The Astrophysical Journal, 591:575–598, 2003 July 10

A FAST GRIDDED METHOD FOR THE ESTIMATION OF THE POWER SPECTRUM OF THE COSMIC MICROWAVE BACKGROUND FROM INTERFEROMETER DATA WITH APPLICATION TO THE COSMIC BACKGROUND IMAGER

S. T. Myers
National Radio Astronomy Observatory, P.O. Box O, Socorro, NM 87801

C. R. Contaldi, J. R. Bond, U.-L. Pen, D. Pogosyan,[1] and S. Prunet[2]
Canadian Institute for Theoretical Astrophysics, 60 St. George Street, Toronto, ON M5S 3H8, Canada

and

J. L. Sievers, B. S. Mason, T. J. Pearson, A. C. S. Readhead, and M. C. Shepherd
Owens Valley Radio Observatory, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125
Received 2002 May 23; accepted 2002 October 8

ABSTRACT

We describe an algorithm for the extraction of the angular power spectrum of an intensity field, such as the cosmic microwave background (CMB), from interferometer data. This new method, based on the gridding of interferometer visibilities in the aperture plane followed by a maximum likelihood solution for band powers, is much faster than direct likelihood analysis of the visibilities and deals with foreground radio sources, multiple pointings, and differencing. The gridded aperture-plane estimators are also used to construct Wiener-filtered images using the signal and noise covariance matrices used in the likelihood analysis. Results are shown for simulated data. The method has been used to determine the power spectrum of the CMB from observations with the Cosmic Background Imager, and the results are given in companion papers.

Subject headings: cosmic microwave background — methods: data analysis

1. INTRODUCTION

The technique of interferometry has been widely used in radio astronomy to image the sky using arrays of antennas. By correlating the complex voltage signals between pairs of antennas, the field of view of a single element can be subdivided into "synthesized beams" of higher angular resolution. In the small-angle approximation, the interferometer forms the Fourier transform of the sky convolved with the autocorrelation of the aperture voltage patterns. In standard radio interferometric data analysis, as described, for example, in the text by Thompson, Moran, & Swenson (1986) and the proceedings of the NRAO Synthesis Imaging School (Taylor, Carilli, & Perley 1999), the correlations or visibilities are inverse Fourier transformed back to the image plane. However, there are applications such as estimation of the angular power spectrum of fluctuations in the cosmic microwave background (CMB) where it is the distribution of and correlation between visibilities in the aperture or (u, v)-plane that is of most interest.

In standard cosmological models, the CMB is assumed to be a statistically homogeneous Gaussian random field (Bond & Efstathiou 1987). In this case, the spherical harmonics of the field are independent and the statistical properties are determined by the power spectrum C_l, where l labels the component of the Legendre polynomial expansion (and is roughly in inverse radians). Bond & Efstathiou (1987) showed that in cold dark matter–inspired cosmological models, there would be features in the CMB power spectrum that reflected critical properties of the cosmology. Recent detections of the first few of these "acoustic peaks" at $l < 1000$ in the power spectrum (Lange et al. 2001; Hanany et al. 2000; Lee et al. 2001; Halverson et al. 2002; Netterfield et al. 2002) have supported the standard inflationary cosmological model with $\Omega_{\rm tot} \approx 1$. Measurement of the higher l peaks and troughs, as well as the damping tail due to the finite thickness of the last scattering surface, is the next observational step. Interferometers are well suited to the challenge of mapping out features in the CMB power spectrum, with a given antenna pair probing a characteristic l proportional to the baseline length in units of the observing wavelength (a 100λ projected baseline corresponds to $l \sim 628$; see § 3).

There are many papers in the literature on the analysis of CMB anisotropy measurements, estimation of power spectra, and the use of interferometry for CMB studies. General issues for analysis of CMB data sets are discussed in Bond, Jaffe, & Knox (1998, 2000). Hobson, Lasenby, & Jones (1995) present a Bayesian method for the analysis of CMB interferometer data, using the three-element Cosmic Anisotropy Telescope data as a test case. A description of analysis techniques for interferometric observations from the Degree Angular Scale Interferometer (DASI) was presented in White et al. (1999a, 1999b), while Halverson et al. (2002) report on the power spectrum results from the first season of DASI observations. Ng (2001) discusses CMB interferometry with application to the proposed AMIBA instrument. Hobson & Maisinger (2002) have recently presented an approach similar to ours and demonstrate their technique on simulated Very Small Array (VSA) data; a brief comparison of their algorithm with ours is given in Appendix C.

In this paper we describe a fast gridded method for the (u, v)-plane analysis of large interferometric data sets. The basis of this approach is to grid the visibilities and perform maximum likelihood estimation of the power spectrum on these

[1] Department of Physics, University of Alberta, Edmonton, AB T6G 2J1, Canada.
[2] Institut d'Astrophysique de Paris, 98 bis Boulevard Arago, F-75014 Paris, France.

compressed data. Our use of gridded estimators is significantly different from what has been done previously. In addition to power spectrum extraction, this procedure has the ability to form optimally filtered images from the gridded estimators and may be of use in interferometric observations of radio sources in general.

We have applied our method to the analysis of data from the Cosmic Background Imager (CBI). The CBI is a planar interferometer array of 13 individual 90 cm Cassegrain antennas on a 6 m pointable platform (Padin et al. 2002). It covers the frequency range 26–36 GHz in 10 contiguous 1 GHz channels, with a thermal noise level of 2 μK in 6 hr and a maximum resolution of 4′ limited by the longest baselines. The CBI baselines probe l in the range of 500–3900. The 90 cm antenna diameters were chosen to maximize sensitivity, but their primary beam width of 45.′2 (FWHM) at 31 GHz limits the instantaneous field of view, which in turn limits the resolution in l. This loss of aperture plane resolution can be overcome by making mosaic observations, i.e., observations in which a number of adjacent pointings are combined (Ekers & Rots 1979; Cornwell 1988; Cornwell, Holdaway, & Uson 1993; Sault, Staveley-Smith, & Brouw 1996). In the CBI observations, mosaicking a field several times larger than the primary beam has resulted in an increase in resolution in l by more than a factor of 3, sufficient to discern features in the power spectrum.

The first CBI results were presented in Padin et al. (2001, hereafter Paper I), using earlier versions of the software that did not make use of (u, v)-plane gridding, and were far too slow to be used on larger mosaicked data sets. It was therefore essential to develop a more efficient analysis method that would be fast enough to carry out extensive tests on the CBI mosaic data. The software package described below has been used to process the first year of CBI data. In the companion papers by Mason et al. (2003, hereafter Paper II) and Pearson et al. (2003, hereafter Paper III), the results from passing CBI deep field data and mosaic data, respectively, through the pipeline are presented. This paper is Paper IV in the series. The output from this pipeline is then used to derive constraints on cosmology (Sievers et al. 2003, hereafter Paper V). Finally, analysis of the excess of power at high l seen in results shown in Paper II in the context of the Sunyaev-Zeldovich effect is carried out, again using the method presented here, in Bond et al. (2003, hereafter Paper VI).

An introduction to the properties of the CMB power spectrum, the response of an interferometer to the incoming radiation, and the computation of the primary beam are given in §§ 2, 3, and 4, respectively. The gridding process is presented in § 5, followed by a description of the likelihood function and construction of the various covariance matrices in § 6. Details on the maximum likelihood solution and the calculation of window functions and component band powers are given in § 7, while § 8 presents our method for making optimally filtered images from the gridded estimators. Finally, a description of the CBI implementation of this method and the performance of the pipeline, including demonstrations using simulated CBI data sets, is given in § 9, followed by a summary and conclusions in § 10.

2. THE CMB POWER SPECTRUM

At small angles, the curvature of the sky is negligible and we can approximate the spherical harmonic transform of the temperature field $T(x)$ in direction $\boldsymbol{x}$ as its Fourier transform $\tilde{T}(\boldsymbol{u})$ (Bond & Efstathiou 1987), where $\boldsymbol{u}$ is the conjugate variable to $\boldsymbol{x}$. We adopt the Fourier convention

$$\tilde{F}(\boldsymbol{u}) = \int d^2\boldsymbol{x}\, F(\boldsymbol{x}) e^{-2\pi i \boldsymbol{u}\cdot\boldsymbol{x}} \Leftrightarrow F(\boldsymbol{x}) = \int d^2\boldsymbol{u}\, \tilde{F}(\boldsymbol{u}) e^{2\pi i \boldsymbol{u}\cdot\boldsymbol{x}} \tag{1}$$

of Bracewell (1986), Thompson et al. (1986), and Taylor et al. (1999). In terms of the multipoles l,

$$\left\langle \tilde{T}(\boldsymbol{u})^2 \right\rangle \approx C_l, \quad l + \tfrac{1}{2} \approx 2\pi|\boldsymbol{u}| , \tag{2}$$

which we simplify to $l = 2\pi|\boldsymbol{u}|$ for the $l > 100$ of interest in this paper. For the low levels of anisotropy seen in the CMB on these scales, it is convenient to give T in units of μK, and thus C_l will be in units of $\mu\mathrm{K}^2$.

Because the CMB is assumed to be a statistically homogeneous Gaussian random field, the components of its Fourier transform are independent Gaussian deviates:

$$\left\langle \tilde{T}(\boldsymbol{u})\tilde{T}^*(\boldsymbol{u}') \right\rangle = C(|\boldsymbol{u}|)\delta^2(\boldsymbol{u} - \boldsymbol{u}') , \tag{3}$$

where $C(|\boldsymbol{u}|) = C_{2\pi|\boldsymbol{u}|}$. Because $T(\boldsymbol{x})$ is real, its transform must be Hermitian, with $\tilde{T}(\boldsymbol{u}) = \tilde{T}^*(-\boldsymbol{u})$, and therefore

$$\left\langle \tilde{T}(\boldsymbol{u})\tilde{T}(\boldsymbol{u}') \right\rangle = \left\langle \tilde{T}(\boldsymbol{u})\tilde{T}^*(-\boldsymbol{u}') \right\rangle = C(|\boldsymbol{u}|)\delta^2(\boldsymbol{u} + \boldsymbol{u}') . \tag{4}$$

Note that it is common to write the CMB power spectrum C_l in a form

$$\mathscr{C}_l = \frac{l(l+1)}{2\pi} C_l \approx \frac{l^2}{2\pi} C_l \Leftrightarrow \mathscr{C}(|\boldsymbol{u}|) \approx 2\pi|\boldsymbol{u}|^2 C(|\boldsymbol{u}|) \tag{5}$$

(White et al. 1999a; Bond et al. 1998, 2000). Constant $\mathscr{C}$ corresponds to equal power in equal intervals of $\log l$.

Although the power spectrum C_l is defined in units of brightness temperature, the interferometer measurements carry the units of flux density S_ν (jansky, 1 Jy $= 10^{-26}$ W m^{-2} Hz^{-1}). In particular, the intensity field on the sky $I_\nu(\boldsymbol{x})$ has units of specific intensity (W m^{-2} Hz^{-1} sr^{-1}, or Jy sr^{-1}), and thus to convert between I_ν and T we use $I_\nu(\boldsymbol{x}) = f_T(\nu)T(\boldsymbol{x})$ with the Planck factor

$$f_T(\nu) = \frac{2\nu^2 k_\mathrm{B} g(\nu, T_0)}{c^2}, \quad g(\nu, T_0) = \frac{x^2 e^x}{(e^x - 1)^2}, \quad x = \frac{h\nu}{k_\mathrm{B} T_0} , \tag{6}$$

where g corrects for the blackbody spectrum. Note that we have treated the temperature T as small fluctuations about the mean CMB temperature $T_0 = 2.725$ K (Mather et al. 1999), and thus the g appropriate to T_0 is used with $g \approx 0.98$ at $\nu = 31$ GHz.

We are not restricted to modeling the CMB. For example, we might wish to determine the power spectrum of fluctuations in a diffuse galactic component such as synchrotron emission or thermal dust emission. In this case, one might wish to express I in Jy sr^{-1} but take out a power-law spectral shape

$$I_\nu = f_0(\nu) I_0, \quad f_0(\nu) = \left(\frac{\nu}{\nu_0}\right)^\alpha , \tag{7}$$

where α is the spectral index and $f_0(\nu)$ is the conversion factor that normalizes to the intensity I_0 at the fiducial frequency ν_0. Note that this normalization is particularly useful for fitting out centimeter-wave foreground emission, which tends to have a power-law spectral index in the range $-1 < \alpha < 1$ that is significantly different from that for the thermal CMB ($\alpha \approx 2$). In addition, foregrounds will also tend to have a power spectrum shape different from that of CMB, which must be included in the analysis (see § 6.4).

3. RESPONSE OF THE INTERFEROMETER

A visibility V_k formed from the correlation of an interferometer baseline between two antennas with projected separation (in the plane perpendicular to the source direction) $\boldsymbol{b}$ meters observed at wavelength λ meters measures (in the absence of noise) the Fourier transform of the sky intensity modulated by the response of the antennas (Thompson et al. 1986)

$$V(\boldsymbol{u}) = \int d^2\boldsymbol{x}\, \mathscr{A}(\boldsymbol{x}) I(\boldsymbol{x}) e^{-2\pi i \boldsymbol{u}\cdot\boldsymbol{x}}, \quad \boldsymbol{u} = \frac{\boldsymbol{b}}{\lambda}, \tag{8}$$

where $\mathscr{A}(\boldsymbol{x})$ is the primary beam and $\boldsymbol{u} = (u, v)$ is the conjugate variable to $\boldsymbol{x}$. For angular coordinates $\boldsymbol{x}$ in radians, $\boldsymbol{u}$ has the dimensions of the baseline or aperture in units of the wavelength. The Fourier domain is also referred to as the (u, v)-plane or aperture plane in interferometry for this reason.

We define the direction cosines

$$\boldsymbol{x}_k = (\Delta x_k, \Delta y_k), \quad \Delta x_k = \cos\delta_k \sin(\alpha_k - \alpha_0), \quad \Delta y_k = \sin\delta_k \cos\delta_0 - \cos\delta_k \times \sin\delta_0 \cos(\alpha_k - \alpha_0) \tag{9}$$

between the point at right ascension and declination α_k, δ_k and the center of the mosaic α_0, δ_0. For the CBI, data are taken keeping the phase center fixed on the pointing center $\boldsymbol{x}_k$ by shifting the phase with the beam and rotating the platform to maintain constant parallactic angle during a scan, so that the response to a point source at the center of the field $I(\boldsymbol{x}) = \delta^2(\boldsymbol{x} - \boldsymbol{x}_k)$ is constant, and thus

$$\mathscr{A}(\boldsymbol{x}) = A_k(\boldsymbol{x} - \boldsymbol{x}_k) e^{2\pi i \boldsymbol{u}_k\cdot\boldsymbol{x}_k} \tag{10}$$

in equation (8), where A_k is the normalized primary beam response at the observing frequency of visibility k. Then, by application of the Fourier shift theorem, it is easy to show that

$$\begin{aligned} V_k &= \int d^2\boldsymbol{x}\, A_k(\boldsymbol{x} - \boldsymbol{x}_k) I_{\nu_k}(\boldsymbol{x}) e^{-2\pi i \boldsymbol{u}_k\cdot(\boldsymbol{x}-\boldsymbol{x}_k)} + e_k \\ &= \int d^2\boldsymbol{v}\, \tilde{A}_k(\boldsymbol{u}_k - \boldsymbol{v}) \tilde{I}_{\nu_k}(\boldsymbol{v}) e^{2\pi i \boldsymbol{v}\cdot\boldsymbol{x}_k} + e_k, \end{aligned} \tag{11}$$

where $\tilde{A}_k$ is the Fourier transform of the primary beam A_k and $I_\nu(\boldsymbol{x})$ is the sky brightness field (expressed in units such as Jy sr^{-1}) with transform $\tilde{I}_\nu(\boldsymbol{v})$. The instrumental noise on the complex visibility measurement is represented by e_k.

The (u, v)-plane resolution of an interferometer in a single pointing is thus limited by the convolution with $\tilde{A}$. However, these subaperture spatial frequencies can be recovered by using the phase gradient in the exponential $\exp(2\pi i \boldsymbol{v}\cdot\boldsymbol{x}_k)$ from a raster of mosaic pointings $\{\boldsymbol{x}_k\}$, provided that the spacing of the pointings is sufficiently small to avoid aliasing as discussed in Appendix A.

To aid us later on, we introduce a convolution kernel

$$P_k(\boldsymbol{v}) = f_k \tilde{A}_k(\boldsymbol{u}_k - \boldsymbol{v}) e^{2\pi i \boldsymbol{v}\cdot\boldsymbol{x}_k} \tag{12}$$

and thus

$$V_k = \int d^2\boldsymbol{v}\, P_k(\boldsymbol{v}) \tilde{T}(\boldsymbol{v}) + e_k, \tag{13}$$

where $f_k = f_T(\nu_k)$ is the Planck conversion factor for the CMB given in equation (6). It is easiest to write these in operator notation, with

$$\boldsymbol{V} = \boldsymbol{P}\tilde{\boldsymbol{T}} + \boldsymbol{e}, \tag{14}$$

where $\boldsymbol{V}$ and $\boldsymbol{e}$ are the visibility and noise vectors, respectively, $\boldsymbol{P}$ is our kernel, and $\tilde{\boldsymbol{T}}$ is the transform of the temperature field. In this representation $\tilde{\boldsymbol{T}}$ can be thought of as a vector of cells in (u, v)-space.

4. THE PRIMARY BEAM

In order to determine the response of the array to the radiation field, we need to know the primary beam $A(\boldsymbol{x})$ of the antenna elements and its Fourier transform $\tilde{A}(\boldsymbol{u})$. In general, for each frequency channel, each baseline has a primary beam that is the Fourier transform of the cross-correlation of the voltage illumination functions across the aperture of each antenna (see Thompson et al. 1986 for a detailed treatment of the interferometer response). For a real and symmetric primary beam that is identical between antennas, then the transforms are symmetric and real, and we can ignore the differences between cross-correlation and convolution and write

$$\tilde{A}(\boldsymbol{u}) = \hat{g} \star \hat{g} \Leftrightarrow A(\boldsymbol{x}) = |\tilde{g}^2| \tag{15}$$

for the voltage illumination function $\hat{g}(r, \nu)$ across the radius of the aperture $r = |\boldsymbol{r}|$ at frequency ν, where $\tilde{g}$ is the Fourier transform of $\hat{g}$. The CBI beams have been measured and are nearly identical and symmetric, and thus we will use a single mean primary beam and its transform for the array. For a heterogeneous array, the individual beams can be used with some added complication.

For most antennas, such as those used in the CBI, the primary beam width scales linearly with observing wavelength, and thus $\hat{g}(r)$ is approximately constant with wavelength. Then, we can define $G(r)$ as the normalized aperture autocorrelation function and write

$$\tilde{A}_k(\boldsymbol{u}) = \frac{1}{A_0} G(|\boldsymbol{u}|\lambda_k) \tag{16}$$

for a channel centered at wavelength λ_k, with

$$A_0 = \int d^2\boldsymbol{u}\, G(|\boldsymbol{u}|\lambda_k) = \frac{2\pi}{\lambda_k^2} \int_0^\infty r\, dr\, G(r) \tag{17}$$

normalizing the response to give unity gain on sky at the beam center [$A(0) \equiv 1$]. If $g(r) = \hat{g}(r)/g(0)$, then $G(r) = g \star g$.

The two-dimensional primary beam response, $A(\boldsymbol{x})$, is determined by means of measurements of a bright radio source over a two-dimensional grid of offset pointings centered on the source. The central lobe of $A(\boldsymbol{x})$ for the CBI is well approximated by a circular Gaussian, which is characterized by its dispersion σ_x, which is related to the FWHM a_x by $\sigma_x = a_x/(8\ln 2)^{1/2}$. The Fourier transform of an infinite circular Gaussian is given by

$$A(\boldsymbol{x}) = e^{-x^2/2\sigma_x^2} \Leftrightarrow \tilde{A}(\boldsymbol{u}) = \frac{1}{2\pi\sigma_u^2} e^{-|\boldsymbol{u}|^2/2\sigma_u^2}, \quad \sigma_u = \frac{1}{2\pi\sigma_x} , \tag{18}$$

where σ_u is the Gaussian dispersion in Fourier space. The function $G(r)$ is therefore

$$G(r) = e^{-r^2/2r_g^2} \tag{19}$$

for Gaussian radius $r_g = \lambda\sigma_u$. For the CBI the measured primary beam (see Paper III) has $a_x = 45'.2(31 \text{ GHz}/\nu)$, so $\sigma_u = 28.50$ at 31 GHz ($\lambda = 0.967$ cm), which corresponds to $r_g = 27.56$ cm.

