and \mathbf{v} could be computed following the definition $\mathbf{u}^\dagger = \mathbf{u}^*/\|\mathbf{u}\|$, where $\|\mathbf{u}\|$ denotes vector norm. Thus we have,

$$\mathbf{u}^\dagger = \frac{\mathbf{f}_1^{0T}(\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger)}{\mathbf{f}_1^{0T}(\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger)\mathbf{f}_1^0}, \mathbf{v}^\dagger = \frac{(\mathbf{I} - \mathbf{A}_2^\dagger \mathbf{A}_2)\mathbf{e}_0}{\mathbf{e}_0^T(\mathbf{I} - \mathbf{A}_2^\dagger \mathbf{A}_2)\mathbf{e}_0}. \quad (\text{E21})$$

All the vectors computed above could be used in eq E20 to express \mathbf{A}_1^\dagger in terms of \mathbf{A}_2^\dagger ,

$$\mathbf{A}_1^\dagger = \mathbf{A}_2^\dagger - \frac{\mathbf{A}_2^\dagger \mathbf{f}_1^0 \mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger)}{\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger) \mathbf{f}_1^0} - \frac{(\mathbf{I} - \mathbf{A}_2^\dagger \mathbf{A}_2) \mathbf{e}_0 \mathbf{e}_0^T \mathbf{A}_2^\dagger}{\mathbf{e}_0^T (\mathbf{I} - \mathbf{A}_2^\dagger \mathbf{A}_2) \mathbf{e}_0} + (1 + \mathbf{e}_0^T \mathbf{A}_2^\dagger \mathbf{f}_1^0) \frac{(\mathbf{I} - \mathbf{A}_2^\dagger \mathbf{A}_2) \mathbf{e}_0}{\mathbf{e}_0^T (\mathbf{I} - \mathbf{A}_2^\dagger \mathbf{A}_2) \mathbf{e}_0} \frac{\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger)}{\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger) \mathbf{f}_1^0} \quad (\text{E22})$$

Now we can easily compute $\mathbf{e}_0^T \mathbf{A}_1^\dagger \mathbf{e}_0$. This involves contraction of both the indices of $A_{1(ij)}^\dagger$. After some algebra all the terms above simplifies a lot and we are left with,

$$\frac{1}{\mathbf{e}_0^T \mathbf{A}_1^\dagger \mathbf{e}_0} = \frac{\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger) \mathbf{f}_1^0}{\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger) \mathbf{e}_0}. \quad (\text{E23})$$

c. *Step 3 : Application of a relation of orthogonal projectors*

Both the numerator and denominator of eq. E23 contains orthogonal projector $\mathbf{A}_2 \mathbf{A}_2^\dagger$ on the column space of \mathbf{A}_2 . At this stage we only need to reexpress this term in terms of $\mathbf{A}_3 \mathbf{A}_3^\dagger$ which is also an orthogonal projector on the column space of \mathbf{A}_3 . We recall that $\mathbf{A}_2 = \mathbf{A}_3 + \mathbf{e}_0 \mathbf{f}_0^T$. The foreground shape vector \mathbf{f}_1^0 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, $\mathbf{e}_0 \notin \mathcal{C}(\mathbf{A}_3)$, $\mathbf{f}_1^0 \in \mathcal{C}(\mathbf{A}_3^*)$. Thus the conditions of case 3 are applicable. Following the notations of the Ref. [29] we obtain,

$$\mathbf{A}_2 \mathbf{A}_2^\dagger = \mathbf{A}_3 \mathbf{A}_3^\dagger - \eta^{-1} \mathbf{e} \mathbf{e}^T + \eta^{-1} \nu^{-1} \mathbf{q} \mathbf{q}^T \quad (\text{E24})$$

To proceed further we need to express $\mathbf{e} \mathbf{e}^T$ and $\mathbf{q} \mathbf{q}^T$ in terms of \mathbf{f}_1^0 and \mathbf{e}_0 . We note that, $\mathbf{e} = \mathbf{A}_3^\dagger \mathbf{f}_1^0$ and $\mathbf{q} = \lambda \mathbf{e} + \eta \mathbf{f}$. Following notations of Ref. [29] 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. $\mathbf{f} = (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{e}_0$. The matrix $\mathbf{e} \mathbf{e}^T$ could immediately be identified as $\mathbf{e} \mathbf{e}^T = \mathbf{A}_3^\dagger \mathbf{f}_1^0 \mathbf{f}_1^{0T} \mathbf{A}_3$. However the other matrix $\mathbf{q} \mathbf{q}^T$ contains several terms,

$$\mathbf{q} \mathbf{q}^T = \lambda^2 \mathbf{e} \mathbf{e}^T + \eta^2 (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{e}_0 \mathbf{e}_0^T (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) + \lambda \eta \left(\mathbf{A}_3^\dagger \mathbf{f}_1^0 \mathbf{e}_0^T (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) + (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{e}_0 \mathbf{f}_1^{0T} \mathbf{A}_3^\dagger \right). \quad (\text{E25})$$

