MPI-PhT/95-62

A METHOD FOR SUBTRACTING FOREGROUNDS FROM MULTI-FREQUENCY CMB SKY MAPS [1]

Max Tegmark

Max-Planck-Institut für Physik, Föhringer Ring 6, D-80805 München; email: max@mppmu.mpg.de

George Efstathiou

Dept. of Physics, University of Oxford, Keeble Road, Oxford, OX1 3RH; email: g.efstathiou@physics.oxford.ac.uk

Abstract

An improved method for subtracting contaminants from Cosmic Microwave Background (CMB) sky maps is presented, and used to estimate how well future experiments will be able to recover the primordial CMB fluctuations. We find that the naive method of subtracting foregrounds (such as dust emission, synchrotron radiation, free-free-emission, unresolved point sources, *etc.*) on a pixel by pixel basis can be improved by more than an order of magnitude by taking advantage of the correlation of the emission in neighboring pixels. The optimal multi-frequency subtraction method improves on simple pixel-by-pixel subtraction both by taking noise-levels into account, and by exploiting the fact that most contaminants have angular power spectra that differ substantially from that of the CMB. The results are natural to visualize in the two-dimensional plane with axes defined by multipole ℓ and frequency ν. We present a brief overview of the geography of this plane, showing the regions probed by various experiments and where we expect contaminants to dominate. We illustrate the method by estimating how well the proposed ESA COBRAS/SAMBA mission will be able to recover the CMB fluctuations against contaminating foregrounds.

[1] Accepted for publication in MNRAS, first submitted June 19, 1995.
Available from *http://www.mpa-garching.mpg.de/~max/wiener.html* (faster from Europe) and from from *http://astro.berkeley.edu/~max/cobepow.html* (faster from the US).
Color figures are also available on these web pages.

1 INTRODUCTION

There has been a surge of interest in the cosmic microwave background radiation (CMB) since the first anisotropies of assumed cosmological origin were detected by the COBE DMR experiment (Smoot *et al.* 1992). On the experimental front, many new experiments have been carried out, and more are planned or proposed for the near future (see White *et al.* 1994, for a review). On the theoretical front, considerable progress has been made in understanding how the CMB power spectrum C_ℓ depends on various cosmological model parameters (see Bond *et al.* 1994, Hu & Sugiyama 1995, and references therein for recent reviews of analytical and quantitative aspects of this problem). It is now fairly clear that an accurate measurement of the angular power spectrum C_ℓ to multipoles $\ell \sim 10^3$ could provide accurate constraints on many of the standard cosmological parameters (Ω, Ω_b, Λ, spectral index n, *etc.*), thus becoming the definitive arbiter between various flavours of the cold dark matter (CDM) cosmogony and other theories of the origin of structure in the Universe. To accurately measure the power spectrum and reach this goal, a number of hurdles must be overcome:

- Technical problems
- Incomplete sky-coverage
- Foregrounds

There is of course a wide variety of technical challenges that must be tackled to attain high resolution, low noise, well-calibrated temperature data over a wide range of frequencies and over most of the sky. However, thanks to the rapid advance in detector technology over the last two decades and the possibilities of ground based interferometers, long duration balloon flights and space-born experiments, there is a real prospect that high sensitivity maps of the CMB will be obtained within the next decade.

The second hurdle refers to the fact that we are unlikely to measure the primordial CMB sky accurately behind the Galactic plane. As is well known, the resulting incomplete sky coverage makes it impossible to exactly recover the spherical harmonic coefficients that would give direct estimates of the angular power spectrum C_ℓ. However, a number of methods for efficiently constraining models, using only partial sky coverage, have now been developed (Górski 1994; Bond 1995; Bunn & Sugiyama 1995; Tegmark & Bunn 1995), and it has recently been shown (Tegmark 1995, hereafter

T95) that even the individual C_ℓ-coefficients can be accurately estimated for all but the very lowest multipoles such as $\ell = 2$, by expanding the data in appropriately chosen basis functions. Rather, the most difficult of the above-mentioned hurdles appears to be the third, which is the topic of the present paper.

The frequency-dependence of the various foregrounds has been extensively studied, in both "clean" and "dirty" regions of the sky (see *e.g.* Brandt *et al.* 1993, Toffolatti *et al.* 1994, for recent reviews). However, these properties alone provide a description of the foregrounds that is somewhat too crude to assess the extent to which they can be separated from the underlying CMB signal, since the foregrounds fluctuations depend on the multipole moment ℓ as well (see Bouchet *et al.*, 1995, for simulations). Most published plots comparing different CMB experiments tend to show ℓ on the horizontal axis and an amplitude (C_ℓ or an r.m.s. $\Delta T/T$) on the vertical axis, whereas most plots comparing different foregrounds show amplitude plotted against frequency ν. Since the fluctuations in the latter tend to depend strongly on both ℓ and ν, *i.e.*, on both spatial and temporal frequency, one obtains a more accurate picture by combining both of these pieces of information and working in a two-dimensional plane as in Figures 1-6. We will indeed find that the ℓ–ν plane arises naturally in the optimized subtraction scheme that we present. Figure 1 shows roughly the regions in this plane probed by various CMB experiments. Each rectangle corresponds to one experiment. Its extent in the ℓ-direction shows the customary 1σ width of the experimental window function (see *e.g.* White & Srednicki 1995), whereas the vertical extent shows the frequency range that is covered. For single-channel experiments, we plot the quoted bandwidth, whereas for multi-channel experiments, the box has simply been plotted in the range between the lowest and highest frequency channel. For a more detailed description of these experiments, see Scott & White (1995) and references therein. Figures 2 though 5, which will be described in Section 4, show the estimated fluctuations of the CMB and various foregrounds in the same plane, and comparing these figures with Figure 1 as in Figure 14, it is easy to understand which experiments are the most affected by the various foregrounds. Moreover, as we will see, familiarity with the geography of this plane provides an intuitive understanding of the advantages and shortcomings of different methods of foreground subtraction.

A number of CMB satellite missions are currently under consideration by various funding agencies, which would offer the excellent sensitivity, resolution and frequency coverage that is needed for accurate foreground subtraction. Throughout this paper, we will use the proposed European Space Agency COBRAS/SAMBA mission (Mandolesi *et al.* 1995) as an illustration of what the next generation of space-borne CMB experiments may be able to achieve. The specifications of competing satellite proposals are similar, though with a more restricted frequency range.

The rest of this paper is organized as follows. After establishing our basic notation in Section 2, we derive the optimized multi-frequency subtraction method in Section 3. In Section 4, we estimate the angular power spectra of the various foreground contaminants. In Section 5, we use these estimates to assess the effectiveness of the subtraction technique and to show how accurately the CMB fluctuations could be recovered from high quality data, such as might be obtained from the proposed COBRAS/SAMBA satellite.

2 NOTATION

Let $B(\widehat{\mathbf{r}},\nu)$ denote the total sky brightness at frequency ν in the direction of the unit vector $\widehat{\mathbf{r}}$. Since we know that B is a sum of contributions of physically distinct origins, we will write it as

$$B(\widehat{\mathbf{r}},\nu)=\sum B_i(\widehat{\mathbf{r}},\nu). \tag{1}$$

In the microwave part of the spectrum, the most important non-CMB components are synchrotron radiation B_{synch}, free-free emission B_{ff}, dust emission B_{dust}, and radiation from point sources B_{ps} (both radio sources and infrared emission from galaxies). We will separate the CMB-contribution into two terms; an isotropic blackbody B_0 and the fluctuations B_{cmb} around it. The former, which is of course independent of $\widehat{\mathbf{r}}$, is given by

$$B_0(\nu)=\frac{2h}{c^2}\frac{\nu^3}{e^x-1}=2\frac{(kT_0)^3}{(hc)^2}\left(\frac{x^3}{e^x-1}\right)\approx 270.2\,\mathrm{MJy/sr}\left(\frac{x^3}{e^x-1}\right), \tag{2}$$

where $x\equiv h\nu/kT_0\approx\nu/56.8\mathrm{GHz}$ and $T_0\approx 2.726\mathrm{K}$ (Mather *et al.* 1994). If the actual CMB temperature across the sky is $T(\widehat{\mathbf{r}})=T_0+\delta T(\widehat{\mathbf{r}})$, then to an excellent approximation, $B_{cmb}(\widehat{\mathbf{r}},\nu)=(\partial B_0(\nu)/\partial T_0)\delta T(\widehat{\mathbf{r}})$. (The quadratic correction will be down by a factor of $\delta T/T\approx 10^{-5}$.) This conversion factor from brightness to temperature is

$$\frac{\partial B_0}{\partial T_0}=\frac{k}{2}\left(\frac{kT_0}{hc}\right)^2\left(\frac{x^2}{\sinh x/2}\right)^2\approx\left(\frac{24.8\mathrm{MJy/sr}}{K}\right)\left(\frac{x^2}{\sinh x/2}\right)^2. \tag{3}$$

Since any other contaminants that we fail to take into account will be mistaken for CMB fluctuations, it is convenient to convert all brightness fluctuations into temperature fluctuations in the analogous way, so we define

$$\delta T_i(\widehat{\mathbf{r}}, \nu) \equiv \frac{B_i(\widehat{\mathbf{r}}, \nu)}{\partial B_0/\partial T_0}. \tag{4}$$

Note that with this definition, $\delta T_{cmb}(\widehat{\mathbf{r}}, \nu)$ is independent of ν (compare Figure 2 with Figure 3), whereas most of the foregrounds will exhibit a strong frequency dependence. We expand the temperature fluctuations in spherical harmonics as usual;

$$\delta T_i(\widehat{\mathbf{r}}, \nu) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} Y_{\ell m}(\widehat{\mathbf{r}})\, a_{\ell m}^{(i)}(\nu). \tag{5}$$

For isotropic CMB fluctuations, we have

$$\begin{aligned} \langle a_{\ell m}^{cmb}(\nu)\rangle &= 0, & (6)\\ \langle a_{\ell m}^{cmb}(\nu)^* a_{\ell' m'}^{cmb}(\nu)\rangle &= \delta_{\ell\ell'}\delta_{mm'}C_\ell, & (7) \end{aligned}$$

where C_ℓ is the frequency independent *angular power spectrum.* For other components, the means $\langle a_{\ell m}^{(i)}(\nu)\rangle$ are not necessarily equal to zero. For instance, most of the foregrounds are by nature non-negative ($B_i \geq 0$), so we expect the monopole to be positive. Also, if there are deviations from isotropy (a secant behavior with Galactic latitude, for instance), we will not have $\langle a_{\ell m}^{(i)}(\nu)^* a_{\ell' m'}^{(i)}(\nu)\rangle \propto \delta_{\ell\ell'}\delta_{mm'}$. For such cases, we simply define

$$C_\ell^{(i)}(\nu) \equiv \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} \left\langle \left| a_{\ell m}^{(i)}(\nu)\right|^2 \right\rangle, \tag{8}$$

since this is the quantity that on average will be added to the estimate of the CMB power spectrum if the contaminant is not removed. As we shall see further on, all contaminants that we have investigated in this paper do become fairly isotropic when we mask out all but the cleanest parts of the sky; the different Fourier components do indeed decouple and so the only important difference from CMB behavior is an additional constant term in the monopole (which is, of course, unmeasurable anyway).

We conclude this Section with a few comments on how to read Figures 2 through 6. If a random field satisfies equations (6) and (7), the addition theorem for spherical harmonics gives the well-known result

$$\langle\delta T_i(\nu)^2\rangle=\sum_{l=0}^{\infty}\left(\frac{2\ell+1}{4\pi}\right)C_\ell^{(i)}(\nu)\approx\int\left[\frac{\ell(2\ell+1)}{4\pi}\right]C_\ell^{(i)}d(\ln\ell). \qquad (9)$$

This is the reason that we plot the quantity $[\ell(2\ell+1)C_l^{(i)}(\nu)/4\pi]^{1/2}$ in the figures of the next section: the resulting r.m.s. temperature fluctuation $\delta T_i(\nu)$ is basically just the r.m.s. height of the curve times a small constant. If we compute the r.m.s. average in a multipole range $\ell_0\leq\ell\leq\ell_1$, then this constant is simply $[(\ln(\ell_1/\ell_0)]^{1/2}$. For the CMB temperature fluctuations of standard CDM, for instance, which are shown in Figure 3 normalized to COBE (Bunn *et al.* 1995), $\langle\delta T^2\rangle^{1/2}\approx 108\mu$K. Convolving the fluctuations with the COBE beam with FWHM $=7.08°$ suppresses C_ℓ for $\ell\gg 10$, giving $\langle\delta T^2\rangle^{1/2}\approx 37\mu$K. The order of magnitude of both of these numbers can be roughly read off by eye with the above prescription. For the CMB brightness fluctuations shown in Figure 2, everything is completely analogous. For instance, the total fluctuations are $\langle B_{cmb}(\nu)^2\rangle^{1/2}\approx 0.05$MJy/sr at the maximum sensitivity frequency $\nu\approx 218$ GHz, and 0.02MJy/sr as seen with the COBE beam.

3 MULTI-FREQUENCY FOREGROUND SUBTRACTION

In this Section, we derive the multi-frequency subtraction scheme mentioned in the introduction. Given data in several frequency channels, the goal is to produce the best maps corresponding to the different physical components, where "best" is taken to mean having the smallest r.m.s. errors. We first derive the method for an idealized case, and then add the various real-world complications one by one.

Before beginning, it is instructive to compare this approach with that of likelihood analysis. Most published analyses of the COBE DMR sky maps (see *e.g.* Tegmark & Bunn, 1995 for a recent review) have used likelihood techniques to constrain models with power spectra described by one or two free parameters, typically a spectral index and an overall normalization. As long as the number of model-parameters is rather small, a useful way to deal with foreground contamination is that described by Dodelson & Stebbins (1994). The basic idea is to include in the likelihood analysis a number of "nuisance parameters" describing the foregrounds, and then marginalize over these parameters to obtain the Bayesian probability distribution for the parameters of interest. Despite its elegance, this method is of course only feasible when the number of parameters, n, to be estimated is small, since the number of grid points in the n-dimensional parameter space (and hence the amount of computer time required for the analysis) grows exponentially with n. The problems addressed in this paper are how to estimate the entire power spectrum C_ℓ (about $n = 10^3$ parameters) and how to reconstruct a high-resolution all-sky map (with perhaps as many as $n \sim 10^7$ parameters (pixels)), which is why a more direct approach other than likelihood analysis is required.

The method we present below is type of linear filtering closely related to Wiener filtering, but with a crucial difference. Linear filtering techniques have recently been applied to a range of cosmological problems. Rybicki & Press (1992) give a detailed discussion of the one-dimensional problem. Lahav *et al.* (1994), Fisher *et al.* (1995) and Zaroubi *et al.* (1995) apply Wiener filtering to galaxy surveys. In particular, Bunn *et al.* (1994) apply Wiener filtering to the COBE DMR maps. We will generalize this treatment to the case of multiple frequencies and more than one physical component. Although it is tempting to refer to this simply as multi-frequency Wiener filtering, we will avoid this term, as it can cause confusion for the following reason. As is discussed in any signal-processing text, regular Wiener filtering modifies the power spectrum of the signal. Specifically, the filtered signal will have a power spectrum

$$P_s'(k) = \frac{P_s(k)}{P_s(k) + P_n(k)}, \tag{10}$$

where P_s and P_n are the power spectra of the true signal and the contaminant, respectively. This of course makes it useless for power-spectrum estimation. As we shall see in Section 3.6, our approach *does not* alter the power spectrum of the signal (the CMB, say), and moreover has the attractive property of being independent of any assumptions about the true CMB power spectrum, requiring only assumptions about the power spectra of the foregrounds. This is possible because more than one frequency channel is available, which allows foreground/background separation even if the two have identical power spectra. The availability of multiple frequencies is absolutely essential to our method: in the special case where m, the number of channels, equals one, it degenerates not to standard Wiener filtering, but to the trivial case of doing no subtraction at all.

The method that we advocate turns out to be related to Wiener filtering by a simple ℓ-dependent rescaling of the weights. It is therefore trivial switch between this power-conserving subtraction scheme and an error-minimizing multi-frequency generalization of Wiener filtering. For pedagogical purposes, we first present the latter, in sections 3.2-3.5, then show how it should be rescaled in Section 3.6.

3.1 The model

Let us make the approximation that we can can write

$$\delta T(\mathbf{r},\nu) = \sum_{i=1}^{n} f_i(\nu) x_i(\mathbf{r}), \tag{11}$$

where each term corresponds to a distinct physical component (such as CMB, dust, synchrotron radiation, free-free emission, radio point sources, *etc.*). Thus we are simply assuming that the contribution from each component is separable into a function that depends only on frequency times a function that depends only on position. For definiteness, let us normalize all the functions f_i so that $f_i(100\,\mathrm{GHz}) = 1$, thus absorbing the physical units into the fields x_i.

