First submitted November 1, 1996

HOW TO MAKE MAPS FROM CMB DATA WITHOUT LOSING INFORMATION [1]

Max Tegmark[2]

Institute for Advanced Study, Olden Lane, Princeton, NJ 08540; Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85740 Garching;

email: max@ias.edu

Abstract

The next generation of CMB experiments can measure cosmological parameters with unprecedented accuracy — in principle. To achieve this in practice when faced with such gigantic data sets, elaborate data analysis methods are needed to make it computationally feasible. An important step in the data pipeline is to make a map, which typically reduces the size of the data set by orders of magnitude. We compare ten map-making methods, and find that for the Gaussian case, both the method used by the COBE DMR team and various forms of Wiener filtering are optimal in the sense that the map retains all cosmological information that was present in the time-ordered data (TOD). Specifically, one obtains just as small error bars on cosmological parameters when estimating them from the map as one could have obtained by estimating them directly from the TOD. The method of simply averaging the observations of each pixel (for total-power detectors), on the contrary, is found to generally destroy information, as does the maximum entropy method and most other non-linear map-making techniques.

Since it is also numerically feasible, the COBE method is the natural choice for large data sets. Other lossless (e.g. Wiener-filtered) maps can then be computed directly from the COBE method map.

---

[1] Published in *ApJ Lett.*, **480**, L87-L90.
Available from *h t t p://www.sns.ias.edu/~max/mapmaking.html* (faster from the US) and from *h t t p://www.mpa-garching.mpg.de/~max/mapmaking.html* (faster from Europe).
Please note that figure 1 will print in color if your printer supports it.

[2] Hubble Fellow.

1 INTRODUCTION

A large number of Cosmic Microwave Background (CMB) experiments that cover extended patches of sky are currently in the phases of planning, design or data analysis, and they all have as partial goals to produce temperature maps. Since there are a plethora of map-making methods available, many experimental groups are currently debating which one(s) to use when reducing their raw data. Indeed, it is indicative that the maps based on COBE (Smoot et al. 1992; Bennett et al. 1996), MAX (White & Bunn 1995), Saskatoon (Tegmark et al. 1996) and Tenerife were made using four different methods, two linear and two non-linear. It is therefore quite timely to compare the various methods and assess their relative merits. The purpose of this *Letter* is to provide such a comparison. Note that we use the term "map-making" to refer to the data *reduction* process — for a discussion of important options in the data *acquisition* process such as scanning and chopping strategy, see e.g. Knox (1996) and Wright (1996).

Which map-making method is preferable clearly depends on what the map is to be used for. Common uses for CMB maps (apart from satisfying a general desire to map the sky in as many frequency bands as possible) are

- to facilitate comparison with other experiments.
- to facilitate comparison with foreground templates such as the DIRBE maps.
- to reveal flaws in the model that are not visible in the power spectrum, such as non-Gaussian CMB features, point sources and spatially distinctive systematic problems.

As CMB experiments collect larger and larger data sets, yet another use for map-making has emerged: as a data-compression step that makes it computationally feasible to constrain cosmological parameters. To a good approximation (Tegmark, Taylor & Heavens 1997), one obtains the smallest possible error bars on estimates of cosmological parameters (such as Ω, Λ, etc.) by performing a likelihood analysis using the entire data set. So far, the small-scale experiments have all produced $n \ll 10^4$ data points, which means that it has been feasible to carry out such a "brute force" analysis. Assuming Gaussianity in the distribution of pixel temperatures and instrument noise, this entails computing determinants of $n \times n$ matrices at a grid of points in parameter space, and the time this takes scales as n^3. For the time-ordered data (TOD) of COBE (with $n \sim 2 \times 10^8$), such brute-force analysis is completely unfeasible at present, not to mention the even larger data sets of the upcoming MAP and COBRAS/SAMBA satellites. Map-making offers a useful way to reduce the data set down to a more manageable size, for instance down to 6144 numbers in the case of COBE and $10^6 - 10^7$ for future satellite missions. The parameters can then be estimated from the maps with the brute force approach (Tegmark & Bunn 1995; Hinshaw et al. 1996) or with some faster and more elaborate scheme. This is schematically illustrated in Figure 1. This purely pragmatic approach to map-making, as a mere time-saving device, offers an objective quantitative way to rank map-making methods: one method is better than another if it retains more of the cosmological information, which operationally means that it will lead to smaller error bars on the parameter estimates.