For the CBI software pipeline, instead of using a Gaussian approximation to $G(r)$, we have chosen to model the antenna illumination $g(r)$ as a Gaussian truncated at both the dish edge and the secondary blockage radius

$$g(r) = \begin{cases} 0, & |r| \le r_{\text{inner}} , \\ e^{-(r/s)^2}, & r_{\text{inner}} < |r| < \dfrac{D}{2} , \\ 0, & |r| \ge \dfrac{D}{2} , \end{cases} \tag{20}$$

where for the CBI antennas $r_{\text{inner}} = 7.75$ cm. Note that if $g(r)$ and $G(r)$ were infinite circular Gaussians, then $s = r_g$. A best-fit taper parameter s is obtained using the measured primary beam A, giving $s = 30.753$ cm or an edge taper of 0.118 (-18.6 dB of power) at the dish edge. We then numerically tabulate the autocorrelation $G(r)$ assuming $s = 30.753$ cm, which is then interpolated in the code when $\tilde{A}$ is required. This model is a better fit to the observed beam than a pure Gaussian beam (see Fig. 1 in Paper III for a plot of this model).

The resolution in (u, v)- or l-space is set by the width of $\tilde{A}_k(\boldsymbol{u})$. For a Gaussian approximation to the beam, the dispersion in multipole l is $\sigma_l = 2\pi\sigma_u = 1/\sigma_x$, and the FWHM is $a_l = 8 \ln 2/a_x$. For $a_x = 45'.2$ at 31 GHz we have FWHM $a_l = 422$ ($\sigma_l = 179$). Given that features are expected in the power spectrum of widths significantly less than this, it is highly desirable to reduce the effective resolution width of the CBI by at least a factor of 3 using mosaicking.

5. GRIDDED ESTIMATORS

The principal problem in using likelihood (see § 6) to determine confidence limits on the power spectrum for CBI data is the large number of visibilities compounded by the large number of mosaic pointings (typically 7×6 or larger). Even a modest reduction in the number of matrix elements passed to the likelihood calculation will greatly aid the computation. This suggests that we grid the visibilities before computing the likelihood function. For an effective resolution in the aperture plane determined by the primary beam and mosaic size, there is little use in sampling below this smearing scale, and we can define an optimum gridding scheme that minimizes the quantity of data and information loss (the gridding is a form of compression).

We implement this by defining *estimators* $\Delta(\boldsymbol{u})$ for the true complex brightness transform that are linear combinations of the measured visibilities. These estimators bin together data from the different frequency bands and mosaic pointings. Thus, a direct sum of visibilities taken at the same $\boldsymbol{u}$ but over the whole mosaic $\boldsymbol{x}$ will result in an estimator that has a higher effective resolution in the (u, v)-plane. The result is that we can speed up the likelihood estimation at the cost of complicating the covariance matrix. In general, this matrix can be computed relatively quickly as it is an N^2 process, and thus this is a worthwhile trade-off versus the N^3 cost of calculating the likelihood. The estimators derived in Appendix A are not orthogonal combinations of the original visibilities, and thus some information loss is expected. However, the tests performed in § 9.1 show that these estimators are unbiased, and comparisons to results obtained using the visibilities directly show that there is no noticeable loss in sensitivity. Thus, our gridding can be considered to be an efficient form of (lossy) compression using the beam as a signal template.

In Appendix A we argue that a Δ_i formed by a linear combination of visibilities will give an estimate of the weighted average of $\tilde{I}$ or $\tilde{T}$ around (u, v) locus $\boldsymbol{u}_i$. We have from equation (A13)

$$\boldsymbol{\Delta} = \boldsymbol{QV} + \bar{\boldsymbol{Q}}\boldsymbol{V}^* , \tag{21}$$

where the kernel $\boldsymbol{Q}$ is defined in equation (A13) and the kernel for the conjugate visibilities $\bar{\boldsymbol{Q}}$ is defined in equation (A17). In particular,

$$\begin{aligned} Q_{ik} &= \frac{\omega_k}{z_i} \tilde{A}_k^*(\boldsymbol{u}_k - \boldsymbol{u}_i) e^{-2\pi i \boldsymbol{u}_i \cdot \boldsymbol{x}_k} , \\ \bar{Q}_{ik} &= \frac{\omega_k}{z_i} \tilde{A}_k^*(-\boldsymbol{u}_k - \boldsymbol{u}_i) e^{-2\pi i \boldsymbol{u}_i \cdot \boldsymbol{x}_k} , \end{aligned} \tag{22}$$

where z_i is the normalization factor given in equation (A21) and $\omega_k = \epsilon_k^{-2}$ is the visibility weight given in equation (A19).

By operating with the gridding kernel on equation (14), we get

$$\boldsymbol{\Delta} = \boldsymbol{R}\tilde{\boldsymbol{T}} + \boldsymbol{n}, \quad \boldsymbol{R} = \boldsymbol{QP} + \bar{\boldsymbol{Q}}\bar{\boldsymbol{P}}, \quad \boldsymbol{n} = \boldsymbol{Qe} + \bar{\boldsymbol{Q}}\boldsymbol{e}^* , \tag{23}$$

where we define $\boldsymbol{R}$ as the convolution kernel that operates on the transform of the intensity (the gridded version of $\boldsymbol{P}$) and $\boldsymbol{n}$ is the gridded noise. The conjugate to $\boldsymbol{P}$ defined in equation (12) is given by

$$\bar{P}_k(\boldsymbol{v}) = f_k \tilde{A}_k(-\boldsymbol{u}_k - \boldsymbol{v}) e^{2\pi i \boldsymbol{v} \cdot \boldsymbol{x}_k} , \tag{24}$$

which gathers the conjugate visibilities under the transformation $\boldsymbol{u}_k \rightarrow -\boldsymbol{u}_k$.

Although it is not necessary to do so, it is convenient to construct the Δ_i on a regular lattice in $\boldsymbol{u}_i$ with a spacing d_u. Thus, the grid "cells" represented by the Δ_i represent an interpolation using $\boldsymbol{Q}$ of the visibilities onto the (u, v)-plane. This will be useful when using the estimators to form filtered images (§ 8).

If it is desired that the visibilities be used directly, for example, when the data sets are small, then the ungridded case can be recovered by setting $Q_{ik} = \delta_{ik}$ and $\bar{Q}_{ik} = 0$, giving $\boldsymbol{\Delta} = \boldsymbol{V}$ and $\boldsymbol{R} = \boldsymbol{P}$, with no loss of generality in the derivations.

6. THE LIKELIHOOD FUNCTION

To carry out the power spectrum estimation, we form the likelihood of the data given covariance matrices for the signal, noise, and foregrounds. Since the estimators are linear combinations of the visibilities, which we assume are made up of Gaussian noise and Gaussian signal components, we can use a multivariate Gaussian probability distribution to describe the estimators also. Because Δ is complex, it is easier to deal with the real and imaginary parts by packing them together in a double-length real vector

$$\boldsymbol{d} = \begin{pmatrix} \mathrm{Re}\boldsymbol{\Delta} \\ \mathrm{Im}\boldsymbol{\Delta} \end{pmatrix} \tag{25}$$

written here as a column vector of length $2N_{\rm est}$.

The log likelihood function for a real multivariate Gaussian probability distribution is

$$\ln L(\boldsymbol{x}|\boldsymbol{q}) = -N_{\rm est}\ln 2\pi - \tfrac{1}{2}\ln(\det \boldsymbol{C}) - \tfrac{1}{2}\boldsymbol{d}^t\boldsymbol{C}^{-1}\boldsymbol{d} , \tag{26}$$

where $\boldsymbol{d}^t$ is the transpose of $\boldsymbol{d}$ and

$$\boldsymbol{C} = \begin{pmatrix} \langle \mathrm{Re}\boldsymbol{\Delta}\mathrm{Re}\boldsymbol{\Delta}^t\rangle & \langle \mathrm{Re}\boldsymbol{\Delta}\mathrm{Im}\boldsymbol{\Delta}^t\rangle \\ \langle \mathrm{Im}\boldsymbol{\Delta}\mathrm{Re}\boldsymbol{\Delta}^t\rangle & \langle \mathrm{Im}\boldsymbol{\Delta}\mathrm{Im}\boldsymbol{\Delta}^t\rangle \end{pmatrix} \tag{27}$$

is a block matrix of the real and imaginary covariances. The vector $\boldsymbol{q}$ represents the parameters of the model or theory against which the data are being measured (see below). These parameters are contained in $\boldsymbol{C}$.

In terms of $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^*$, we can write

$$\mathrm{Re}\boldsymbol{\Delta} = \frac{1}{2}\left(\boldsymbol{\Delta} + \boldsymbol{\Delta}^*\right), \quad \mathrm{Im}\boldsymbol{\Delta} = \frac{1}{2i}\left(\boldsymbol{\Delta} - \boldsymbol{\Delta}^*\right) \tag{28}$$

and therefore

$$\begin{aligned} \langle \mathrm{Re}\boldsymbol{\Delta}\mathrm{Re}\boldsymbol{\Delta}^t\rangle &= \tfrac{1}{2}\mathrm{Re}\left(\langle \boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger\rangle + \langle \boldsymbol{\Delta}\boldsymbol{\Delta}^t\rangle\right) , \\ \langle \mathrm{Im}\boldsymbol{\Delta}\mathrm{Im}\boldsymbol{\Delta}^t\rangle &= \tfrac{1}{2}\mathrm{Re}\left(\langle \boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger\rangle - \langle \boldsymbol{\Delta}\boldsymbol{\Delta}^t\rangle\right) , \\ \langle \mathrm{Re}\boldsymbol{\Delta}\mathrm{Im}\boldsymbol{\Delta}^t\rangle &= -\tfrac{1}{2}\mathrm{Im}\left(\langle \boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger\rangle - \langle \boldsymbol{\Delta}\boldsymbol{\Delta}^t\rangle\right) , \\ \langle \mathrm{Im}\boldsymbol{\Delta}\mathrm{Re}\boldsymbol{\Delta}^t\rangle &= \tfrac{1}{2}\mathrm{Im}\left(\langle \boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger\rangle + \langle \boldsymbol{\Delta}\boldsymbol{\Delta}^t\rangle\right) , \end{aligned} \tag{29}$$

where $\boldsymbol{\Delta}^\dagger = (\boldsymbol{\Delta}^*)^t$ is the Hermitian transpose of $\boldsymbol{\Delta}$ (a row vector containing the complex conjugate of a column vector) and $\boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger$ is the tensor or outer product of $\boldsymbol{\Delta}$ and $\boldsymbol{\Delta}^\dagger$, which is a matrix with elements $\Delta_i\Delta_j^*$.

It is important to include the covariances of $\boldsymbol{\Delta}\boldsymbol{\Delta}^t$, as well as those for $\boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger$. Normally, only one of a given visibility V_k or its conjugate V_k^* will correlate with $V_{k'}$. However, for short baselines $b < \sqrt{2}D$ (less than 127.3 cm for the 90 cm CBI dishes), there may be overlap between the support for a given visibility and both another visibility and its conjugate, as shown in Figure 1, and thus both may be nonzero. Note that the correlation between distant conjugate pairs is small, since the overlap occurs far out in the antenna response $\tilde{A}$, although it is nonnegligible on the shortest CBI baselines where the overlap occurs at the $0.57D$ point (illustrated in Fig. 1) for perpendicular 1 m baselines with the beam response $\sim$30%. Outside the baseline range $b > \sqrt{2}D$ one of $\langle V_k^* V_{k'}\rangle$ or $\langle V_k V_{k'}\rangle$ will be zero.

The covariance matrix $\boldsymbol{C}$ can be split into a sum of independent contributions from instrumental noise $\boldsymbol{C}^N$, the CMB signal $\boldsymbol{C}^S$, and foreground signals $\boldsymbol{C}^{\rm src}$ and $\boldsymbol{C}^{\rm res}$. We further split $\boldsymbol{C}^S$ into a sum of terms $\boldsymbol{C}_B^S$ from separate l bands of the power spectrum,

$$\boldsymbol{C} = \boldsymbol{C}^N + \sum_B q_B \boldsymbol{C}_B^S + q_{\rm src}\boldsymbol{C}^{\rm src} + q_{\rm res}\boldsymbol{C}^{\rm res} . \tag{30}$$

The factors $\{q_B, B = 1, \ldots, N_B\}$ are the "band powers" (Bond et al. 1998) for bins with centers at $l = l_B$ and are the model parameters to be determined by maximizing the likelihood. The factor $q_{\rm res}$ is the amplitude of the covariance due to a residual Gaussian foreground, and $q_{\rm src}$ is the amplitude of the covariance contributed by known point sources; there may be more than one of each of these types of foreground covariance matrices. The $q_{\rm src}$ and $q_{\rm res}$ can be treated as adjustable parameters to be determined by maximum likelihood, or they can be held fixed at a priori values, in which case $\boldsymbol{C}^{\rm src}$ and $\boldsymbol{C}^{\rm res}$ are constraint matrices with their corresponding terms in equation (30) behaving like additional noise terms.

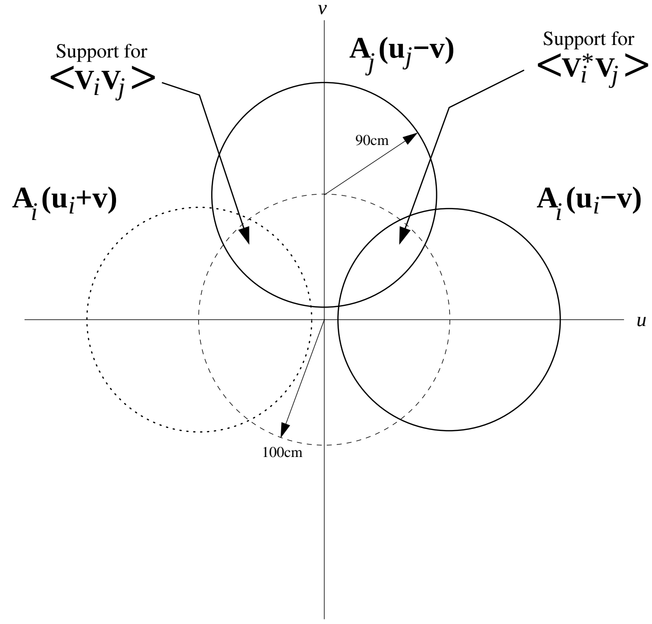

FIG. 1.—Graphical representation of the regions of support in the aperture plane for the correlation between visibilities on short baselines $B < \sqrt{2}D$. Note that both V_i and its conjugate $V_i^* = V(-\boldsymbol{u}_i)$ have overlapping support for visibility V_j, and this must be taken into account in computing the covariance matrix element.

In the following sections we consider each of the terms $\boldsymbol{C}^N$, $\boldsymbol{C}_B^S$, $\boldsymbol{C}^{\rm src}$, and $\boldsymbol{C}^{\rm res}$ in turn. If we write

$$\boldsymbol{M} = \langle \boldsymbol{\Delta}\boldsymbol{\Delta}^\dagger\rangle, \quad \bar{\boldsymbol{M}} = \langle \boldsymbol{\Delta}\boldsymbol{\Delta}^t\rangle , \tag{31}$$

then in each case we calculate the contributions to the covariance matrix for the real and imaginary parts of the estimators using equations (27) and (29):

$$\boldsymbol{C} = \begin{bmatrix} \tfrac{1}{2}\mathrm{Re}\left(\boldsymbol{M} + \bar{\boldsymbol{M}}\right) & -\tfrac{1}{2}\mathrm{Im}\left(\boldsymbol{M} - \bar{\boldsymbol{M}}\right) \\ \tfrac{1}{2}\mathrm{Im}\left(\boldsymbol{M} + \bar{\boldsymbol{M}}\right) & \tfrac{1}{2}\mathrm{Re}\left(\boldsymbol{M} - \bar{\boldsymbol{M}}\right) \end{bmatrix} , \tag{32}$$

with the individual covariance matrices given by insertion of the appropriate contribution to $\boldsymbol{M}$ and $\bar{\boldsymbol{M}}$ for that component, e.g., $\boldsymbol{M}_B^S$ and $\bar{\boldsymbol{M}}_B^S$ to compute the block elements of $\boldsymbol{C}_B^S$.

6.1. *The Noise Covariance Matrix*

The instrumental noise correlations are assumed to be Gaussian and independent between different baselines and frequency channels. For the CBI, tests have been carried out on the data that show this to be true to a high level of accuracy. In this case, the noise contributions to the real and imaginary parts of the visibilities are independent zero-mean normal deviates with

$$\langle \mathrm{Re}e_k \mathrm{Re}e_{k'} \rangle = \langle \mathrm{Im}e_k \mathrm{Im}e_{k'} \rangle = \epsilon_k^2 \, \delta_{kk'} , \quad \langle \mathrm{Re}e_k \mathrm{Im}e_{k'} \rangle = 0 , \tag{33}$$

and thus we can write

$$\langle \boldsymbol{e}\boldsymbol{e}^\dagger \rangle = \boldsymbol{E}, \quad \langle \boldsymbol{e}\boldsymbol{e}^t \rangle = 0 \tag{34}$$

for real noise matrix $\boldsymbol{E}$, where $E_{kk'} = 2\epsilon_k^2 \, \delta_{kk'}$.

It can be shown that the noise contributions $\boldsymbol{n}$ to the estimators defined in equation (23) have the contributions to the covariance elements $\boldsymbol{M}$ and $\bar{\boldsymbol{M}}$ defined in equation (31) given by

$$\begin{aligned} \boldsymbol{M}^N &= \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle = \boldsymbol{Q}\boldsymbol{E}\boldsymbol{Q}^\dagger + \bar{\boldsymbol{Q}}\boldsymbol{E}\bar{\boldsymbol{Q}}^\dagger , \\ \bar{\boldsymbol{M}}^N &= \langle \boldsymbol{n}\boldsymbol{n}^t \rangle = \boldsymbol{Q}\boldsymbol{E}\bar{\boldsymbol{Q}}^t + \bar{\boldsymbol{Q}}\boldsymbol{E}\boldsymbol{Q}^t , \end{aligned} \tag{35}$$

using the covariances of $\boldsymbol{e}$ given in equation (34). This is assembled into the covariance matrix $\boldsymbol{C}^N$ using equation (32). In general, the gridding kernel $\boldsymbol{Q}$ will map a given visibility to more than one estimator, and thus $\boldsymbol{C}^N$ will have nonzero off-diagonal elements. Furthermore, if there are noise correlations between baselines or channels, then the structure of $\boldsymbol{C}^N$ will be even more complicated.

6.2. *The CMB Signal Covariance Matrix*

The CMB contribution to the visibility covariance matrix is given by the covariance of the $\boldsymbol{R}\tilde{\boldsymbol{T}}$ term in equation (23),

$$\boldsymbol{M}^S = \boldsymbol{R}\langle \tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}^\dagger \rangle \boldsymbol{R}^\dagger, \quad \bar{\boldsymbol{M}}^S = \boldsymbol{R}\langle \tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}^t \rangle \boldsymbol{R}^t , \tag{36}$$

where $\langle \tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}^\dagger \rangle$ and $\langle \tilde{\boldsymbol{T}}\tilde{\boldsymbol{T}}^t \rangle$ are given in equations (3) and (4), respectively. Then, the elements of $\boldsymbol{M}^S$ and $\bar{\boldsymbol{M}}^S$ for estimators i and j are

$$\begin{aligned} M_{ij}^S &= \int d^2\boldsymbol{v}\, C(|\boldsymbol{v}|) R_i(\boldsymbol{v}) R_j^*(\boldsymbol{v}) = 2\pi \int d\varpi\, \varpi C(\varpi) W_{ij}(\varpi) , \\ \bar{M}_{ij}^S &= \int d^2\boldsymbol{v}\, C(|\boldsymbol{v}|) R_i(\boldsymbol{v}) R_j(-\boldsymbol{v}) = 2\pi \int d\varpi\, \varpi C(\varpi) \bar{W}_{ij}(\varpi) , \end{aligned} \tag{37}$$

with

$$\begin{aligned} W_{ij}(\varpi) &= \frac{1}{2\pi} \int_0^{2\pi} d\theta\, R_i(\varpi, \theta) R_j^*(\varpi, \theta) , \\ \overline{W}_{ij}(\varpi) &= \frac{1}{2\pi} \int_0^{2\pi} d\theta\, R_i(\varpi, \theta) R_j(\varpi, \theta - \pi) , \end{aligned} \tag{38}$$

where to aid in breaking up the CMB covariance matrices into bands we write the integrations in terms of polar Fourier coordinates $(u, v) \rightarrow (\varpi, \theta)$ $(\varpi = |\boldsymbol{v}|)$.