Suppose that we observe the sky in m different frequency channels, so that at each point $\mathbf{r}$ on the sky, we measure m different numbers $y_i(\mathbf{r})$, $i = 1, ..., m$. (These need of course not be different channels observed by the same experiment — we may for instance want to use the IRAS 100 micron survey as an additional "channel".) Assuming that we know the spectra $f_i(\nu)$ of all the components, we can write

$$y(\mathbf{r}) = Fx(\mathbf{r}) + \varepsilon(\mathbf{r}). \tag{12}$$

Here the vector $\varepsilon(\mathbf{r})$ corresponds to the instrumental noise in the various channels, and F is a fixed $m \times n$ matrix, the *frequency response matrix*, given by

$$F_{ij} \equiv \int_0^\infty w_i(\nu) f_j(\nu) d\nu, \tag{13}$$

$w_i(\nu)$ being the frequency response of the i^{th} channel.

3.2 The idealized flat case

We now turn to the highly idealized case where the sky is flat rather than spherical, there is no pixelization, no Galactic zone of avoidance, *etc.* This simple case is directly applicable only to a small patch of sky sampled at high resolution. In the subsequent sections, we will show how to tackle the numerous real-world complications. It will be seen that none of these complications change the basic matrix prescription of the simple case described here.

The second term in equation (12), the vector ε, contains the instrumental noise in the different frequency channels. We model this by

$$\begin{aligned} \langle\varepsilon_i(\mathbf{r})\rangle &= 0, & (14)\\ \langle\widehat{\varepsilon}_i(\mathbf{k})^*\widehat{\varepsilon}_j(\mathbf{k}')\rangle &= (2\pi)^2\delta_{ij}\delta(\mathbf{k}'-\mathbf{k})P_i^{(n)}(k), & (15)\end{aligned}$$

thereby allowing for the possibility that the noise within each channel may exhibit some correlation (hats denote Fourier transforms). Uncorrelated noise simply corresponds to the case where the noise power spectra are given by $P_i^{(n)}(\mathbf{k}) = \sigma_i^2$, where the σ_i are constants.

Analogously, we assume that the physical components satisfy

$$\begin{aligned} \langle x_i(\mathbf{r})\rangle &= 0, & (16)\\ \langle\widehat{x}_i(\mathbf{k})^*\widehat{x}_j(\mathbf{k}')\rangle &= (2\pi)^2\delta_{ij}\delta(\mathbf{k}'-\mathbf{k})P_i^{(s)}(k). & (17)\end{aligned}$$

Note that we are not assuming the random fields to be Gaussian. Equation (17) follows directly from equation (16) and homogeneity (translational invariance), together with the obvious assumption that the different components are independent.

Our goal is to make a reconstruction, denoted $\mathbf{x}'$, of the physical fields $\mathbf{x}$ from the observed data $\mathbf{y}$. We will set out to find the best *linear* reconstruction. Because of the translational invariance, the most general linear estimate $\mathbf{x}'$ of $\mathbf{x}$ can clearly be written as

$$\mathbf{x}'(\mathbf{r}) = \int W(\mathbf{r}-\mathbf{r}')\mathbf{y}(\mathbf{r}')d^2r', \qquad (18)$$

for some matrix-valued function W that we will refer to as the *reconstruction matrix.* We now proceed to derive the best choice of the reconstruction matrix. Defining the *reconstruction errors* as $\Delta_i(\mathbf{r}) \equiv x_i'(\mathbf{r}) - x_i(\mathbf{r})$, a straightforward calculation gives

$$\begin{aligned} \langle\Delta_i(\mathbf{r})\rangle &= 0, & (19)\\ \langle\Delta_i(\mathbf{r})^2\rangle &= \int\left[\sum_{j=1}^{n}\left|\left(\widehat{W}(\mathbf{k})F - I\right)_{ij}\right|^2 P_j^{(s)}(\mathbf{k})\right. \end{aligned}$$

$$+\sum_{j=1}^{m}\left|\widehat{W}_{ij}(\mathbf{k})\right|^2 P_j^{(n)}(\mathbf{k})\Bigg] d^2k, \quad (20)$$

independent of $\mathbf{r}$. Equation (19) merely tells us that our estimators of the fields are unbiased. Pursuing the analogy of ordinary Wiener filtering, we select the reconstruction matrix that minimizes the r.m.s. errors, *i.e.*, minimizes $\langle\Delta_i(\mathbf{r})^2\rangle$. We thus require $\delta\langle\Delta_i(\mathbf{r})^2\rangle = 0$, where the variation is carried out with respect to $\widehat{W}_{ij}$, and obtain

$$\sum_{k=1}^{n}\left(\widehat{W}(\mathbf{k})F - I\right)_{ik} F_{jk}P_k^{(s)}(\mathbf{k}) + \sum_{k=1}^{m}\widehat{W}_{ik}(\mathbf{k})P_k^{(n)}(\mathbf{k})\delta_{kj} = 0. \quad (21)$$

Defining the noise and signal matrices N and S by

$$N_{ij}(\mathbf{k}) \equiv \delta_{ij}P_i^{(n)}(\mathbf{k}) = \mathrm{diag}\{P_i^{(n)}(\mathbf{k})\}, \quad (22)$$
$$S_{ij}(\mathbf{k}) \equiv \delta_{ij}P_i^{(s)}(\mathbf{k}) = \mathrm{diag}\{P_i^{(s)}(\mathbf{k})\}, \quad (23)$$

equation (21) reduces to simply $\widehat{W}[FSF^t + N] - SF^t = 0$, which has the solution

$$\widehat{W}(\mathbf{k}) = S(\mathbf{k})F^t[FS(\mathbf{k})F^t + N(\mathbf{k})]^{-1}. \quad (24)$$

These are the appropriate formulae to use when limiting the attention to a rectangular patch of sky whose sides are small enough ($\ll$ one radian $\approx 60°$) that its curvature can be neglected. With all sky coverage, we need to solve the corresponding subtraction problem on a sphere instead, which is the topic of the next subsection.

3.3 The idealized spherical case

Above we saw that the optized subtraction became much simpler in Fourier space, where it became diagonal. In other words, although the linear combination mixed the m different frequencies, it never mixed Fourier coefficients corresponding to different wave vectors $\mathbf{k}$. The generalization of the derivation above to the case of fields on the celestial sphere is trivial, and not surprisingly, the corresponding natural basis functions in which the subtraction becomes diagonal are the spherical harmonics. Expanding all fields in spherical harmonics as in Section 2, and combining the observed $a_{\ell m}$-coefficients for the various frequency channels in the vector $\mathbf{a}_{\ell m}$, we can thus write our estimate of the true coefficients for the various components, denoted $\mathbf{a}'_{\ell m}$, as

$$\mathbf{a}'_{\ell m} = W^{(\ell)}\mathbf{a}_{\ell m}. \quad (25)$$

The analogues of equations (22)-(24), giving the reconstruction matrix $W^{(\ell)}$, become

$$\begin{aligned} W^{(\ell)} &= S^{(\ell)}F^t[FS^{(\ell)}F^t + N^{(\ell)}]^{-1}, & (26)\\ N^{(\ell)}_{ij} &\equiv \delta_{ij}C_\ell^{noise,i}, & (27)\\ S^{(\ell)}_{ij} &\equiv \delta_{ij}C_\ell^{(i)}. & (28)\end{aligned}$$

The corresponding reconstruction errors $\Delta a'^{(i)}_{lm}$ have their mean square value $\Delta C_\ell^{(i)} \equiv \langle|\Delta a'^{(i)}_{lm}|^2\rangle$ given by

$$\Delta C_\ell^{(i)} = \left[\sum_{j=1}^{n}\left|\left(W^{(\ell)}F - I\right)_{ij}\right|^2 C_\ell^{(j)} + \sum_{j=1}^{m}\left|W^{(\ell)}_{ij}\right|^2 C_\ell^{noise,j}\right]. \tag{29}$$

As we will see in subsection 4.3.1, the relevant power spectrum of the noise in channel i is simply $C_\ell^{noise,i} = 4\pi\sigma_i^2/N_i$, where σ_i is the r.m.s. pixel noise and N_i is the number of pixels. $C_\ell^{(i)}$ denotes the power spectrum of the i^{th} component at 100 GHz. In summary, the subtraction procedure is as follows: first the maps from all frequency channels are expanded in spherical harmonics, then the a_{lm}-coefficients of the various physical components are estimated as above, and finally the filtered maps are obtained by summing over these estimated coefficients, as in equation (5) with ν =100 GHz.

3.4 Pixelization and incomplete sky coverage

All real-world CMB maps are pixelized, *i.e.*, smoothed by some experimental beam and sampled only at a finite number of points. In addition, the presence of "dirty" regions such as the Galactic plane, the Large Magellanic Cloud, bright point sources, *etc*, means that we may want to throw away some of the pixels, leaving us with a map with a topology reminiscent of a Swiss cheese.

In Section 5, we will see that the subtraction technique is quite efficient in removing the various foregrounds from a CMB map. The reason that it works so well is that it takes advantage of the fact that the foregrounds have quite different power spectra, as summarized in Figure 14, by doing the subtraction multipole by multipole. With incomplete sky coverage, one cannot do quite this well, since it is impossible to compute the exact coefficients $a_{\ell m}$ using merely part of the sky. Instead of the spherical harmonics, we must expand our maps in some other set of basis functions, functions that vanish in all the "holes" in our maps. In contrast to the spherical harmonics, each of these functions will inevitably probe a range of ℓ-values, rather than just a single multipole, specified by a *window function* as described in T95. To exploit the fact that the various foregrounds have different power spectra C_ℓ, we clearly want these window functions to be as narrow as possible.

A prescription for how to calculate such basis functions, taking incomplete sky coverage, pixelization, and position-dependent noise into account, is given in T95, and it is found that given a patch of sky whose smallest angular dimension is $\Delta\theta$, each basis function will probe an ℓ-band of width $\Delta\ell \approx 60^\circ/\Delta\theta$. For instance, if we restrict our analysis to a $10^\circ \times 10^\circ$ square, then $\Delta\ell \approx 6$. This is very good news. It means that the only performance degradation in the subtraction technique will stem from the fact that it is unable to take advantage of sharp features in the power spectra of width $\Delta\ell \approx 6$ or smaller. This is essentially no loss at all, since as discussed in Section 4, we expect all the foregrounds to have fairly smooth power spectra, without any sharp spikes or discontinuities.

Whatever set of orthonormal basis functions is chosen for the analysis, our multi-frequency subtraction prescription is to expand all maps in these functions and do the estimation separately for each of the expansion coefficients. For any one basis function (corresponding to a set of weights w_k, one for each pixel k), there will be such a coefficient a_i for each of the m frequency channels, and we combine these into the m-dimensional vector $\mathbf{a}$. The subtraction now decomposes into the following steps:

1. Compute σ_i^2, the variance in a_i that is due to pixel noise (this variance is simply a weighted sum of the noise variance in each pixel, the weights being w_k^2).

2. Compute Δ_i^2, the 100 GHz variance in $\mathbf{a}$ that is due to the i^{th} physical component, as in T95 (this variance depends only on the power spectra and the weights w_k).

3. Compute the estimated coefficients for the n different components, denoted $\mathbf{a}'$.

The last step is of course analogous to the cases we discussed above, *i.e.*, $\mathbf{a}' = W\mathbf{a}$, where

$$W = SF^t[FSF^t + N]^{-1}, \tag{30}$$

$$N_{ij} \equiv \delta_{ij}\sigma_i^2, \tag{31}$$

$$S_{ij} \equiv \delta_{ij}\Delta_i^2. \tag{32}$$

Let us illustrate this with the simple example of a small square region, sampled in a square grid of $N\times N$ points with say $N = 512$. A convenient set of basis functions is then the discrete Fourier basis, for which our subtraction would reduce to the following steps:

1. Fast Fourier transform (FFT) the data.
2. Filter as above, separately for each of the N^2 Fourier coefficients.
3. Perform an inverse FFT to recover the filtered maps.

To do still better, we can use the optimal basis functions of T95. For this simple case, they turn out to be simply the Fourier basis functions, but weighted by a two-dimensional cosine "bell" $\cos x\cos y$ so that they go smoothly to zero at the boundary of the square. Thus the prescription becomes

1. Multiply by cosine bell
2. FFT
3. Filter
4. Inverse FFT
5. Divide by cosine bell

The resulting map will be very accurate in the central parts of the square, but the noise levels will explode towards the edges where the cosine bell goes to zero. Thus the way to make efficient use of this technique is to tile the sky into a mosaic of squares with considerable overlap, so that one can produce a low noise composite map using only the central regions of each square.

3.5 Non-Gaussianity and lack of translational invariance

In the above treatment, we assumed that the statistical properties of all random fields were translationally invariant, so that our only a priori knowledge about them was their power spectra. In reality, this is of course not the case. A flagrant counterexample is the Galactic plane, where we expect much larger fluctuations in the dust, synchrotron and free-free components than at high galactic latitude. In addition, most of the foregrounds exhibit non-Gaussian behavior. We wish to emphasize that for the purposes of estimating the underlying CMB-fluctuations, all of these features work to our advantage. If we know the power spectrum of a contaminant, then translational invariance and Gaussianity means that we have no additional knowledge whatsoever about the contaminant, since the power spectrum defines the random field completely. Clearly, the more we know about our enemy, the greater our ability will be to tackle it and distinguish it from CMB fluctuations.

The type of non-Gaussianity that we encounter in both diffuse components and point sources manifests itself in that the trouble is more spatially localized than it would be for a Gaussian random field with the same power spectrum. A bright point source affects only a very small region of the celestial sphere, of the order of the beam width, which can simply be removed (perhaps by using higher resolution observations at lower sensitivities, see *e.g.* O'Sullivan *et al.* 1995). Dust emission, free-free emission and synchrotron radiation tends to be localized to "dirty regions", with the fluctuation levels in "clean regions" in some cases being orders of magnitude lower. Again, we can take advantage of this non-Gaussianity (and lack of translational invariance when we know the cause of the emission, such as the Galactic plane), to simply remove these regions from the data set. After this initial step, the analysis proceeds as described in the previous subsection, using the power spectra that are appropriate for the clean regions. Thus we sift out the CMB fluctuations from the foregrounds in a two-step process, by exploiting the fact that their statistical properties are different both in real space and in Fourier space:

1. We place most of the weight on the clean regions in real space.

2. We place most of the weight on the clean regions in Fourier space, as illustrated in Figure 14.

3.6 How to avoid distorting the CMB power spectrum

Although the community has displayed considerable interest in map-making, there are of course many cases where one is merely interested in measuring the CMB power spectrum, for instance to constrain the parameters of theories of the formation of structure (*e.g.* the amplitude and spectral index of the initial irregularities). Rather than first generate a map with the method presented above and then use it to estimate the power spectrum, the latter can of course be obtained directly by aborting the reconstruction "half way through". Thus we can estimate C_{ℓ^*} by first estimating the $(2\ell^*+1)$ coefficients $a_{\ell m}$ that have $\ell=\ell^*$, as described above, and then taking some appropriate weighted average of the estimated $|a_{\ell m}|^2$, to reduce cosmic variance. For technical details on the choice of basis functions, the best weights to use for the averaging, *etc.*, see T95.

When using our subtraction technique to estimate power spectra, the normalization must be modified as described below. The reason for this is the above-mentioned fact that Wiener filtering tends to "suck power" out of the data, so that the power spectrum of the filtered map is smaller than the true power spectrum. Moreover, this power deficit normally depends on scale, as indicated by equation (10).

Let us use the notation of subsection 3.2 and investigate how the quantity $|a'_{\ell m}|^2$ is related to the power spectrum C_ℓ. To avoid unnecessary profusion of indices, let us focus on one single multipole, say $\ell = 17$, $m = 5$, and suppress the indices ℓ and m throughout. Thus the vector $\mathbf{a}$ contains the multipole coefficients from the m different frequency channels, and $\mathbf{a}'$ the coefficients for the n different physical components, as before. A straightforward calculation shows that (no summation implied)

$$\langle |a'_i|^2 \rangle = |V_{ii}|^2 C^{(i)} + b_i, \tag{33}$$

where $V \equiv WF$, and the b_i, the additive *bias*, is given by

$$b_i \equiv \sum_{j \neq i} |V_{ij}|^2 C^{(j)} + \sum_{j=1}^{m} |W_{ij}|^2 C^{noise,j}. \tag{34}$$

The power estimator

$$C^{(i)\prime} \equiv |a'_i|^2 - b_i \tag{35}$$

will thus be an unbiased estimator of the true power $C^{(i)}$, *i.e.*, $\langle C^{(i)\prime} \rangle = C^{(i)}$, if we impose the normalization constraint $V_{ii} = 1$ for all $i = 1, ..., n$. As is seen in equation (34), b_i incorporates the power leakage from the other physical components ($j \neq i$) and from the pixel noise. Note that when $V_{ii} = 1$, b_i equals $\Delta C^{(i)}$, the reconstruction errors of equation (29).