The rest of this *Letter* is organized as follows. We describe ten map-making methods in Section 2, compare them according to this criterion in Section 3 and summarize our conclusions in Section 4.

2 A LIST OF METHODS

2.1 The mapping problem

Suppose we have measured n numbers $y_1, ..., y_n$, which we will refer to as the *raw data* or the *time-ordered data* (TOD), and wish to use this TOD to estimate a set of m numbers $x_1, ..., x_m$ which we will refer to as a *map*. Typically, our map would be pixelized and x_i would denote the temperature in pixel i. We will limit our treatment to the case where the time-ordered data (TOD) depends *linearly* on the map. Grouping the TOD and the map into an n-dimensional vector $\mathbf{y}$ and an m-dimensional vector $\mathbf{x}$, respectively, this means that we can write

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \tag{1}$$

for some known matrix $\mathbf{A}$ and some random noise vector $\mathbf{n}$. Despite the linearity limitation, this formalism is still very general. The numbers in the vector $\mathbf{x}$ need not be restricted to CMB temperatures in various pixels, but can also include any other unknown parameters upon which the TOD depends linearly. For instance, the COBE analysis included a fit for three magnetic susceptibility coefficients (Wright et al. 1996), and in many cases, it may also be convenient to include various calibration-related parameters in $\mathbf{x}$. To remove foregrounds, the TOD vector $\mathbf{y}$ can be expanded to include the temperatures measured at several different frequencies. In this case, $\mathbf{x}$ would

No.	Method	Specification
1	Generalized COBE	$\mathbf{W}=[\mathbf{A}^t\mathbf{M}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{M}$
2	Bin averaging	$\mathbf{W}=[\mathbf{A}^t\mathbf{A}]^{-1}\mathbf{A}^t$
3	COBE	$\mathbf{W}=[\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}$
4	Wiener 1	$\mathbf{W}=\mathbf{S}\mathbf{A}^t[\mathbf{A}\mathbf{S}\mathbf{A}^t+\mathbf{N}]^{-1}$
5	Wiener 2	$\mathbf{W}=[\mathbf{S}^{-1}+\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}$
6	Saskatoon	$\mathbf{W}=[\eta\mathbf{S}^{-1}+\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1}$
7	TE96	$\mathbf{W}=\mathbf{\Lambda}\mathbf{S}\mathbf{A}^t[\mathbf{A}\mathbf{S}\mathbf{A}^t+\mathbf{N}]^{-1},\ (\mathbf{W}\mathbf{A})_{ii}=1$
8	TE97	$\mathbf{W}=\mathbf{\Lambda}[\eta\mathbf{S}^{-1}+\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{A}^t\mathbf{N}^{-1},\ (\mathbf{W}\mathbf{A})_{ii}=1$
9	Maximum probability	Nonlinear method if non-Gaussian
10	Maximum entropy	Nonlinear method

Table 1: Map-making methods

be augmented to include the brightness of various foreground components in each pixel, and the matrix $\mathbf{A}$ would encompass the assumptions made about their frequency dependence.

Without loss of generality, we can take the noise vector to have zero mean, i.e., $\langle\mathbf{n}\rangle=0$, so the noise covariance matrix is

$$\mathbf{N}\equiv\langle\mathbf{n}\mathbf{n}^t\rangle. \tag{2}$$

In some of the methods described below (methods 4-9), the following prior assumptions are made about the map: it is assumed to be a realization of random vector with zero mean, i.e., $\langle\mathbf{x}\rangle=0$, with some known covariance matrix

$$\mathbf{S}\equiv\langle\mathbf{x}\mathbf{x}^t\rangle \tag{3}$$

and uncorrelated with the noise, i.e., $\langle\mathbf{n}\mathbf{x}^t\rangle=0$.