As an illustration, consider the case without gridding. Then, $\boldsymbol{R} = \boldsymbol{P}$, and using equation (12) in equation (37), we get

$$\begin{aligned} M_{kk'}^S &= f_k f_{k'} \int d^2\boldsymbol{v}\, C(|\boldsymbol{v}|) \tilde{A}_k(\boldsymbol{u}_k - \boldsymbol{v}) \tilde{A}_{k'}^*(\boldsymbol{u}_{k'} - \boldsymbol{v}) e^{2\pi i \boldsymbol{v} \cdot (\boldsymbol{x}_k - \boldsymbol{x}_{k'})} , \\ \bar{M}_{kk'}^S &= f_k f_{k'} \int d^2\boldsymbol{v}\, C(|\boldsymbol{v}|) \tilde{A}_k(\boldsymbol{u}_k - \boldsymbol{v}) \tilde{A}_{k'}(\boldsymbol{u}_{k'} + \boldsymbol{v}) e^{2\pi i \boldsymbol{v} \cdot (\boldsymbol{x}_k - \boldsymbol{x}_{k'})} \end{aligned} \tag{39}$$

for the covariance matrix element between visibilities V_k and $V_{k'}$.

We furthermore write the radial integral over $\varpi = l/2\pi$ as a sum with respect to $\mathscr{C}_l$ of equation (5):

$$\begin{aligned} M_{ij}^S &= \sum_l \frac{W_{lij}}{l} \mathscr{C}_l W_{lij} = W_{ij}\left(\frac{l}{2\pi}\right) , \\ \bar{M}_{ij}^S &= \sum_l \frac{\bar{W}_{lij}}{l} \mathscr{C}_l \bar{W}_{lij} = \bar{W}_{ij}\left(\frac{l}{2\pi}\right) , \end{aligned} \tag{40}$$

where W_{lij} is the *variance window function* (e.g., Knox 1999).

We define the band powers $\{q_B, B = 1, \ldots, N_B\}$ by constructing $\mathscr{C}_l$ piecewise with respect to a fiducial shape $\mathscr{C}_l^{\mathrm{shape}}$,

$$\mathscr{C}_l = \sum_B q_B \mathscr{C}_l^{\mathrm{shape}} \chi_{Bl} , \tag{41}$$

where

$$\chi_{Bl} = \begin{cases} 1 & l \in B \\ 0 & l \notin B \end{cases} \tag{42}$$

breaks the power spectrum into non-overlapping bands. The standard choice for the shape is $\mathscr{C}_l^{\mathrm{shape}} = 1$ for equal power per log l interval, with q_B then giving the band powers in units of T^2. Then, to calculate $\boldsymbol{C}_B^S$, we construct band versions of the covariance matrix elements in equation (36),

$$\boldsymbol{M}_B^S = \sum_l \frac{\boldsymbol{W}_l}{l} \mathscr{C}_l^{\mathrm{shape}} \chi_{Bl}, \quad \bar{\boldsymbol{M}}_B^S = \sum_l \frac{\bar{\boldsymbol{W}}_l}{l} \mathscr{C}_l^{\mathrm{shape}} \chi_{Bl} , \tag{43}$$

where $\boldsymbol{M}^S = \sum_B q_B \boldsymbol{M}_B^S$ and $\bar{\boldsymbol{M}}^S = \sum_B q_B \bar{\boldsymbol{M}}_B^S$. These are then combined following the prescription in equation (32) to assemble the $\boldsymbol{C}_B^S$.

The variance window function $W_{ij}(\boldsymbol{v})$ is the convolution of the $\tilde{A}_i(\boldsymbol{v})$ and $\tilde{A}_j(\boldsymbol{v})$, and thus its width is characteristic of the square of the Fourier transform primary beam, or FWHM $\Delta l \approx a_l/\sqrt{2}$. Thus, we would expect in a single field to be able to achieve a limiting resolution of $\Delta l \approx 300$ for $a_l = 422$ at 31 GHz. This will be increased by the mosaicking by a factor roughly equal to the extent of the half-power width of the mosaic relative to that of a single field. In practice, the limiting useful width for the l bins for the band powers will be set by the band-band correlations introduced in the maximum likelihood estimation procedure (see §§ 7 and 9.1 for further discussion and examples).

6.3. *Known Point-Source Constraint Matrices*

Consider a set of N_c point sources at positions $\boldsymbol{x}_c$ with flux densities $S_c(\nu)$ $(c = 1, \ldots, N_c)$. The intensity field at frequency ν is then given by

$$I_\nu(\boldsymbol{x}) = \sum_c S_c(\nu) \delta^2(\boldsymbol{x} - \boldsymbol{x}_c) , \tag{44}$$

which is assumed to be uncorrelated with other intensity components like the CMB. The effect V_k^{src} on the visibilities V_k (e.g., eq. [11]) is then given by the sum over sources

$$V_k^{\mathrm{src}} = \sum_c V_{ck}, \quad V_{ck} = S_c(\nu_k) A_k(\boldsymbol{x}_c - \boldsymbol{x}_k) e^{-2\pi i \boldsymbol{u}_k \cdot (\boldsymbol{x}_c - \boldsymbol{x}_k)} , \tag{45}$$

where V_{ck} is the contribution to visibility k of source c. We assume that the positions of the sources can be determined with negligible uncertainty through radio surveys and that the errors are due to uncertainties in the measurements of the flux densities. Then, the covariance between the source contributions to visibilities k and k' is

$$\left\langle V_k^{\rm src} V_{k'}^{\rm src*}\right\rangle = \sum_c \sum_{c'} \langle S_c(\nu_k) S_{c'}(\nu_{k'})\rangle A_k(\boldsymbol{x}_c - \boldsymbol{x}_k) A_{k'}^* \times (\boldsymbol{x}_{c'} - \boldsymbol{x}_{k'}) e^{-2\pi i \boldsymbol{u}_k\cdot(\boldsymbol{x}_c-\boldsymbol{x}_k)} e^{2\pi i \boldsymbol{u}_{k'}\cdot(\boldsymbol{x}_{c'}-\boldsymbol{x}_{k'})} , \quad (46)$$

where $\langle S_c(\nu_k) S_{c'}(\nu_{k'})\rangle$ is the flux density covariance matrix between sources c and c' at frequencies ν_k and $\nu_{k'}$, respectively. There is a similar covariance matrix $\langle V_k^{\rm src} V_{k'}^{\rm src}\rangle$. These can be passed through the gridding procedure using equation (21) to make

$$\boldsymbol{\Delta}^{\rm src} = \boldsymbol{Q}\boldsymbol{V}^{\rm src} + \bar{\boldsymbol{Q}}\boldsymbol{V}^{\rm src*} \quad (47)$$

and used to construct the covariance elements

$$\boldsymbol{M}^{\rm src} = \left\langle \boldsymbol{\Delta}^{\rm src}\boldsymbol{\Delta}^{{\rm src},\dagger}\right\rangle, \quad \bar{\boldsymbol{M}}^{\rm src} = \langle \boldsymbol{\Delta}^{\rm src}\boldsymbol{\Delta}^{{\rm src},t}\rangle \quad (48)$$

using equation (31).

This covariance matrix can be greatly simplified if we can subtract off the mean source flux densities, leaving a zero-mean residual error. Let the true source flux density $S_c(\nu)$ be the sum of the measured flux density $S_c^{\rm obs}(\nu)$ and an error $\delta S_c(\nu)$. If our measurements of these foreground sources are accurate, then the residuals $\delta S_c(\nu)$ should be uncorrelated between sources (they are due to measurement errors) and have zero mean. In this case, we can make corrected visibilities $V_k^{\rm cor}$,

$$V_k^{\rm cor} = V_k - \sum_c V_{ck}^{\rm obs} = \sum_c S_c^{\rm obs}(\nu_k) \times A_k(\boldsymbol{x}_c - \boldsymbol{x}_k) e^{-2\pi i \boldsymbol{u}_k\cdot(\boldsymbol{x}_c-\boldsymbol{x}_k)} , \quad (49)$$

to be used in place of $\boldsymbol{V}$ in subsequent analysis. Then, we are left with the fluctuating component

$$\delta V_k^{\rm src} = V_k^{\rm src} - \sum_c V_{ck}^{\rm obs} ,$$
$$\delta V_{ck} = \delta S_c(\nu_k) A_k(\boldsymbol{x}_c - \boldsymbol{x}_k) e^{-2\pi i \boldsymbol{u}_k\cdot(\boldsymbol{x}_c-\boldsymbol{x}_k)} , \quad (50)$$

which we must deal with statistically. The covariance between the source error contributions to the visibilities, assuming that the flux density errors are independent between sources (but not between frequency channels for the same source), is given by

$$\left\langle \delta V_k^{\rm src} \delta V_{k'}^{\rm src*}\right\rangle = \sum_c \langle \delta S_c(\nu_k) \delta S_c(\nu_{k'})\rangle A_k(\boldsymbol{x}_c - \boldsymbol{x}_k) A_{k'}^* \times (\boldsymbol{x}_c - \boldsymbol{x}_{k'}) e^{-2\pi i \boldsymbol{u}_k\cdot(\boldsymbol{x}_c-\boldsymbol{x}_k)} e^{2\pi i \boldsymbol{u}_{k'}\cdot(\boldsymbol{x}_c-\boldsymbol{x}_{k'})} \quad (51)$$

and similarly for $\langle \delta V_k^{\rm src} \delta V_{k'}^{\rm src}\rangle$. Finally, if the covariance is separable, e.g.,

$$\langle \delta S_c(\nu) \delta S_c(\nu')\rangle = \sigma_{Sc}(\nu)\sigma_{Sc}(\nu') , \quad (52)$$

then we can write

$$\left\langle \delta V_k^{\rm src} \delta V_{k'}^{\rm src*}\right\rangle = \sum_c \sigma_{ck}^{\rm src} \sigma_{ck'}^{\rm src*} ,$$
$$\sigma_{ck}^{\rm src} = \sigma_{Sc}(\nu_k) A_k(\boldsymbol{x}_c - \boldsymbol{x}_k) e^{-2\pi i \boldsymbol{u}_k\cdot(\boldsymbol{x}_c-\boldsymbol{x}_k)} . \quad (53)$$

The other covariance $\langle \delta V_k^{\rm src} \delta V_{k'}^{\rm src}\rangle$ can be computed in the same way. Because we have assumed that the covariance is separable, we can speed up the covariance calculation as only the vector $\boldsymbol{\sigma}_c^{\rm src}$ for each source is needed. We can grid this onto the estimators

$$\boldsymbol{\Delta}_c^{\rm src} = \boldsymbol{Q}\boldsymbol{\sigma}_c^{\rm src} + \bar{\boldsymbol{Q}}\boldsymbol{\sigma}_c^{\rm src*} \quad (54)$$

and then

$$\boldsymbol{M}^{\rm src} = \sum_c \boldsymbol{\Delta}_c^{\rm src}\boldsymbol{\Delta}_c^{{\rm src},\dagger}, \quad \bar{\boldsymbol{M}}^{\rm src} = \sum_c \boldsymbol{\Delta}_c^{\rm src}\boldsymbol{\Delta}_c^{{\rm src},t} , \quad (55)$$

which are used to build $\boldsymbol{C}^{\rm src}$.

There are two components to the source flux density uncertainties $\sigma_{Sc}(\nu_k)$, one from the uncertainties on the source frequency spectrum, and the other from the uncertainties on the flux density measurements and any extrapolation of the measured flux densities to the observing frequencies ν_k (using the estimated source spectrum). As an example, consider a source with a flux density $S_c(\nu_0)$ measured with standard deviation σ_{S0} at frequency ν_0 and a known power-law frequency spectrum with spectral index α,

$$S_c(\nu_k) = S_c(\nu_0) f\left(\frac{\nu_k}{\nu_0}, \alpha\right), \quad f\left(\frac{\nu_k}{\nu_0}, \alpha\right) = \left(\frac{\nu_k}{\nu_0}\right)^\alpha . \quad (56)$$

Then, it is easy to show that

$$\frac{\sigma_{Sc}(\nu_k)}{S_c(\nu_k)} = \frac{\sigma_{S0}}{S_c(\nu_0)} , \quad (57)$$

with the fractional uncertainty in the flux density σ_{Sc}/S_c remaining independent of the frequency.

On the other hand, consider the case in which there is now an uncertainty σ_α in the spectral index between ν_k and ν_0. Then, our extrapolation factor $f(\nu/\nu_0, \alpha)$, which we write as

$$f\left(\frac{\nu}{\nu_0}, \alpha\right) = e^{\alpha \ln(\nu/\nu_0)} , \quad (58)$$

propagates to the extrapolated flux density as

$$\frac{\sigma_{Sc}(\nu_k)}{S_c(\nu_k)} = \ln\left(\frac{\nu}{\nu_0}\right)\sigma_\alpha , \quad (59)$$

which can be negative: for two channels flanking the fiducial frequency (e.g., $\nu < \nu_0 < \nu'$) the errors will be anticorrelated. Note that we have approximated the resulting distribution as Gaussian. In general it is not, e.g., for a Gaussian distribution in α we find a lognormal distribution in $S(\nu)$.

Although the dominant spectral error is due to the extrapolation from a frequency ν_0 outside the range of the CMB instrument, there is an additional error due to an error in the spectral index over the frequency channels ν_k of the visibilities. This is as if you extrapolated using one spectrum appropriate for the band center $\bar{\nu}$ of the instrument, but when the flux densities $S_c^{\rm obs}(\nu_k)$ are extrapolated from band center $S_c^{\rm obs}(\bar{\nu})$, there is an error from using the wrong α over the band. This is handled using equation (59) with another σ_α appropriate to the uncertainty in the spectral index over the ν_k.

For the CBI analysis, we have approximated both the flux density error and the spectral extrapolation error as a single equivalent flux density error. For the CBI, the frequency span (26–36 GHz, or $d\nu/\nu = \pm 16\%$) is small enough that we can approximate the spectral index uncertainty as an effective flux density uncertainty σ_c extrapolated to band center $\bar{\nu}$ from ν_0 using α_0,

$$\sigma_{Sc}(\nu_k) \approx f\left(\frac{\nu}{\bar{\nu}}, \alpha\right)\sigma_c,$$

$$\sigma_c^2 = f^2\left(\frac{\bar{\nu}}{\nu_0}, \alpha_0\right)\left\{\sigma_{S0}^2 + S_0^2\left[\ln\left(\frac{\bar{\nu}}{\nu_0}\right)\right]^2 \sigma_\alpha^2\right\}, \quad (60)$$

where α need not equal α_0 and should reflect the spectral index over the observing band, not the one used for extrapolation from ν_0.

In principle, if the true mean flux densities for the sources are correctly subtracted from the visibilities and the covariance matrix $\boldsymbol{C}^{\rm src}$ is built using the correct elements $\langle \delta S_c(\nu_k)\delta S_{c'}(\nu_{k'})\rangle$, then inclusion of this as a noise term in $\boldsymbol{C}$ using $q_{\rm src} = 1$ would remove the effects of these sources from our power spectrum estimation in a statistical sense. However, there are a number of factors that make this difficult. If the source flux density measurements have a calibration error, then the errors will not be independent between sources. In addition, the fainter sources (which are still significant contributors to the signal) have flux densities that are often extrapolated from much lower frequencies (e.g., the "NVSS" sources in Papers II and III). Furthermore, since there are a relatively small number of discrete sources contributing to a given field, it is not clear that we are in the statistical limit where the flux density covariance is an accurate description of what is happening to the data. For these reasons, for the CBI analysis we treat the covariance matrix $\boldsymbol{C}^{\rm src}$ constructed using the approximation in equation (60) as a constraint matrix for the nuisance parameters due to the sources and set $q_{\rm src}$ to a high enough amplitude to *project out* the contaminated modes in the data. Because the source modes are spread out by the effect of the synthesized beam (the "point-spread function" [PSF] in imaging terms), setting $q_{\rm src}$ to too high a value will start to down-weight modes that are not significantly contaminated, while too low a value will eat into the noise and CMB signal power in those modes without down-weighting them sufficiently, thus biasing the affected band powers low. The exact values to be used thus depend on the signal and noise levels in the data; we refer the reader to Papers II and III for descriptions of what was chosen for the CBI analysis. See Bond et al. (1998) for a description of the constraint matrix formalism and the technique of projection.

6.4. *Gaussian Foregrounds and Residual Point Sources*

In § 2 it was mentioned that a single foreground component could be modeled with a modified covariance matrix, power spectrum shape, and frequency dependence. As long as these foregrounds can be treated as a Gaussian random field, they can be processed in the same manner as the CMB. Therefore, once the amplitude and shape of the foreground fluctuation power spectrum $C_{\rm res}(v)$ are input, we compute the foreground covariance matrix elements

$$\boldsymbol{M}^{\rm res} = \sum_l \frac{\boldsymbol{W}_l^{\rm res}}{l}\mathscr{C}_l^{\rm res}, \quad \bar{\boldsymbol{M}}^{\rm res} = \sum_l \frac{\bar{\boldsymbol{W}}_l^{\rm res}}{l}\mathscr{C}_l^{\rm res}, \quad (61)$$

where the variance window functions $\boldsymbol{W}^{\rm res}$ and $\bar{\boldsymbol{W}}^{\rm res}$ are given by substituting for $\boldsymbol{R}$ in equation (38) a new $\boldsymbol{R}^{\rm res}$ built from a kernel

$$P_k^{\rm res}(\boldsymbol{v}) = f_k^{\rm res}\tilde{A}_k(\boldsymbol{u}_k - \boldsymbol{v})e^{2\pi i\boldsymbol{v}\cdot\boldsymbol{x}_k} \quad (62)$$

using a frequency factor $f_k^{\rm res} = f^{\rm res}(\nu_k)$ appropriate to the foreground in question. The matrix $\boldsymbol{C}^{\rm res}$ is then obtained by substitution of $\boldsymbol{M}^{\rm res}$ and $\bar{\boldsymbol{M}}^{\rm res}$ as usual using equation (32). Although it is possible to break up the Gaussian foreground component into bands as we did the CMB, it is preferable to compute the foreground covariance matrix in a single band using its shape $\mathscr{C}_l^{\rm res}$, to reduce the degeneracy with the CMB; you cannot distinguish between the two in narrow l bands where the shapes are unimportant.

An example of a foreground that strongly affects the CBI data is that from point sources below the limit for subtraction contaminating the CBI fields. This *residual* statistical background, in the limit where there are many sources per field, can be modeled as a white-noise Gaussian field with constant angular power spectrum and power-law frequency spectrum. Each individual source has a flux density drawn from a differential number count distribution $dN(S_\nu)/dS$, which represents the number of sources per steradian with flux densities between S_ν and $S_\nu + dS$ at observing frequency ν. The angular clustering in these sources is very small and can be neglected.