Let us minimize b_i subject to the constraint that $V_{ii} = 1$. Introducing the Lagrange multipliers λ_i, we thus differentiate $L_i \equiv b_i - \lambda_i V_{ii}$ (no summation) with respect to the components of the matrix W and require the result to vanish. After a straightforward calculation, we obtain the solution

$$W = \Lambda F^t [FSF^t + N]^{-1}, \tag{36}$$

where the matrix $\Lambda \equiv \mathrm{diag}\{\lambda_1, ..., \lambda_n\}$. Imposing the normalization constraints $V_{ii} = 1$ now gives $\lambda_i = 1/(F^t[FSF^t + N]^{-1}F)_{ii}$. Comparing this to equation (26), we draw the following conclusion:

- *Our optimized power spectrum estimate uses the same reconstruction matrix W that we derived previously, except that the row vectors should be rescaled so that $(WF)_{ii} = 1$.*

(The extra matrix S in equation (26) is of course irrelevant here, as it is diagonal and can be absorbed into Λ.) This is the normalization that has been used in Figure 13. Since one of the main purposes of CMB sky maps is to serve as an easy-to-visualize compliment to the power spectrum, we strongly advocate using the above normalization convention $(WF)_{ii} = 1$ when generating sky maps as well. As we saw above, this will ensure that CMB fluctuations in the map will retain their true power spectrum, rather than suffer the ℓ-dependent suppression characteristic of Wiener filtering.

Let us compare this with the situation in standard Wiener filtering, which corresponds to $m = n = 1$. In this simple case, F and W are merely scalars, so the normalization condition gives $V = WF = 1$. Thus W equals a constant $1/F$ which is independent of ℓ, corresponding to no subtraction at all. In other words, if there is only one frequency channel, our subtraction method is of no use for power spectrum estimation. In the general case, there are $m \times n$ components in W and n constraints $V_{ii} = 1$, so the subtraction method will help whenever m, the number of channels, exceeds one.

An attractive feature of the reconstruction method given by equation (36) is that it gives a reconstructed CMB map that is completely independent of our assumptions about the CMB power spectrum. More formally, W_{ij} is independent of $S_{ii} = C^{(i)}$. This might seem surprising, since S enters in the right-hand side of equation (36). The easiest way to prove this result is to note that since the optimization problem is independent of the assumed CMB power spectrum (both the target function b_i and the constraint equation $(WF)_{ii}$ are independent of $C^{(i)}$), its solution (the i^{th} row of W) must be as well.

Above we chose our filter to minimize b_i, the total contribution from the other physical components and pixel noise. This of course produces a robust power spectrum estimator, since if our estimate of the power spectrum of some contaminant (or our estimate of the noise level of some channel) is off by some number of percent, the resulting error will scale as the corresponding term in b_i, *e.g.*, as $|V_{ij}|^2 C^{(j)}$ (or as $|W_{ij}|^2 C^{noise,j}$). If one is confident that there are no such systematic errors, one may instead opt to minimize the variance of our estimator C_i', which is equivalent to minimizing the variance of b_i. This would lead to a system of cubic (rather than linear) equations for the components of W, to be solved numerically.

4 POWER SPECTRA OF THE FOREGROUNDS

In this Section, we make estimates of the angular power spectra $C_\ell(\nu)$ for the various foregrounds. The results are plotted in Figures 3 though 6, and summarized in Figure 14. The former are truncated at $[\ell(2\ell+1)C_\ell/4\pi]^{1/2} = \sqrt{2}\times 20\mu\mathrm{K} \approx 28\mu\mathrm{K}$, which approximately corresponds to COBE-normalized scale-invariant temperature fluctuations. Thus the shaded regions in Figure 14 are simply the top contours of Figures 3 though 6. It should be emphasized that these estimates are *not* intended to be very accurate, especially when it comes to normalization. Rather, the emphasis is on their *qualitative* features, especially those that differentiate them from one another. Despite the fact that we currently lack accurate high-resolution data in many important frequency bands, we will see that quite robust qualitative conclusions can be drawn about which regions of the $\ell-\nu$-plane will be most suitable for estimating various parts of the CMB power spectrum.

4.1 Point sources

In this Section, we make estimates of the angular power spectrum $C_\ell(\nu)$ for point sources. Here the ℓ-dependence is well known, but the ν-depencence quite uncertain. However, despite these uncertainties, we will see that radio point sources will be contribute mainly to the lower right corner of Figure 14, whereas infrared point sources will contribute mainly to the upper right.

If at some frequency there are N point sources Poisson distributed over the whole sky, all with the same flux ϕ, is is easy to show that

$$\langle a_{\ell m}\rangle = \begin{cases} \sqrt{4\pi}\bar{n}\phi & \text{if } \ell = 0, \\ 0 & \text{if } \ell \neq 0, \end{cases} \tag{37}$$

where $\bar{n} \equiv N/4\pi$ is the average number density per steradian, and

$$C_\ell \equiv \langle |a_{\ell m}|^2\rangle - |\langle a_{\ell m}\rangle|^2 = \bar{n}\phi^2. \tag{38}$$

In other words, this would produce a simple white-noise power spectrum, with the same power in all multipoles, together with a non-zero monopole caused by the fact that no fluxes are negative. If there are two independent Poisson populations, with densities $\bar{n}_1$ and $\bar{n}_2$ and fluxes ϕ_1 and ϕ_2, both the means and the variances will of course add, giving a monopole $\sqrt{4\pi}(\bar{n}_1\phi_1 + \bar{n}_2\phi_2)$ and a power spectrum $C_\ell = \bar{n}_1\phi_1^2 + \bar{n}_2\phi_2^2$. Taking the limit of infinitely many populations, we thus obtain

$$\langle a_{00}\rangle = \sqrt{4\pi}\int_0^{\phi_c}\frac{\partial\bar{n}}{\partial\phi}\phi\,d\phi, \tag{39}$$

$$C_\ell = \int_0^{\phi_c}\frac{\partial\bar{n}}{\partial\phi}\phi^2 d\phi, \tag{40}$$

where $\frac{\partial\bar{n}}{\partial\phi}$ is the *differential source count*. In other words, we have defined $\bar{n}(\phi)$ as the number density per steradian of sources with flux less than ϕ. In real life, we are of course far from powerless against these point sources, and can either attempt to subtract them by using spectral information from point source catalogues, or simply choose to throw away all pixels containing a bright point source. In either case, the end result would be that we eliminate all sources with a flux exceeding some flux cut ϕ_c, which then becomes the upper limit of integration in equations (39) and (40). We have estimated the source counts at 1.5 GHz from a preliminary point source catalog from the VLA FIRST all sky survey (Becker *et al.* 1995). This catalog contains 16272 radio sources in a narrow strip $110° < \mathrm{ra} < 195°$, $28.5° < \mathrm{dec} < 31.0°$, complete down to a flux limit of 0.75 mJy. A flux histogram is plotted in Figure 7, together with a simple double power law fit

$$\frac{\partial\bar{n}}{\partial\phi} \approx \frac{524000}{\mathrm{mJy\,sr}}\left(\frac{\phi}{0.75\mathrm{mJy}}\right)^{-1.65}\left(1+\frac{\phi}{100\mathrm{mJy}}\right)^{-1} \tag{41}$$

that will be quite adequate for our purposes. We can obviously never eliminate *all* radio sources, as there is for all practical purposes an infinite number of them, the integral of the differential source count diverging at the faint end. There is also a rather obvious lower limit to ϕ_c in practice. Since the highest resolution COBRAS/SAMBA channels have a FWHM of 4.5 arcminutes (see Table 1, Section 4.3.3 below), there are only about 10^7 independent pixels in the sky. Assuming that the above-mentioned FIRST data is representative of the entire sky, there are about 6 million sources brighter than 0.75 mJy, so if we choose ϕ_c much lower than this and reject all data that is contaminated at this level, we would have to throw away almost all our pixels. The subtraction strategy also has its limits, quite apart from the large amount of work that would be involved: if we try to model and subtract the sources, it appears unlikely that we will be ever to do this with an accuracy exceeding 1%, and even 10% could prove difficult given complications such as source variability. Since the choice of flux cut will depend on the level of ambition of future observing projects, we simply compute how the resulting power spectrum depends on the flux cut ϕ_c. To give a rough idea of what flux cuts may be practically feasible in the near future, the number of radio sources in the entire sky are about 4×10^6 above 1 mJy, 8×10^5 above 10 mJy, 7×10^4 above 100 mJy and 800 above 1 Jy, all at 1.5 GHz. The result, computed using equations (3), (39), (40), and (41), is shown in Figure 8. Notice that the fluctuations have quite a different magnitude and ϕ_c-scaling than the monopole, since the two are dominated by quite different parts of the differential source count function. Since the slope is close to -2, the monopole, the total brightness, gets similar contributions from several different decades of flux, whereas the fluctuations are strongly dominated by the brightest sources. Thus we need not worry about not knowing the exact differential source count at the faint end, as all that really matters is its behavior immediately below our flux cut.

Some authors (*e.g.*, Franceschini *et al.* 1989) have raised the possibility that point source clustering could create more large-scale power that the Poisson assumption would indicate. We have tested this by computing the power spectrum of the FIRST data, and find no evidence for any departure from Poisson noise (see also Benn & Wall 1995). This conclusion, which of course simplifies the issue considerably, is not surprising, because most of the sources are located at very large distances. Correlations in the two-dimensional galaxy distribution that we observe (and which is the relevant quantity when it comes to CMB contamination) are therefore diluted by projection to negligible levels.

Although there is good reason to believe that the power spectrum will remain Poissonian at the higher frequencies that are relevant to the CMB, the issue of its normalization is of course quite complex, given the uncertainties about the spectra and the evolution of the various galaxy and AGN populations (see Franceschini *etal* 1991). In Figure 5, we have simply made a 100 mJy flux cut at 1.5 GHz (for an all-sky survey, this corresponds to removing about 70000 sources) and extrapolated to higher frequencies with a power law $B(\nu)\propto\nu^{-\alpha}$, thus obtaining

$$\left[\frac{2\ell+1}{4\pi}\ell C_\ell^{ps}(\nu)\right]^{1/2}\approx 0.30\mathrm{K}\left(\frac{\sinh^2(x/2)}{[\nu/1.5\,\mathrm{GHz}]^{4+\alpha}}\right)\ell, \qquad (42)$$

where $x=h\nu/kT_0$ as before. For $\ell=100$, this corresponds to $1.5\,\mu$K at 100 GHz if $\alpha=0$. Lowering the flux cut to 10 mJy (removing about 900000 sources) reduces this by about a factor 4, and a 1 mJy cut (removing about 5 million sources) gains us another factor of four. Obviously, ambitious flux cuts become feasible if only a small fraction of the sky is surveyed. Flat-spectrum sources with spectral index $\alpha \approx 0.3$ are likely to dominate at higher frequencies (Franceschini *et al.* 1989), but this is of course only to be used as a crude first approximation, as the emission at higher frequencies is likely to be dominated by sources whose spectra rise and peak near those frequencies, and very little is known about the abundances of such objects. We make the rather cautious assumption of an effective spectral index $\alpha = 0.0$ for the population as a whole. This approximation is of course quite unsatisfactory at the high-frequency end, where infrared emission from high redshift galaxies could play an important role. For instance, if this emission is dominated by dust in these galaxies with emissivity $\beta = 2$ (see the following Section), we would expect $\alpha = -4$ to be a better description at the higher microwave frequencies. Unfortunately, the differential source counts of such infrared point sources around 100 GHz is still completely unknown. For a recent review of these issues, see Franceschini *et al.* (1991).

4.2 Diffuse Galactic sources

In this Section, we discuss the qualitative features we expect for the angular power spectra of the diffuse Galactic contaminants, namely dust, free-free emission and synchrotron radiation.

4.2.1 Power spectrum

We have estimated the power spectrum of Galactic dust from a large number of $12.5° \times 12.5°$ fields of the 100 micron IRAS all-sky survey (Neugebauer *et al.* 1984), which have an angular resolution of two arcminutes (about twice as good as the best COBRAS/SAMBA channels). Although the amplitude varies greatly with Galactic latitude, the overall shape is strikingly independent of latitude, and typically declines as $C_\ell \propto 1/\ell^3$ for ℓ between 100 and a few thousand, steepening slightly on the smallest scales. This agrees well with previous findings (Low & Cutri 1994; Guarini *et al.* 1995). To estimate the power spectrum of synchrotron radiation, we used the Haslam 408 GHz map (Haslam *et al.* 1982). Although the angular resolution of this map is only of order $0.85°$, *i.e.*, far too low to provide information for $\ell \gg 100$, the logarithmic slope was found to be consistent with that for dust in the overlapping multipole range; around -3. These results are hardly surprising, since even without analyzing observational data, one may be able to guess the qualitative features of the power spectra of the three diffuse components.

Since they are all caused by emission from diffuse blobs, one might expect their power spectra to exhibit the following characteristic features:

- C_ℓ independent of ℓ for small ℓ, corresponding to scales much greater than the coherence length of the blobs (this is the standard Poisson behavior, and follows if one assumes that well separated blobs are uncorrelated).

- C_ℓ falls off at least as fast as $1/\ell^4$ for very large ℓ, corresponding to scales much smaller than typical blob sizes (this follows from the simple assumption that the brightness is a *continuous* function of position).

- If $\ell^2 C_\ell$ thus decreases both as ℓ gets small and as ℓ gets large, it must peak at some scale, a scale which we refer to as the coherence scale.

The behavior of the contaminant power spectrum for very small ℓ (whether there is indeed a coherence scale, *etc*), is of course quite a subtle one, as the presence of the Galactic plane means that the answer will be strongly dependent on which patches of sky we choose to mask out during the analysis. We will return to this issue in the subsection about non-Gaussianity below. In the figures, we have simply assumed that all three components have a coherence scale of about 10°, corresponding to $\ell \approx 10$, and used power spectra of the simple form $C_\ell \propto (5+l)^{-3}$.

4.2.2 Frequency dependence

The frequency dependence of the three components has been extensively discussed in the literature (see *e.g.* Reach *et al.* 1995 and references therein). For synchrotron radiation and free-free emission, we use simple power laws $B(\nu) \propto \nu^{-\beta}$. For synchrotron emission, $\beta \approx 0.75$ below 10 GHz (de Bernardis *et al.* 1991), steepening to $\beta \sim 1$ above 10 GHz (Banday & Wolfendale 1991), so we simply assume $\beta = 1$ here. For free-free emission, we make the standard assumption $\beta = 0.1$. For dust, we assume a spectrum of the standard form

$$C_\ell^{dust} \propto \frac{\nu^{3+\beta}}{e^{h\nu/kT}-1}. \tag{43}$$

Although an emissivity index $\beta = 2$ is found to be a good fit in the Galactic plane (Wright *et al.* 1991), we use instead the more conservative parameters $T = 20.7$K, $\beta = 1.36$, which are found to better describe the data at high

Galactic latitudes (Reach *et al.* 1995), since it is of of course the cleanest regions of the sky that are the most relevant ones for measurement of CMB fluctuations.

4.2.3 Non-Gaussianity and inhomogeneity

The spatial distributions of synchrotron radiation, free-free and dust emission of course exhibit strong non-Gaussian features, and also a strong departure from translational invariance because of the Galactic plane. As discussed in Section 3, this is good news regarding our ability to estimate CMB fluctuations. However, it forces us to be careful when presenting plots of estimated power spectra. Thus plots showing foreground contributions to the CMB, such as those presented in this paper, should be read with the following two caveats in mind.