It is easy to see that, the numerator of eq. (E23) involves computation of the term $\mathbf{f}_1^{0T} \mathbf{A}_2 \mathbf{A}_2^\dagger \mathbf{f}_1^0$. At this point we have all the necessary expressions to compute this term. First we note that, following eqs. (E24), (E25) the projector on the column space of \mathbf{A}_2 could be simplified as,

$$\begin{aligned} \mathbf{A}_2 \mathbf{A}_2^\dagger &= \mathbf{A}_3 \mathbf{A}_3^\dagger - \eta^{-1} \mathbf{A}_3^\dagger \mathbf{f}_1^0 \mathbf{f}_1^{0T} \mathbf{A}_3 + \eta^{-1} \nu^{-1} (\lambda^2 \mathbf{A}_3^\dagger \mathbf{f}_1^0 \mathbf{f}_1^{0T} \mathbf{A}_3 + \eta^2 (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{e}_0 \mathbf{e}_0^T (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) + \\ &\quad \lambda \eta \left(\mathbf{A}_3^\dagger \mathbf{f}_1^0 \mathbf{e}_0^T (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) + (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{e}_0 \mathbf{f}_1^{0T} \mathbf{A}_3^\dagger \right)). \end{aligned} \quad (\text{E26})$$

Now we can easily get an expression for $\mathbf{f}_1^{0T} \mathbf{A}_2 \mathbf{A}_2^\dagger \mathbf{f}_1^0$. Using eq. (E26) we obtain,

$$\begin{aligned} \mathbf{f}_1^{0T} \mathbf{A}_2 \mathbf{A}_2^\dagger \mathbf{f}_1^0 &= \mathbf{f}_1^{0T} \mathbf{A}_3 \mathbf{A}_3^\dagger \mathbf{f}_1^0 - \eta^{-1} (\mathbf{f}_1^{0T} \mathbf{A}_3^\dagger \mathbf{f}_1^0)^2 + \eta^{-1} \nu^{-1} (\lambda^2 (\mathbf{f}_1^{0T} \mathbf{A}_3^\dagger \mathbf{f}_1^0)^2 + \eta^2 (\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{e}_0)^2 + \\ &\quad 2 \lambda \eta \left(\mathbf{f}_1^{0T} \mathbf{A}_3^\dagger \mathbf{f}_1^0 \mathbf{e}_0^T (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{e}_0 \right)). \end{aligned} \quad (\text{E27})$$

Though there are several terms on the right hand side of the above equation as we will see some of them drop. We note that, $\mathbf{f}_1^0 \in \mathcal{C}(\mathbf{A}_3)$, so that $(\mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{f}_1^0 = \mathbf{f}_1^0$. Hence $\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{e}_0 = \mathbf{e}_0^T (\mathbf{I} - \mathbf{A}_3 \mathbf{A}_3^\dagger) \mathbf{f}_1^0 = \mathbf{0}$. If we assume $X = \mathbf{f}_1^{0T} \mathbf{A}_3 \mathbf{A}_3^\dagger \mathbf{f}_1^0$ and $Y = \mathbf{f}_1^{0T} \mathbf{A}_3^\dagger \mathbf{f}_1^0$ then using eq. (E27) and $\nu = \lambda^2 + \eta \phi$, as stated earlier, we may easily obtain,

$$\mathbf{f}_1^{0T} \mathbf{A}_2 \mathbf{A}_2^\dagger \mathbf{f}_1^0 = X + \frac{Y^2 \phi}{\nu}. \quad (\text{E28})$$

We now identify $X = \mathbf{f}_1^{0T} \mathbf{A}_3 \mathbf{A}_3^\dagger \mathbf{f}_1^0$ and obtain numerator of the eq. (E23) as

$$\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger) \mathbf{f}_1^0 = -\frac{Y^2 \phi}{\nu}. \quad (\text{E29})$$

The denominator of eq. (E23) can be computed in a similar fashion. We do not elaborate the mathematical details any more for this part. The final result is

$$\mathbf{f}_1^{0T} (\mathbf{I} - \mathbf{A}_2 \mathbf{A}_2^\dagger) \mathbf{e}_0 = \frac{1}{\nu} Y \phi. \quad (\text{E30})$$

Thus using eqs. (E23), (E28), (E30) we obtain

$$\frac{1}{\mathbf{e}_0^T \mathbf{A}_1^\dagger \mathbf{e}_0} = -Y = -\mathbf{f}_1^{0T} \mathbf{A}_3^\dagger \mathbf{f}_1^0. \quad (\text{E31})$$

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

d. Simplification of the bias expression

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

$$A_3^{ij} = \sum_{pp'}^{n_f} f_p^i f_{p'}^j C_l^{(pp')0}. \quad (\text{E32})$$

Also the elements of the chance correlation vector between CMB and all the foregrounds

$$\mathbf{f}_1^{0i} = f_l^{0i} = \sum_p^{n_f} \hat{C}_l^{(cp)0} f_p^i. \quad (\text{E33})$$