The contribution of a source c to visibility V_k was given by V_{ck} in equation (45). The sources are independently distributed in flux density and on the sky, so

$$\begin{aligned}\langle V_k V_{k'}^*\rangle &= \Big\langle \sum_c S_c(\nu_k)S_c(\nu_{k'})A_k(\boldsymbol{x}_c - \boldsymbol{x}_k) \\ &\quad \times A_{k'}^*(\boldsymbol{x}_c - \boldsymbol{x}_{k'})e^{2\pi i\boldsymbol{u}_k\cdot(\boldsymbol{x}_c-\boldsymbol{x}_k)}e^{-2\pi i\boldsymbol{u}_{k'}\cdot(\boldsymbol{x}_c-\boldsymbol{x}_{k'})}\Big\rangle \\ &= \frac{1}{\Omega}\Big\langle \sum_c S_c(\nu_k)S_c(\nu_{k'})\Big\rangle B_{kk'} \\ &= C^{\rm res}(\nu_k,\nu_{k'})B_{kk'}, \end{aligned} \quad (63)$$

where the angular average can be written as an integral over

$$B_{kk'} = \int d^2\boldsymbol{x}\, A_k(\boldsymbol{x} - \boldsymbol{x}_k)A_{k'}^*(\boldsymbol{x} - \boldsymbol{x}_{k'})e^{2\pi i\boldsymbol{u}_k\cdot(\boldsymbol{x}-\boldsymbol{x}_k)} \times e^{-2\pi i\boldsymbol{u}_{k'}\cdot(\boldsymbol{x}-\boldsymbol{x}_{k'})}, \quad (64)$$

with Ω as the normalizing solid angle. This integral is just a Fourier transform, and so

$$B_{kk'} = \int d^2\boldsymbol{v}\, \tilde{A}_k(\boldsymbol{u}_k - \boldsymbol{v})\tilde{A}_{k'}^*(\boldsymbol{u}_{k'} - \boldsymbol{v})e^{2\pi i\boldsymbol{v}\cdot(\boldsymbol{x}_k-\boldsymbol{x}_{k'})}, \quad (65)$$

which is the same as the CMB visibility covariance matrix $M_{kk'}$ in equation (39) with $f_k = f_{k'} = 1$ and $C(v) = 1$. Similarly, the other covariance $\langle V_k V_{k'}\rangle$ reduces to $\bar{M}_{kk'}$. Thus, in the stochastic limit the residual source background behaves as a Gaussian random field and can thus be treated as we do the CMB signal in § 6.2 but with a power spectrum shape $C_l = C^{\rm res}(\nu_k, \nu_{k'})$, which is constant over l for a given pair of frequency channels.

The amplitude of the covariance matrix is the ensemble average of the source power per solid angle, which is obtained by integration over the flux density and spectral index distributions

$$C^{\rm res}(\nu_k,\nu_{k'}) = \int_{S_{\rm min}}^{S_{\rm max}} dS\, S^2\frac{dN(S)}{dS}\int_{-\infty}^{\infty} d\alpha\, p(\alpha|S,\nu_0) \times \left(\frac{\nu_k\nu_{k'}}{\nu_0^2}\right)^\alpha, \quad (66)$$

where we have again assumed that the spectrum is a power law with spectral index α over the range of interest for the ν_k and integrate over the number counts over the flux density range from $S_{\rm min}$ to $S_{\rm max}$. We also assume that there is a large number of these faint sources over the solid angles of interest (e.g., the CBI primary beam), and thus the Poisson contribution to the probability can be ignored and we can use the mean source density given by the number counts dN/dS at the fiducial frequency ν_0 for which S is given. The spectral index distribution as a function of flux density $p(\alpha|S, \nu_0)$ must be estimated from radio surveys, although it can be uncertain at the high frequencies and faint levels at which the CMB experiments are carried out. If $p(\alpha|S, \nu_0) = p(\alpha|\nu_0)$ and thus is independent of flux density, then it can be shown (e.g., Appendix B) that the integrals in equation (66) can be evaluated at a single frequency ν in the band and scaled using an effective spectral index $\alpha_{\rm eff}$,

$$C^{\rm res}(\nu_k, \nu_{k'}) = C^{\rm res}_\nu f^{\rm eff}_k f^{\rm eff}_{k'} , \quad f^{\rm eff}_k = \left(\frac{\nu_k}{\nu}\right)^{\alpha_{\rm eff}} , \qquad (67)$$

where $C^{\rm res}_\nu$ is the amplitude of the fluctuation power per solid angle (in units of Jy^2 sr^{-1}) at ν. In terms of the logarithmic average $\mathscr{C}$ for the CMB, $\mathscr{C}^{\rm res}_\nu = l^2 C^{\rm res}_\nu / 2\pi$, which rises at high l with respect to the CMB. See Appendix B for an example analytic calculation using power-law source counts and a Gaussian spectral index distribution.

The frequency range of the CBI is insufficient to distinguish nonthermal foreground emission from the thermal CMB, and thus this is treated as a constraint matrix (i.e., $q_{\rm res}$ is not solved for as a parameter). Therefore, in the CBI analysis we construct the covariance matrix $\boldsymbol{C}^{\rm res}$ using the matrix elements in equation (61) built assuming unit power (1 Jy^2 sr^{-1}) and the frequency dependence $f_k = f^{\rm eff}_k$ from equation (67). The value used for $q_{\rm res}$ is equal to the source fluctuation power $\mathscr{C}^{\rm res}_\nu$ calculated as an a priori estimate based on knowledge of the residual foreground source populations (see Appendix B).

6.5. *Other Signal Components*

We are not restricted to CMB, Gaussian foreground, and discrete point sources as the components of our signal or noise in the covariance matrix $\boldsymbol{C}$ in equation (30). This approach can be generalized to deal with other signals of interest. For example, extended sources with a known profile, such as the Sunyaev-Zeldovich effect from clusters of galaxies, could be modeled either analytically or numerically given a power spectrum shape (e.g., Bond & Myers 1996). In the case of a signal with a known distribution on the sky, a template can be used. Examples of this include dust emission in the millimeter-wave bands or the anomalous centimeter-wave emission observed at the Owens Valley Radio Observatory (OVRO; Leitch et al. 1997) and by *COBE* (Kogut et al. 1996). In particular, the latter foreground, which is posited as due to spinning dust grains by Draine & Lazarian (1999), correlates with the 100 μm dust emission as measured by *IRAS* and DIRBE, and thus a template of emission can be constructed.

6.6. *Differencing*

Unfortunately, with its low intrinsic fringe rates and extremely short ($<90\lambda$) spacings, the CBI is susceptible to ground pickup. To remove this, we observe for each field a trailing field displaced 8$^{\rm m}$ in right ascension 8$^{\rm m}$ later and difference the corresponding visibilities. Therefore, we must take this differencing into account in our correlation analysis. This effectively modifies the window function, quenching low spatial frequencies further.

Let us write

$$V^{\rm sw}_k = V^{\rm main}_k - V^{\rm trail}_k , \quad \boldsymbol{x}^{\rm trail}_k = \boldsymbol{x}_k + \Delta\boldsymbol{x}_k \qquad (68)$$

for switching offset $\Delta\boldsymbol{x}_k$ (e.g., 8$^{\rm m}$ in R.A. $\approx 2°$ for the CBI fields near the celestial equator). Then, from equation (11) we find

$$V^{\rm sw}_k = f_k \int d^2\boldsymbol{v}\, \tilde{A}_k(\boldsymbol{u}_k - \boldsymbol{v}) \tilde{T}(\boldsymbol{v}) e^{2\pi i \boldsymbol{v}\cdot\boldsymbol{x}_k} \left(1 - e^{2\pi i \boldsymbol{v}\cdot\Delta\boldsymbol{x}_k}\right) + e^{\rm sw}_k , \qquad (69)$$

where the switched noise $e^{\rm sw}_k = e^{\rm main}_k - e^{\rm trail}_k$. In terms of the kernel of equation (12),

$$V^{\rm sw}_k = \int d^2\boldsymbol{v}\, P^{\rm sw}_k(\boldsymbol{v}) \tilde{T}(\boldsymbol{v}) + e^{\rm sw}_k ,$$
$$P^{\rm sw}_k(\boldsymbol{v}) = P_k(\boldsymbol{v})\left(1 - e^{2\pi i \boldsymbol{v}\cdot\Delta\boldsymbol{x}_k}\right) , \qquad (70)$$

and thus for our switched visibilities we compute everything as before, but substituting $P^{\rm sw}_k$ for P_k.

Note that if the trail field offsets $\Delta\boldsymbol{x}$ were constant in arc length rather than in right ascension (this is approximately true since the declination range of the mosaic is limited), we could write the convolution kernel as

$$P^{\rm sw}_k(\boldsymbol{v}) P^{\rm sw*}_{k'}(\boldsymbol{v}) = P_k(\boldsymbol{v}) P^*_{k'}(\boldsymbol{v}) [2 - 2\cos(2\pi \boldsymbol{v}\cdot\Delta\boldsymbol{x})] , \qquad (71)$$

where the leading factor of 2 dominates (you essentially get twice the CMB power). Note that a noticeable effect of the differencing is that the window function will have a ripple of "wavelength" Δx^{-1} superimposed on its envelope. For example, the 8$^{\rm m}$ switching in right ascension that the CBI uses corresponds to $\Delta x = 2°$ at the celestial equator, and thus the ripple has $\Delta x^{-1} = 28.6$ in u. This corresponds to 180 in l but is azimuthally averaged in the (u, v)-plane, and thus the peak-to-peak amplitude is reduced.

7. SOLVING THE LIKELIHOOD EQUATION

We have expressed the estimators as a real vector $\boldsymbol{d}$ and obtained expressions for the components of its covariance matrix, and we now turn to the problem of solving for the maximum likelihood estimators for the band powers using equation (26). As shown below, we will be carrying out a large number of matrix operations using $\boldsymbol{C}$ and its component matrices ($\boldsymbol{C}^N$, $\boldsymbol{C}^S_B$, etc.), and thus these will need factorization. Because $\boldsymbol{C}$ is positive definite, we use optimized Cholesky decomposition routines[3] (DCHDC from LINPACK, or DPOTRF from LAPACK) to carry out the required factorizations.

The large number of visibilities *times* the number of mosaic pointings makes this computation extremely costly (the matrix inversions and/or solution of systems of equations are order N^3 processes!), especially for a large number of bands N_B. Clever perturbative or gradient search methods can help to reduce the overhead in finding the maximum in parameter space. One such method is the quadratic relaxation technique of Bond et al. (1998). To summarize here, if one Taylor expands the log likelihood around the maximum

[3] Available at http://www.netlib.org.

likelihood band powers $\hat{\boldsymbol{q}} = \{\hat{q}_B, B = 1, \ldots, N_B\}$ to second order,

$$\ln L(\hat{\boldsymbol{q}} + \delta\boldsymbol{q}) = \ln L(\hat{\boldsymbol{q}}) + \sum_B \frac{\partial \ln L(\hat{\boldsymbol{q}})}{\partial q_B} \delta q_B + \frac{1}{2} \sum_{B'} \sum_{B''} \frac{\partial^2 \ln L(\hat{\boldsymbol{q}})}{\partial q_B \partial q_{B'}} \delta q_B \, \delta q_{B'} , \quad (72)$$

then we can move toward the maximum using the quadratic approximation

$$\delta q_B = -\sum_{B'} \left[\frac{\partial^2 \ln L(\boldsymbol{q})}{\partial q_B \partial q_{B'}} \right]^{-1} \frac{\partial \ln L(\boldsymbol{q})}{\partial q_{B'}} . \quad (73)$$

The first derivative (gradient) is given by

$$\frac{\partial \ln L(\boldsymbol{q})}{\partial q_{B'}} = \frac{1}{2} \mathrm{Tr} \left[(\boldsymbol{d}\boldsymbol{d}^t - \boldsymbol{C}) \left(\boldsymbol{C}^{-1} \frac{\partial C}{\partial q_{B'}} \boldsymbol{C}^{-1} \right) \right] , \quad (74)$$

and the second derivative (curvature matrix) is given by

$$\begin{aligned} \mathscr{F}_{BB'} &= -\frac{\partial^2 \ln L(\boldsymbol{q})}{\partial q_B \partial q_{B'}} \\ &= \mathrm{Tr} \Big[(\boldsymbol{d}\boldsymbol{d}^t - \boldsymbol{C}) \Big(\boldsymbol{C}^{-1} \frac{\partial C}{\partial q_B} \boldsymbol{C}^{-1} \frac{\partial C}{\partial q_{B'}} \boldsymbol{C}^{-1} \\ &\quad - \frac{1}{2} \boldsymbol{C}^{-1} \frac{\partial^2 C}{\partial q_B \partial q_{B'}} \boldsymbol{C}^{-1} \Big) \Big] \\ &\quad + \frac{1}{2} \mathrm{Tr} \left(\boldsymbol{C}^{-1} \frac{\partial C}{\partial q_B} \boldsymbol{C}^{-1} \frac{\partial C}{\partial q_{B'}} \right) . \end{aligned} \quad (75)$$

Note that the partial derivatives of the covariance matrix are just the band signal covariance matrices $\partial \boldsymbol{C} / \partial q_B = \boldsymbol{C}_B^S$ defined above.

The final approximation is to replace the curvature matrix with its expectation value, which is the Fisher information matrix

$$F_{BB'} = \langle \mathscr{F}_{BB'} \rangle = \tfrac{1}{2} \mathrm{Tr} \left(\boldsymbol{C}^{-1} \boldsymbol{C}_B^S \boldsymbol{C}^{-1} \boldsymbol{C}_{B'}^S \right) . \quad (76)$$

This yields

$$\delta q_B = \tfrac{1}{2} \sum_{B'} \left[F^{-1} \right]_{BB'} \mathrm{Tr} \left[(\boldsymbol{d}\boldsymbol{d}^t - \boldsymbol{C}) \left(\boldsymbol{C}^{-1} \boldsymbol{C}_{B'}^S \boldsymbol{C}^{-1} \right) \right] \quad (77)$$

for the iterative correction to the band powers. This amounts to making a quadratic approximation to the shape of the likelihood function around the maximum and iteratively approaching it. At each step, the total covariance matrix $\boldsymbol{C}$ must be updated using the new band powers $\boldsymbol{q} + \delta\boldsymbol{q}$. A convergence criterion based on the magnitude of the corrections δq_B will allow approach to the true $\{\hat{q}_B\}$ to be controlled.

The inverse of the Fisher matrix $[F^{-1}]_{BB'}$ evaluated at maximum likelihood is the covariance matrix of the parameters (Bond et al. 1998). The diagonals $[F^{-1}]_{BB}$ give an estimated Gaussian error bar for the derived band powers $\{\hat{q}_B\}$, although the full Fisher matrix must be used to take the (usually significant) band-band correlations into account. As the width of the l bins for the bands B is reduced, anticorrelation between adjacent bands increases as a result of the intrinsic l-space resolution of the data.

The presence of known or residual point-source foregrounds in equation (30) is dealt with either by fixing the amplitudes $q_{\rm src}$ or $q_{\rm res}$ and treating $q_{\rm src}\boldsymbol{C}^{\rm src}$ or $q_{\rm res}\boldsymbol{C}^{\rm res}$ as additions to the noise matrix $\boldsymbol{C}^N$, or by solving for the $q_{\rm src}$ or $q_{\rm res}$ and treating them as extra band powers q_B with associated entries in the Fisher matrix. In practice, for the CBI, it is necessary to hold fixed the $q_{\rm res}$ because the contribution from the source foreground with a white-noise power spectrum and appropriate frequency spectrum is largely indistinguishable from an overall offset of the CMB power spectrum. In addition, the uncertainties on the individual known source contributions to an aggregate $\boldsymbol{C}^{\rm src}$ will be substantial, and thus solving for a single amplitude $q_{\rm src}$ will not be as useful as it might appear. In this case, the $\boldsymbol{C}^{\rm src}$ acts as a constraint matrix and the $q_{\rm src}$ can be set to an arbitrarily high value, which will effectively *project* out the modes corresponding to the known sources by down-weighting the relevant combinations of the estimators in the likelihood (Bond et al. 1998; Bond & Crittenden 2001).

7.1. *Combination of Independent Data Sets*

Consider observations taken of separate sets of single fields or mosaics f where there is effectively no correlation between fields from separate f and the fields within a given set f are related by the mosaic covariance given in the previous sections. In this case, we can assemble a giant data vector

$$\boldsymbol{D} = (\, \boldsymbol{d}_1 \quad \ldots \quad \boldsymbol{d}_M \,)^t \quad (78)$$

from the M individual field vectors $\boldsymbol{d}_f$ (e.g., eq. [25]), with the block diagonal covariance matrix

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1 & & \\ & \ldots & \\ & & \boldsymbol{C}_M \end{pmatrix} , \quad (79)$$

which in turn can be written as sums of block-diagonal noise and signal covariance matrices $\boldsymbol{C}_f^N$ and $\boldsymbol{C}_{Bf}^S$, etc., with blocks given by $\boldsymbol{C}_f^N$ and $\boldsymbol{C}_{Bf}^S$, etc. Because they are block diagonal, we can write the log likelihood in equation (26) as the sum over data sets

$$\begin{aligned} \ln L &= -\ln(2\pi) \sum_f N_f - \tfrac{1}{2} \ln(\det \boldsymbol{C}) - \tfrac{1}{2} \boldsymbol{D}^t \boldsymbol{C}^{-1} \boldsymbol{D} \\ &= -\ln(2\pi) \sum_f N_f - \tfrac{1}{2} \sum_f \ln\left(\det \boldsymbol{C}_f\right) - \tfrac{1}{2} \sum_f \boldsymbol{d}_f^t \boldsymbol{C}_f^{-1} \boldsymbol{d}_f . \end{aligned} \quad (80)$$

We proceed as before, with the same band powers $\{q_B\}$ and with the block-diagonal band covariance matrix

$$\boldsymbol{C}_B^S = \frac{\partial \boldsymbol{C}}{\partial q_B} = \begin{pmatrix} \boldsymbol{C}_{B1}^S & & \\ & \ldots & \\ & & \boldsymbol{C}_{BM}^S \end{pmatrix} , \quad (81)$$

and thus all matrices are block-diagonal and composed of the individual single field or mosaic matrices. Therefore,

$$\begin{aligned} F_{BB'} &= \tfrac{1}{2} \sum_f \mathrm{Tr}\left(\boldsymbol{C}_f^{-1} \boldsymbol{C}_{Bf}^S \boldsymbol{C}_f^{-1} \boldsymbol{C}_{B'f}^S \right) , \\ \delta q_B &= \tfrac{1}{2} \sum_{B'} \left[F^{-1} \right]_{BB'} \sum_f \mathrm{Tr}\left[\left(\boldsymbol{d}_f \boldsymbol{d}_f^t - \boldsymbol{C}_f \right) \left(\boldsymbol{C}_f^{-1} \boldsymbol{C}_{B'f}^S \boldsymbol{C}_f^{-1} \right) \right] , \end{aligned} \quad (82)$$

which is used to iteratively approach the $\{\hat{q}_B\}$ using the Bond et al. (1998) scheme as in the single data set case.

7.2. *The Band Power Window Function*

To compare the band powers obtained from the data to model power spectra, we need to define a set of filter functions that project models $\mathcal{C}_l$ into a set of band powers C_B,

$$C_B = \sum_l \frac{W_l^B}{l} \mathcal{C}_l \,, \tag{83}$$

as in Bond et al. (1998). In the ensemble limit, the expectation value $\langle (\boldsymbol{x}\boldsymbol{x}^t - \boldsymbol{C}^N) \rangle$ will approach the underlying signal covariance matrix $\boldsymbol{C}^S$. We can then use the expression for the minimum variance estimate of the band power to derive the filter functions W_l^B (Knox 1999). Since

$$\langle q_B \rangle = \tfrac{1}{2} \sum_{B'} \left[F^{-1}\right]_{BB'} \mathrm{Tr}\left[\left(\boldsymbol{C}^{-1}\boldsymbol{C}_{B'}^S \boldsymbol{C}^{-1}\right)\boldsymbol{C}^S\right] \tag{84}$$

and

$$\boldsymbol{C}^S \equiv \sum_B \boldsymbol{C}_B^S = \sum_l \frac{\partial \boldsymbol{C}^S}{\partial \mathcal{C}_l} \mathcal{C}_l \,, \tag{85}$$

the normalized filter functions can be computed using the band-averaged Fisher matrix (e.g., eq. [76])

$$\frac{W_l^B}{l} = \frac{1}{2} \sum_{B'} \left[F^{-1}\right]_{BB'} \mathrm{Tr}\left[\left(\boldsymbol{C}^{-1}\boldsymbol{C}_{B'}^S \boldsymbol{C}^{-1}\right) \frac{\partial \boldsymbol{C}^S}{\partial \mathcal{C}_l}\right] \tag{86}$$

with respect to the $\mathcal{C}_l^{\rm shape} = 1$ that is built into the $\boldsymbol{C}^S$. Because of the χ_{Bl} used in the construction of the $\boldsymbol{C}_B^S$ in equation (43),

$$\sum_l \chi_{B'l} \frac{W_l^B}{l} = \delta_{BB'} \,, \tag{87}$$

and thus W_l^B/l is orthonormal with respect to the bands defined by χ_{Bl}.