First, one of the manifestations of the type of non-Gaussianity that these components display is the presence of "clean regions" and "dirty regions". For instance, a raw histogram of the brightness per pixel in the DIRBE 240 micron map shows that although 3% of the pixels have a brightness exceeding 100 MJy/sr, the mean is only about 6 MJy/sr. The extremely bright pixels are of course mainly located in the Galactic plane, but the level of "cleanness" also exhibits strong variations with Galactic longitude, caused both by known objects such as the Large Magellanic Cloud and the North Galactic Spur, and by the non-Gaussian clumpiness of the dust component itself. Similar conclusions follow from an analysis of the IRAS 100 micron maps. The result of this is that although the power spectrum may have a similar shape in clean and dirty regions, the normalization will vary considerably, much more than it would due to sample variance in a Gaussian field. It is thus important that plots of C_ℓ-estimates are supplemented with a description of what type of region they refer to. In the figures in this paper, all such power spectra refer to averages for *the cleanest 20% of the sky.*

Secondly, when estimating the lowest multipoles, it is important to use as much of the celestial sphere as possible, to keep the window functions in ℓ-space narrow (T95). Of course, the contribution of Galactic foregrounds increases as one includes more sky, but this is unlikely to be a serious problem. The cleanest two thirds of the DIRBE 240 micron pixels have an average brightness about three times that of the cleanest 20% and this is a large enough area to recover all multipoles fairly accurately except the quadrupole and octupole (T95). Thus if we increase the sky coverage to estimate the lowest multipoles more accurately, the Galactic foregrounds are likely to be within a factor of a few of the contributions from the cleanest regions of the sky, and perhaps less if the contaminant power $\ell^2 C_\ell$ falls of on scales larger than some coherence scale.

4.3 The effects of discreteness, pixel noise and beam smoothing

Although we usually think of pixel noise as a problem of a different nature than the other contaminants, it can be described by an angular power spectrum $C_\ell^{noise}(\nu)$ and so be treated on an equal footing. One may ask what is the point of doing this, since the statistical impact of the noise on the subtraction described in this paper is straightforward to calculate anyway. The answer is that it provides better physical intuition. Real world brightness data is of course discretely sampled as "pixels" rather than smooth functions known at every point $\widehat{\mathbf{r}}$, but as long as the sampling is sufficiently dense (the typical pixel separation being a few times smaller than the beamwidth), this discreteness is merely a rather irrelevant technical detail. It enters when we do the analysis in practice, but our results are virtually the same as if we had continuous sampling.

4.3.1 Pixel noise

If we estimate the angular power spectrum from a sky map containing only isotropic pixel noise, we find that all the C_ℓ-coefficients are equal (the white noise power spectrum), at least down to the angular scale corresponding to the inter-pixel separation. This well-known result simply reflects the fact that the noise in the different pixels is uncorrelated. We will now elaborate on this in slightly greater detail.

For a CMB sky map at frequency ν, pixelized into N pixels pointing in the directions $\widehat{\mathbf{r}}_i$ and with noise n_i, $i = 1, 2, ..., N$, one typically has to a good approximation that the n_i are Gaussian random variables satisfying $\langle n_i \rangle = 0$ and

$$\langle n_i n_j \rangle = \delta_{ij} \sigma_i^2, \tag{44}$$

for some known numbers σ_i. We want to eliminate this discreteness from the problem, and describe the noise as a continuous field $B_{noise}(\widehat{\mathbf{r}}, \nu)$ instead, a random field that gets added to the actual brightness $B(\widehat{\mathbf{r}}, \nu)$. More specifically, for any weight function $\psi(\widehat{\mathbf{r}})$ that we use in our analysis, we want the result to be the same whether we sum over the pixels or integrate over the field, so we require

$$\frac{1}{N}\sum_{i=1}^{N}\psi(\widehat{\mathbf{r}}_i)n_i \approx \frac{1}{4\pi}\int\psi(\widehat{\mathbf{r}})B_{noise}(\mathbf{r},\nu)d\Omega. \tag{45}$$

Fortunately, this is easy to arrange: when the pixels are placed according to an equal-area method (as in the COBE pixelization system), a simple choice that works is to choose $B_{noise}(\widehat{\mathbf{r}})$ to be a white noise field satisfying

$$\langle B_{noise}(\widehat{\mathbf{r}})B_{noise}(\widehat{\mathbf{r}}')\rangle = \delta(\widehat{\mathbf{r}},\widehat{\mathbf{r}}')\frac{4\pi}{N}\sigma(\widehat{\mathbf{r}})^2, \tag{46}$$

where δ is the angular Dirac delta function, and $\sigma(\widehat{\mathbf{r}})$ denotes the σ_i corresponding to the pixel position closest to $\widehat{\mathbf{r}}$. Since white noise by definition has no correlations on any scale, it is easy to see why this reproduces the basic feature of the pixel noise, *i.e.*, no correlation between neighbouring pixels. The fact that white noise fluctuates wildly on sub-pixel scales does not invalidate equation (45), since any weighting function ψ that we use in practice cannot vary on sub-pixels scales (since we will after all apply it to pixels), and thus smoothes out this substructure.

For the purposes of analyzing future experiments, let us assume that the pixel noise is independent of position, so that σ_i simply equals some constant, σ. This means that the white noise is isotropic, and has a well-defined angular power spectrum C_ℓ^{noise} that we will now compute. For white noise, the power spectrum is independent of ℓ, and thus all we need to do is find the overall normalization, expressed in terms of the pixel noise and number of pixels. We do this by examining the simplest multipole, $\ell = 0$. Since $Y_{00} = 1/\sqrt{4\pi}$, equation (5) gives

$$a_{00}^{noise} = \sqrt{4\pi}\left(\frac{1}{4\pi}\int\delta T_{noise}(\widehat{\mathbf{r}})d\Omega\right), \tag{47}$$

where we have factored out $\sqrt{4\pi}$ so that the expression in parenthesis is the sky-averaged temperature. Replacing this by the pixel-averaged temperature (simply using equation (45) with $\psi = 1$), we obtain

$$a_{00}^{noise} = \sqrt{4\pi}\left(\frac{1}{N}\sum_{i=1}^{N}\delta T_{noise}(\widehat{\mathbf{r}}_i)\right). \tag{48}$$

Using equation (44), this leaves us with our desired result,

$$C_\ell^{noise} = C_0^{noise} = \left\langle|a_{00}|^2\right\rangle = \frac{4\pi}{N}\sigma^2. \tag{49}$$

It is straightforward to verify that this normalization agrees with that in equation (46).

4.3.2 Beam smoothing

In this subsection, we point out that the effects of beam smoothing can be completely absorbed into the description of the noise.

In a single-beam experiment, the brightness field B^{obs} is the true field B convolved with a beam function w,

$$B^{obs}(\widehat{\mathbf{r}}) = \int w(\theta)B(\widehat{\mathbf{r}}')d\Omega', \tag{50}$$

where $\theta \equiv \cos^{-1}(\widehat{\mathbf{r}}\cdot\widehat{\mathbf{r}}')$ is the angle between the two vectors. As long as the beam width is much less than a radian, expanding this in spherical harmonics gives the familiar result

$$C_\ell^{obs} \approx |\widehat{w}(\ell)|^2 C_\ell, \tag{51}$$

where $\widehat{w}$ is the Fourier transform of w. In the common approximation that the beam profile is a Gaussian,

$$w(\theta) = \frac{1}{\sqrt{2\pi}\theta_b}e^{-\frac{1}{2}\frac{\theta^2}{\theta_b^2}}, \tag{52}$$

equation (51) reduces to

$$C_\ell^{obs} \approx e^{-\theta_b^2\ell(\ell+1)}C_\ell. \tag{53}$$

The standard way to quote the beam width is to give the full-width-half-max (FWHM) width of w, so the correspondence is $\theta_b = \mathrm{FWHM}/\sqrt{8\ln 2} \approx 0.425\,\mathrm{FWHM}$.
This suppression of high multipoles by beam smoothing of course affects the fields B_i for all the different components (B_{cmb}, B_{dust}, *etc.*) except one. The exception is B_{noise}, since the pixel noise gets added after the beam smoothing, and thus has nothing to do with the beam width. Although it is easy to convert between actual and observed fields (using equation (51) for the power spectrum or a simple deconvolution for the fields $B_i(\widehat{\mathbf{r}})$), it is quite a nuisance to always have to distinguish between the two. Since B_{noise} is the only field that is simpler to describe "as observed", we adopt the following convention:

Channel	1	2	3	4	5	6	7	8	9
Center frequency ν [GHz]	31.5	53	90	125	143	217	353	545	857
Bandwidth $\Delta\nu/\nu$	0.15	0.15	0.15	0.15	0.35	0.35	0.35	0.35	0.35
FWHM beam size [arcmin]	30	18	12	10	10.5	7.5	4.5	4.5	4.5
Pixel noise $\sigma(\nu)$, 2 years [μK]	17.7	18.3	23.7	69.5	5.7	6.0	36.0	271	62700

Table 1: COBRAS/SAMBA channel specification

- We let all fields $B_i(\widehat{\mathbf{r}},\nu)$ refer to the *actual* fields, as they would appear if there were no beam smoothing.

We thus define the "unsmoothed" noise field as

$$C_\ell^{noise}(\nu) \equiv \frac{4\pi\sigma(\nu)^2/N}{|\widehat{w}(\ell)|^2}. \tag{54}$$

This is what is plotted in Figure 6. In other words, the advantages of this convention are that

- this figure can be directly compared with those for the various foregrounds, and
- the latter figures are independent of any assumptions about the beam width.

4.3.3 COBRAS/SAMBA specifics

The proposed COBRAS/SAMBA satellite mission (Mandolesi *et al.* 1995) is currently being evaluated by the European Space Agency, and if approved, is scheduled for launch in 2003. As currently proposed, it would have nine frequency channels, as summarized in Table 1. Although we will use the exact numbers from the this table in the analysis in Section 5, we have made a few simplifying approximations in generating Figure 6, since we want this plot to be approximately applicable to any satellite mission using similar technology.

Sensitivity: The reason that $\sigma(\nu)$ explodes for large ν is of course the exponential fall-off of the Planck-spectrum. Converting σ to brightness fluctuations using equation (3), one finds that to a crude approximation, $\sigma(\nu) \propto \nu$. The main deviation from this power law occurs around 130 GHz, where the

the technology transition from HEMT to bolometer detectors between channels 4 and 5 can be seen to cause the noise to drop by an order of magnitude in Table 1. In generating figures 6 and 14, we have simply interpolated the tabulated values with a cubic spline.

Number of pixels: Typically, the number of pixels is chosen so that neighbouring pixels overlap slightly. Since the combined area of all pixels is proportional to $\mathrm{FWHM}^2 N$, it is therefore convenient to define the dimensionless quantity $\eta \equiv \mathrm{FWHM}^2 N/4\pi$. In terms of η, which is thus a measure of the rate of oversampling, we can write the numerator of equation (54) as

$$\frac{4\pi\sigma^2}{N} = \frac{\mathrm{FWHM}^2\sigma^2}{\eta}, \tag{55}$$

e.g., in terms of the quantities that are usually quoted in experimental specifications. The COBE DMR experiment had FWHM = $7.08°$ and N=6144 pixels, which gives an oversampling factor $\eta \approx 7.47$. In this paper, we will assume that the COBRAS/SAMBA experiment will use this same degree of oversampling, in all channels.

Beam width: From Table 1, we see that the resolution is diffraction limited ($\mathrm{FWHM} \propto 1/\nu$) at low frequencies, corresponding to a mirror size of order 1 meter, whereas it is constant at the highest frequencies. In generating figures 6 and 14, we have simply interpolated the tabulated values with a cubic spline.

5 RESULTS

We have computed the signal-to-noise ratio obtainable with the technique presented in Section 3 assuming that the foregrounds behave as conjectured in Section 4. We have taken $m = 9$ and $n = 5$, corresponding to the nine COBRAS/SAMBA channels specified in Table 1 and the five components CMB, dust, radio point sources, synchrotron radiation and free-free emission.

5.1 Multi-frequency subtraction versus no subtraction

The resulting reconstruction errors ΔC_l computed from equation (29) are shown in Figure 13 (bottom heavy line, delimiting the double-hatched region), and are seen to lie about three orders of magnitude beneath a CDM power spectrum, corresponding to a signal-to-noise ratio of about 10^3 for CMB map reconstruction (referring to all foregrounds collectively as "noise" in this context). For estimating the power spectrum C_ℓ, a quadratic quantity, the corresponding signal-to-noise ratio would be an outrageous 10^6 for $\ell \lesssim 1000$. Clearly, this sort of accuracy will not be attainable in practice, because we do not know the frequency dependence of the various contaminants accurately enough to trust the extrapolations involved in the subtraction. We will return to this issue below. Thus interpreting this as an upper limit to how well we can hope to do, we also wish to place a lower limit on how poorly we can do. The most simplistic approach possible is of course making no attempts whatsoever at foreground subtraction, and using merely a single channel. The resulting errors for the 143 GHz channel are given by the uppermost curve in the figure. A slight improvement would be the solid curve below it, which corresponds to choosing the single best COBRAS/SAMBA channel separately for each multipole. The frequency where the combined foreground contribution is minimized is plotted as a function of ℓ in Figure 14 (heavy dashed line). The reason for its positive slope is of course that when ℓ increases, the power spectrum of dust decreases whereas the power spectrum of radio point sources, attacking from below, increases. This same effect is illustrated in Figure 12, which shows the total contribution from all foregrounds to the observed C_l as a function of frequency. It is seen that although the most promising frequency for estimating low multipoles is $\nu \sim 50$GHz, the central COBE channel, the optimal frequency for searching for CDM Doppler peaks ($\ell \gtrsim 200$) is around 100 GHz. The situation at three of the COBRAS/SAMBA frequencies is summarized in

Figures 9 through 11.

5.2 Why pixel-by-pixel subtraction performs so much worse

An alternative subtraction method would be to combine the smaller pixels of the high channels into 30 arcminute pixels, and then simply apply our subtraction scheme on a pixel-by-pixel basis, using equations (30)-(32) as follows:

1. Compute the pixel variance Δ_i^2 contributed by each component, as $\Delta_i^2 = \sum_\ell (2\ell+1) C_l^{(i)} |\widehat{w}(\ell+1/2)|^2/4\pi$.
2. Use the pixel variance of Table 1 for σ_i^2 in the matrix N.
3. Use the resulting matrix W to reconstruct the CMB temperature for each pixel.
4. Estimate C_ℓ by expanding the resulting CMB map in spherical harmonics.

Loosely speaking, this method corresponds to subtracting before Fourier transforming, rather than vice versa as in the optimized method. We found that this method produced a signal-to-noise ratio of about 15 for each pixel. The resulting reconstruction errors are shown by the heavy dashed line in Figure 13, and are seen to be more than an order of magnitude worse than the optimized method for the small ℓ-values where the low ($30'$) resolution is not a problem. This degradation in performance may seem surprising, since the same (overly optimistic) assumption that we know the the frequency dependence of all components is made with this method. However, the cause is exactly the same as that of the discrepancy between the two uppermost curves in the figure: the optimal channel weighting is different for high and low multipoles. Thus the pixel-by-pixel approach selects one single channel weighting that does not perform unduly badly for any multipole, at the price of not doing especially well at any multipole either. The optimized method, on the other hand, tailors the channel weighting separately for each multipole.

5.3 Why "direct subtraction" does so poorly

Another natural method would be to simply select as many channels as there are components, and reconstruct the true fields by merely inverting the matrix F, *i.e.*, by choosing $\mathbf{x}' = F^{-1}\mathbf{y}$. Although this gives exactly the right answer in the absence of pixel noise, this method performs very poorly when the pixel noise is non-negligible. The noise in the resulting reconstruction will typically be dominated by the most noisy of all the channels, and when the matrix F is poorly conditioned (such as when two components like synchrotron radiation and free-free emission have similar spectral indices), there will be additional noise amplification. If the number of frequencies m equals the number of components n, then in the limit of no noise ($N = 0$), our general subtraction scheme reduces to

$$W = SF^t(FSF^t)^{-1} = F^{-1}, \tag{56}$$

i.e., to simple "direct subtraction", independently of our assumptions about the various power levels (which are contained in S). Since for this special case, $WF = F^{-1}F = I$, we see that the power-preserving subtraction we advocate also reduces to direct subtraction when the noise is ignored, since the normalization condition $(WF)_{ii} = 1$ is satisfied. Since the optimized method is computationally trivial anyway, involving merely the inversion of a small matrix, there appears to be no advantage in neglecting the noise and using "direct subtraction".

It should also be emphasized that whereas "direct subtraction" and least-squares generalizations thereof require at least as many channels as components, our more general method has no such restriction, allowing $m < n$, $m = n$ and $m > n$.

5.4 Modeling spectral uncertainties

The main caveat to bear in mind when using any of the above-mentioned subtraction schemes is that, in reality, we do not know the matrix F with perfect accuracy. For instance, if the spectral index of free-free emission is $\beta = 0.2$ rather than $\beta = 0.1$, a subtraction attempt based on an extrapolation from 31 GHz to 217 GHz will be off by about 20%. We obviously expect the spectral indices to vary slightly from one sky region to another, for example, unsubtracted radio sources will not all have the same spectra and so their average spectrum will vary with position.

