SUBMITTED TO APJ NOVEMBER 30, 1997, ACCEPTED FEBRUARY 19
Preprint typeset using LATEX style emulateapj

REMOVING REAL-WORLD FOREGROUNDS FROM CMB MAPS*

MAX TEGMARK[1,2]

Submitted to ApJ November 30, 1997, accepted February 19

ABSTRACT

Most work on foreground removal has treated the case where the frequency dependence of all components is perfectly known and independent of position. In contrast, real-world foregrounds are generally not perfectly correlated between frequencies, with the spectral index varying slightly with position and (in the case of some radio sources) with time. A method incorporating this complication in presented, and illustrated with an application to the upcoming satellite missions MAP and Planck. We find that even spectral index variations as small as $\Delta\alpha \sim 0.1$ can have a substantial impact on how channels should be combined and on attainable accuracy.

1. INTRODUCTION

Future Cosmic Microwave Background (CMB) experiments can measure many key cosmological parameters to great precision (Jungman *et al.* 1996; Bond *et al.* 1997; Zaldarriaga *et al.* 1997) — in principle. To achieve this in practice, foreground contamination must be removed with comparable accuracy. Tegmark & Efstathiou (1996, hereafter "TE96"), derived the foreground subtraction method that minimized the residual variance from foregrounds and noise under the assumption that the frequency dependence of all components was perfectly known and independent of position. This method, which was independently derived by Bouchet & Gispert (unpublished), has now been extensively tested with simulations (see *e.g.* Bouchet *et al.* 1995; Bersanelli *et al.* 1996) where each frequency channel was the appropriate linear combination of a simulated CMB map, foreground templates such as the Haslam, DIRBE and IRAS maps, radio sources, and random noise. The inversion was found to accurately recover the input maps even though the foreground templates exhibited strong non-Gaussianity.

To further improve such modeling, one must incorporate the complication that real-world foregrounds are generally not perfectly correlated between frequencies, with the spectral index varying slightly with position and (in the case of some radio sources) with time. Such spatial variations of the spectral index have been observed for both dust (*e.g.*, Reach *et al.* 1996; Schlegel *et al.* 1997) and synchrotron radiation (Banday & Wolfendale 1991; Platania *et al.* 1997), and are of course even more pronounced for point sources (*e.g.*, Franceschini *et al.* 1989, 1991; Toffolatti *et al.* 1997). As we will see, neglect of this complication can cause severe underestimates of the residual foreground level in the cleaned CMB map. It can also produce foreground residuals substantially higher than can be obtained with the method we derive below.

2. METHOD

As in TE96, we assume that we have sky maps at m frequencies $\nu_1, ..., \nu_m$ (these maps may be internal channels of a CMB experiment, but can include external templates such as the DIRBE maps as well), and that these maps receive contributions from n different physical components (CMB, dust, *etc.*).

2.1. *Pixel by pixel or wavelet by wavelet?*

The general problem treated in this paper is how to take linear combinations of these m maps to produce accurate maps of individual components. The traditional approach (*e.g.*, Brandt *et al.* 1994) has been to perform this multi-frequency subtraction separately for each pixel (direction in the sky). However, this does not take advantage of the differences in smoothness between CMB and the various foregrounds. The correlation between neighboring pixels is typically stronger for diffuse galactic foregrounds (dust, synchrotron and free-free emission) than for CMB, where it is in turn stronger than for point sources. One can therefore do better by performing the linear combinations mode by mode rather than pixel by pixel, using some variant of a Fourier expansion of the maps (TE96). The optimal weights for the linear combination then differ on large angular scales (where diffuse foregrounds are important) and small angular scales (dominated by point sources), as illustrated in Figure 1. In addition to this scale dependence, there is also a direction dependence, since some pixels typically have higher average levels of noise (from receiving less observing time) or foreground contamination (from being closer to the galactic plane, say), than others. The suggestion of this author is therefore that foreground subtraction be performed using modes that are fairly localized both in real space and in Fourier space, for instance some form of wavelets.

2.2. *Notation*

All the methods discussed below can be applied regardless of which of the above-mentioned approaches is taken. Let y_j denote the is temperature δT measured at the j^{th} frequency in a given direction (or in a given mode — in that case, y_j is simply the corresponding multipole, Fourier, or wavelet coefficient). Let $y_j^{(i)}$ denote the contribution to y_j from the i^{th} physical component. Grouping

[1]Institute for Advanced Study, Princeton, NJ 08540; max@ias.edu
[2]Hubble Fellow