2.2 Ten mapping methods

We will now summarize some map-making methods that have recently been used or advocated in the CMB context. All *linear* methods can clearly be written in the form

$$\tilde{\mathbf{x}}=\mathbf{W}\mathbf{y}, \tag{4}$$

where $\tilde{\mathbf{x}}$ denotes the estimate of the map $\mathbf{x}$ and $\mathbf{W}$ is some $m\times n$ matrix that specifies the method. Table 1 shows the choices of $\mathbf{W}$ that define the linear methods we will discuss.

Method 1 has the attractive property that $\mathbf{WA} = \mathbf{I}$, which means that the reconstruction error ε, defined as

$$\varepsilon \equiv \tilde{\mathbf{x}} - \mathbf{x} = [\mathbf{WA} - \mathbf{I}]\mathbf{x} + \mathbf{Wn} \tag{5}$$

becomes independent of $\mathbf{x}$. In other words, the recovered map $\tilde{\mathbf{x}}$ is simply the true map $\mathbf{x}$ plus some noise that is independent of the signal one is trying to measure. Here $\mathbf{M}$ is an arbitrary $n \times n$ matrix.

Method 2 is the special case of method 1 for which $\mathbf{M} = \mathbf{I}$. It can be derived by minimizing $|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}|$, the mismatch between the observed and expected data sets (Dodelson 1996). If the data set consists of "total power" (undifferenced) observations of the sky, then the i^{th} row of $\mathbf{A}$ will vanish except for a 1 in the column corresponding to the pixel observed at time step i, and it is easy to see that Method 2 corresponds to simply averaging the measurements of each pixel. As we will see, this is an inferior method when noise correlations (due to for instance $1/f$-noise) are present.

Method 3, the method used by the COBE/DMR team (Jansen & Gulkis 1992), is the special case of Method 1 where $\mathbf{M} = \mathbf{N}^{-1}$. It is straightforward to prove that it has the following three desirable properties:

1. It minimizes $\chi^2 \equiv (\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}})^t\mathbf{N}^{-1}(\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}})$.
2. It minimizes $\langle|\varepsilon|^2\rangle$ subject to the constraint $\mathbf{WA} = \mathbf{I}$.
3. It is the maximum-likelihood estimate of $\mathbf{x}$ if the probability distribution for $\mathbf{n}$ is Gaussian.

For this method, the noise covariance matrix in the map is $\mathbf{\Sigma} \equiv \langle\varepsilon\varepsilon^t\rangle = [\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}$.

Method 4, known as Wiener filtering (Wiener 1949), can be derived in two ways (see e.g. Bunn et al. 1994, Zaroubi et al. 1995):

1. It minimizes $\langle|\varepsilon|^2\rangle$.
2. It is the maximum posterior probability estimate of $\mathbf{x}$ in a Bayesean analysis if the probability distributions for $\mathbf{n}$ and $\mathbf{x}$ are Gaussian.

It is stable even for "poorly connected" observations where $[\mathbf{A}^t\mathbf{MA}]$ is singular or ill-conditioned. Although **Method 5** looks different, it is in fact identical to Method 4. This can be proven using the same geometric series trick that is employed in equation (15) below. It is computationally preferable over Method 4 if the matrix to be inverted is smaller, i.e., if $m < n$. **Method 6** lets the user choose a desired signal-to-noise ratio in the reconstructed map by means of the parameter η, and was used in generating the maps from the Saskatoon experiment (Tegmark et al. 1996). The COBE method clearly corresponds to the special case $\eta \to 0$.

Wiener filtering generally gives less noisy maps at the price of suppressing the power in different pixels unequally. This is remedied by **Method 7**, which simply multiplies $\mathbf{W}$ by a diagonal matrix $\mathbf{\Lambda}$ (rescales each pixel) so that $(\mathbf{WA})_{ii} = 1$ for all i. This method can also be derived by minimizing $\langle|\varepsilon|^2\rangle$ subject to the constraint $(\mathbf{WA})_{ii} = 1$ (Tegmark & Efstathiou 1996). In that paper, $\mathbf{x}$ did not denote a map but the CMB and foreground fluctuations in a given mode, but the mathematics is of course identical. **Method 8** simply combines the features of 6 and 7, and is relevant to the foreground problem (Tegmark & Efstathiou 1997).