Calculating the filter functions at each l becomes somewhat prohibitive in both processor time (the problem scales as an extra $N^3 + 2l_{\rm max}N^2$ operations) and storage since the calculation of equation (86) can only proceed once we have relaxed to the maximum likelihood solution. For this reason, in practice we sample the full filter functions in bands at intervals B_f where the B_f are narrower than the bands B, with

$$\frac{W_{B_f}^B}{l_{B_f}} = \frac{1}{2} \sum_{B'} \left[F^{-1}\right]_{BB'} \mathrm{Tr}\left[\left(\boldsymbol{C}^{-1}\boldsymbol{C}_{B'}^S \boldsymbol{C}^{-1}\right)\boldsymbol{C}_{Bf}^S\right] \,. \tag{88}$$

In principle, this is equivalent to assuming a flat window over the "fine" band B_f, and as long as the curvature of the exact window function is small over the intervals B_f, this should provide an adequate sampling of the continuous limit.

To obtain model band powers, we can then either interpolate the samples $W_{B_f}^B$ to obtain an approximate form for W_l^B for use in equation (83) or *preaverage* the model spectrum over the fine bands B_f as

$$C_B = \sum_{B_f} \left(\frac{W_{B_f}^B}{l_{B_f}}\right) \mathcal{C}_{B_f}^{(\rm flat)} \,, \tag{89}$$

where $\mathcal{C}_{B_f}^{(\rm flat)}$ are band powers calculated using flat filters ($W_l^{B_f} = 1$). We find that a fine band width $\Delta l_{B_f} \sim 20$ is sufficient to adequately sample the window functions and ensure normality and orthogonality to within 0.5% with respect to integration over the bands (e.g., eq. [87]). Example window functions calculated in this manner for mock deep fields and mosaics are shown in the bottom panels of Figures 2 and 3, respectively.

7.3. *Component Band Powers*

A further complication at the parameter end of the process is that the likelihood of the band powers cannot be assumed to be a Gaussian. This is especially so in cases in which the error in the band powers is sample or cosmic variance limited. Assuming the band powers to be Gaussian distributed can lead to the well-known problem of *cosmic bias* where the likelihood of low-power models can be overestimated and conversely that of high-power models can be underestimated. Bond et al. (2000) have shown how one can avoid this problem while still retaining Gaussianity in the χ^2 analysis by treating certain functions of the band powers as Gaussian distributed. Very good fits to the non-Gaussian distribution of the band powers can be obtained by use of the *offset lognormal* and *equal variance* approximations to the likelihood.

Both approximations use offsets x_B in the band powers that describe the contributions to the error in the band powers due to components other than the CMB. For the range of scales probed by instruments such as CBI these components will include the foregrounds such as point sources in addition to the usual noise "on the sky" offset $x_B^N \sim \sum_l \chi_{Bl} x_l$, where x_l is the offset due to the noise contribution to the error such that the quantity $Z_l = \ln(\mathcal{C}_l + x_l)$ has a normal distribution (Bond et al. 2000). For accurate parameter fits we therefore require estimates of band powers for all the components making up the total covariance $\boldsymbol{C}$. An approximation for these can be obtained by modifying the minimum variance estimator for the band powers q_B at the maximum likelihood

$$q_B^X = \tfrac{1}{2} \sum_{B'} [F_{BB'}]^{-1} \mathrm{Tr}\left[\left(\boldsymbol{C}^{-1}\boldsymbol{C}_{B'}^S \boldsymbol{C}^{-1}\right)\boldsymbol{C}^X\right] \,, \tag{90}$$

where we have substituted $\boldsymbol{C}^X$ in equation (77) for the observed measure for the signal covariance $(\boldsymbol{d}\boldsymbol{d}^t - \boldsymbol{C}^N)$. We then set $\boldsymbol{C}^X$ to the noise $\boldsymbol{C}^N$, foreground source $q_{\rm src}\boldsymbol{C}^{\rm src}$, or Gaussian residual foreground $q_{\rm res}\boldsymbol{C}^{\rm res}$ covariance components as desired (or use the maximum likelihood values $\hat{q}_{\rm src}$ and $\hat{q}_{\rm res}$ if these are included as parameters in the solution rather than being fixed). Examples of these are shown in Papers II and III for the deep field data and mosaic data, respectively. The offset to the lognormal is then obtained by summing the q_B^X over the components, $x_B \approx q_B^N + q_B^{\rm src} + q_B^{\rm res}$ (e.g., Bond & Crittenden 2001). This formalism is used in Paper V to approximate the shapes of the likelihood functions in order to derive limits on the cosmological parameters.

8. IMAGING FROM THE GRIDDED ESTIMATORS

Although not the primary goal of this method, an image can be constructed by Fourier transforming back to the sky plane using equation (1). If the estimators Δ_i are constructed on a regular lattice in $\boldsymbol{u}_i$ with spacing d_u and (u, v) extent $L\, d_u$, then the resulting image will have an extent on the sky

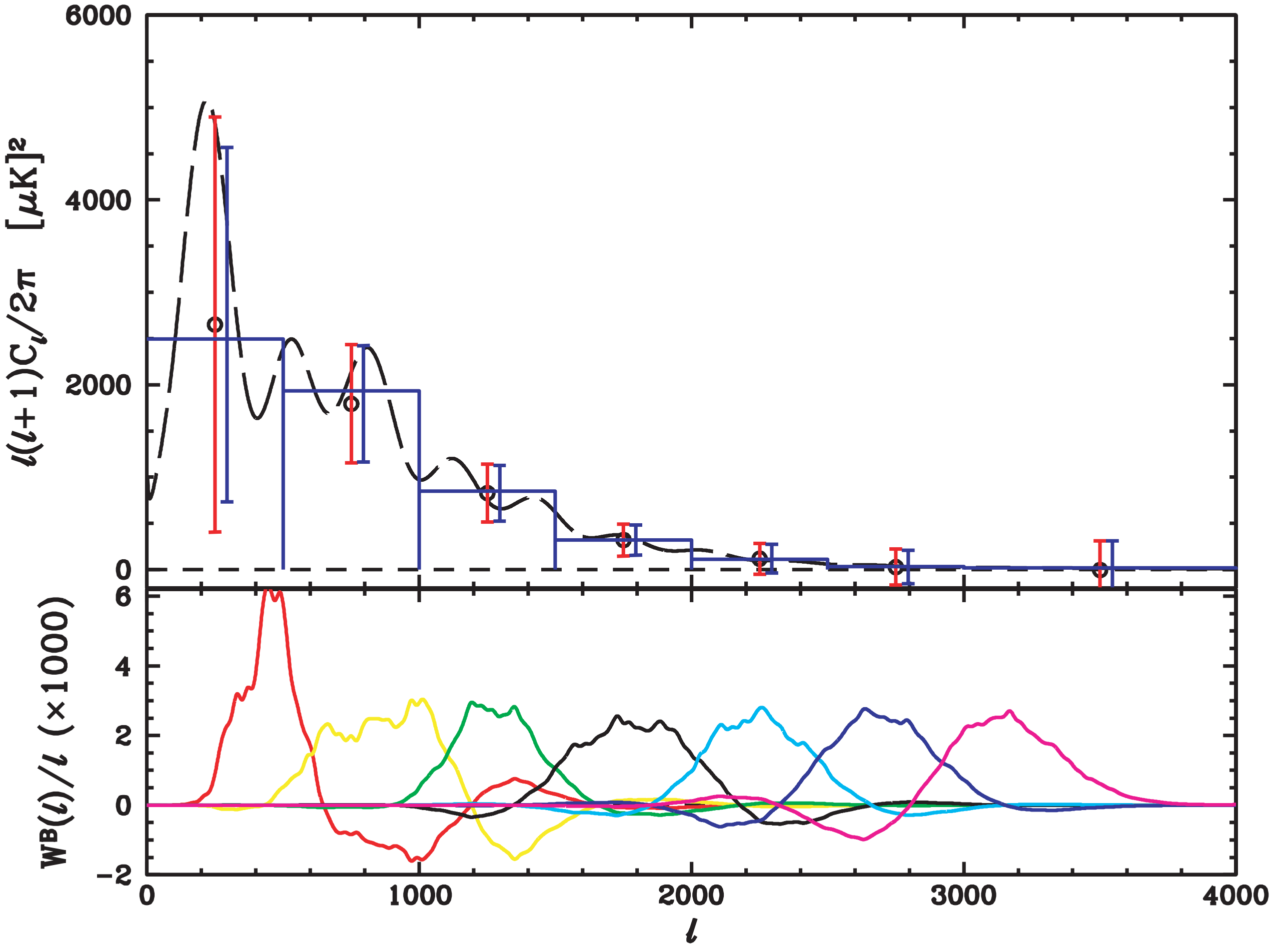

FIG. 2.—Results of the gridded method plus quadratic relaxation for 387 mock CBI 08^{h} deep field data sets (*top*), with each mock observation drawn from an independent realization of the sky given the model power spectrum (*dashed curve*) and the instrumental noise with the appropriate rms. The points (*black circles*) are placed at the band centers, at the mean of the reconstructed band powers with the red error bars given by the scatter among the realizations. The blue error bars to the right of the points show the average of the inverse Fisher matrix diagonals. The histograms show the width of each band and the level expected by integrating the model $\mathscr{C}_l$ over the window functions W_l^B, which are shown in the bottom panel. The mean of the realizations converges to the expected value within the Poisson uncertainty, taking into account the correlations between adjacent bands.

given by the inverse of the spacing d_u^{-1} and a resolution given by d_u^{-1}/L. In the continuum limit (see Appendix A), we can define an estimator $\hat{T}(\boldsymbol{x})$ for the temperature field $T(\boldsymbol{x})$,

$$\hat{T}(\boldsymbol{x}) = \int d^2\boldsymbol{u}\, \Delta(\boldsymbol{u}) e^{2\pi i \boldsymbol{u}\cdot\boldsymbol{x}} \;, \tag{91}$$

where $\Delta(\boldsymbol{u})$ is the continuous functional form (e.g., eq. [A2]) for the estimators, with $\Delta_i = \Delta(\boldsymbol{u}_i)$. In practice, the lattice of estimators Δ_i is embedded in a wider grid padded with zero in the unsampled cells, and a fast Fourier transform is carried out.

For our standard gridding normalization $z_i = z_i^{(1)}$ given in equation (A21), the units of $\boldsymbol{\Delta}$ will be flux density units (Jy), and thus its inverse Fourier transform will produce a map in units of Jy beam^{-1}, where the beam area is given by the PSF of the image. For a single field, the PSF is just the image generated using equation (91) using estimators $\Delta_i^{\rm PSF}$ computed by introducing unit "visibilities" into equation (21),

$$\Delta_i^{\rm PSF} = \sum_k \{Q_{ik} + \bar{Q}_{ik}\} \;. \tag{92}$$

The situation for the mosaics is somewhat more complicated, as the mosaic offsets must be taken into account in constructing equivalent visibilities for point sources

$$\Delta_i^{\rm PSF}(\hat{\boldsymbol{x}}) = \sum_k \{Q_{ik} V_k^{\rm PSF}(\hat{\boldsymbol{x}}) + \bar{Q}_{ik} V_k^{*\rm PSF}(\hat{\boldsymbol{x}})\} \;,$$
$$V_k^{\rm PSF}(\hat{\boldsymbol{x}}) = A_k(\hat{\boldsymbol{x}} - \boldsymbol{x}_k) e^{-2\pi i \boldsymbol{u}_k\cdot(\hat{\boldsymbol{x}}-\boldsymbol{x}_k)} \tag{93}$$

obtained by evaluating equation (11) with no noise and $I(\boldsymbol{x}) = \delta^2(\boldsymbol{x} - \hat{\boldsymbol{x}})$. In this case one would evaluate the PSF at various positions $\hat{\boldsymbol{x}}$ in the map.

Because our estimators use the kernel $\boldsymbol{Q}$ as given in equation (22), which includes the beam transform $\tilde{A}$, we are effectively multiplying the image on the sky by the primary beam squared: once in the kernel, and once due to the instrument itself (e.g., eq. [12]). Images made directly from the $\boldsymbol{\Delta}$ will therefore be strongly attenuated in the (noisy) outskirts.

As mentioned in Appendix A, the optimal weighting for the imaging of the CMB component is to use the Planck factor in equation (6) to correct for the thermal frequency spectrum (e.g., eq. [A20]), while our standard intensity weighting given in equation (A19) is optimized for a flat nonthermal power-law spectrum with spectral index $\alpha = 0$.

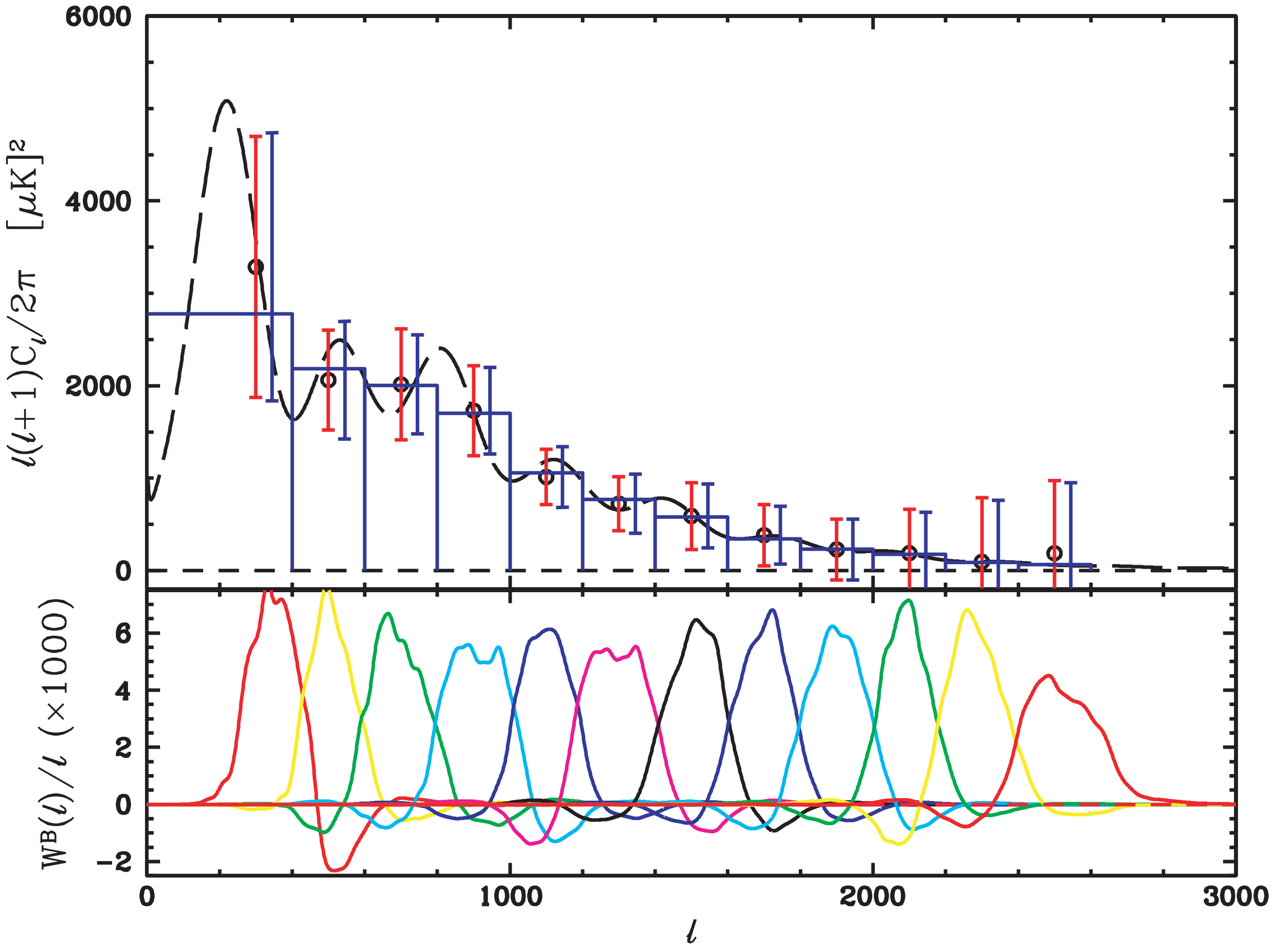

Fig. 3.—*Top:* Results of the gridded method plus quadratic relaxation for 117 mock CBI (7 × 6 field) mosaic data sets with a CDM-based power spectrum (*dashed curve*). As in Fig. 2, the points (*black circles*) are centered in each bin in l with red error bars giving rms scatter of about the mean for the band powers from the processed realizations, with the blue error bar from average inverse Fisher diagonals to the right of the points. The window functions are plotted in the bottom panel.

We can also filter the gridded estimators in such a way as to enhance or down-weight certain signals or noise. We can do this with optimal or Wiener filtering (e.g., Bond & Crittenden 2001),

$$\boldsymbol{\Delta}^{\phi} = \Phi\boldsymbol{\Delta} , \tag{94}$$

where the choice of the filter Φ depends on the application.

For example, the covariances $\boldsymbol{C}^X$ calculated from equation (32) can be used to construct optimal filters for each component contributing to the observations. For the signal component described by the covariances $\boldsymbol{C}^X$ we can construct a Wiener filter to be applied to the gridded (u, v) estimators

$$\boldsymbol{\Delta}^X = \boldsymbol{C}^X\boldsymbol{C}^{-1}\boldsymbol{\Delta} . \tag{95}$$

The amplitudes for the signal models such as the band powers q_B or the source amplitudes $q_{\rm src}$ can be set to their maximum likelihood values or to fiducial model amplitudes. The Wiener-filtered image is then recovered by Fourier transforming.

Examples of images created using this method are described in the next section and shown in Figure 6 below. Wiener-filtered images are also used in Paper VI to explore the possibility of detection of the Sunyaev-Zeldovich effect in the data at high l.

9. IMPLEMENTING THE METHOD

The algorithm described above was coded as a scalar FORTRAN (f77 compatible) program designated CBIGRIDR, with a parallelized FORTRAN 90 version using OpenMP[4] directives also available for use on multiprocessor machines. In addition, the Bond et al. (1998) likelihood relaxation was coded in a second parallelized FORTRAN 90 program called MLIKELY using parallel versions of LAPACK matrix algebra routines (e.g., § 7). Together, these two programs make up the CBI analysis pipeline. This pipeline has undergone numerous tests and development since its inception in 2001 April and has been used to produce the power spectra and to provide the band powers as input to the cosmological parameter analysis given in the companion papers. We now give a brief description of our implementation.

[4] Available at http://www.openmp.org.

In order to carry out the numerical integrations, a fine-grain rectangular lattice in (u, v)-space was used. The fine-grain grid size $\Delta u_{\rm fine}$ (in units of wavelength) was chosen to adequately sample the phase turns in the (u, v)-plane as a result of the mosaic size and differencing; for a standard 7×6 CBI mosaic with $20'$ spacing, the maximum field separation along the grid direction is $x_{\rm max} = 2°$, which gives oscillations in the (u, v)-plane with wavelength of $x_{\rm max}^{-1} = 28.65$, giving $\Delta u_{\rm fine} \leq 14.3$ for two samples per cycle. To evaluate the projection operators $R_i(\boldsymbol{v})$ (e.g., eq. [23]), we store a small fine-grain lattice around each $\boldsymbol{u}_i$. The maximum radius in (u, v)-space needed for the support of $\boldsymbol{R}$ is $r_u = 2D/\lambda_{\rm min}$. For CBI $D = 90$ cm, and $\lambda_{\rm min} = 0.844$ cm at 35.5 GHz, we get $r_u = 213.1$. A grid of size 53×53 cells with $\Delta u_{\rm fine} = 8.526$ will fit both the sampling and radius requirements.

The estimators are evaluated on a coarse-grain lattice of $\boldsymbol{u}_i$, with a spacing of $\Delta u_{\rm coarse}$. The fine-grain lattices on which we accumulate the $R_i(\boldsymbol{v})$ will be cross-correlated to form the covariance elements $\boldsymbol{M}_B$ and $\bar{\boldsymbol{M}}_B$ (e.g., eq. [37]); it is desirable to have the coarse-grid size locked to integer multiples of the fine-grid cells. This coarse grid does not have to sample the highest mosaic frequencies, but only the effective width of $\boldsymbol{R}$. Tests were carried out using the mock data (see below) using different fine-grain cell sizes and coarse-grain spacings, looking for changes in the derived band powers as these were varied. We find that for single CBI fields, $\Delta u_{\rm coarse} = 3\Delta u_{\rm fine}$ is adequate. For CBI mosaics, a hybrid lattice with $\Delta u_{\rm coarse} = \Delta u_{\rm fine}$ in the inner part ($l < 800$) and $\Delta u_{\rm coarse} = 2\Delta u_{\rm fine}$ in the outer part was found to work well.