Available in color from* **h t t p://www.sns.ias.edu/~max/foregrounds.html

the measurements y_j into an m-dimensional vector $\mathbf{y}$, we can thus write

$$\mathbf{y} = \sum_{i=0}^{n} \mathbf{y}^{(i)}. \tag{1}$$

As in TE96, we assume that the different components have zero mean ($\langle \mathbf{y}^{(i)} \rangle = 0$)[3] and are uncorrelated, which means that the data covariance matrix is simply given by

$$\mathbf{C} \equiv \langle \mathbf{y}\mathbf{y}^t \rangle = \sum_{i=0}^{n} \mathbf{C}^{(i)}, \tag{2}$$

where $\mathbf{C}^{(i)} \equiv \langle \mathbf{y}^{(i)} \mathbf{y}^{(i)^t} \rangle$ is the covariance matrix of the i^{th} component. It is convenient to factor this as $\mathbf{C}^{(i)}_{jk} = \mathbf{R}^{(i)}_{jk} \sigma^{(i)}_j \sigma^{(i)}_k$, where the standard deviation and correlation is defined by $\sigma^{(i)}_j \equiv [\mathbf{C}^{(i)}_{jj}]^{1/2}$ and $\mathbf{R}^{(i)}_{jk} = \mathbf{C}^{(i)}_{jk} / \sigma^{(i)}_j \sigma^{(i)}_k$, respectively. For definiteness, let us take component 0 to be the CMB. Since the CMB temperature is the same in all channels, equal to x, say, we thus have $\mathbf{y}^{(0)} = \mathbf{e}x$, where the constant vector $\mathbf{e}$ is defined by $e_j = 1$. We therefore have $\sigma^{(0)}_j = \langle x^2 \rangle^{1/2}$, independent of j, and the correlation matrix $\mathbf{R}^{(0)} = \mathbf{E}$, where $\mathbf{E} \equiv \mathbf{e}\mathbf{e}^t$, a matrix consisting entirely of ones. Let us take component 1 to be the instrumental noise. Then $\sigma^{(1)}_j$ is simply the r.m.s. noise level in the j^{th} channel, and if the noise is uncorrelated between channels, we have $\mathbf{R}^{(1)} = \mathbf{I}$, the identity matrix. The remaining components (the various foregrounds) will typically have correlation matrices $\mathbf{R}^{(i)}$ that are intermediate between these two extreme cases of perfect correlation ($\mathbf{R} = \mathbf{E}$) and no correlation ($\mathbf{R} = \mathbf{I}$).

Tegmark (1997, hereafter "T97") compared ten different methods for making CMB maps from time-ordered data. The CMB foreground removal problem is quite analogous to the mapmaking problem in that one seeks a linear inversion given certain assumptions about the "noise". Indeed, all of the inversion methods described in T97 can be used for foreground removal as well, and we will repeatedly return to these connections below.

2.3. *A signal-to-noise eigenvalue problem*

Let us consider an arbitrary linear combination of the channels,

$$\tilde{x} \equiv \mathbf{w} \cdot \mathbf{y}, \tag{3}$$

specified by some m-dimensonal weight vector $\mathbf{w}$. If we want $\tilde{x}$ to estimate the i^{th} component, then $\mathbf{y}^{(i)}$ is our signal and all the other components act as noise. Let $\mathbf{N}$ denote the covariance matrix of this generalized "noise", *i.e.*,

$$\mathbf{N} \equiv \sum_{i \neq j}^{n} \mathbf{C}^{(i)}. \tag{4}$$

The contribution to the variance $\langle \tilde{x}^2 \rangle$ of our estimator $\tilde{x}$ from signal and noise is $\mathbf{w}^t \mathbf{C}^{(i)} \mathbf{w}$ and $\mathbf{w}^t \mathbf{N} \mathbf{w}$, respectively. Maximing the signal-to-noise ratio (maximing $\mathbf{w}^t \mathbf{C}^{(i)} \mathbf{w}$ with $\mathbf{w}^t \mathbf{N} \mathbf{w}$ held fixed), we find that $\mathbf{w}$ is a solution to the generalized eigenvalue problem

$$\mathbf{C}^{(i)} \mathbf{w} = \lambda \mathbf{N} \mathbf{w}. \tag{5}$$

This is analogous to the signal-to-noise eigenmode method (Bond 1995; Bunn & Sugiyama 1995; Tegmark *et al.* 1997) used in CMB power spectrum analysis, except that the data set $\mathbf{y}$ is now the measurement at different frequencies rather than at different positions in the sky. The m different eigenvectors $\mathbf{w}$ give m uncorrelated estimators $\tilde{x}$, the least noisy one being that corresponding to the largest eigenvalue λ.

2.4. *TE96 as a special case*

Throughout the rest of this paper, we limit our attention to estimating component 0, the CMB. Since $\mathbf{C}^{(0)} \propto \mathbf{E}$, a matrix of rank 1 (with only one non-zero eigenvalue, which corresponds to the eigenvector $\mathbf{e}$), the eigenvalue problem reduces to a simple matrix inversion for this case: equation (5) gives $\mathbf{N}\mathbf{w} \propto \mathbf{E}\mathbf{w} = \mathbf{e}(\mathbf{e}^t\mathbf{w}) \propto \mathbf{e}$, so $\mathbf{w} \propto \mathbf{N}^{-1}\mathbf{e}$. Normalizing $\mathbf{w}$ so that $\mathbf{w}^t \mathbf{E} \mathbf{w} = 1$, we obtain

$$\mathbf{w} = \frac{\mathbf{N}^{-1}\mathbf{e}}{\mathbf{e}^t \mathbf{N}^{-1} \mathbf{e}}. \tag{6}$$

This normalization corresponds to $\sum w_i = \mathbf{e} \cdot \mathbf{w} = 1$, so we can interpret $\tilde{x}$ as simply a weighted average of the m channels, with w giving the weights (some weights may be negative).