As mentioned above, **Method 9** (the maximum posterior probability method) reduces to Wiener filtering when all probability distributions are Gaussian. When this is not the case, $\tilde{\mathbf{x}}$ is a non-linear function of $\mathbf{y}$ which must generally be determined numerically. A special case of this is **Method 10**, the Maximum Entropy Method (MEM) (see e.g. Press et al. 1992; White & Bunn 1995), which is also non-linear. Here the prior probability distribution involves the entropy of the map, a measure of how smooth and featureless it is.

3 WHICH METHODS DESTROY INFORMATION?

Which of the above-mentioned map-making methods is preferable clearly depends on what the map is to be used for. However, if the map is to be used for constraining cosmological parameters, we can make quite strong statements as to which methods are better and which are worse. Specifically, we will consider a method to be better than another if the map it produces allows the cosmological parameters to be measured with smaller error bars.

3.1 The Fisher Information Matrix

Let $\mathbf{\Theta}$ denote a vector consisting of the parameters we wish to estimate. For instance, Jungman et al. (1996) assess attainable accuracies by choosing

$$\mathbf{\Theta} = (\Omega, \Omega_b, h, \Lambda, n_S, r, n_T, T/S, \tau, Q, N_\nu), \tag{6}$$

the density parameter, the baryon density, the Hubble parameter, the cosmological constant, the spectral index of scalar fluctuations, the "running" of this index, the spectral index of tensor fluctuations, the quadrupole tensor-to-scalar ratio, the optical depth to reionization, and the number of light neutrino species, respectively. As described in detail in Tegmark, Taylor & Heavens (1997), the best possible unbiased estimates of these parameters will have a covariance matrix that is well approximated by $\mathbf{F}^{-1}$, where $\mathbf{F}$ is known as the *Fisher Information Matrix*. For the case where the data has a Gaussian distribution with zero mean and a covariance matrix $\mathbf{C}$, $\mathbf{F}$ is given by

$$\mathbf{F}_{ij} = \frac{1}{2}\mathrm{tr}\,\mathbf{G}_i\mathbf{G}_j, \tag{7}$$

where

$$\mathbf{G}_i \equiv \mathbf{C}^{-1}\mathbf{C}_{,i} \tag{8}$$

and the comma notation $\mathbf{C}_{,i}$ is shorthand for $d\mathbf{C}/d\theta_i$. This means that if all parameters except θ_i are known, the data set contains enough information to determine θ_i with error bar $\Delta\theta_i = 1/\mathbf{F}_{ii}$, whereas if we need to determine all parameters jointly, we can obtain $\Delta\theta_i = (\mathbf{F}^{-1})_{ii}$. It is in this sense that $\mathbf{F}$ is a measure of how much information the data contains about the parameters, and loosely speaking, the larger $\mathbf{F}$ is, the better.

3.2 The notion of a lossless map

Since the time-ordered data (TOD) contains all the information we have, computing $\mathbf{F}$ directly from the TOD places a rock-bottom lower limit on the error bars we can hope to attain. Although these minimal error bars can generally be attained with a brute-force likelihood analysis of the TOD, this unfortunately tends to be computationally unfeasible in practice, since even in the Gaussian case that we are considering, this involves repeated determinant calculations (essentially Cholesky decompositions) of $n\times n$ matrices. For COBE, we had $n \sim 2\times 10^8$, as compared to $m = 6144$. This is why map-making is such a useful intermediate step, reducing the data set to a more manageable size. By computing $\mathbf{F}$ from the map, we can assess the effectiveness of the map-making method. If $\mathbf{F}^{\mathrm{map}} = \mathbf{F}^{\mathrm{tod}}$, the map is lossless in the sense that it contains all the cosmological information that the TOD did, in a distilled form. Conversely, if $\mathbf{F}^{\mathrm{map}} \neq \mathbf{F}^{\mathrm{tod}}$, some useful information has been destroyed in the map-making process.