As stated in § 2, the choice of sign of the exponential of the Fourier transform in equation (1) is a convention. This choice varies throughout the literature on the subject, but in practice it depends on the way the baseline vectors are defined in the data and how the correlation products are made (e.g., which antenna gets the quadrature phase shift). We note that in coding our algorithm to conform to the imaging standards of the AIPS[5] and DIFMAP[6] (Shepherd 1997) packages using the CBI data, we had to use the opposite sign convention from the one presented in equation (1).

To process a data set, a spectral weighting $f(\nu)$ and shape function $C_l^{\rm shape}$ are chosen. The visibilities V_k are looped over, and any source subtraction (§ 6.3) is applied. For each estimator i that V_k contributes to either directly or as a conjugate, its contribution to the fine-grain lattice q for $R_i(\boldsymbol{v}_q)$ is accumulated, e.g.,

$$R_i(\boldsymbol{v}_q) = \sum_k \left\{ Q_{ik} P_k(\boldsymbol{v}_q) + \bar{Q}_{ik} \bar{P}_k(\boldsymbol{v}_q) \right\}, \quad \boldsymbol{v}_q = \boldsymbol{u}_i + \Delta u_{\rm fine} \hat{\boldsymbol{v}}_q \ , \tag{96}$$

where $\hat{\boldsymbol{v}}_q$ is a 53×53 unit (fine-grain) lattice. This means that for $n_{\rm est}$ estimators, the storage required for $\boldsymbol{R}$ is only $2809 n_{\rm est}$ double-precision complex numbers. If the data were differenced (as for CBI data), then $P_k^{\rm sw}(\boldsymbol{v})$ from equation (70) is used. The contributions to the noise covariance elements $\boldsymbol{M}^N$ and $\bar{\boldsymbol{M}}^N$ (eq. [35]) and the $\boldsymbol{\Delta}_c^{\rm src}$ (eq. [54]) are also accumulated at this time. Finally, this visibility's contributions are added to estimator Δ_i and normalization z_i.

[5] See http://www.cv.nrao.edu/aips.
[6] See ftp://ftp.astro.caltech.edu/pub/difmap.

The storage for $\boldsymbol{R}$, $\boldsymbol{M}^N$, and $\bar{\boldsymbol{M}}^N$ dominates the memory requirements. For example, the CBI mosaics use around $n_{\rm est} = 2500$ estimators, and thus storage for $53 \times 53 \times 2500 \approx 7 \times 10^6$ double-precision complex numbers is needed. A single packed $n_{\rm est}^2 \approx 6 \times 10^6$ array is needed to hold $\boldsymbol{M}^N$ and $\bar{\boldsymbol{M}}^N$. The $\boldsymbol{C}^X$ matrices are calculated in place and written out row by row, and thus they need not be stored. There are no instances where matrices of dimension $N_{\rm vis}^2$ are stored; the storage for a matrix of this size would be prohibitive as our largest CBI mosaics have $N_{\rm vis} > 2 \times 10^5$.

When all the visibilities have been processed, the estimators are normalized by z_i, split into real and imaginary parts, and written out to disk. The covariance matrices C_{ij}^N, C_{Bij}^S, and any $C_{ij}^{\rm src}$ are constructed by looping over pairs of rows corresponding to the real and imaginary parts of each estimator, e.g., rows i and $i + N_{\rm est}$ for estimator i. For each $j \leq i$, the stored R_i and R_j are cross-correlated along with the shape function $\mathscr{C}(|\boldsymbol{v}|)$ to form the band power covariance elements M_{Bij} and $\bar{M}_{Bij}$ of equation (37), combined to make C_{Bij}^S using equation (32), and stored. The relevant columns of these rows of C_{ij}^N are formed from the stored M_{ij}^N and $\bar{M}_{ij}^N$. At this point, for each $\boldsymbol{C}^{\rm src}$ desired (there may be more than one; in the CBI analysis we use three), the relevant $\Delta_c^{\rm src}$ are combined using equation (55) to form $M_{ij}^{\rm src}$ and $\bar{M}_{ij}^N$, which in turn are used to make $C_{ij}^{\rm src}$. After all columns for these rows of the covariance matrices are computed, they are written to disk, and this process is repeated for the pair of rows corresponding to the next estimator i. When all rows are complete, the output file is complete. Note that different binnings of the $\boldsymbol{C}_B^S$ can be run without regridding using the original $\boldsymbol{R}$, saving significant time.

If a residual foreground covariance matrix $\boldsymbol{C}^{\rm res}$ is desired, the procedure outlined above is repeated in its entirety using the description in § 6.4. The spectrum and shape appropriate for the source population or foreground emission are used during the gridding and covariance matrix construction. Other than these factors, the same gridding as in the CMB and noise estimators must be used.

The output files from CBIGRIDR are then used as input to MLIKELY. These can be for single fields or mosaics, or for combinations of independent fields or mosaics (§ 7.1). At this point, the prefactors $q_{\rm src}$ and $q_{\rm res}$ for any $\boldsymbol{C}^{\rm src}$ and $\boldsymbol{C}^{\rm res}$ covariance matrices are chosen and fixed. Relaxation to the likelihood maximum is carried out as described in § 7, and the resulting band powers $\{q_B\}$ and inverse Fisher matrix elements $[F^{-1}]_{BB'}$ are written out. If desired, the band power window functions $W_{B_f}^B$ (§ 7.2) can be computed if CBIGRIDR was run to produce narrow-bin $\boldsymbol{C}_{B_f}^S$. The component band powers q_B^N, $q_B^{\rm src}$, and $q_B^{\rm res}$ (§ 7.3) can also be computed at this time.

Finally, filtered images using the formalism of § 8 can be computed from the estimators, the $\boldsymbol{C}$ (at maximum likelihood), and the component covariance matrices. Results from this are shown below and in Paper VI.

The timing for CBIGRIDR depends on the degree of parallelization, processor speed on a given machine, number of visibilities gridded, number of foreground sources, and number of bins B for the band powers. As an example, the processing of the $14^{\rm h}$ mosaic field of Paper III (the largest of the data sets) involved gridding 228,819 visibilities from 65 separate nights of data in 41 fields to 2352 complex estimators. A total of 916 sources were gridded into three source covariance matrices. A total of seven different binnings for

$\boldsymbol{C}_B^S$ were run at this time from the same gridding. The execution time using the parallel version of CBIGRIDR was $2^{\rm h}40^{\rm m}$ running on 22 processors on a 32 processor Alpha GS320 workstation at the Canadian Institute for Theoretical Astrophysics (CITA). It then took $3^{\rm h}22^{\rm m}$ on the same computer for MLIKELY to process 4704 double-precision real estimators in 16 $\boldsymbol{C}_B^S$ bands, with three $\boldsymbol{C}^{\rm src}$ matrices, one $\boldsymbol{C}^{\rm res}$, and one $\boldsymbol{C}^N$. This included the time needed to calculate the component band powers $\boldsymbol{C}_B^X$, but not the window functions. The speed of this fast gridded method has allowed us to carry out numerous tests on both real and simulated data sets, which would not have been possible carrying out maximum likelihood (e.g., using even the optimized MLIKELY) on the 200,000 plus visibilities.

9.1. *Method Performance Tests Using Mock CBI Data*

The performance of the method was assessed by applying it to mock CBI data sets. Simulated CBI data sets were obtained by replacing the actual visibilities from the data files containing real CBI observations of the various fields used in Papers II and III with the response expected for a realization of the CMB sky drawn from a representative power spectrum, plus uncorrelated Gaussian instrumental noise with the same variance as given by the scatter in the actual CBI visibilities. The differencing of the lead and trail fields used in CBI observations was included (e.g., § 6.6). This mock data set had the same (u, v) distribution as the real data and gives an accurate demonstration of expected sensitivity levels and the effect of cosmic variance. The power spectrum chosen for these simulations was for a model that fitted the *COBE* and BOOMERANG data (Netterfield et al. 2002).

Figure 2 shows the power spectrum estimation derived following the procedure detailed above. The mock data sets were drawn as realizations for the $08^{\rm h}$ CBI deep field from Paper II. The binning of the signal covariance matrix $\boldsymbol{C}_B^S$ was chosen to be uniform in l with bin width $\Delta l = 500$. Because a single realization of the sky drawn from the model power spectrum will have individual mode powers that deviate from the mean given by the power spectrum as a result of this intrinsic so-called cosmic variance plus the effect of the thermal instrumental noise, we analyze 387 realizations, each taken from a different realization of the sky and a different set of instrumental noise deviates. The mean q_B for each band B converge to $\langle C_B \rangle$, which is obtained by integrating the model $\mathscr{C}_l$ over the window functions W_l^B (e.g., eq. [89]), within the sample uncertainty for the realizations. Furthermore, the standard deviation of the q_B from the mean for each band agrees with the value obtained from the diagonals of the inverse of the Fisher matrix.

The choice of the l bin size is driven by the trade-off between the desired narrow bands for localizing features in the power spectrum and the correlations between bins introduced by the transform of the primary beam. There is an anticorrelation between adjacent bands seen in $[F^{-1}]_{BB'}$ at the level of -13% to -23% for $\Delta l = 500$ with a single field. We have found that correlations up to about -25% give plots of the q_B that are more visually appealing than those made with narrower band and higher correlation levels as a result of the increasing scatter in the band powers about the mean values. Bins of this size do not achieve the best possible l resolution, and thus our cosmological parameter runs use finer binned bands since the correlations are taken into account in the analyses.

The band window functions W_l^B are shown in the bottom panel of Figure 2 and were computed using narrow binnings $W_{B_f}^B$ (e.g., eq. [88]) with $\Delta l = 20$. The small-scale structures seen in the window functions, particularly visible around the peaks, are due to the differencing that introduces oscillations (see § 6.6). As shown in equation (87), a window function W_l^B is normalized to sum to unity within the given band B and to sum to zero in the other bands, and thus there must be compensatory positive and negative "sidelobes" of the window function outside the band.

Figure 3 shows the power spectrum derived for a simulated mosaic of 7×6 fields separated by $20'$ using the actual CBI $20^{\rm h}$ mosaic fields from Paper III as a template. This mosaic field was chosen as it had incomplete mosaic coverage and thus would be the most difficult test for the method. The binning for $\boldsymbol{C}_B^S$ shown used $\Delta l = 200$, which gave adjacent band anticorrelations of -13% to -18% in the $F_{BB'}$. Again the mean of the 117 realizations converges to the value expected within the error bars, showing that there is no bias introduced by the method, even in the presence of substantial holes in the mosaic (see Paper III for the mosaic weight map). Furthermore, the rms scatter in the realizations converges to the mean of the inverse Fisher error bars, as in the single-field case. As in the previous figures, the band power window functions are shown in the lower panels.

In Figure 4 are shown three randomly chosen realizations from the ensemble, plotted along with the input power spectrum. This shows the level of field-to-field variations that we might expect to see in CBI data. There are noticeable deviations from the expected band powers in individual realizations, particularly at low l where cosmic variance and the highly correlated bins conspire to increase the scatter. These differences are within the expected scatter when bin-bin correlations and limited sample size are taken into account, but care must be exercised in interpreting single-field power spectra. In particular, the acoustic peak structures are obscured by the sample variations. However, the average band powers for the three runs (shown in Fig. 4 as open black circles) are better representations of the underlying power spectrum. Although this is not a proper "joint" maximum likelihood solution (e.g., § 7.1) as is done for the real CBI mosaic fields, the improvement seen using the three-field average leads us to expect that the combination of even three mosaic fields damps the single-field variations sufficiently to begin to see the oscillatory features in the CMB power spectrum. While we do not show the equivalent plots of the deep fields from Figure 2, the same behavior is seen (with even larger field-to-field fluctuations in the relatively unconstrained first bin, although still consistent with the error bars).

The effect of adding point sources to the mock fields and then attempting power spectrum extraction is shown in Figure 5. A set of 200 realizations were made in the same manner as in the runs in Figure 2, but the list of point-source positions, flux densities and uncertainties, and spectral indices from lower frequency used in the analysis in Paper II (the "NVSS" sources) was used to add mock sources to the data. The flux densities of the sources actually added to the data were perturbed using the stated uncertainties as 1 σ standard deviations. The errors used were 33% of the flux density except for a few of the brighter sources that were put

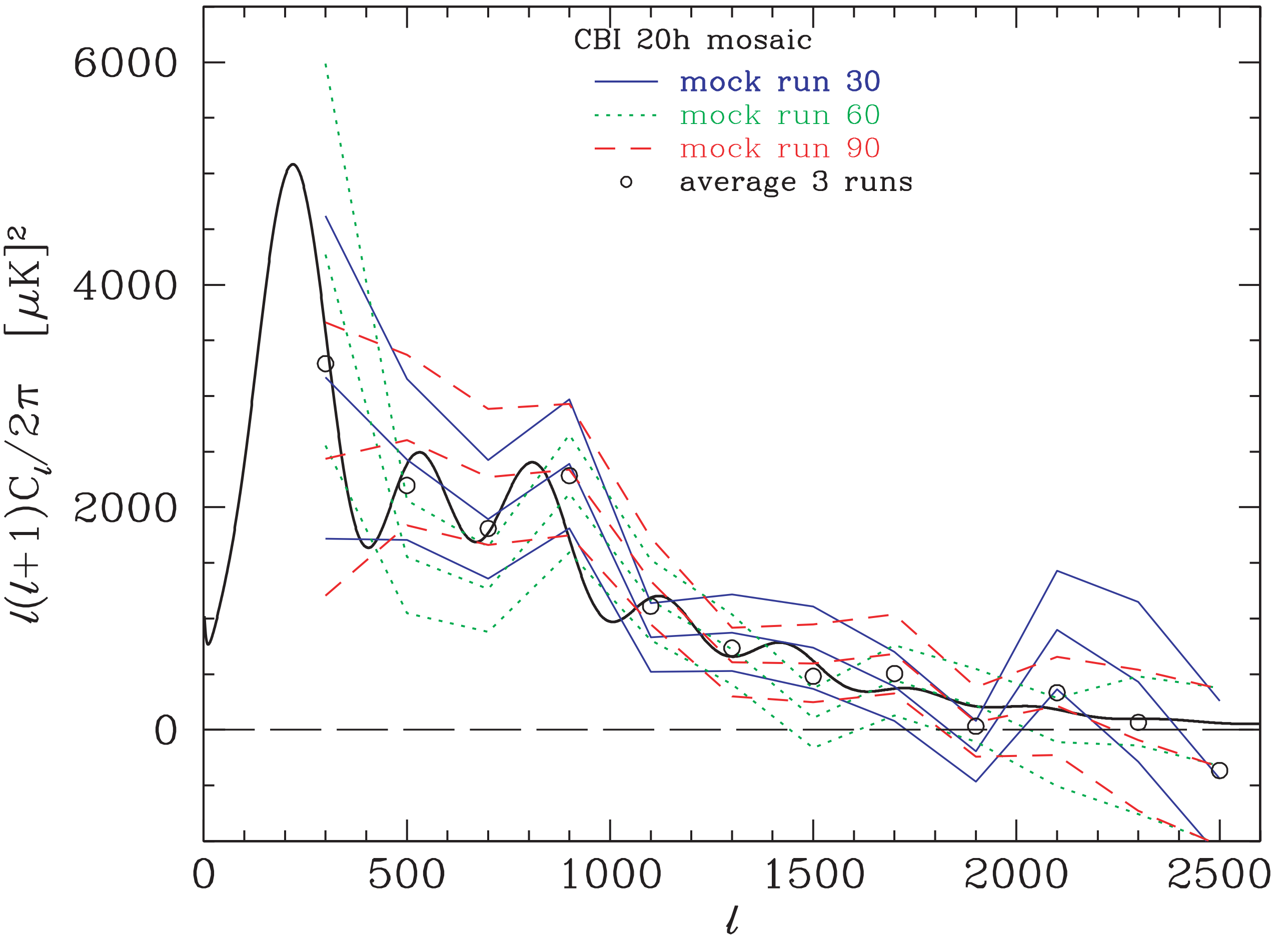

FIG. 4.—Three randomly selected realizations from the mosaic field simulations shown in Fig. 3 plotted against the model power spectrum (*solid black curve*). The three lines for the band powers reconstructed from the three realizations correspond to the band powers (*central lines*) and the $\pm 1\ \sigma$ excursions using the inverse Fisher error bars. The scatter in band powers between realizations is within the expected range. Also shown is the unweighted scalar average of the band powers for the three realizations, which is an approximation to a true joint likelihood solution. The average is a better fit to the model, as is expected.

in with 100% uncertainties. We then used the methodology described in § 6.3 to compute the constraint matrices. The first method of correction used was to subtract the (unperturbed) flux densities from the visibilities and build the $\boldsymbol{C}^{\rm src}$ from $\boldsymbol{\Delta}^{\rm src}$ built using the uncertainties (shown as the red triangles). In addition, we also did no subtraction but built $\boldsymbol{C}^{\rm src}$ from $\boldsymbol{\Delta}^{\rm src}$ using the full (unperturbed) flux densities (shown as the blue squares). This is equivalent to assuming a 100% error on the source flux densities and thus canceling the average source power in those modes. In both cases the factor $q_{\rm src} = 1$ was used. The simulations show that both methods are effective, with no discernible bias in the reconstructed CMB band powers.

Finally, the production of images using the gridded estimators described in § 8 is demonstrated in Figure 6. The series of plots show the effect of Wiener filtering using the noise and various signal components on an image derived from one of the mock $08^{\rm h}$ CBI deep field realizations with sources from the ensemble shown in Figure 5. The Planck factor weighting of equation (A20) was used during gridding to optimize for the thermal CMB spectrum, although in practice this makes little difference as a result of the restricted frequency range of the CBI. The estimators for this realization were computed by subtracting the mean values of the source flux densities and putting the standard deviations into $\boldsymbol{C}^{\rm src}$ with $q_{\rm src} = 1$ (the red triangles in Fig. 5). The filtering down-weights the high spatial frequency noise seen in the unfiltered image and effectively separates the CMB and source components as shown by comparing the bottom panels of Figure 6 to the total signal in the top right-hand panel. The signal in this realization is dominated by the residuals from two bright point sources that had 100% uncertainties put in for their flux densities and thus escaped subtraction. The effectiveness of $\boldsymbol{C}^{\rm src}$ in picking out the sources in the image plane underlines its utility as a constraint matrix in the power spectrum estimation.