In TE96, we assumed that the frequency dependence of each component was independent of position and time. Since this means that the map of a component looks the same at all frequencies, apart from an overall frequency-dependent scale factor, this assumption is equivalent to saying that each component is perfectly correlated between frequencies, *i.e.*, that $\mathbf{R}^{(i)} = \mathbf{E}$ except for $i = 1$, the instrumental noise component. Following the notation of TE96, let $\mathbf{S}$ denote the diagonal covariance matrix of the different components at some fiducial frequency ν_* (say 100 GHz), and let F_{ji} specify the r.m.s. of the i^{th} component at the j^{th} frequency relative to the value at ν_*. The correspondence between TE96 and our equations is then given by $\sigma^{(i)}_j = F_{ji} S^{1/2}_{ii}$. Defining $\mathbf{\Sigma} \equiv \mathbf{C}^{(1)}$ and $\mathbf{N}_{fg} \equiv \sum_{i=2}^{n} \mathbf{C}^{(i)}$ as the contributions to the covariance matrix $\mathbf{N}$ from receiver noise and foregrounds, respectively, we can thus write

$$\mathbf{N} = \mathbf{N}_{fg} + \mathbf{\Sigma} = \mathbf{F}\mathbf{S}\mathbf{F}^t + \mathbf{\Sigma}, \tag{7}$$

TE96 reconstructed all components, not merely the CMB, with an estimator of the form $\tilde{\mathbf{x}} = \mathbf{W}\mathbf{y}$. Since $\mathbf{R}^{(i)} = \mathbf{E}$ for all the foregrounds, equation (5) will give a single eigenvector with $\lambda > 0$ for estimating each one, just as for the CMB component. Arranging these vectors $\mathbf{w}$ as the rows of the matrix $\mathbf{W}$ and performing the relevant algebra, we obtain

$$\mathbf{W} = \mathbf{\Lambda} \mathbf{F}^t [\mathbf{F}\mathbf{S}\mathbf{F}^t + \mathbf{\Sigma}]^{-1}, \tag{8}$$

with $\Lambda_{jk} \equiv \delta_{jk} / (\mathbf{W}\mathbf{F})_{jj}$, *i.e.*, equation (36) of TE96. We have thus generalized the TE96 result, and found that it corresponds to the special case of perfect foreground correlations, $\mathbf{R}^{(i)} = \mathbf{E}$. Conversely, it is easy to show that our method (6) can be derived from the TE96 method by replacing each foreground component with many subcomponents, each with a slightly different frequency dependence, as described in §5.4 of TE96.

[3] Another advantage of working with modes rather than pixels is that it can eliminate the nuisance of a non-zero mean (most foregrounds are strictly non-negative). In a Fourier or multipole expansion, all modes but the (irrelevant) monopole will have a vanishing average.

3. THE TRADEOFF BETWEEN FOREGROUNDS AND NOISE

Our discussion above has illustrated that foregrounds are very much like detector noise — they are simply more correlated between channels. When chosing $\mathbf{w}$ to make a CMB map, there is generally a tradeoff between the amount of residual noise $\sigma_n^2 \equiv [\mathbf{w}^t\mathbf{\Sigma}\mathbf{w}]^{1/2}$ and residual foreground contamination $\sigma_{fg}^2 \equiv [\mathbf{w}^t\mathbf{N}_{fg}\mathbf{w}]^{1/2}$. This is clearly seen if we minimize σ_{fg}^2 for some fixed level of noise σ_{fg}^2, maintaining our normalization constraint $\mathbf{e}\cdot\mathbf{w} = 1$. Solving this constrained minimization problem by introducing Lagrange multipliers γ and λ, this corresponds to minimizing the expression $\mathbf{w}^t[\mathbf{N}_{fg}+\gamma\mathbf{\Sigma}]\mathbf{w}-\lambda\mathbf{e}\cdot\mathbf{w}$, which gives $\mathbf{w}\propto[\mathbf{N}_{fg}+\gamma\mathbf{\Sigma}]^{-1}\mathbf{e}$. We recognize this as our solution in equation (6), but with $\mathbf{N}=\mathbf{N}_{fg}+\mathbf{\Sigma}$ replaced by $\mathbf{N}_{fg}+\gamma\mathbf{\Sigma}$, *i.e.*, with the noise level rescaled relative to its true value by a factor γ. For the TE96 case, this gives

$$\mathbf{W} = \mathbf{\Lambda}\mathbf{F}^t[\mathbf{F}\mathbf{S}\mathbf{F}^t+\gamma\mathbf{\Sigma}]^{-1}, \tag{9}$$

which corresponds to "Method 8" in the method table of T97. Below we will see that this free parameter γ can be chosen to indicate how concerned we are about σ_n relative to σ_{fg}.