Are any of the above-mentioned methods lossless? First of all, note that $\mathbf{F}$ remains unchanged if we multiply our data set by an invertible matrix $\mathbf{B}$: if the new data set is $\mathbf{x}' = \mathbf{B}\mathbf{x}$, then $\mathbf{C}' = \mathbf{B}\mathbf{C}\mathbf{B}^t$, $\mathbf{G}_i' = \mathbf{C}'^{-1}\mathbf{C}'_{,i} = \mathbf{B}^{-t}\mathbf{G}_i\mathbf{B}^t$, and $\mathbf{F}' = \mathbf{F}$. This is simple to understand intuitively: $\mathbf{x}'$ must clearly contain the same information that $\mathbf{x}$ does, since $\mathbf{x}$ can be computed from $\mathbf{x}'$. This elementary observation immediately tells us that methods 3-8 are are information-theoretically equivalent, giving the same $\mathbf{F}$, since each of these six $\mathbf{W}$-matrices can be obtained from each of the other by multiplying by some invertible matrix from the left. For instance, we can compute a Wiener-filtered map $\mathbf{x}'$ from a COBE map $\mathbf{x}$ by multiplying it by $\mathbf{B} = [\mathbf{S}^{-1} + \mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}[\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]$ as was done by Bunn et al. (1994) and Bunn et al. (1996).

3.3 A proof that methods 3-8 lose no information

We will now compute the Fisher matrices $\mathbf{F}^{\mathrm{map}}$ from the maps made with methods 3-8. As mentioned above, they are all identical, and do not change if we multiply $\mathbf{W}$ from the left by an arbitrary invertible matrix. Let us take advantage of this by making the simple choice $\mathbf{W} = \mathbf{A}^t\mathbf{N}^{-1}$ in our calculation (for instance, method 3 can be put in this form by multiplying its $\mathbf{W}$ by $\mathbf{\Sigma}^{-1} = [\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]$). This gives

$$\begin{aligned}
\mathbf{C}^{\mathrm{map}} &= \langle\tilde{\mathbf{x}}\tilde{\mathbf{x}}^t\rangle = \mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A} + \mathbf{A}^t\mathbf{N}^{-1}\mathbf{N}\mathbf{N}^{-1}\mathbf{A} \\
&= \mathbf{\Sigma}^{-1}[\mathbf{I} + \mathbf{S}\mathbf{\Sigma}^{-1}], && (9) \\
\mathbf{C},_i^{\mathrm{map}} &= \mathbf{\Sigma}^{-1}\mathbf{S},_i\mathbf{\Sigma}^{-1}, && (10) \\
\mathbf{G}_i^{\mathrm{map}} &= [\mathbf{I} + \mathbf{S}\mathbf{\Sigma}^{-1}]^{-1}\mathbf{S},_i\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}. && (11)
\end{aligned}$$

For the time-ordered data, the corresponding expressions are

$$\begin{aligned}
\mathbf{C}^{\mathrm{tod}} &= \langle\mathbf{y}\mathbf{y}^t\rangle = \mathbf{A}\mathbf{S}\mathbf{A}^t + \mathbf{N}, && (12) \\
\mathbf{C},_i^{\mathrm{tod}} &= \mathbf{A}\mathbf{S},_i\mathbf{A}^t, && (13) \\
\mathbf{G}_i^{\mathrm{tod}} &= [\mathbf{A}\mathbf{S}\mathbf{A}^t + \mathbf{N}]^{-1}\mathbf{A}\mathbf{S},_i\mathbf{A}^t = [\mathbf{I} + \mathbf{N}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}^t]^{-1}\mathbf{N}^{-1}\mathbf{A}\mathbf{S},_i\mathbf{A}^t && (14)
\end{aligned}$$

Since matrices of the form $[\mathbf{I} + \mathbf{M}]^{-1}$ can be expanded as a geometric series $\mathbf{I} - \mathbf{M} + \mathbf{M}^2 - \mathbf{M}^3 + ...$, we obtain

$$\begin{aligned}
\mathbf{G}_i^{\mathrm{tod}} &= [\mathbf{I} - \mathbf{N}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}^t + \mathbf{N}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}^t - ...]\mathbf{N}^{-1}\mathbf{A}\mathbf{S},_i\mathbf{A}^t \\
&= \mathbf{N}^{-1}\mathbf{A}[\mathbf{I} - \mathbf{S}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A} + \mathbf{S}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}\mathbf{S}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A} - ...]\mathbf{S},_i\mathbf{A}^t \\
&= \mathbf{N}^{-1}\mathbf{A}[\mathbf{I} + \mathbf{S}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A}]^{-1}\mathbf{S},_i\mathbf{A}^t \\
&= \mathbf{N}^{-1}\mathbf{A}[\mathbf{I} + \mathbf{S}\mathbf{\Sigma}^{-1}]^{-1}\mathbf{S},_i\mathbf{A}^t. && (15)
\end{aligned}$$