For us to be able to trust the error bars produced by a foreground removal method, we clearly need a way of quantifying the impact of such uncertainties in spectral indices. A simple way to do this is of course to assume some probability distributions for the spectral indices, make Monte-Carlo realizations, and compute the average of the actual reconstruction errors obtained when (erroneously) assuming the the spectral indices are exactly known. However, this is merely a method of diagnostics, and would not in itself provide a more robust subtraction scheme. Fortunately, there is quite a simple way of incorporating such uncertainties into the foreground model itself, which we will now describe.

Suppose we have reason to believe that about half of the synchrotron emission is characterized by $\beta \approx 1.1$ and half by $\beta \approx 0.7$. We can simply incorporate this into our model as two separate components with the same power spectrum C_ℓ but with different spectral dependencies $f_i(\nu)$. More realistically, we may wish to include an allowed range of spectral indices, reflecting either our lack of knowledge of their precise values or the presence of several physically distinct sub-components. In either case, the way to incorporate these ranges into the analysis is the same. If we wish to put into the model that $\beta = 0.9 \pm 0.2$, for instance, we can simply insert two synchrotron components, one with $\beta = 0.7$ and one with $\beta = 1.1$ (or a range of components, if one prefers) and the multi-frequency subtraction formalism will automatically reflect our uncertainty in β by associating appropriately large reconstruction errors with synchrotron subtraction.

5.5 Satellite specifics

We conclude this Section with a few more technical results relating to the impact of COBRAS/SAMBA specifics on the reconstruction errors, the lowermost heavy curve in Figure 13.

Removing channel 9 makes almost no difference. Removing the three highest frequency channels produces the thin solid curve in the same figure, *i.e.*, a worsening by a factor of a few over the whole range of multipoles. Removing the all the HEMT channels (channels 1-4), yields the thin dashed line, and most of this loss of accuracy is incurred even if only channel 1 is removed — basically because this greatly reduces the ability to subtract out synchrotron radiation from the higher frequency channels. When all channels are included, the multi-frequency subtraction scheme is found to place most of the weight on channels 4, 5 and 6. Although the other channels receive considerably smaller weights, they still help considerably, as these weights are tuned so as to subtract out the residual foregrounds.

It should be emphasized that although removing a few channels as described above made a relatively minor difference when the spectral indices were assumed to be perfectly known, broad spectral coverage is of course of paramount importance to ensure that the removal process is robust and can handle uncertainties in spectral slopes as described in the previous subsection.

6 DISCUSSION

We have presented a new method for subtracting foreground contamination from multi-frequency microwave sky maps, which can be used to produce accurate CMB maps and estimates of the CMB power spectrum C_ℓ.

6.1 Relation to other subtraction methods

The method incorporates the noise levels and subtracts the foregrounds in the Fourier, or multipole, domain rather than in real space, thereby exploiting the fact that contaminants such as dust, point sources, *etc.*, tend to have power spectra that differ substantially from that of the CMB. Compared to a standard subtraction procedure, it thus improves the situation in two ways:

1. The noise levels of the input maps are taken into account.
    - In a simple subtraction scheme, the cleaned map is just a particular linear combination of the input maps at various frequencies, where the weights take no account of the noise levels. For instance, if both the IRAS 100 micron map and the DIRBE 140 micron map were used as dust templates, the relative weight assigned to the two would be arbitrary in a simple subtraction scheme but optized in the approach presented here.
    - Simple subtraction is merely a special case of our method applied on a pixel-by-pixel basis, obtained in the limit of zero noise when $m = n$.
    - Since any spatial template whatsoever can be included as a "channel", the special case of simple subtraction thus has no advantages whatsoever over the more general method presented here.
2. The subtraction is done in Fourier space.
    - Since no phase information is lost by passing to the Fourier (multipole) domain and back, this preserves all the information about the various spatial templates, for instance the location of the galactic plane.

- Thus this approach is superior to working in real space whenever the contaminants have power spectra that differ substantially from that of the CMB.
- This will be the case for *regardless* of what the true CMB power spectrum turns out to be, since certain contaminants are known to have power spectra that differ from one another (for instance, the dust power scales roughly as ℓ^{-3}, whereas the point source power scales as ℓ^0.
- Although normal Wiener-filtering modifies the power spectrum of the measured signal, our prescription in equation (36) (where the row vectors of W are rescaled so that $(WF)_{ii} = 1$) does *not* not have this undesirable property, which means that our subtraction scheme is suitable for power spectrum estimation as well as map-making (we showed that this is only possible when more than one frequency is available).
- With this prescription, the optimal weighting coefficients in the W-matrix are independent of any assumptions about the CMB power spectrum (in stark contrast to conventional one-channel Wiener filtering). The weights only depend on the assumptions about the foreground power spectra.
- It is of course advantageous to perform the subtraction in Fourier space even if one does not wish to make any assumptions about the foreground levels. This model-independent special case of our method corresponds to simply setting the noise levels to zero.

Any of these two improvements can of course be implemented without the other. Traditional subtraction, but mode by mode in Fourier space, will perform better than if done pixel-by-pixel in real space. Likewise, if the overall foreground levels are known in many channels, the result will be better if noise is taken into account than if it is ignored in the subtraction.

6.2 Foreground estimates

To provide a qualitative understanding of how well the method will be able to tackle the next generation of CMB data, we made rough estimates of the power spectra of the relevant foregrounds in Section 4. The results are summarized in Figure 14. Although these estimates are not intended to be very accurate, the following qualitative conclusions appear to be quite robust:

- Galactic dust poses a problem mainly at the upper left, corresponding to large scales and high frequencies.
- Synchrotron radiation and free-free emission pose a problem mainly at the lower left, corresponding to large scales and low frequencies.
- Radio point sources are a problem mainly at the lower right, corresponding to small scales and low frequencies.
- If infrared emission from high redshift galaxies pose a significant problem, it will be at the upper right, corresponding to small scales and high frequencies.
- Experimental noise and beam dilution are mainly a limitation to the right and above.
- The most favorable frequencies for measuring high multipoles such as the CDM Doppler peaks are larger (around 100 GHz and above) than the best ones for probing the largest scales (around 50 GHz).

6.3 An example: COBRAS/SAMBA

In Section 5, we assessed the effectiveness of the method using the specifications of the proposed COBRAS/SAMBA satellite mission and the above-mentioned foreground modeling. As is seen in Figure 13, our method provides a gain of more than a factor of ten over the entire multipole range compared to subtracting the backgrounds on a pixel-by-pixel basis. If the frequency dependencies of the foregrounds were perfectly known and independent of position in the sky (which they are not), then the method could recover the CMB multipoles $a_{\ell m}$ out to about $\ell = 10^3$ to an accuracy of about a tenth of a percent. We also found that our uncertainty about the spectral indices could be incorporated into the formalism in straightforward way, by simply replacing a poorly understood component by several components, all with the same power spectra, but with slightly different spectral indices. We saw that even with extremely pessimistic assumptions about our ability to model the foregrounds, it is likely that an estimate of the CMB power spectrum C_ℓ can be made where the residual foreground contamination is merely a few percent. Since the power spectrum of dust falls off and that of point sources rises with ℓ (relative to that of the CMB), the most favorable situation occurs around $\ell = 200$, where even power accuracies of 1% would be an extremely pessimistic prediction. This is fortunate, as this scale coincides with the location of the first Doppler peak of standard CDM, and accurate determination of its position (if some variant of CDM is correct) would provide a direct measurement of Ω, the cosmological density parameter (see *e.g.* Kamionkowski & Spergel 1994).

6.4 The effect of non-Gaussianity

The case where the foregrounds are Gaussian is in a sense the worst possible case when it comes to subtracting them. Since Gaussian random fields are completely specified by their power spectra, this means that we have no additional information that we can take advantage of in our attempts at foreground removal. Since the method we derived was optimized for the Gaussian case, a natural question to ask is whether one can do still better by making use of the fact that many foregrounds do in fact exhibit non-Gaussian features. One simple way to do this is to simply remove spatially localized contaminants (*e.g.*, the galactic plane and bright point sources), as was discussed in Section 3.5. Making optimal use of non-Gaussian features, however, is quite difficult in practice, as it generally leads to a non-linear optimization problem. Even if a numerical solution were found, one might be left wondering to what extent poor assumptions about the precise type of non-Gaussianity might have degraded the final results.

Fortunately, the method we have presented appears to be quite adequate in practice. As part of the scientific case for the COBRAS/SAMBA satellite, detailed simulations have been made where real foreground maps (extrapolated to the relevant frequencies using the data from IRAS, the Haslam map, *etc.*) were added to simulated CMB data. The simulated real-world maps were then "cleaned" with various subtraction methods. The results (Bouchet *et al.* 1996) show that our method works extremely well, even though the IRAS dust maps are known to exhibit strong non-Gaussian features.

6.5 Outlook: what more do we need to know?

To be able to better quantify the effectiveness of future CMB missions and answer questions such as "what design changes would improve the ability to remove foregrounds the most?", it is important that more accurate measurements of the various foreground power spectra be made. Experiments are now becoming so sensitive that it no longer suffices to summarize each foreground by a single number ΔT (which usually refers to its average intensity, the monopole) — its entire power spectrum is needed, to chart out the foreground landscape of Figure 14. Some of the most urgent outstanding questions are the following:

- What is the differential source count of radio point sources between 10 GHz and 100 GHz, *i.e.*, how do we normalize their power spectrum for various flux cuts?
- What is the differential source count of infrared point sources between 50 and 300 GHz, *i.e.*, how do we normalize their power spectrum for various flux cuts?
- What is the power spectrum of synchrotron radiation and free-free emission on small angular scales?

If none of the foreground contaminants turn out to be much worse than we assumed in Section 4, then the next generation of CMB experiments may indeed allow us to measure the key cosmological parameters with hitherto unprecedented accuracy.

ACKNOWLEDGMENTS: We thank Ted Bunn, François Bouchet, Anthony Lasenby, Jean-Lup Puget and Ned Wright for many useful comments and Gavin Dalton for his help with the IRAS 100μm and DIRBE maps. MT acknowledges partial support from European Union contract CHRX-CT93-0120 and Deutsche Forschungsgemeinschaft grant SFB-375.

7 REFERENCES

Banday, A. J. & Wolfendale, A. W. 1991, *MNRAS*, **248**, 705.
Becker, R. H., White, R. L. & Helfand, D. J. 1995, ApJ, in press.
Benn, C.R. & Wall, J.V. 1995, *MNRAS*, **272**, 481.
Bond, J.R. & Efstathiou, G., 1987, *MNRAS*, **226**, 655.
Bond, J. R. *et al.* 1994, *Phys. Rev. Lett.*, **72**, 13.
Bond, J. R. 1995, *Phys. Rev. Lett.*, **74**, 4369..
Bouchet, F. *et al.* 1995, preprint astro-ph/9410004.
Bouchet, F. *et al.* 1996, in preparation.
Brandt, W. N. *et al.* 1994, *ApJ*, **424**, 1.
Bunn, E. F. *et al.* 1994, *ApJ*, **432**, L75.

Bunn, E. F., Scott, D. & White, M. 1995, *ApJ*, **441**, L9.
Bunn, E. F. & Sugiyama, N. 1995, *ApJ*, **446**, 49.
de Bernardis, P., Masi, S. & Vittorio, N. 1991, *ApJ*, **382**, 515.
Dodelson, S. & Stebbins, A. 1994, *ApJ*, **433**, 440.
Fisher, K. B. *et al.* 1995, *MNRAS*, **272**, 885.
Franceschini, A. *et al.* 1989, *ApJ*, **344**, 35.
Franceschini, A. *et al.* 1991, *A&A Supp.*, **89**, 285.
Górski, K. M. 1994, *ApJ*, **430**, L85.
Guarini, G., Melchiorri, B. & Melchiorri, F. 1995, *ApJ*, **442**, 23.
Haslam, C. G. T. *et al.* 1982, *A&A Supp.*, **47**, 1.
Hu, W. & Sugiyama, N. 1995, *ApJ*, **436**, 456.
Kamionkowski, M. & Spergel, D.N., 1994, *ApJ*, **431**, 1.
Lahav, O. *et al.* 1994, *ApJ*, **423**, L93.
Low, F. J. & Cutri, R. M. 1994, *Infrared Phys. & Technol.*, **35**, 291.
Mandolesi, N. *et al.* 1995, *COBRAS/SAMBA: the ESA Medium Size Mission for Measurement of CBR Anisotropy*, in *1993 Capri Workshop on the cosmic microwave background*, *Astrophys. Lett & Comm.*, in press.
Mather *et al.* 1994, *ApJ*, **420**, 439.
Neugebauer, G. *et al.* 1984, *ApJ*, **278**, L1.
O'Sullivan, C. *et al.* 1995, *MNRAS*, **274**, 861.
Reach, W. T. *et al.* 1995, *ApJ*, **451**, 188.
Rybicki, G. B. & Press, W. H. 1992, *ApJ*, **398**, 169.
Scott D, Silk J and White M 1995, *Science*, **268**, 829.
Smoot, G.F. *et al.* 1992, *ApJ*, **396**, L1.
Tegmark, M. & Bunn, E. F. 1995, *ApJ*, **455**, 1.
Tegmark, M. 1995, MNRAS, in press.
Toffolatti, L. *et al.* 1995, in *1993 Capri Workshop on the cosmic microwave background*, *Astrophys. Lett & Comm.*, in press.
White, M. & Srednicki, M, 1995, *ApJ*, **443**, 6.
White, M., Scott, D. and Silk, J. 1994, *ARA&A*, **32**, 319.
Wright, E. L. *et al.* 1991, *ApJ*, **381**, 200.
Zaroubi, S. *et al.*, preprint astro/ph 9410080.

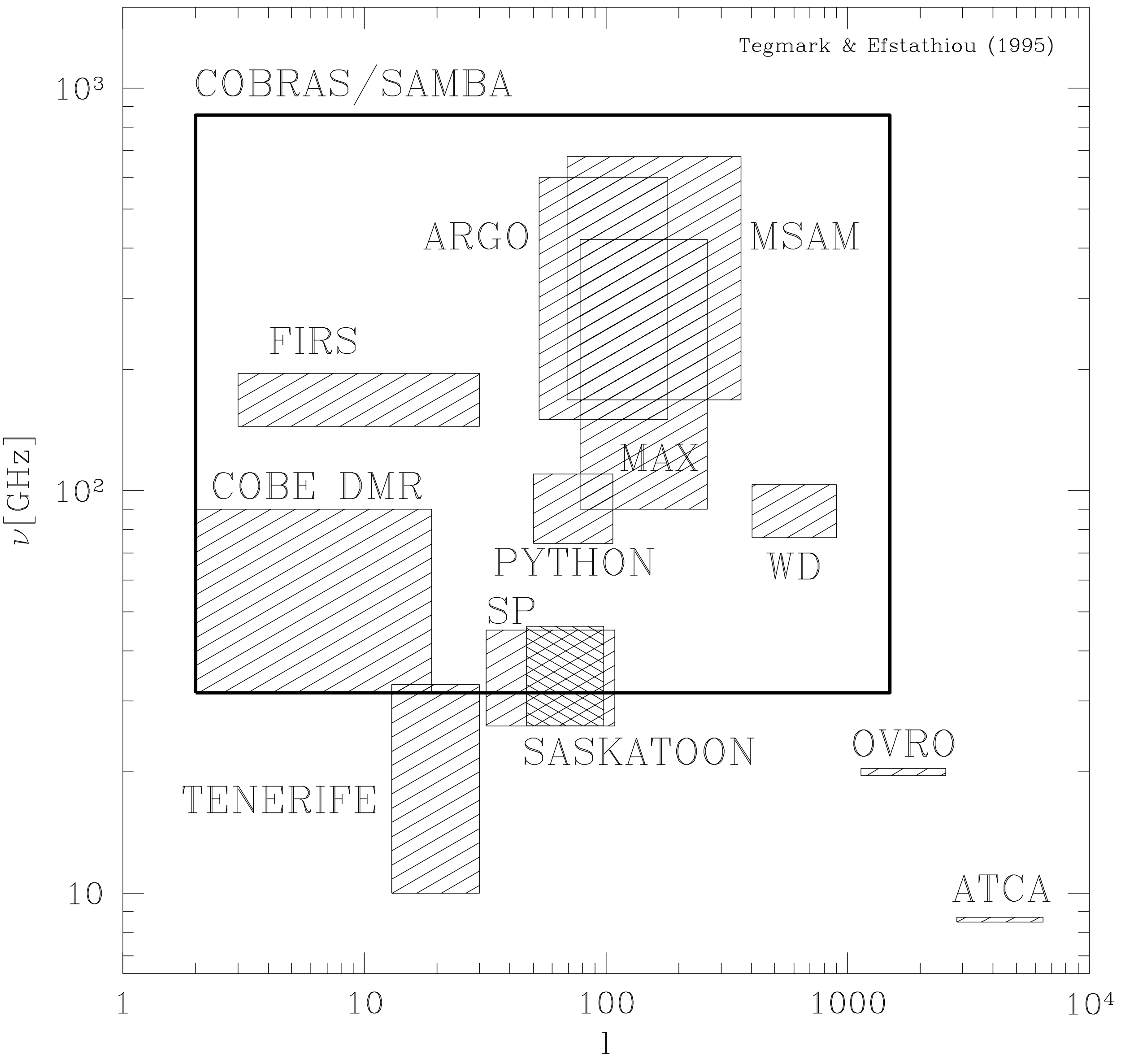

Figure 1: Where various CMB experiments are sensitive.
The boxes roughly indicate the range of multipoles ℓ and frequencies ν probed by various CMB experiments. The large heavy unshaded box corresponds to the proposed COBRAS/SAMBA satellite.