Channel spec.	COBE	MAP	LFI	HFI
ν [GHz]		22		100
FWHM [arcmin]		57		10.6
$10^6\Delta T/T$		10		1.81
ν [GHz]	31.5	30	30	143
FWHM [arcmin]	425	41	33	7.4
$10^6\Delta T/T$	77	12	1.6	2.1
ν [GHz]	51	40	44	217
FWHM [arcmin]	425	28	23	4.9
$10^6\Delta T/T$	31	10	2.4	4.6
ν [GHz]		60	70	353
FWHM [arcmin]		21	14	4.5
$10^6\Delta T/T$		13	3.6	15.0
ν [GHz]	90	90	100	545
FWHM [arcmin]	425	13	10	4.5
$10^6\Delta T/T$	34	13	4.3	144.0
ν [GHz]				857
FWHM [arcmin]				4.5
$10^6\Delta T/T$				4630

Table 1. — Specifications used for COBE, MAP and Planck.

Figure 2 shows the result of applying equation (9) to the three satellite experiments COBE, MAP and Planck, with the lines corresponding to γ ranging from 0 to ∞. Here we have used four foreground components ($n=5$): dust, free-free emission, synchrotron radiation and point sources. These are modeled as in TE96, with some minor updates as shown on Figure 1 to reflect recent foreground measurements (Bersanelli *et al.* 1996; Kogut *et al.* 1996ab; de Oliveira-Costa *et al.* 1997; Toffolatti *et al.* 1997). We have used the experimental specifications shown in Table 1, taken from Bennett *et al.* (1996) and the MAP and Planck web sites (*http://map.gsfc.nasa.gov* and *http://tonno.tesre.bo.cnr.it/research/planck/tab_sens.htm*). LFI and HFI refers to the low and high frequency instruments on board Planck.

Our derivation above showed that no method can give a point (σ_n,σ_{fg}) below or to the left of this line, *i.e.*, that equation (9) minimizes the foreground residual for any given noise level σ_{fg}. The original TE96 method ($\gamma=1$, indicated by a solid square) corresponds to minimizing the *total* residual variance, *i.e.*, $\sigma_n^2+\sigma_{fg}^2$, so in a linear-linear plot, the TE96 point lies where the line is closest to the origin. As we increase γ, the algorithm cares more about reducing noise and less about foregrounds. The upper endpoint of the line, with $\gamma=\infty$, corresponds to $\mathbf{w}\propto\mathbf{\Sigma}^{-1}\mathbf{e}$, which if the detector noise is uncorrelated ($\mathbf{R}^{(1)}=\mathbf{I}$) is a simple minimum-variance weighting, ignoring the foregrounds. For COBE, this extreme case is seen to minimize the total residuals as well — indeed, most published analyses of the COBE data made this choice, abstaining from foreground subtraction.

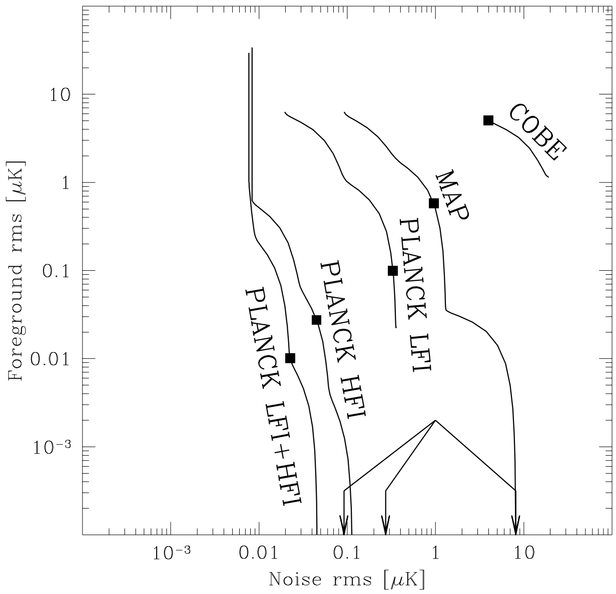

Figure 2. The curves show the smallest residual foreground level attainable for a given noise level, assuming that the frequency dependence of the foregrounds is perfectly known. The total r.m.s. residual $(\sigma_n^2+\sigma_{fg}^2)^{1/2}$ is minimized at the solid squares (the TE96 method). Arrows point to (σ_n,σ_{fg}) for the marginalization method, which is seen to give $\sigma_{fg}=0$ (off the scale), but occasionally at a higher noise cost than necessary. These curves are for a mode $a_{\ell m}$ with $\ell=10$ — due to their differences in angular resolution, the experiments differ more dramatically for larger ℓ as well as on a pixel-by-pixel basis.