10. CONCLUSIONS

We have outlined a maximum likelihood approach to determining the power spectrum of fluctuations from interferometric CMB data. This fast gridded method is able to handle the large amounts of data produced in large mosaics such as those observed by the CBI. Software encoding this algorithm was written and tested using mock CBI data drawn from a realistic power spectrum. The results of the code were shown to converge as expected to the input power spectrum with no discernible bias. For small data sets, this

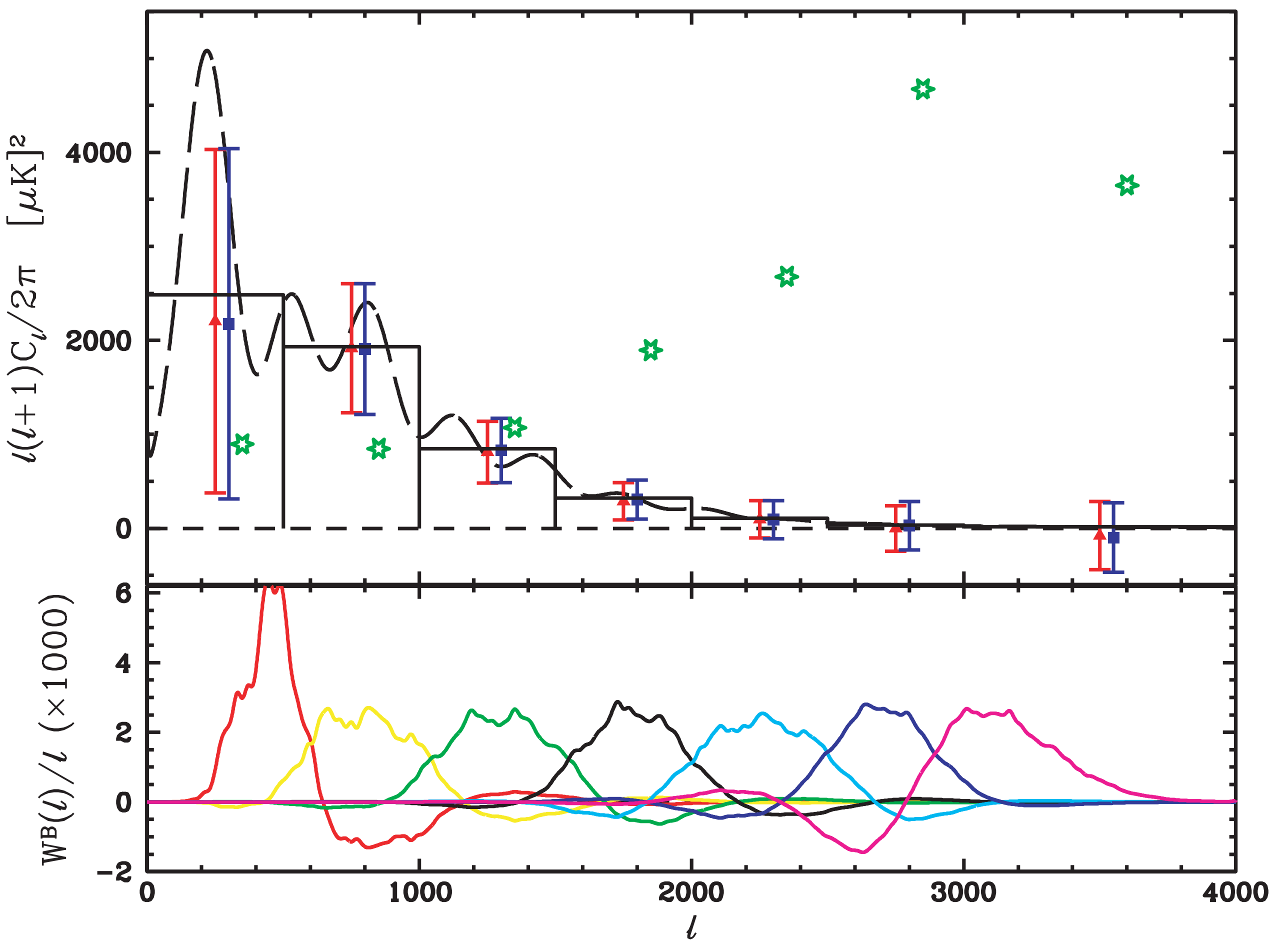

Fig. 5.—Results using mock deep field data sets including foreground point sources based on the actual list used in the CBI data (*top*) along with the window functions (*bottom*). The input power spectrum and expected band powers are as in Fig. 2. The green stars show the average for 200 realizations where no source subtraction or projection was done, with the powers divided by a factor of 2 to fit on the plot. The expected increase with l^2 is seen, along with a falloff in the last bin due to the source frequency spectrum. The points with error bars at the center of each bin (*red triangles*) were computed from 200 realizations processed with subtraction of the mean flux density from the visibilities and construction of $C^{\rm src}$ using the uncertainties, while the points with error bars to the right of these (*blue squares*) are from 200 realizations where no source subtraction was done, but we built $C^{\rm src}$ using the full (unperturbed) flux densities that projects out the sources with only a slight increase in noise. Despite the large power from sources at high l, our method successfully removes the foreground power from the spectrum with no sign of bias.

code was also tested against independently written software that worked directly on the visibilities. In addition, the pipeline was run with gridding turned off as described in § 6, again for small test data sets. No bias or significant loss in sensitivity was seen in these comparisons.

This software pipeline was used to analyze the actual CBI data, producing the power spectra presented for the deep fields and mosaics in Papers II and III, respectively. The output of the pipeline also was used as the input for the cosmological parameter analysis in Paper V and the investigation of the Sunyaev-Zeldovich effect in Paper VI.

This method is of interest for carrying out power spectrum estimation for interferometer experiments that produce a large number of visibilities but with a significantly smaller number of independent samples of the Fourier plane (such as close-packed arrays like VSA or DASI). The CBI pipeline analysis is carried out in two parts, the gridding and covariance matrix construction from input uv-FITS files in CBIGRIDR and the maximum likelihood estimation of band powers using quadratic relaxation in MLIKELY. The software for the pipeline is available by contacting the authors.

We close by noting that our formalism can be extended to deal with polarization data. In the case of CMB polarization, there are as many as six different signal covariance matrices of interest in each band, with estimators (or visibilities) for parallel-hand and cross-hand polarization products, and thus development of a fast method such as this is critical. In 2001 September polarization-capable versions of CBIGRIDR and MLIKELY were written and tested. We describe the method, the polarization pipeline, and results in an upcoming paper (S. T. Myers et al. 2003, in preparation).

S. T. M. was supported during the early years of the CBI by an Alfred P. Sloan Fellowship from 1996 to 1999 while at the University of Pennsylvania. Genesis of this method by S. T. M. greatly benefited by a stay in 2000 July at the ITP in Santa Barbara, supported in part by the National Science Foundation under grant PHY 99-07949. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The CBI was funded under NSF grants AST 94-13934, AST 98-02989, and

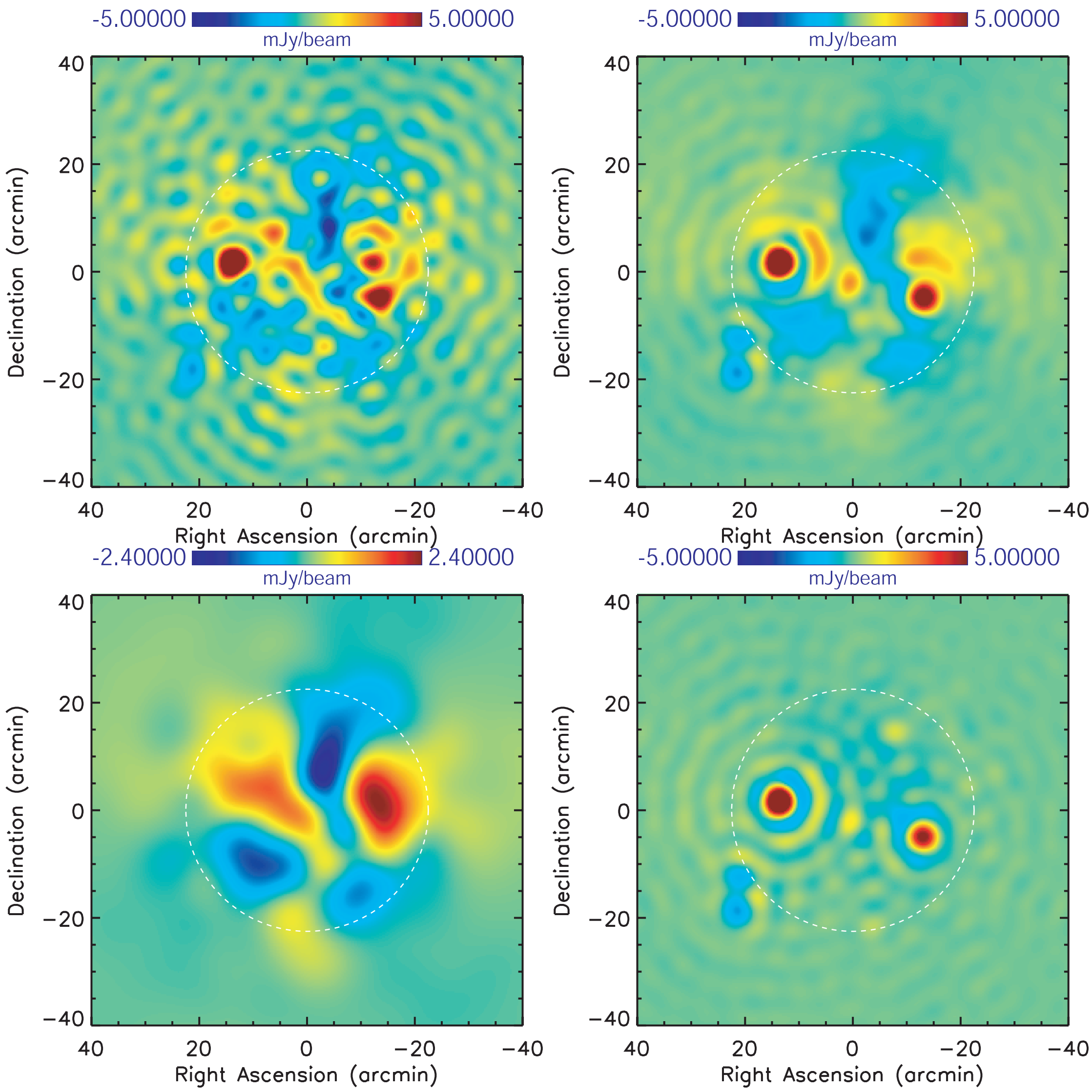

FIG. 6.—Images reconstructed from the gridded estimators using the formalism of § 8. Data are for one of the mock 08[h] deep field realizations with sources used in Fig. 5. *Top left:* Image computed without any filtering. *Top right:* Image derived by setting $\boldsymbol{C}^X = \boldsymbol{C}^{\rm src} + \sum_B q_B \boldsymbol{C}_B^S$ the sum of the signal terms. *Bottom left:* Image using $\boldsymbol{C}^X = \sum_B q_B \boldsymbol{C}_B^S$ for the CMB only. *Bottom right:* Image for $\boldsymbol{C}^X = \boldsymbol{C}^{\rm src}$ to pick out the point sources. The filter clearly dampens the noise and separates the CMB and source components. The residuals from several bright sources dominate the signal in all but the CMB-filtered image (the brightness scale in that image is enhanced). The white dashed circle shows the 45!2 FWHM of the CBI at 31 GHz. The attenuation of the signal brightness due to the square of the primary beam is clearly seen.

AST 00-98734, with contributions by Maxine and Ronald Linde and Cecil and Sally Drinkward and the strong support of the California Institute of Technology, without which this project would not have been possible. In addition, this project has benefited greatly from the computing facilities available at CITA and from discussions with other members of the group at CITA not represented as authors on this paper.

APPENDIX A

FORM OF THE LINEAR ESTIMATOR

Suppose we were to construct a simple linear "dirty" mosaic on the sky obtained by a linear combination of the dirty (not deconvolved) images of the individual fields (e.g., Cornwell et al. 1993). In the (u, v)-plane, this reduces to summing (integrating) up the visibilities from each mosaic "tile" with some weighting function, e.g.,

$$\Delta_i = \sum_k Q_{ik} V_k \;, \tag{A1}$$

where for the time being we ignore the contribution from the complex conjugates of the visibilities (see below). For illustrative purposes, let us consider only a single frequency channel and write the estimator as a function $\Delta(\boldsymbol{u})$, where $\Delta_i = \Delta(\boldsymbol{u}_i)$, which in the absence of instrumental noise is given by

$$\Delta(\boldsymbol{u}) = \int d^2\boldsymbol{v} \int d^2\boldsymbol{x}\, \mathscr{F}(\boldsymbol{x}, \boldsymbol{v}) \mathscr{Q}(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{v}) \langle V(\boldsymbol{x}, \boldsymbol{v}) \rangle \ , \tag{A2}$$

with kernel $\mathscr{Q}$, sky and aperture plane sampling given by $\mathscr{F}$, and where

$$V(\boldsymbol{x}, \boldsymbol{v}) = \int d^2\boldsymbol{v}'\, \tilde{I}(\boldsymbol{v}')\tilde{A}(\boldsymbol{v} - \boldsymbol{v}')e^{2\pi i \boldsymbol{v}' \cdot \boldsymbol{x}} \tag{A3}$$

is the visibility at pointing position $\boldsymbol{x}$ and (u, v) locus $\boldsymbol{v}$ from equation (11). In practice, the sampling function is just a series of delta functions

$$\mathscr{F}(\boldsymbol{x}, \boldsymbol{v}) = \sum_k \omega_k \delta^2(\boldsymbol{x} - \boldsymbol{x}_k)\delta^2(\boldsymbol{v} - \boldsymbol{u}_k) \tag{A4}$$

over the measured visibilities $k = 1, \ldots, N_{\rm vis}$ each with weight ω_k.

As an *Ansatz*, we let the mosaicking kernel have the form

$$\mathscr{Q}(\boldsymbol{u}, \boldsymbol{x}, \boldsymbol{v}) = Q(\boldsymbol{v} - \boldsymbol{u})e^{-2\pi i \boldsymbol{u} \cdot \boldsymbol{x}} \ , \tag{A5}$$

where Q is the interpolating kernel. Furthermore, let us assume that the (u, v)-plane coverage is the same for all mosaic pointings, and thus $\mathscr{F}(\boldsymbol{x}, \boldsymbol{v})$ is separable,

$$\mathscr{F}(\boldsymbol{x}, \boldsymbol{v}) \equiv F(\boldsymbol{v})G(\boldsymbol{x}) \ , \tag{A6}$$

where $F(\boldsymbol{v})$ and $G(\boldsymbol{x})$ are the sampling and weighting in the two domains. Combining these and rearranging terms, we get

$$\Delta(\boldsymbol{u}) = \int d^2\boldsymbol{v}'\, \tilde{I}(\boldsymbol{v}') \int d^2\boldsymbol{v}\, F(\boldsymbol{v})Q(\boldsymbol{v} - \boldsymbol{u})\tilde{A}(\boldsymbol{v} - \boldsymbol{v}') \int d^2\boldsymbol{x}\, G(\boldsymbol{x})e^{-2\pi i(\boldsymbol{u} - \boldsymbol{v}') \cdot \boldsymbol{x}} \tag{A7}$$

$$= \int d^2\boldsymbol{v}'\, \tilde{I}(\boldsymbol{v}')\tilde{G}(\boldsymbol{u} - \boldsymbol{v}') \int d^2\boldsymbol{v}\, F(\boldsymbol{v})Q(\boldsymbol{v} - \boldsymbol{u})\tilde{A}(\boldsymbol{v} - \boldsymbol{v}') \ , \tag{A8}$$

where in equation (A8) we used the fact that the final right-hand side integral in equation (A7) is the Fourier transform $\tilde{G}$ of the mosaic function G.

For an infinite continuous mosaic, $\tilde{G}(\boldsymbol{v}' - \boldsymbol{u}) = \delta^2(\boldsymbol{v}' - \boldsymbol{u})$ and thus

$$\Delta(\boldsymbol{u}) = \tilde{I}(\boldsymbol{u}) \int d^2\boldsymbol{v}\, F(\boldsymbol{v})Q(\boldsymbol{v} - \boldsymbol{u})\tilde{A}(\boldsymbol{v} - \boldsymbol{u}) \ . \tag{A9}$$

If we wish to recover $\Delta(\boldsymbol{u}) = \tilde{I}(\boldsymbol{u})$ in this limit, then

$$Q(\boldsymbol{v} - \boldsymbol{u}) = \frac{1}{z(\boldsymbol{u})}\tilde{A}^*(\boldsymbol{v} - \boldsymbol{u}) \tag{A10}$$

with normalization

$$z(\boldsymbol{u}) = \int d^2\boldsymbol{v}\, F(\boldsymbol{v})\tilde{A}^2(\boldsymbol{v} - \boldsymbol{u}) \tag{A11}$$

will fulfill our requirements. We have chosen $\tilde{A}^*(\boldsymbol{v} - \boldsymbol{u})$ as the (u, v) kernel as it reproduces the least-squares estimate of the sky brightness in the linear mosaic (Cornwell et al. 1993). Then, equation (A8) becomes

$$\Delta(\boldsymbol{u}) = \frac{1}{z(\boldsymbol{u})} \int d^2\boldsymbol{v}'\, \tilde{I}(\boldsymbol{v}')\tilde{G}(\boldsymbol{u} - \boldsymbol{v}') \int d^2\boldsymbol{v}\, F(\boldsymbol{v})\tilde{A}^*(\boldsymbol{v} - \boldsymbol{u})\tilde{A}(\boldsymbol{v} - \boldsymbol{v}') \ , \tag{A12}$$

which has a width controlled by the narrower of the width of $\tilde{A}^2$ or the width of $\tilde{G}$. Thus, by widening the mosaic $G(\boldsymbol{x})$ to a larger area than the beam A, we will fill in the desired information inside the $\tilde{A}$ smeared patches in the (u, v)-plane. Thus, a properly sampled M^2 mosaic will fill in an M^2 subgrid within each (u, v) cell you would have normally had for a single pointing, and thus an M^2 mosaic consisting of N^2 "images" each is equivalent to a (u, v) supergrid of size $(M \times N)^2$ (e.g., Ekers & Rots 1979).

Note that for a noncontinuous mosaic, there will be "aliases" in the (u, v)-plane separated by the inverse of the mosaic spacing in the sky (Cornwell 1988). Ideally, we would like the separation between aliased copies to be larger than the extent of the beam transform, which is satisfied for $\Delta x \leq \lambda/2D$, which for $D = 90$ cm corresponds to 21.′6 at 26.5 GHz and only 16.′1 at 35.5 GHz, the centers of the extremal CBI bands. The spacing used in the CBI mosaics is a compromise between the aliasing limits over the bands and the desire to have a fewer number of pointings on a convenient grid. We chose to observe with pointing centers separated by 20′, which is suboptimal above 27.5 GHz. However, the effect of aliasing is small, with the overlap point $a^{-1} - D\lambda^{-1}$ occurring at the 0.61% point of $\tilde{A}$ at 31 GHz and the 6.5% point for the highest frequency channel at 35.5 GHz.

We obtain the gridding kernel Q_{ik} of equation (A1) corresponding to equation (A10) by using the discrete sampling in equation (A4),

$$Q_{ik} = \frac{\omega_k}{z_i} \tilde{A}_k^*(\boldsymbol{u}_k - \boldsymbol{u}_i) e^{-2\pi i \boldsymbol{u}_i \cdot \boldsymbol{x}_k} , \tag{A13}$$

with visibility weights ω_k and normalization factor z_i. The discrete form of the normalization derived in equation (A11) is

$$z_i = \sum_k \omega_k \tilde{A}_k^2(\boldsymbol{u}_k - \boldsymbol{u}_i) . \tag{A14}$$

Then,

$$\Delta_i = \frac{1}{z_i} \sum_k \omega_k \tilde{A}_k^*(\boldsymbol{u}_k - \boldsymbol{u}_i) V_k e^{-2\pi i \boldsymbol{u}_i \cdot \boldsymbol{x}_k} \tag{A15}$$

is the weighted sum of visibilities used for the estimators. Note that because $V(\boldsymbol{u}) = V^*(-\boldsymbol{u})$, there are also visibilities $V_{k'}$ for which $-\boldsymbol{u}_{k'}$ lies within the support range around $\boldsymbol{u}_i$, i.e., $|\tilde{A}_{k'}^*(-\boldsymbol{u}_{k'} - \boldsymbol{u}_i)| > 0$. Thus, we should add in the complex conjugates V_k^*

$$\Delta_i = \frac{1}{z_i} \sum_k \omega_k \left[\tilde{A}_k^*(\boldsymbol{u}_k - \boldsymbol{u}_i) V_k + \tilde{A}_k^*(-\boldsymbol{u}_k - \boldsymbol{u}_i) V_k^*\right] e^{-2\pi i \boldsymbol{u}_i \cdot \boldsymbol{x}_k} . \tag{A16}$$

To do this, we construct another kernel $\bar{Q}_{ik}$,

$$\bar{Q}_{ik} = \frac{\omega_k}{z_i} \tilde{A}_k^*(-\boldsymbol{u}_k - \boldsymbol{u}_i) e^{-2\pi i \boldsymbol{u}_i \cdot \boldsymbol{x}_k} , \tag{A17}$$

which will gather the appropriate V_k^*, giving

$$\Delta_i = \sum_k \left\{ Q_{ik} V_k + \bar{Q}_{ik} V_k^* \right\} \tag{A18}$$

for the final form of our linear estimator.