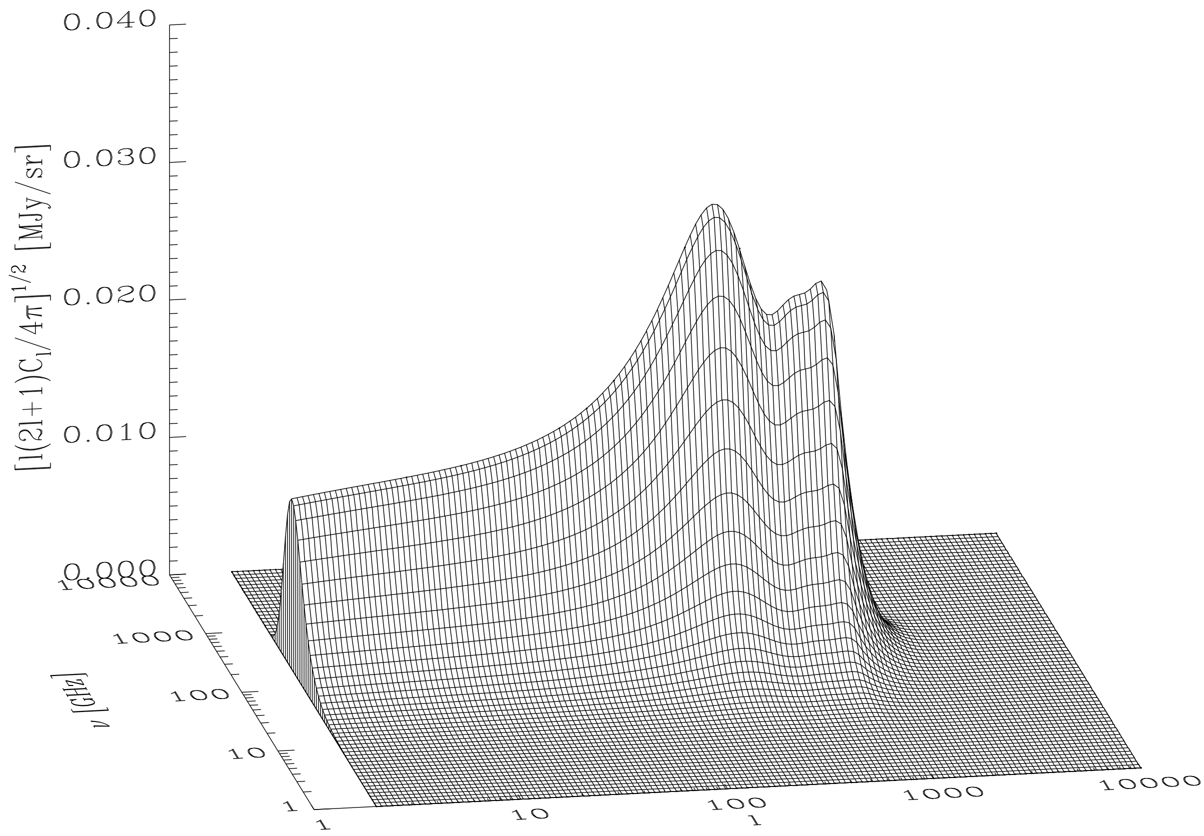

Figure 2: How the CMB brightness fluctuations depend on multipole and frequency in the standard CDM model (assuming scale-invariant scalar fluctuations in a critical density, $\Omega = 1$, universe, with Hubble constant $H_0 = 50 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, and baryon density $\Omega_b = 0.05$). The CDM power spectrum was computed as described by Bond & Efstathiou (1987).

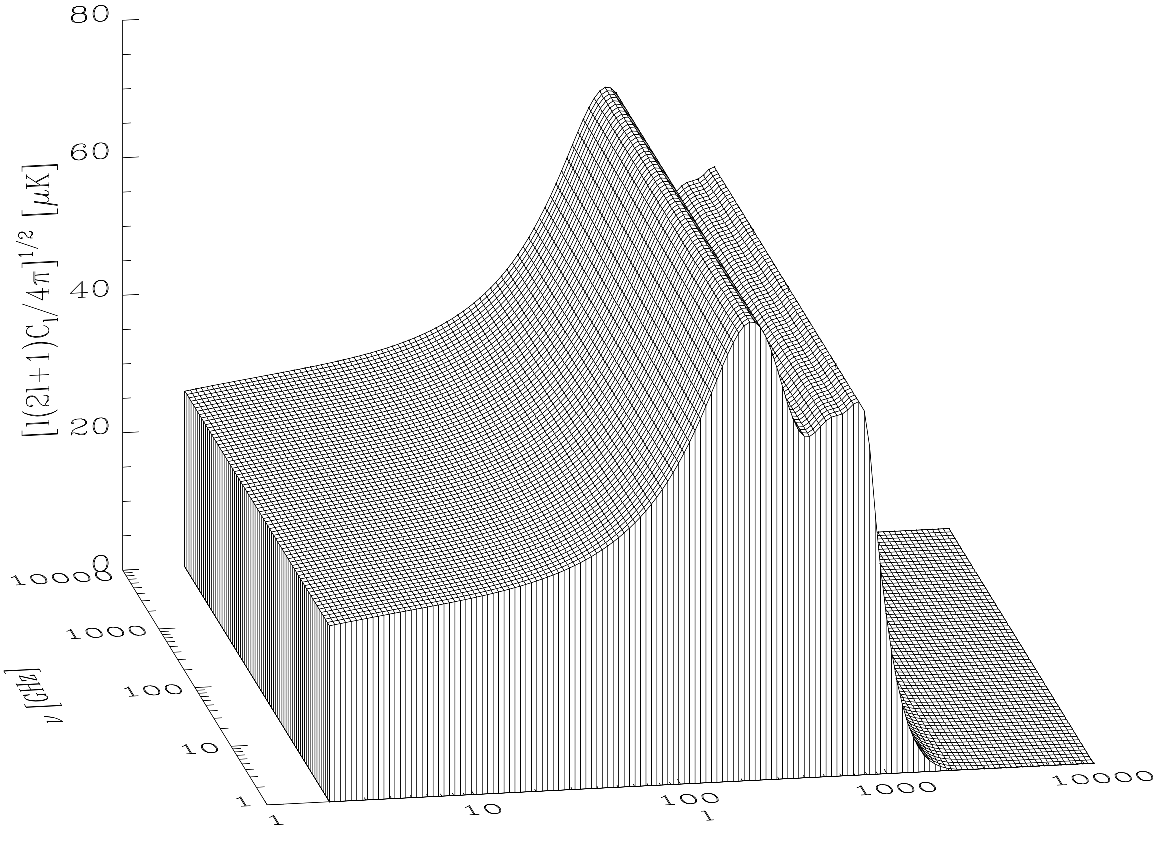

Figure 3: How the CMB temperature fluctuations depend on multipole and frequency for the CDM model with parameters as in Figure 2.

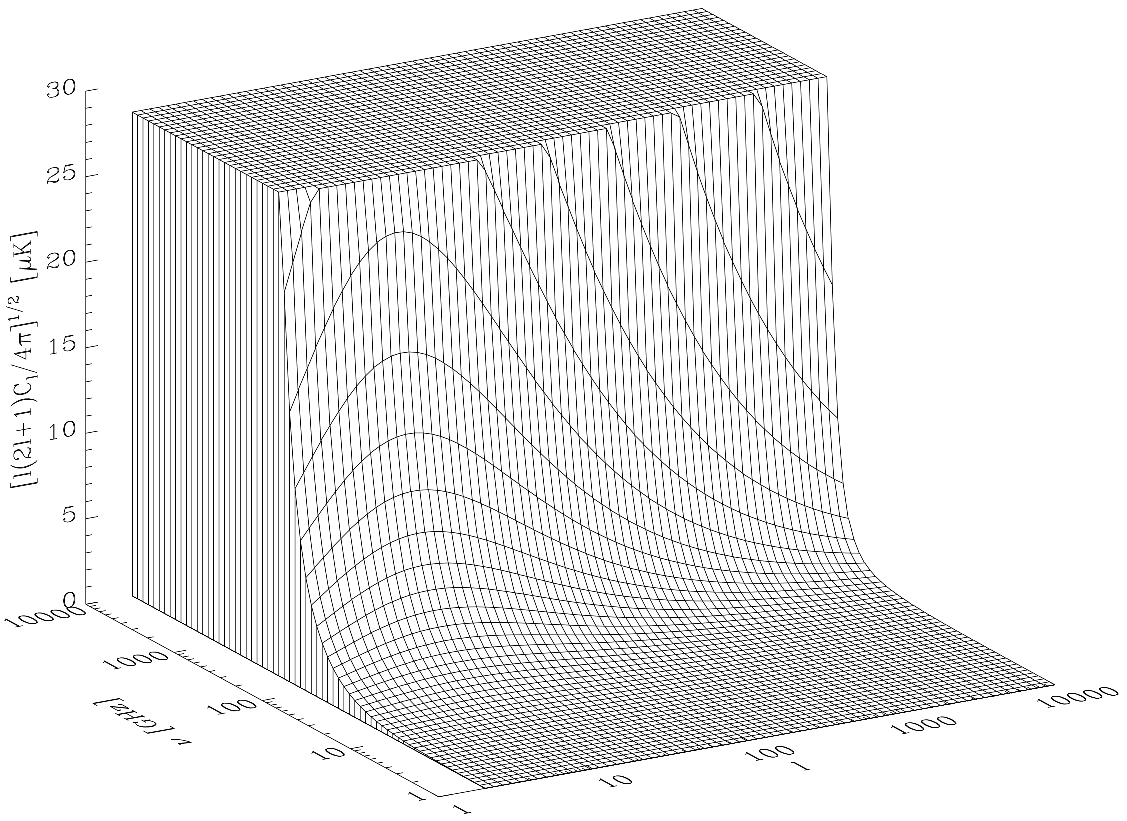

Figure 4: Model for how the dust power spectrum depends on multipole and frequency.

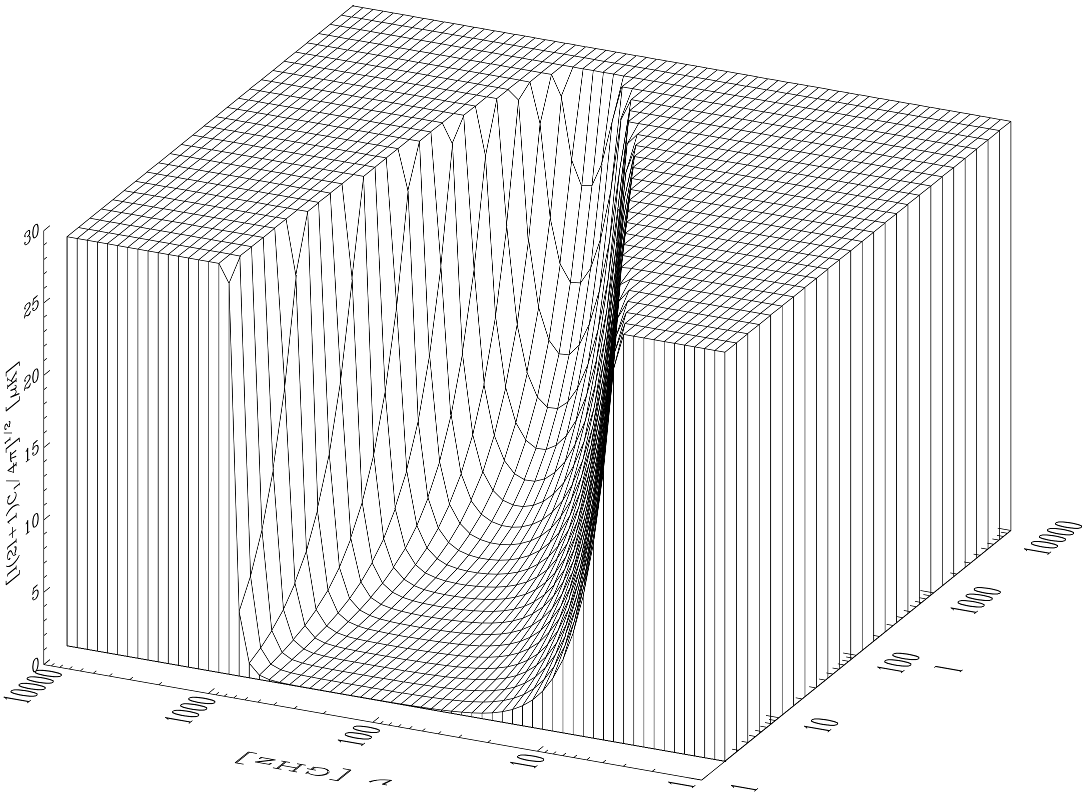

Figure 5: Model for how the power spectrum of point sources depends on multipole and frequency.

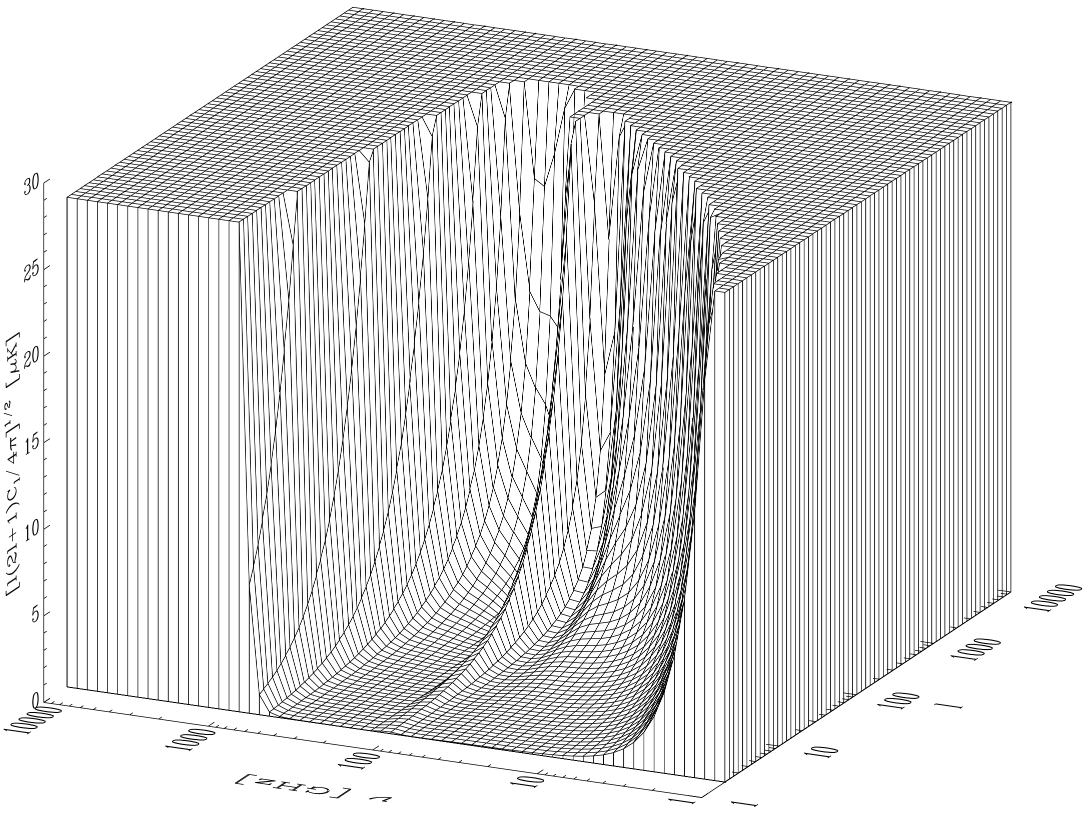

Figure 6: How the COBRAS/SAMBA pixel noise power depends on multipole and frequency.