There may be good reasons to be more concerned about foregrounds than detector noise. For instance, they tend to be non-Gaussian and we are usually unable to model their frequency and scale dependence as accurately as for detector noise. If we in this vein decrease γ, we move downward along the curve. For cases when the number of channels equals or exceeds the number of components (as for HFI and MAP), complete foreground removal is possible: $\sigma_{fg}\to 0$ as $\gamma\to 0$, corresponding to a weighting where $\mathbf{w}$ is the first row of $[\mathbf{F}^t\mathbf{\Sigma}^{-1}\mathbf{F}]^{-1}\mathbf{F}^t\mathbf{\Sigma}^{-1}$ ("Method 3" of T97). For these cases, the factor by which σ_n increases as we go from no foreground removal (upper endpoint) to complete foreground removal (lower endpoint) is the *Foreground Degradation Faction* (FDF) introduced by Dodelson (1996). The residual noise is also shown (by arrows at bottom) for the marginalization method derived by Dodelson (1996), in which $\mathbf{w}$ is the first row of $[\mathbf{F}^t\mathbf{F}]^{-1}\mathbf{F}^t$ ("Method 2" in T97). This is seen to give an FDF that is about a factor of 2 larger for the HFI (m=6) and HFI-LFI (m=10) cases, but identical to the TE96 method for MAP (m=5). The reason is that when there are more channels than components ($m > n$), there are $m-n$ degrees of freedom left in $\mathbf{w}$ after we have required that the foregrounds be eliminated and imposed the normalization constraint. The TE96 formula uses these extra degrees of freedom to minimize σ_n.

Although the above-mentioned reasons for trying to reduce σ_{fg} below σ_n (which might require $\gamma < 1$) may be valid, they are not grounds for outright foreground paranoia. Attempts to push σ_{fg} down say a factor of ten below σ_n are probably overkill and not worth the heavy cost in terms of increased noise. Most importantly, as we will see in the next section, such attempts are likely to be misleading, since even tiny departures from perfect correlations can reintroduce non-negligible foreground residuals. Although the $\gamma = 0$ method has been advocated as conservative (Dodelson & Stebbins 1994; Dodelson 1996), since it requires no assumptions about the *amplitude* of the foreground fluctuations, we will see that it is quite sensitive to assumptions about their frequency dependence.

4. THE EFFECT OF FREQUENCY COHERENCE

Figure 2 assumed perfect foreground correlations, $\mathbf{R}^{(i)} = \mathbf{E}$ for $i > 1$. We will now relax this assumption.

4.1. *A toy model*

To illustrate the qualitative changes that occur, let us derive a simple toy model in which we can relate the correlation matrices $\mathbf{R}^{(i)}$ to more familiar quantities. Given some foreground component i and two frequencies ν_j and ν_k, we define $\phi_- \equiv y_j^{(i)}$, $\phi_+ \equiv y_k^{(i)}$, $\phi \equiv (\phi_-\phi_+)^{1/2}$, $\eta \equiv \nu_k/\nu_j$ and $\alpha \equiv \ln(\phi_+/\phi_-)/\ln\eta$. Thus ϕ_- and ϕ_+ denote the brightness of a pixel at the two frequencies, ϕ is the (geometric) mean brightness, and α, the "color", is the spectral index for which a power law spectrum $\phi(\nu) \propto \nu^\alpha$ would connect ϕ_- with ϕ_+. With this notation, we have

$$\phi_\pm = \phi\eta^{\pm\alpha/2}. \tag{10}$$

Let us make the simplifying assumption that the brightness ϕ and the color α are statistically independent. Although probably not very accurate, this approximation is motivated by the fact that ϕ depends strongly on color-independent quantities such as the distance (in the case of radio sources) and on the amount of emitting material along the line of sight (in the case of the diffuse foreground components). Using this independence gives

$$\begin{aligned}
\langle\phi_\pm\rangle &= \langle\phi\rangle\langle\eta^{\pm\alpha/2}\rangle, &(11)\\
\langle\phi_\pm^2\rangle &= \langle\phi^2\rangle\langle\eta^{\pm\alpha}\rangle, &(12)\\
\langle\phi_-\phi_+\rangle &= \langle\phi^2\rangle. &(13)
\end{aligned}$$