We may 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,

$$\left\langle \mathbf{f}_1^{\mathbf{0T}} \mathbf{A}_3^\dagger \mathbf{f}_1^{\mathbf{0}} \right\rangle = \sum_{ij} A_3^{\dagger ij} \left\langle \sum_{pp'} \hat{C}_l^{(cp)0} \hat{C}_l^{(cp')0} \right\rangle f_p^i f_{p'}^j. \quad (\text{E34})$$

Using $\hat{C}_l^{(cp)0} = \sum_m (a_{lm}^c a_{lm}^{p0*}) / (2l+1)$ and the fact that CMB anisotropies are statistically isotropic, i.e. $\langle a_{lm} a_{l'm'} \rangle = C_l^c \delta_{ll'} \delta_{mm'}$ we may obtain,

$$\left\langle \sum_{pp'} \hat{C}_l^{(cp)0} \hat{C}_l^{(cp')0} \right\rangle = \frac{C_l^c}{2l+1} \sum_{pp'} C_l^{(pp')0}. \quad (\text{E35})$$

Using eq. (E35) and $A_3^{ij} = \sum_{pp'}^{n_f} f_p^i f_{p'}^j C_l^{(pp')0}$ we can rewrite eq. (E34) in terms of following expression consisting of CMB and foreground covariance matrices,

$$\left\langle \mathbf{f}_1^{\mathbf{0T}} \mathbf{A}_3^\dagger \mathbf{f}_1^{\mathbf{0}} \right\rangle = \frac{C_l^c}{2l+1} \sum_{ij} A_3^{\dagger ij} A_3^{ij}. \quad (\text{E36})$$

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

$$\left\langle \hat{C}_l^{\text{Clean}} \right\rangle = \left\langle \hat{C}_l^c \right\rangle - n_f \frac{\left\langle \hat{C}_l^c \right\rangle}{2l+1}. \quad (\text{E37})$$

which is eq. (25). The other formulae given by eq. (26) and eq. (27) can be obtained similarly but need much less algebra.

-
- [1] A. Benoit et al., *Astropart. Phys.*, **17**, 101 (2002).
 - [2] G. F. Smoot et al., *Astrophys. J. Lett.*, **316**, L1, (1992).
 - [3] W. C. Jones et al., *Astrophys. J.*, **647**, 823 (2006).
 - [4] S. Padin, et al., *Astrophys. J.*, **549**, 1 (2001).
 - [5] C. L. Bennett et al. , *Astrophys. J., Suppl. Ser.*, **148**, 1 (2003a).
 - [6] C. L. Bennett et al. , *Astrophys. J.*, **583**, 1 (2003c).
 - [7] M. Limon et al., Wilkinson Microwave Anisotropy Probe (WMAP): Explanatory Supplement, version 1.0, at the LAMBDA website.
 - [8] G. Hinshaw et al., *Astrophys. J., Suppl. Ser.*, **148**, 63 (2003a).
 - [9] F. R. Bouchet and R. Gispert, *New Astronomy* **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 (2003b).
 - [12] G. Hinshaw et al., *astro-ph/0603451*, (2006).
 - [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.*, in press, *astro-ph/0603452*.
 - [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); *ibid.* *Astrophys. J.*, **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, P. Jain, *New Astron. Rev.* **50**, 854 (2006), *astro-ph/0608199*.
 - [23] H. K. Eriksen *et al.*, *Astrophys. J.* **656**, 641-652, (2007).
 - [24] P. Fosalba and I. Szapudi, *Astrophys. J.*, **617**, 95 (2004).
 - [25] G. Patanchon, J. F. Cardoso, J. Delabrouille and P. Vielva, *astro-ph/0410280* (2004).
 - [26] M. Tegmark, D. J. Eisenstein, W. Hu and , A. Oliveira-Costa, *Astrophys. J.*, **530**, 133 (2000).
 - [27] Eriksen et al., *Astrophys. J.*, **612** 633 (2004).
 - [28] C. D. Meyer, JR, *SIAM J. APPL. MATH.*, vol 24, No 3, May 1973.
 - [29] J. K. Baksalary *et al.* *Linear Algebra and its Applications* **372**,207-224, (2003).

- [30] K. M Gorski et al., astro-ph/9905275 (1999a).
- [31] K. M. Gorski et al., astro-ph/9812350 (1999b).
- [32] M. R. Calabretta, astro-ph/0412607 (2004).
- [33] R. Saha, A. Hajian, P. Jain and T. Souradeep, *in preparation*.
- [34] E. Hivon et al., *Astrophys. J.*, **567**, 2 (2002).
- [35] M. Tristram, J. F. Macs-Pez, C. Renault and D. Santos, *Mon. Not. R. Astron. Soc.*, **358**, 833 (2005).
- [36] M. Tegmark and A. de Oliveira-Costa, *Astrophys. J. Lett.*, **500**, L83 (1998).
- [37] E. Komatsu, et al., *Astrophys. J., Suppl. Ser.*, **148**, 119 (2003).
- [38] K. M. Huffenberger et al., *Astrophys. J. Lett.*, **651** 81 (2006).
- [39] L. Page et al, *Astrophys. J., Suppl. Ser.* , **148**, 233 (2003).
- [40] <http://www.netlib.org/> .