For estimated visibility variances ϵ_k^2, the optimal weighting factor (in the least-squares estimation sense) is given by

$$\omega_k = \frac{1}{\epsilon_k^2} \tag{A19}$$

but may also include factors based on position in the mosaic or frequency channel. For example, up until now we have neglected the frequency dependence of the observed visibilities. If we are reconstructing an intensity field with a uniform flux density spectrum, then no changes need be made. If there is a frequency dependence, such as that for a power-law foreground (e.g., eq. [7]) or the thermal spectrum of the CMB (e.g., eq. [6]), then the visibilities should be scaled and weighted by the appropriate factor f_k when gridded in order to properly estimate $\tilde{I}_0(\boldsymbol{u}_k)$ or $\tilde{T}(\boldsymbol{u}_k)$, respectively. For example, for the CMB using equation (6) for the spectrum, we find

$$\begin{aligned} Q_{ik}^T &= \frac{f_T^{-1}(\nu_k)\omega_k}{z_i} \tilde{A}_k^*(\boldsymbol{u}_k - \boldsymbol{u}_i) , \\ \omega_k &= \frac{f_T^2(\nu_k)}{\epsilon_k^2} . \end{aligned} \tag{A20}$$

In practice for the CBI, the frequency range of the data is not great enough for the spectral weighting factor to matter, and we therefore use the default weighting given in equation (A19). This will therefore be slightly nonoptimal in the signal-to-noise sense (it will not be the minimum-variance estimator), but it will not introduce a bias in the band powers.

The choice of the normalization z_i is somewhat arbitrary, as it only determines the units of the Δ_i and not the correlation properties. However, this can be important if we wish to use the estimators to make images using the formalism of § 8. For instance, the normalization given in equation (A14) has the drawback of diverging in cells where all the $\tilde{A}_k^2$ are vanishingly small [such as the innermost and outermost supported parts of the (u, v)-plane] and will produce images with heightened noise on short and long spatial wavelengths. It is therefore more convenient to use the alternate normalization

$$z_i = \sum_k \omega_k \tilde{A}_k^*(\boldsymbol{u}_k - \boldsymbol{u}_i) , \tag{A21}$$

which when inserted into equation (A15) will properly normalize the weighted sums of visibilities. This will then produce images with the desired units of Jy beam^{-1} (see § 8). We therefore use equation (A21) for the normalization in our CBI pipeline.

APPENDIX B

SOURCE COUNTS AND THE RESIDUAL COVARIANCE MATRIX

We wish to calculate $C_\nu^{\rm res}$ (see eq. [67]) using equation (66) with $\nu_k = \nu_{k'} = \nu$. If $p(\alpha|\nu_0)$ is independent of flux density, then

$$C_\nu^{\rm res} = \int_{-\infty}^{\infty} d\alpha\, p(\alpha|\nu_0) \left(\frac{\nu}{\nu_0}\right)^{2\alpha} \int_0^{S_{\rm max}(\alpha)} dS\, S^2 \frac{dN(S)}{dS} , \tag{B1}$$

where we have left in the possibility that the upper flux density cutoff will depend on spectral index (see below) and set the lower flux density cutoff to zero (the results for realistic power-law counts with $\gamma > -2$ are insensitive to the lower cutoff, but one can easily be included). As an example for the calculation of the fluctuation power due to residual sources in the Gaussian limit, consider power-law integral source counts

$$N(>S) = N_0 \left(\frac{S}{S_0}\right)^{\gamma} \Rightarrow \frac{dN(S)}{dS} = -\frac{\gamma N_0}{S_0}\left(\frac{S}{S_0}\right)^{\gamma-1} , \tag{B2}$$

where $N(>S)$ is the mean number density of sources with flux density *greater* than S at frequency ν_0 and a Gaussian spectral index distribution at frequency ν_0,

$$p(\alpha|S, \nu_0) = p(\alpha|\nu_0) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} e^{-(\alpha-\alpha_0)^2/2\sigma_\alpha^2} . \tag{B3}$$

First consider the case in which there is a fixed flux density upper cutoff $S_{\rm max}$ at the frequency where the number counts are defined. The two parts of equation (67) separate easily, where the source count part of the integral is

$$\int_0^{S_{\rm max}} dS\, S^2 \frac{dN(S)}{dS} = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{S_{\rm max}}{S_0}\right)^{\gamma+2} . \tag{B4}$$

For the distribution in equation (B3), the integral over α becomes

$$\begin{aligned}\int_{-\infty}^{\infty} d\alpha\, p(\alpha|\nu_0)\left(\frac{\nu}{\nu_0}\right)^{2\alpha} &= \int_{-\infty}^{\infty} \frac{d\alpha}{\sqrt{2\pi\sigma_\alpha^2}} e^{-(\alpha-\alpha_0)/2\sigma_\alpha^2} e^{2\alpha\beta} \\ &= e^{-(\bar\alpha^2-\alpha_0^2)^2/2\sigma_\alpha^2} \int_{-\infty}^{\infty} \frac{d\alpha}{\sqrt{2\pi\sigma_\alpha^2}} e^{-(\alpha-\bar\alpha)^2/2\sigma_\alpha^2} \\ &= e^{2\alpha_{\rm eff}\beta} ,\end{aligned} \tag{B5}$$

where $\beta = \ln(\nu/\nu_0)$ and

$$\begin{aligned}\alpha_{\rm eff} &= \frac{\bar\alpha^2 - \alpha_0^2}{4\beta\sigma_\alpha^2} = \alpha_0 + \beta\sigma_\alpha^2 , \\ \bar\alpha &= \alpha_0 + 2\beta\sigma_\alpha^2 ,\end{aligned} \tag{B6}$$

where $\bar\alpha$ is the mean of the extrapolated spectral index distribution, which remains a Gaussian, and the effective spectral index $\alpha_{\rm eff}$ for the spectral component is shifted from the mean spectral index of the input distribution α_0 by the combination of the scatter in the α and the lever arm β from the frequency extrapolation. Putting these together, we get

$$C_\nu^{\rm res} = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{S_{\rm max}}{S_0}\right)^{\gamma+2} e^{2\alpha_{\rm eff}\beta} . \tag{B7}$$

One can also deal with the case in which there is an upper flux density cutoff $\hat S_{\rm max}$ imposed at a frequency $\hat\nu$ other than ν_0 where the $N(S)$ distribution is defined. In this case, the flux density cutoff in equation (B1) is

$$S_{\rm max}(\alpha) = \bar S_{\rm max} e^{(\alpha_0-\alpha)\hat\beta} \bar S_{\rm max} = \hat S_{\rm max} e^{-\alpha_0\hat\beta} , \tag{B8}$$

where $\hat\beta = \ln(\hat\nu/\nu_0)$ and $\bar S_{\rm max}$ is the cutoff $\hat S_{\rm max}$ extrapolated to ν_0 using α_0. Then,

$$\int_0^{S_{\rm max}(\alpha)} dS\, S^2 \frac{dN(S)}{dS} = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{\bar S_{\rm max}}{S_0}\right)^{\gamma+2} e^{(\alpha_0-\alpha)\hat\beta(\gamma+2)} , \tag{B9}$$

and thus

$$C_\nu^{\rm res} = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{\bar{S}_{\rm max}}{S_0}\right)^{\gamma+2} \int_{-\infty}^{\infty} d\alpha\, p(\alpha|\nu_0) e^{2\alpha\beta} e^{(\alpha_0-\alpha)\hat{\beta}(\gamma+2)}$$
$$= -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{\bar{S}_{\rm max}}{S_0}\right)^{\gamma+2} e^{2\hat{\alpha}_{\rm eff}\beta} , \quad \text{(B10)}$$

with

$$\hat{\alpha}_{\rm eff} = \alpha_0 + \beta\sigma_\alpha^2\kappa^2 ,$$
$$\bar{\alpha} = \alpha_0 + 2\beta\sigma_\alpha^2\kappa ,$$
$$\kappa = 1 - \frac{\hat{\beta}}{2\beta}(\gamma+2) , \quad \text{(B11)}$$

where κ gives the modification of the effective spectral index due to the change in the frequency at which the cutoff is done.

One often has an upper flux density cutoff at two different frequencies. For example, sources that are extrapolated to be bright at the CMB observing frequency will have been detected and subtracted. If there is a flux density cutoff of $\hat{S}_{\rm max}$ imposed at a frequency $\hat{\nu}$ as before, but an additional upper cutoff of $\hat{S}'_{\rm max}$ at another frequency $\hat{\nu}'$, then there is a critical spectral index

$$\alpha_{\rm crit} = \frac{\ln\left(\hat{S}'_{\rm max}/\hat{S}_{\rm max}\right)}{\ln(\hat{\nu}'/\hat{\nu})} \quad \text{(B12)}$$

above which the effective cutoff $\bar{S}_{\rm max}$ of equation (B8) changes from that appropriate to $\hat{\nu}$ to that at $\hat{\nu}'$ (assuming $\hat{\nu}' > \hat{\nu}$). Thus, the integral over α in equation (B10) will be broken into two pieces,

$$J_1 = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{\bar{S}_{\rm max}}{S_0}\right)^{\gamma+2} e^{2\hat{\alpha}_{\rm eff}\beta} \int_{-\infty}^{\alpha_{\rm crit}} \frac{d\alpha}{\sqrt{2\pi\sigma_\alpha^2}} e^{-(\alpha-\bar{\alpha})^2/2\sigma_\alpha^2} ,$$
$$J_2 = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{\bar{S}'_{\rm max}}{S_0}\right)^{\gamma+2} e^{2\hat{\alpha}'_{\rm eff}\beta} \int_{\alpha_{\rm crit}}^{\infty} \frac{d\alpha}{\sqrt{2\pi\sigma_\alpha^2}} e^{-(\alpha-\bar{\alpha}')^2/2\sigma_\alpha^2} , \quad \text{(B13)}$$

where $C_\nu^{\rm res} = J_1 + J_2$. The quantities in J_1 are as defined in equations (B8)–(B11), and the parameters in J_2 are defined in the same way but using the higher frequency $\hat{\nu}'$. The truncated Gaussian integrals are just the integrated probabilities for the normal distribution

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} dt\, e^{-t^2/2} = \frac{1}{2} + \frac{1}{2}{\rm erf}\left(\frac{x}{\sqrt{2}}\right) , \quad \text{(B14)}$$

with erf (z) the error function. Then,

$$J_1 = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{\bar{S}_{\rm max}}{S_0}\right)^{\gamma+2} e^{2\hat{\alpha}_{\rm eff}\beta} F(x_{\rm crit}) , \quad \text{(B15)}$$

$$J_2 = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{\bar{S}'_{\rm max}}{S_0}\right)^{\gamma+2} e^{2\hat{\alpha}'_{\rm eff}\beta} [1 - F(x'_{\rm crit})] , \quad \text{(B16)}$$

where $x_{\rm crit} = (\alpha_{\rm crit} - \bar{\alpha})/\sigma_\alpha$ and $x'_{\rm crit} = (\alpha_{\rm crit} - \bar{\alpha}')/\sigma_\alpha$.

As an example, consider the source counts presented in Paper II (§ 4.3.2), with $N_0 = 9.2 \times 10^3$ sr^{-1} above $S_0 = 10$ mJy at $\nu_0 = 31$ GHz and $\gamma = -0.875$, which gives

$$-\frac{\gamma}{\gamma+2} N_0 S_0^2 = 0.715 \text{ Jy}^2 \text{ sr}^{-1} \quad \text{(B17)}$$

as the raw source power. In the analysis described there, Mason et al. find that a Gaussian 1.4–31 GHz spectral index distribution with $\alpha_0 = -0.45$ and $\sigma_\alpha = 0.37$ fits the observed data. The CBI and OVRO direct measurements have a cutoff of $S_{\rm max} = 25$ mJy at 31 GHz (sources brighter than this have been subtracted from the CBI data and have residual uncertainties placed in a source covariance matrix), and sources above $\hat{S}_{\rm max} = 3.4$ mJy at $\hat{\nu} = 1.4$ GHz have already been accounted for in a second source matrix. Therefore, the critical spectral index is $\alpha_{\rm crit} = 0.644$ from equation (B12). For $\alpha \geq \alpha_{\rm crit}$, the 31 GHz cutoff holds. Since the cutoff and source distribution are at the same frequency as the observations $\nu = \nu_0$, there is no extrapolation factor $\beta = 0$ and the spectral index distribution is unchanged ($\bar{\alpha}' = \alpha_0$). Then, $x'_{\rm crit} = 2.957$ and $1 - F(x'_{\rm crit}) \approx 1.56 \times 10^{-3}$, so

$$J_2 = -\frac{\gamma}{\gamma+2} N_0 S_0^2 \left(\frac{S_{\rm max}}{S_0}\right)^{\gamma+2} [1 - F(x'_{\rm crit})] = 0.003 \text{ Jy}^2 \text{ sr}^{-1} \quad \text{(B18)}$$

for the flat-spectrum tail of the spectral index integral. The rest of the integral uses the 1.4 GHz cutoff, which we extrapolate using the mean spectrum to 31 GHz using equation (B8),

$$\bar{S}_{\max} = \hat{S}_{\max} e^{-\alpha_0 \hat{\beta}} = 0.843 \text{ mJy}, \qquad \hat{\beta} = -3.098 \; . \tag{B19}$$

Because $\beta = 0$, we have to modify the quantities in equation (B11) by explicitly expanding the terms in κ and canceling remaining terms in β, giving

$$\bar{\alpha} = \alpha_0 - \hat{\beta}\sigma_\alpha^2(\gamma + 2) = 0.027, \qquad 2\hat{\alpha}_{\text{eff}}\beta = \tfrac{1}{2}\hat{\beta}^2\sigma_\alpha^2(\gamma + 2)^2 = 0.831 \; , \tag{B20}$$

which can then be inserted into equation (B15), giving

$$J_1 = -\frac{\gamma}{\gamma + 2} N_0 S_0^2 \left(\frac{\bar{S}_{\max}}{S_0}\right)^{\gamma+2} e^{2\hat{\alpha}_{\text{eff}}\beta} F(x_{\text{crit}}) = 0.100 \text{ Jy}^2 \text{ sr}^{-1} \tag{B21}$$

for $x_{\text{crit}} = 1.667$, $F(x_{\text{crit}}) \approx 0.952$, and thus we expect

$$C_\nu^{\text{res}} = 0.10 \text{ Jy}^2 \text{ sr}^{-1} \tag{B22}$$

for the amplitude of the residual sources in the CBI fields. In Paper II, it is noted that there is a 25% uncertainty on N_0, and more importantly the power-law slope of the source counts could conceivably be as steep as $\gamma = -1$. Taking the extreme of $\gamma = -1$, we get

$$C_\nu^{\text{res}} = 0.15 \text{ Jy}^2 \text{ sr}^{-1} \tag{B23}$$

using the above procedure. We thus conservatively estimate a 50% uncertainty on the value of C_ν^{res} derived in this manner. Note that in Paper II we actually use the value of $C_\nu^{\text{res}} = 0.08 \text{ Jy}^2 \text{ sr}^{-1}$ derived using a Monte Carlo procedure emulating the integrals in equation (B13) but using the actual observed distribution of source flux densities and spectral indices. The agreement between these two estimates shows the efficacy of this procedure in practice.

APPENDIX C

COMPARISON WITH THE HOBSON & MAISINGER (2002) METHOD

Recently, Hobson & Maisinger (2002) have independently proposed a binned (u, v)-plane method that is somewhat similar to ours, although it is more directly related to the "optimal maps" of Bond & Crittenden (2001). Hobson & Maisinger (2002) use a gathering mapping $\boldsymbol{H}$ ($\boldsymbol{M}$ in their notation),

$$\boldsymbol{V} = \boldsymbol{H}\boldsymbol{s} + \boldsymbol{e} \; , \tag{C1}$$

rather than our scattering kernel $\boldsymbol{Q}$ of equation (21). In the Hobson & Maisinger (2002) method, the vector $\boldsymbol{s}$ can be thought of as a set of ideal pixels in the (u, v)-plane. They show that the likelihood depends on binned visibilities $\boldsymbol{v}$ and noise $\boldsymbol{n}$,

$$\begin{aligned} \boldsymbol{v} &= \left(\boldsymbol{H}^\dagger \boldsymbol{E}^{-1} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^\dagger \boldsymbol{E}^{-1} \boldsymbol{V} = \boldsymbol{s} + \boldsymbol{n} \; , \\ \boldsymbol{n} &= \left(\boldsymbol{H}^\dagger \boldsymbol{E}^{-1} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^\dagger \boldsymbol{E}^{-1} \boldsymbol{e} \; , \end{aligned} \tag{C2}$$

where

$$\boldsymbol{C}^N = \langle \boldsymbol{n}\boldsymbol{n}^\dagger \rangle = \left(\boldsymbol{H}^\dagger \boldsymbol{E}^{-1} \boldsymbol{H}\right)^{-1} \; . \tag{C3}$$

The Hobson & Maisinger (2002) kernel H_{jk} is chosen to equal 1 if the $\boldsymbol{u}_j$ of visibility V_j lies in cell k, although other more complicated kernels could be imagined. The Hobson & Maisinger (2002) method will also give a calculational speedup through the reduction in number of independent gridded estimators, and the use of the method is demonstrated using simulated VSA data in their paper.

REFERENCES

Bond, J. R., & Crittenden, R. G. 2001, in Structure Formation in the Universe, ed. R. G. Crittenden & N. G. Turok (NATO ASIC Proc. 565 Dordrecht: Kluwer), 241
Bond, J. R., & Efstathiou, G. 1987, MNRAS, 226, 655
Bond, J. R., Jaffe, A. H., & Knox, L. 1998, Phys. Rev. D, 57, 2117
———. 2000, ApJ, 533, 19
Bond, J. R., & Myers, S. T. 1996, ApJS, 103, 63
Bond, J. R., et al. 2003, ApJ, submitted (Paper VI)
Bracewell, R. N. 1986, The Fourier Transform and Its Applications (New York: McGraw-Hill)
Cornwell, T. J. 1988, A&A, 202, 316
Cornwell, T. J., Holdaway, M. A., & Uson, J. M. 1993, A&A, 271, 697
Draine, B. T., & Lazarian, A. 1999, ApJ, 512, 740
Ekers, R. D., & Rots, A. H. 1979, in IAU Colloq. 49, Image Formation from Coherence Functions in Astronomy, ed. C. van Schooneveld (Dordrecht: Reidel), 61
Halverson, N. W., et al. 2002, ApJ, 568, 38
Hananay, S., et al. 2000, ApJ, 545, L5
Hobson, M. P., Lasenby, A. N., & Jones, M. 1995, MNRAS, 275, 863
Hobson, M. P., & Maisinger, K. 2002, MNRAS, 334, 569
Knox, L. 1999, Phys. Rev. D, 60, 103516
Kogut, A., Banday, A. J., Bennett, C. L., Gorski, K. M., Hinshaw, G., & Reach, W. T. 1996, ApJ, 460, 1
Lange, A. E., et al. 2001, Phys. Rev. D, 63, 42001

Lee, A. T., et al. 2001, ApJ, 561, L1
Leitch, E. M., Readhead, A. C. S., Pearson, T. J., & Myers, S. T. 1997, ApJ, 486, L23
Mason, B. S., et al. 2003, ApJ, 591, 540 (Paper II)
Mather, J. C., Fixsen, D. J., Shafer, R. A., Mosier, C., & Wilkinson, D. T. 1999, ApJ, 512, 511
Netterfield, C. B., et al. 2002, ApJ, 571, 604
Ng, K.-W. 2001, Phys. Rev. D, 63, 123001
Padin, S., et al. 2001, ApJ, 549, L1 (Paper I)
———. 2002, PASP, 114, 83
Pearson, T. J., et al. 2003, ApJ, 591, 556 (Paper III)
Sault, R. J., Staveley-Smith, L., & Brouw, W. N. 1996, A&AS, 120, 375
Shepherd, M. C. 1997, in ASP Conf. Ser. 125, Astronomical Data Analysis Software and Systems VI, ed. G. Hunt & H. E. Payne (San Francisco: ASP), 77
Sievers, J. S., et al. 2003, ApJ, 591, 599 (Paper V)
Taylor, G. B., Carilli, C. L., & Perley, R. A., eds. 1999, ASP Conf. Ser. 180, Synthesis Imaging in Radio Astronomy II (San Francisco: ASP)
Thompson, A. R., Moran, J. M., & Swenson, G. W., Jr. 1986, Interferometry and Synthesis in Radio Astronomy (New York: Wiley)
White, M., Carlstrom, J. E., Dragovan, M., & Holzapfel, W. L. 1999a, ApJ, 514, 12
———. 1999b, preprint (astro-ph/9912422)