We define the means and standard deviations $\bar\alpha \equiv \langle\alpha\rangle$, $\bar\phi \equiv \langle\phi\rangle$, $\Delta\alpha \equiv (\langle\alpha^2\rangle - \bar\alpha^2)^{1/2}$, $\Delta\phi \equiv (\langle\phi^2\rangle - \bar\phi^2)^{1/2}$. Let us also assume that the quantity $\Delta\alpha\ln\eta \ll 1$, so that a fairly definite spectral index will be apparent in a scatter plot of $\ln\phi_+$ against $\ln\phi_-$. Taylor expanding the exponential, this allows us to make the approximations $\langle\eta^{\pm\alpha}\rangle = \eta^{\pm\bar\alpha}\langle e^{\pm(\alpha-\bar\alpha)\ln\eta}\rangle \approx \eta^{\pm\bar\alpha}e^{(\Delta\alpha\ln\eta)^2/2}$ and $\langle\eta^{\pm\alpha/2}\rangle \approx \eta^{\pm\bar\alpha/2}e^{(\Delta\alpha\ln\eta)^2/8}$. Substituting this into equations (11)–(13), we can compute the standard deviations $\Delta\phi_\pm \equiv (\langle\phi_\pm^2\rangle - \langle\phi_\pm\rangle^2)^{1/2}$ and the correlation. We find that $\phi_+/\phi_- \approx \Delta\phi_+/\Delta\phi_- \approx \eta^{\bar\alpha}$, so the mean brightness and the r.m.s. fluctuations scale in the same way with frequency, as expected. The correlation coefficient is given by

$$\mathbf{R}_{jk}^{(i)} \equiv \frac{\langle\phi_-\phi_+\rangle - \langle\phi_-\rangle\langle\phi_+\rangle}{\Delta\phi_+\Delta\phi_-} \approx e^{-(\ln\nu_j - \ln\nu_k)^2/2\xi^2}, \tag{14}$$

where

$$\xi \equiv \frac{1}{\Delta\alpha\,(1+\beta^2)^{1/2}}. \tag{15}$$

and $\beta \equiv \bar\phi/\Delta\phi$ is the ratio of the mean brightness to the r.m.s. fluctuations. We will call the parameter ξ the *frequency coherence*, since it determines how many powers of e we can change the frequency by before the correlation starts breaking down. The two limits $\xi \to 0$ and $\xi \to \infty$ correspond to the two extreme cases $\mathbf{R}^{(i)} = \mathbf{I}$ and $\mathbf{R}^{(i)} = \mathbf{E}$ that we encountered above. Since the temperature in a foreground map typically range from its maximum down to values near zero, with the r.m.s. fluctuations $\Delta\phi$ being of the same order of magnitude as the mean $\bar\phi$, β is usually of order unity and we arrive at the following useful rule of thumb: *The frequency coherence is of the order of the inverse spectral index dispersion.*

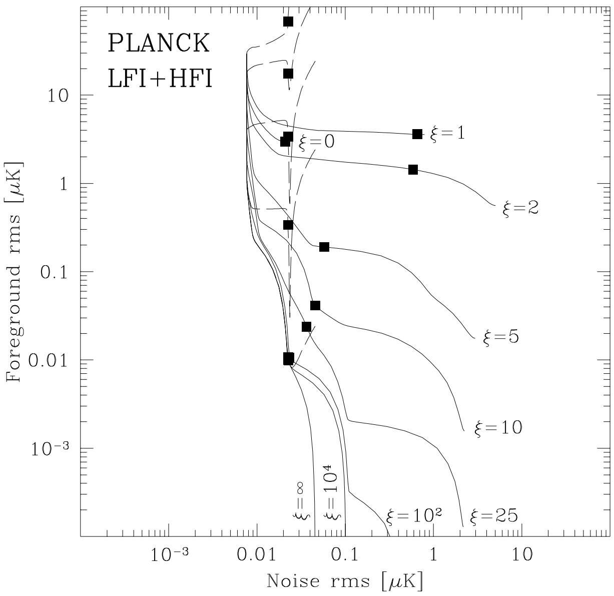

Figure 3. Same as figure 2, but for Planck HFI+LFI only, varying the frequency coherence ξ. The five dashed (unlabeled) curves show the result of assuming $\xi = \infty$ when in fact $\xi = 0$, 2, 10, 100 and 10^4 (from top to bottom).

Figure 3 shows how the Planck results from figure 2 ($\xi = \infty$) change when ξ is reduced. The solid squares correspond to equation (6), and the tradeoff curves are generated by rescaling the receiver noise contribution to $\mathbf{N}$ in equation (6) by different constants. We see that these curves follow the $\xi = \infty$ curve down from the top, then branch off to the right at a foreground level that depends on ξ. To reduce the foreground residual below this level becomes extremely costly in terms of extra noise (giving a large FDF, in Dodelson's terminology). Complete foreground removal is of course impossible when $\xi < \infty$.

The dashed curves show that if the subtraction method assumes ideal ($\xi=\infty$) foregrounds, it is disastrous to be too greedy and try to push the σ_{fg} way below σ_n, since this can actually make things worse!