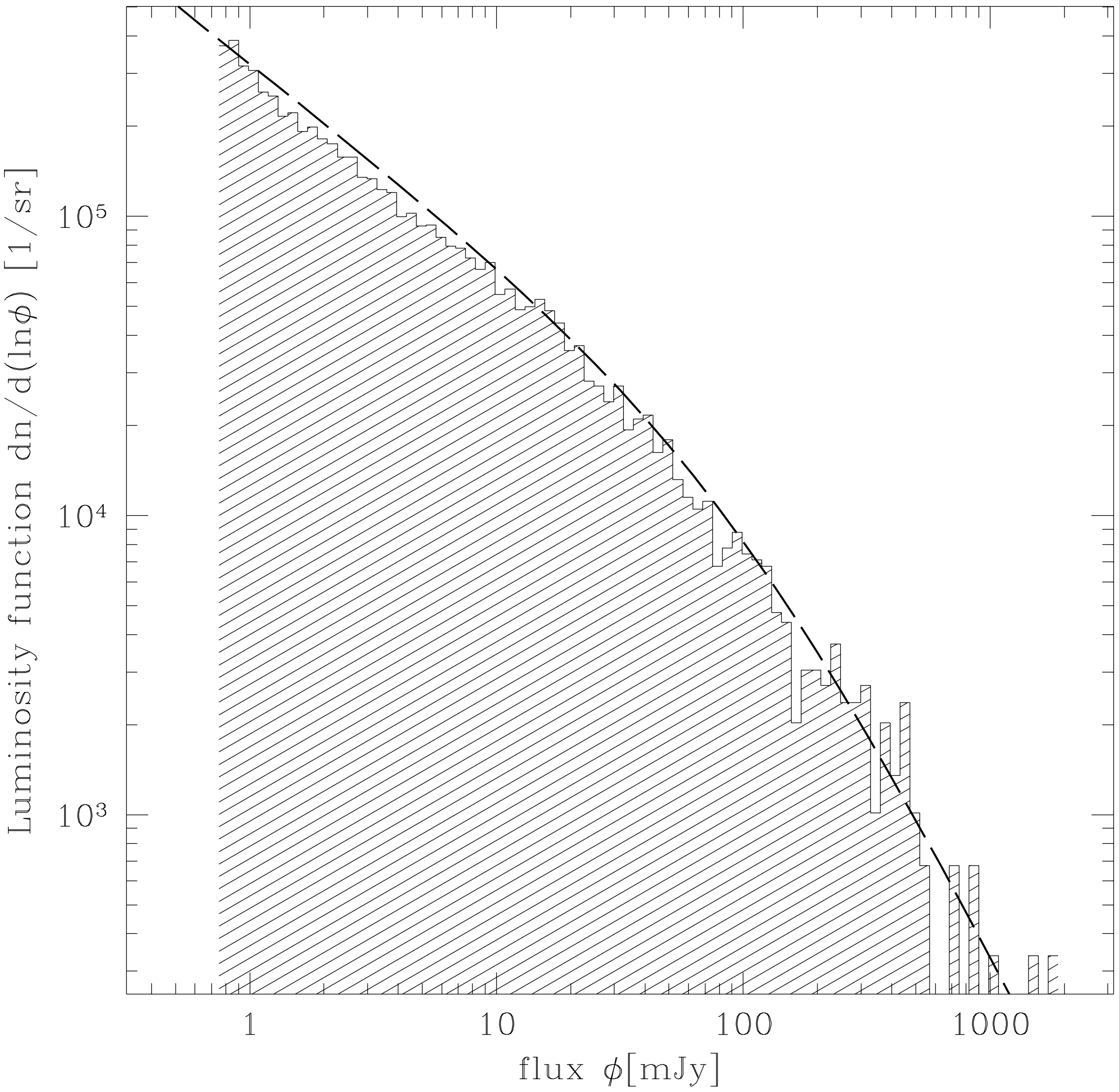

Figure 7: The VLA FIRST differential source count for radio sources at 1.5 GHz.

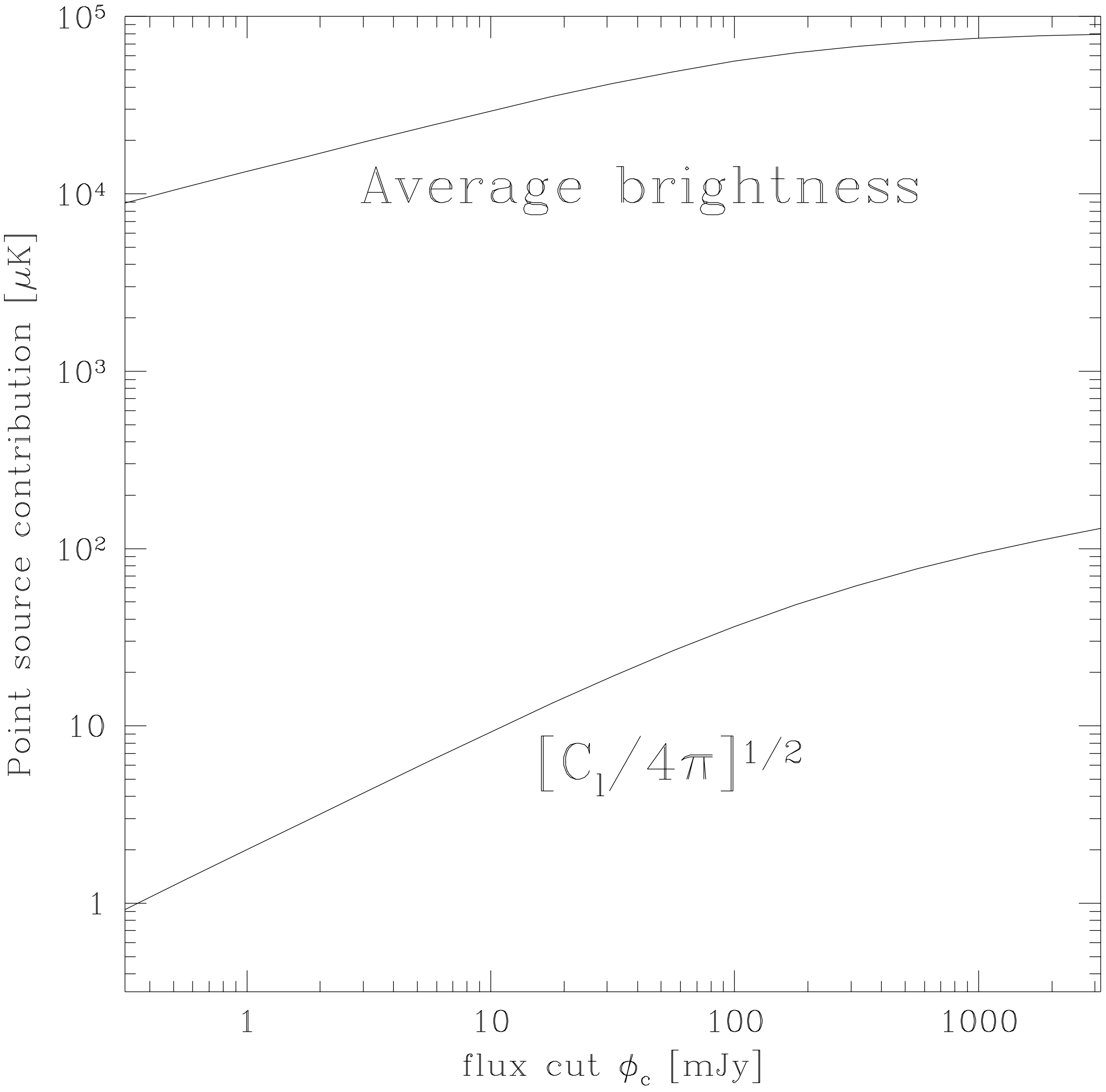

Figure 8: Dependence of radio source fluctuations on flux cut. The average brightness (monopole, upper curve) and the power spectrum normalization (lower curve) at 1.5 GHz from the VLA FIRST survey is plotted as a function of flux cut.

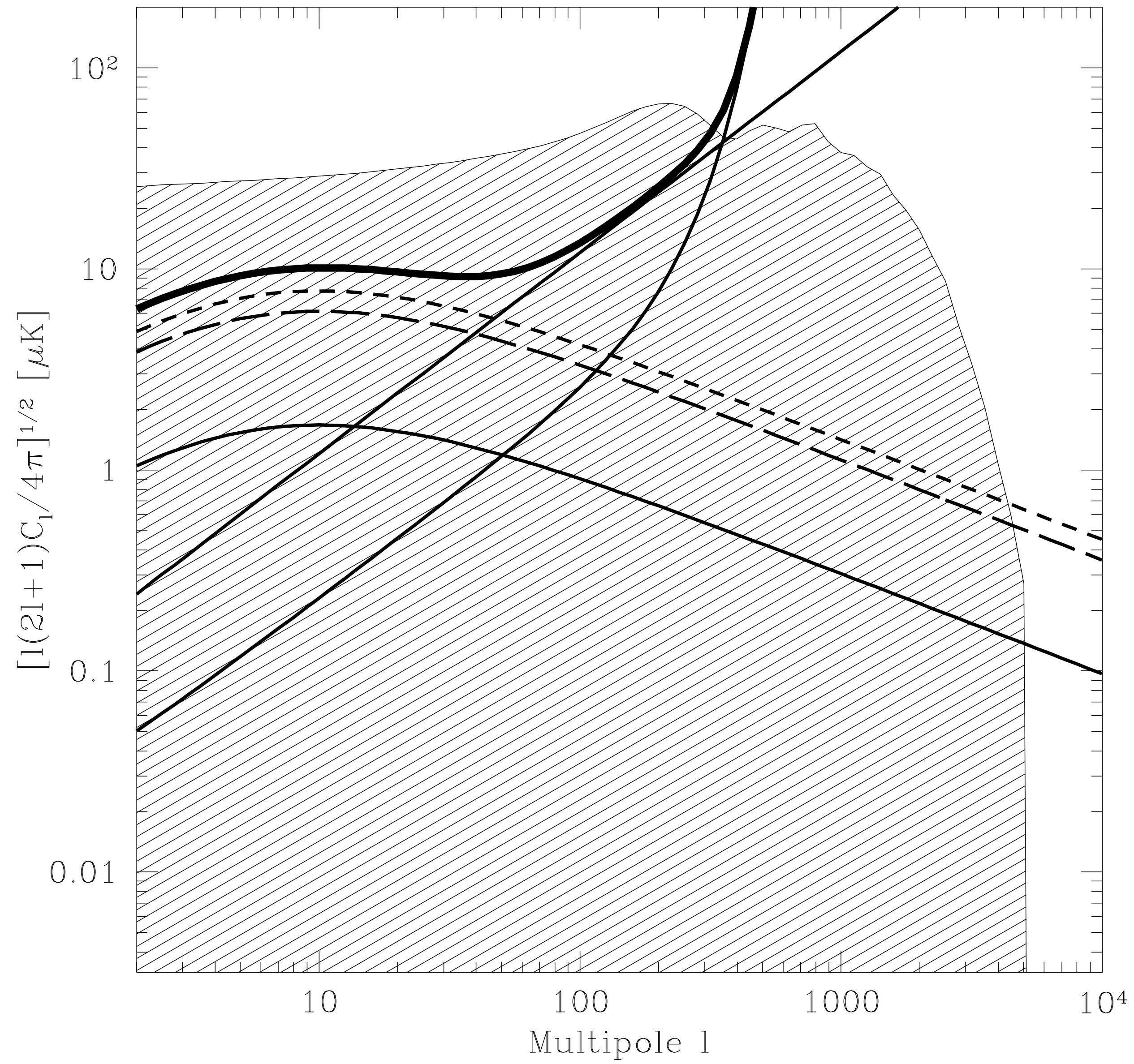

Figure 9: Model power spectra of various components at 31 GHz. From the bottom up, on the left hand side, the curves correspond to pixel noise, radio point sources, dust, synchrotron radiation, free-free emission, all sources combined (heavy), and CMB (shaded) according to the CDM model plotted in Figures 2 and 3.

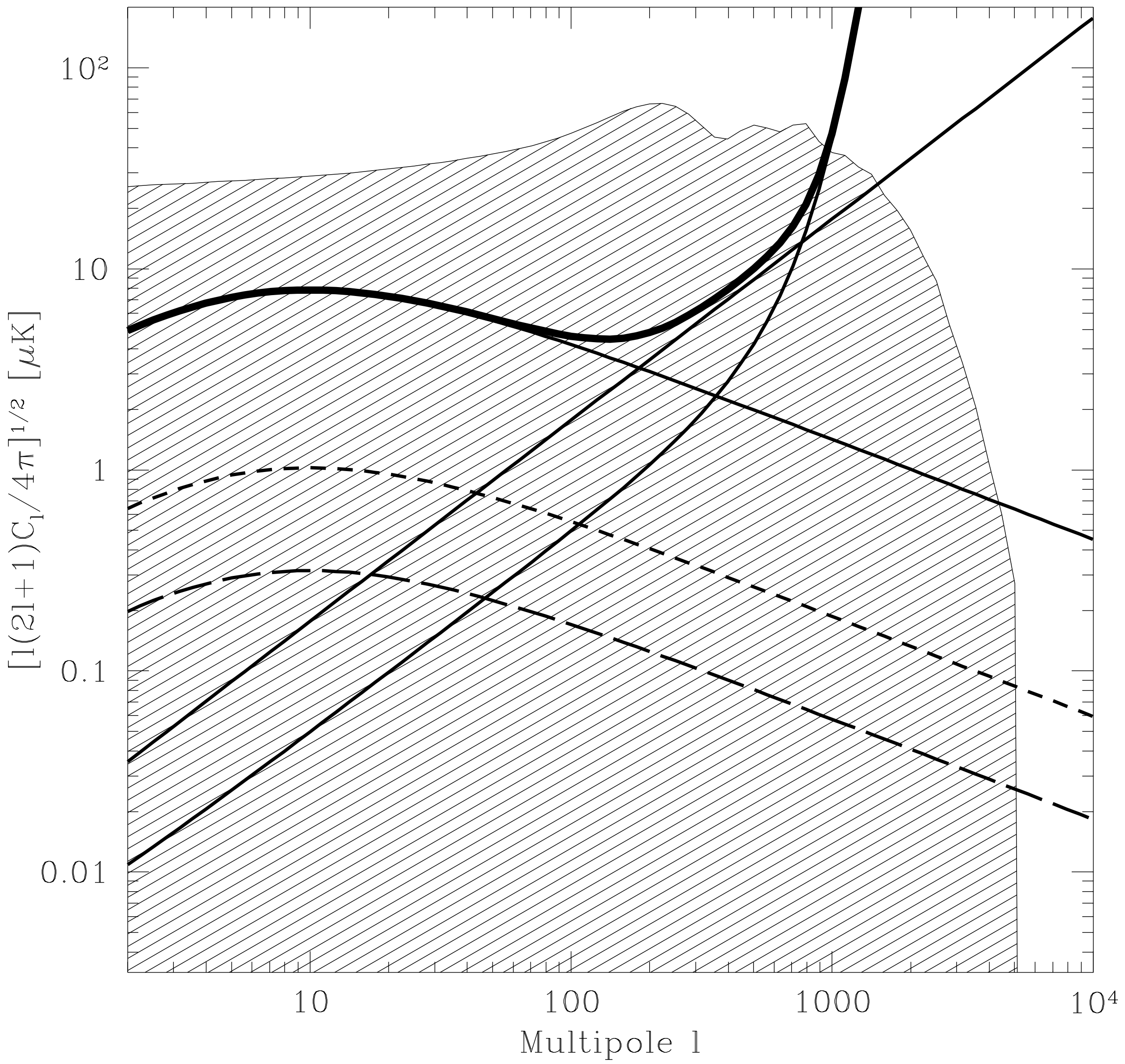

Figure 10: Model power spectra of various components at 90 GHz. From the bottom up, on the left hand side, the curves correspond to pixel noise, radio point sources, synchrotron radiation, free-free emission, dust, all sources combined (heavy), and CMB (shaded).

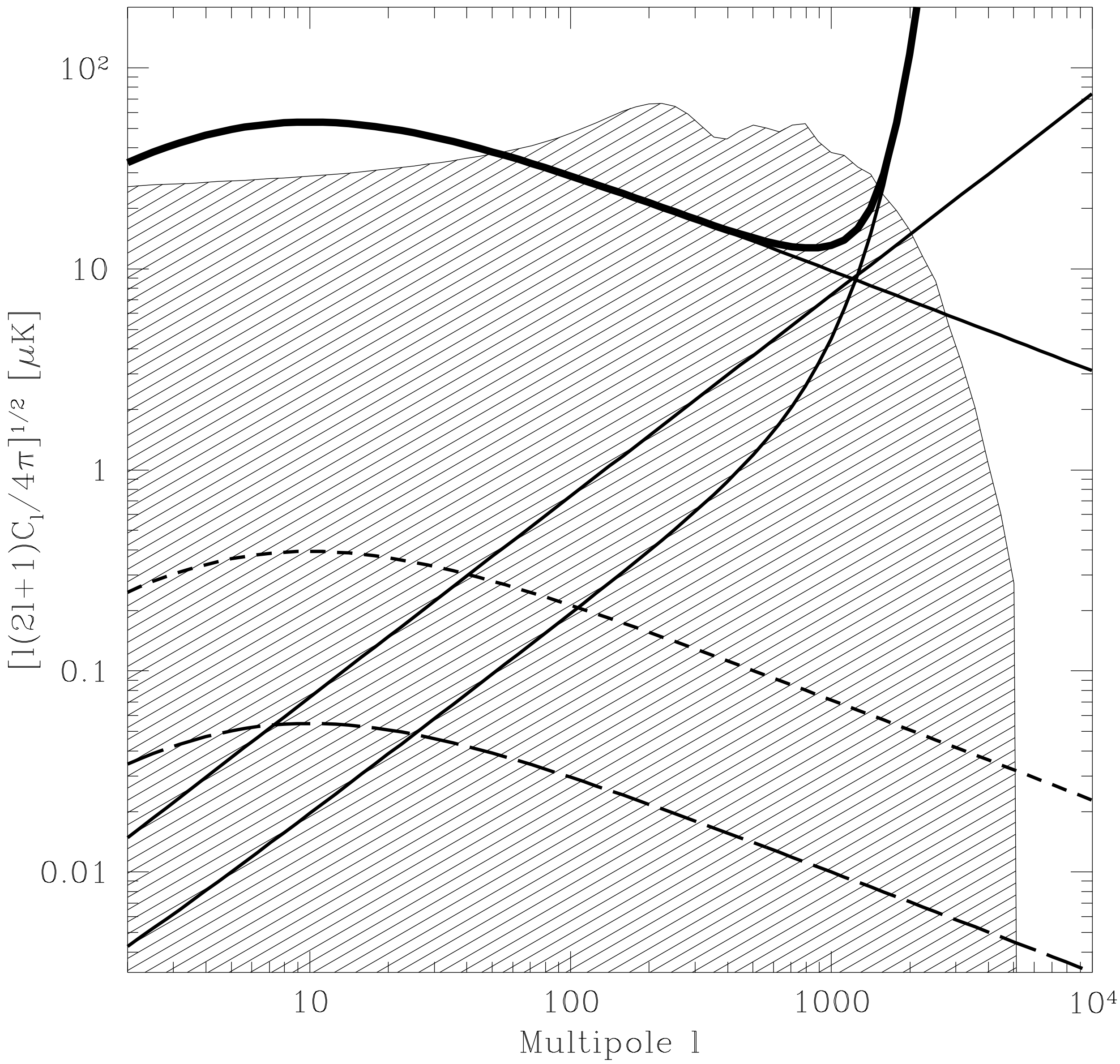

Figure 11: Model power spectra of various components at 217 GHz. From the bottom up, on the left hand side, the curves correspond to pixel noise, radio point sources, synchrotron radiation, free-free emission, CMB (shaded), dust, and all sources combined (heavy).

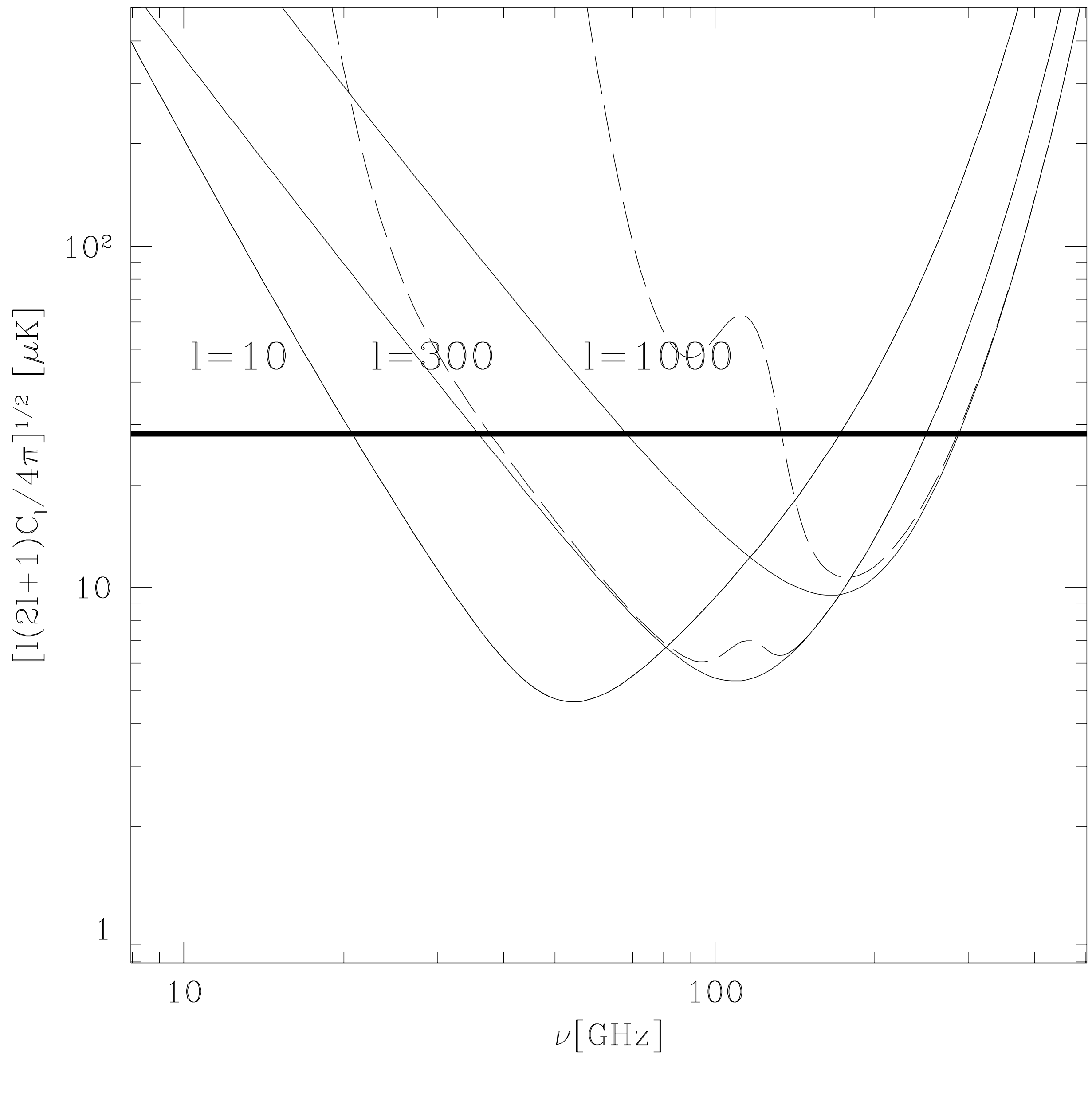

Figure 12: Total contamination at selected multipoles.
The total fluctuations of all foregrounds excluding (solid lines) and including (dashed lines) COBRAS/SAMBA pixel noise are plotted for the multipoles 10, 300 and 1000. The heavy horizontal line corresponds to COBE-normalized Sachs-Wolfe fluctuations in the CMB.

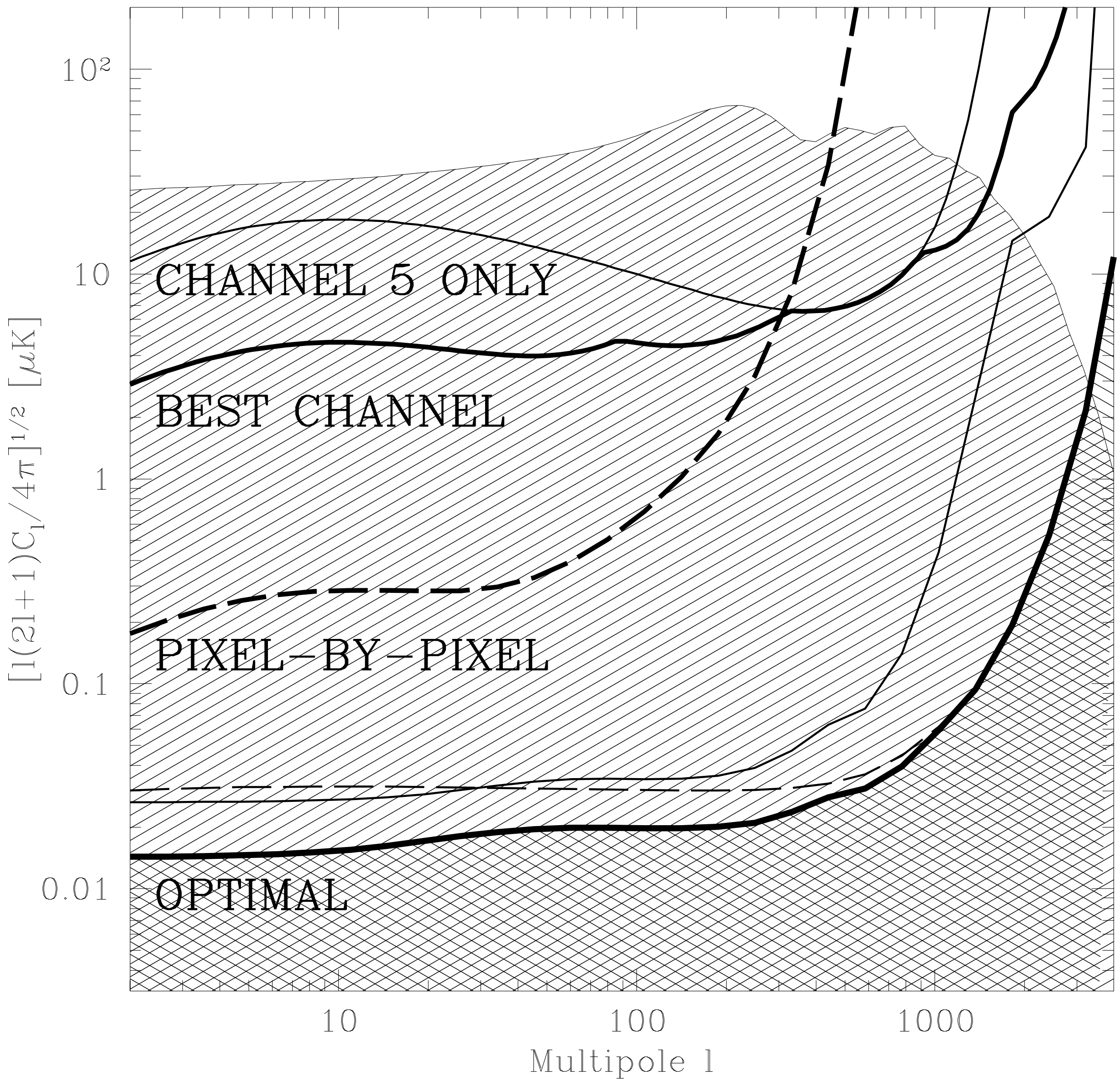

Figure 13: Comparison of methods

The combined residual contribution of all foregrounds and noise is plotted for different approaches to foreground subtraction. From top to bottom, the four labeled curves correspond to (1) use of the 143 GHz channel with no subtraction, (2) use of the best COBRAS/SAMBA channel at each multipole with no subtraction, (3) optimized subtraction on a pixel-by-pixel basis, and (4) the optimal subtraction technique. The uppermost curve (shaded) is a standard CDM power spectrum as plotted in earlier figures. The two thin curves at the bottom correspond to reducing the number of channels as described in the text.

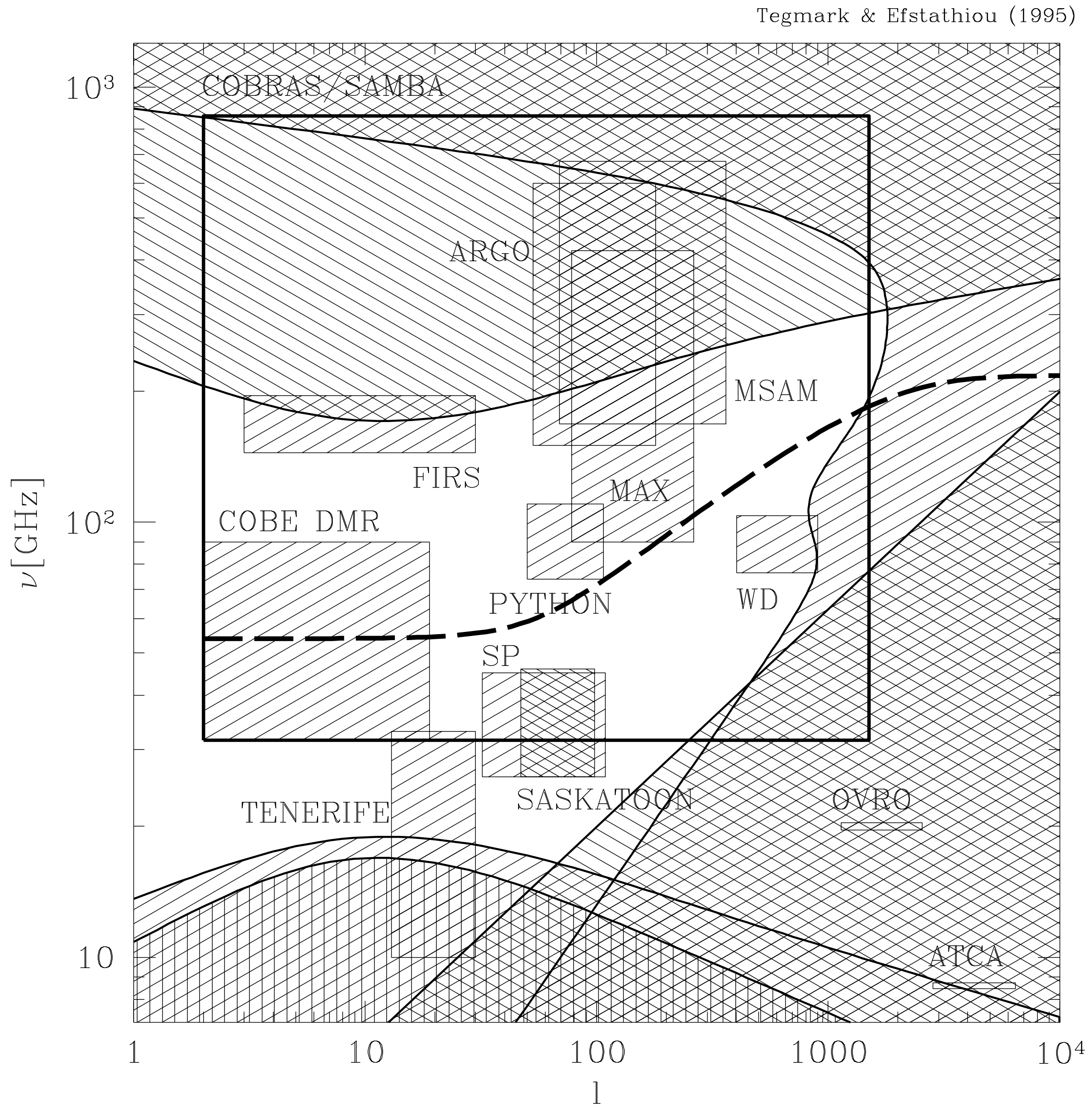

Figure 14: Where various foregrounds dominate
The shaded regions indicate where the various foregrounds cause fluctuations exceeding those of COBE-normalized scale-invariant fluctuations ($\approx \sqrt{2} \times 20\mu$K), thus posing a substantial challenge to estimation of genuine CMB fluctuations. They correspond to dust (top), free-free emission (lower left), synchrotron radiation (lower left, vertically shaded), radio point sources (lower right) and COBRAS/SAMBA instrumental noise and beam dilution (right). The heavy dashed line shows the frequency where the total foreground contribution to each multipole is minimal. The boxes indicate roughly the range of multipoles ℓ and frequencies ν probed by various CMB experiments, as in Figure 1.