Note that since we assumed that $\Delta\alpha\ln\eta \lll 1$, our derivation of equation (14) only applies to the first 3 terms in a Taylor expansion, showing that $R^{(i)}_{jk}=f(\ln\eta/\xi)$ where $f(x)=1-x^2/2+....$ We recomputed Figure 3 for a variety of such function of the form $f(x)=(1+x^2/2n)^{-n}$ ($n=\infty$ gives the Gaussian of equation (14), $n=1$ gives a Lorentzian, *etc.*), and found that the shape of the far wings of f is only of secondary importance — the main question is how correlated neighboring channels are, which for $\xi\gg 1$ depends mainly on the curvature of f near the origin. Narrower wings (larger n) can occasionally help slightly, just as $\xi=0$ is better than $\xi=1$ in Figure 3.

5. CONCLUSIONS

When removing CMB foregrounds, one can take advantage of all ways in which they differ from CMB fluctuations.

1. Non-Gaussian behavior can be exploited to throw out severely contaminated regions (*e.g.*, bright point sources, the Galactic plane).

2. Their frequency dependence can be exploited to subtract them out as we have described above.

3. Knowledge of their power spectra can be used by including residual foreground fluctuation amplitudes as additional free parameters when fitting the measured power spectrum to theoretical models.

The TE96 subtraction method (for step 2) has been shown (T97) to be lossless (retain all the cosmological information) if the foregrounds are Gaussian with $\xi=\infty$, and if the subtraction is performed mode by mode (as suggested by TE96 and implemented by Bouchet *et al.* 1995) rather than pixel by pixel — the latter destroys information by not taking advantage of correlations between neighboring pixels. In this *Letter*, we have studied the more realistic case $\xi<\infty$, and found that even spectral index variations as small as $\Delta\alpha=0.1$ make a substantial difference for the choice of method and for attainable results. Complete foreground removal becomes impossible, and attempting this nonetheless by assuming $\xi=\infty$ can even be worse than no foreground removal at all. It is easy to show that the method of equation (6) is lossless with the same assumptions, for any ξ. This means that one more property needs to be determined for each foreground component, in addition to its dependence on frequency and scale: its frequency correlations $\mathbf{R}^{(i)}$. With a simple toy model, we illustrated that this is directly linked to the spectral index dispersion $\Delta\alpha$. $\Delta\alpha$ could easily be as large as 0.1 for synchrotron radiation, 0.3 for dust, 0.01 for free-free emission and 0.5 for radio sources if we neglect sources of prior information about α. It has been argued that the spectral index for dust depends on galactic latitude (*e.g.*, Reach *et al.* 1995), whereas that for synchrotron emission is correlated with both the spectral index that can be measured at lower frequencies (Brandt *et al.* 1995) and with the degree of synchrotron polarization (Bernstein 1992). By improving our understanding and modeling of how the foreground spectral indices vary with position, it may thus be possible to reduce the effective $\Delta\alpha$, thereby improving our foreground removal and the accuracy with which cosmological parameters can be measured with the CMB.

The author wishes to thank Angélica de Oliveira-Costa, Scott Dodelson, George Efstathiou, Lyman Page and David Wilkinson for useful discussions and suggestions. This work was supported by NASA grant NAG5-6034 and Hubble Fellowship #HF-01084.01-96A, awarded by the Space Telescope Science Institute, which is operated by AURA, Inc. under NASA contract NAS5-26555.

REFERENCES

Banday, A. J. & Wolfendale, A. W. 1991, *MNRAS*, **248**, 705
Bennett, C. L. *et al.* 1996, *ApJ*, **464**, L1
Bernstein, R. 1992, B.A. Thesis, Princeton University
Bersanelli, M.. *et al.* 1996, *COBRAS/SAMBA, Phase A Study for an ESA M3 Mission*, ESA Report D/SCI(96)3, http://astro.estec.esa.nl/sa-general/projects/cobras/cobras.html
Bond, J. R. 1995, *Phys. Rev. Lett.*, **74**, 4369
Bond, J. R., Efstathiou, G., & Tegmark, M. 1997, *MNRAS*, **291**, L33
Bouchet, F. *et al.* 1995, *Space Science Rev.*, **74**, 37
Brandt, W. N. *et al.* 1994, *ApJ*, **424**, 1
Bunn, E. F., & Sugiyama, N. 1995, *ApJ*, **446**, 49
de Oliveira-Costa, A. *et al.* 1997, *ApJ*, **482**, L17
Dodelson, S., & Stebbins A 1994, *ApJ*, **433**, 440
Dodelson, S. 1996, *ApJ*, **482**, 577
Franceschini, A., Toffolatti, L., Danese, L., & de Zotti, G. 1989, *ApJ*, **344**, 35
Franceschini, A. *et al.* 1991, *A&A Supp.*, **89**, 285
Jungman, G., Kamionkowski, M., Kosowsky, A., & Spergel, D. N. 1996, *Phys. Rev. D*, **54**, 1332
Kogut, A. *et al.* 1996a, *ApJ*, **460**, 1
Kogut, A. *et al.* 1996b, *ApJ*, **464**, L5
Reach, W. T.. *et al.* 1995, *ApJ*, **451**, 188
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1997, astro-ph/9710327
Platania, P. *et al.* 1997, astro-ph/9707252
Tegmark, M., & Efstathiou, G. 1996, *MNRAS*, **281**, 1297 ("TE96")
Tegmark, M., Taylor, A., & Heavens, A. F. 1997, *ApJ*, **480**, 22
Tegmark, M. 1997, *ApJ*, **480**, L87 ("T97")
Toffolatti, M. *et al.* 1997, astro-ph/9711085
Zaldarriaga, M., Spergel, D., & Seljak, U. 1997, *ApJ*, **488**, 1

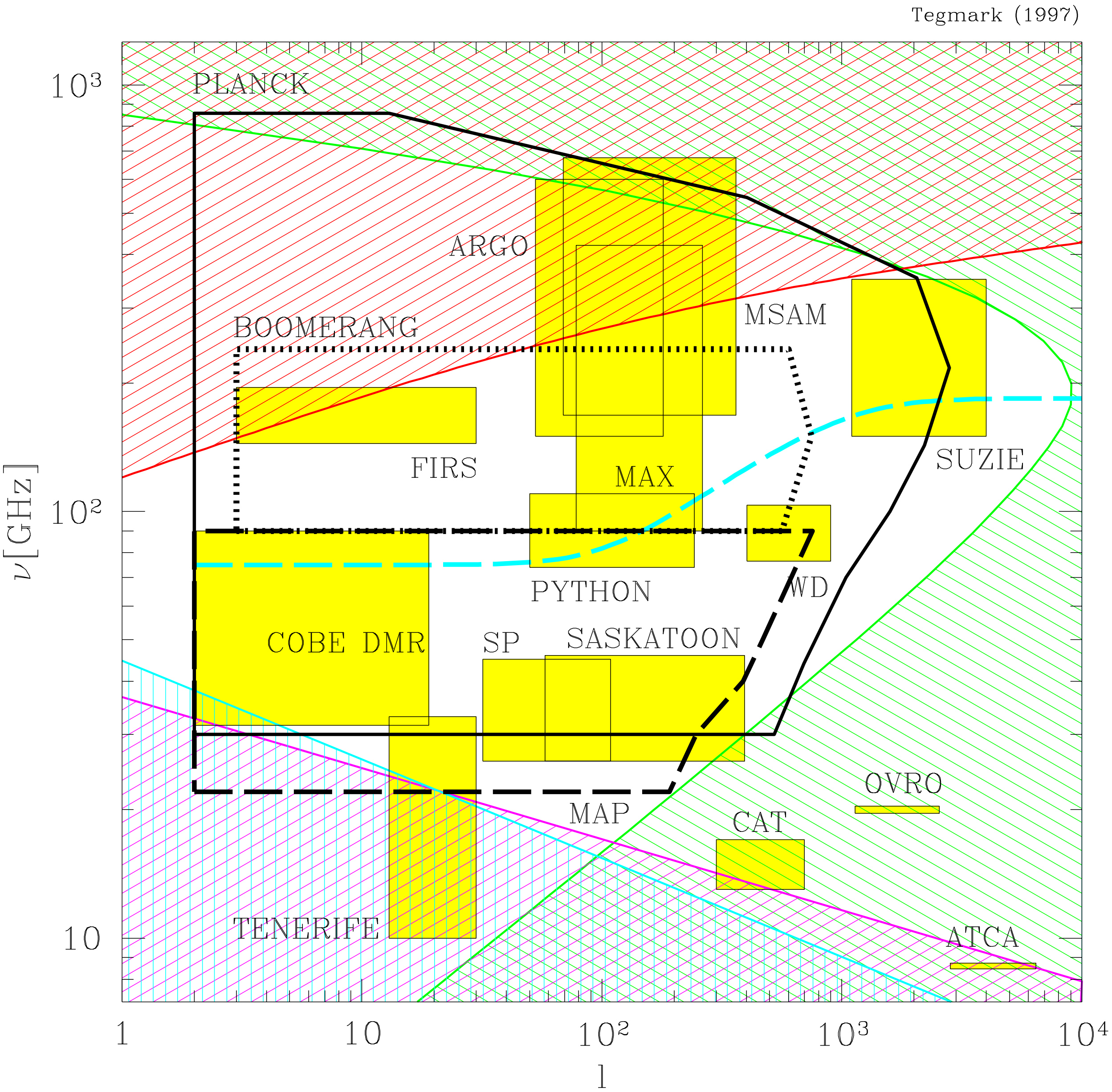

Figure 1. Where various foregrounds dominate. The shaded regions indicate where the various foregrounds cause fluctuations exceeding those of COBE-normalized scale-invariant fluctuations, thus posing a substantial challenge to estimation of genuine CMB fluctuations. They correspond to dust (top), free-free emission (lower left), synchrotron radiation (lower left, vertically shaded) and point sources (lower and upper right). The heavy dashed line shows the frequency where the total foreground contribution to each multipole is minimal. The boxes roughly indicate the range of multipoles ℓ and frequencies ν probed by various CMB experiments, as in TE96.