Comparing equations (11) and (15), we see that the matrices $\mathbf{G}_i^{\mathrm{map}}$ and $\mathbf{G}_i^{\mathrm{tod}}$ differ only by a cyclic permutation, moving the factor $\mathbf{N}^{-1}\mathbf{A}$ from one side to the other. Since a trace of a product of matrices is invariant under cyclic permutations, we obtain our desired result:

$$\mathbf{F}_{ij}^{\text{map}} = \frac{1}{2}\text{tr}\,\mathbf{G}_i^{\text{map}}\mathbf{G}_j^{\text{map}} = \frac{1}{2}\text{tr}\,\mathbf{G}_i^{\text{tod}}\mathbf{G}_j^{\text{tod}} = \mathbf{F}_{ij}^{\text{tod}}. \tag{16}$$

In other words, methods 3-8 are all lossless, regardless of what parameters we choose to estimate.

4 CONCLUSIONS

We have compared ten methods for making maps from CMB data. We found that for the Gaussian case, both the COBE method and assorted variants of Wiener filtering are optimal in the sense that they retain all the cosmological information that was present in the time-ordered data. The choice between them is mainly one of numerical convenience, since these six maps (and indeed any lossless maps) can all be computed from one another without going back to the TOD. The linear methods 1 and 2, on the other hand, destroy information whenever they differ from Method 3, i.e., unless $\mathbf{M} = \mathbf{N}^{-1}$ in method 1 or $\mathbf{N} \propto \mathbf{I}$ in Method 2. Among other things, this means that in the presence of $1/f$-noise, we should not simply average the observation in each pixel, since we can do better. The non-linear methods 9 and 10 also destroy information unless they can be inverted to reproduce say map 3, the map from the COBE method.

Our proof that methods 3-8 are lossless was strictly valid only if both signal and noise are Gaussian. However, as long as the noise in the TOD is Gaussian (after appropriate removal of glitches, known systematics etc.), the same results hold even if the sky pattern is non-Gaussian. Letting $f_x(\mathbf{x};\mathbf{\Theta})$ denote the (not necessarily Gaussian) probability distribution for the map $\mathbf{x}$, the likelihood function for the parameter vector $\mathbf{\Theta}$ is

$$L(\mathbf{\Theta}) = \int f_n(\mathbf{y}-\mathbf{A}\mathbf{x})f_x(\mathbf{x};\mathbf{\Theta})d^m x, \tag{17}$$

where f_n is the Gaussian noise probability distribution $f_n(\mathbf{n}) \propto \exp[-\mathbf{n}^t\mathbf{N}^{-1}\mathbf{n}/2]$. Proportionality constants that are independent of $\mathbf{\Theta}$ are of course irrelevant in a likelihood analysis, so since

$$\begin{aligned} L(\mathbf{\Theta}) &= e^{-\frac{1}{2}\mathbf{y}^t\mathbf{N}^{-1}\mathbf{y}}\int e^{-\frac{1}{2}\mathbf{x}^t\mathbf{A}^t\mathbf{N}^{-1}[\mathbf{A}\mathbf{x}-2\mathbf{y}]}f_x(\mathbf{x};\mathbf{\Theta})d^m x \\ &\propto \int e^{-\frac{1}{2}\mathbf{x}^t\mathbf{\Sigma}^{-1}[\mathbf{x}-2\tilde{\mathbf{x}}]}f_x(\mathbf{x};\mathbf{\Theta})d^m x, \end{aligned} \tag{18}$$

where $\tilde{\mathbf{x}} \equiv \mathbf{\Sigma}\mathbf{A}^t\mathbf{N}^{-1}\mathbf{y}$ is the map made with Method 3 and $\mathbf{\Sigma} = (\mathbf{A}^t\mathbf{N}^{-1}\mathbf{A})^{-1}$ as before, we see that our likelihood function depends on the the data $\mathbf{y}$ only indirectly, via $\tilde{\mathbf{x}}$. In other words, we can compute the full TOD likelihood function directly from the map made with the COBE method. This shows that even if the CMB fluctuations are non-Gaussian, Method 3 (and consequently also 4-8) are lossless, so that we will get the strongest possible constraints on cosmological models by splitting the data processing into two steps, as in Figure 1:

1. Use one of the simple linear methods 3-8 to compress the TOD into a map.
2. Use this map as the starting point for any non-linear data processing (for removing point sources, for detecting topological defects, etc.).

The fact that Methods 9 and 10 destroy information is of course not an argument against nonlinearly processed maps *per se* even in the Gaussian case, since maps have other uses than parameter estimation. The point is simply that these methods are inferior (slower and not lossless) in the process of *data compression* from TOD to a map, so if one wants for instance a maximum entropy map from a huge data set, it is better to split the data processing into the above-mentioned two steps.

This is quite good news for the CMB community, since it has recently been demonstrated (Wright et al. 1996; Wright 1996) that clever algorithms make it numerically feasible to make maps with the COBE method (Method 3) even when millions of pixels are involved. This makes it the natural choice as the first step in the data compression pipeline, since the other lossless methods can be computed directly from this map if desired, without using the TOD. Two additional desirable properties of the COBE method reenforce this conclusion:

- It is independent of $\mathbf{S}$, i.e., of cosmological model assumptions.
- With a well chosen observational strategy, the covariance matrix $\mathbf{\Sigma}$ of the map is approximately diagonal (Wright 1996), simplifying subsequent analysis.

In conclusion, although much work remains to be done on other aspects of CMB data analysis, the map-making problem now appears to be under control, since we are armed with methods that are both optimal and feasible.

Support for this work was provided by NASA through a Hubble Fellowship, #HF-01084.01-96A, awarded by the Space Telescope Science Institute, which is operated by AURA, Inc. under NASA contract NAS5-26555.

5 REFERENCES

Bennett, C. L. 1996, *ApJ*, **464**, L1.

Bunn, E. F. et al. 1994, *ApJ*, **432**, L75.

Bunn, E. F., Hoffman, Y & Silk, J 1996, *ApJ*, **464**, 1.

Dodelson, S. 1996, preprint astro-ph/9512021.

Hinshaw, G. et al. 1996, *ApJL*, **464**, L17.

Jansen, D. J. & Gulkis, S. 1992, "Mapping the Sky With the COBE-DMR", in "The Infrared and Submillimeter Sky after COBE", eds. M. Signore & C. Dupraz (Dordrecht:Kluwer).

Jungman, G.. Kamionkowski, M., Kosowsky, A & Spergel, D. N. 1996, *Phys. Rev. D*, **54**, 1332.

Knox, L. 1996, preprint astro-ph/9606066.

Press, W. H., Flannery, B. P., Teukolski, S. A. & Vetterling, W. T. 1992, *Numerical Recipes*, 2nd ed. (New York, Cambridge Univ. Press).

Smoot, G. F. et al. 1992, *ApJ*, **396**, L1.

Tegmark, M. & Bunn, E. F. 1995, *ApJ*, **455**, 1.

Tegmark, M., de Oliveira-Costa, A., Devlin, M. J., Netterfield, C. B, Page, L. & Wollack, E. J. 1996, *ApJL*, **474**, L77.

Tegmark, M. & Efstathiou, G. 1996, *MNRAS*, **281**, 1297.

Tegmark, M. & Efstathiou, G. 1997, in preparation.

Tegmark, M., Taylor, A. & Heavens, A. F. 1997, preprint astro-ph/9603021

White, M. & Bunn, E. F. 1995, *ApJ*, **443**, L53.

Wiener, N. 1949, *Extrapolation and Smoothing of Stationary Time Series* (NY: Wiley).

Wright, E. L. 1996, preprint astro-ph/9612006.

Wright, E. L., Hinshaw, G. & Bennett, C. L. 1996, *ApJL*, **458**, L53.

Zaroubi, S. et al. 1995, *ApJ*, **449**, 446.

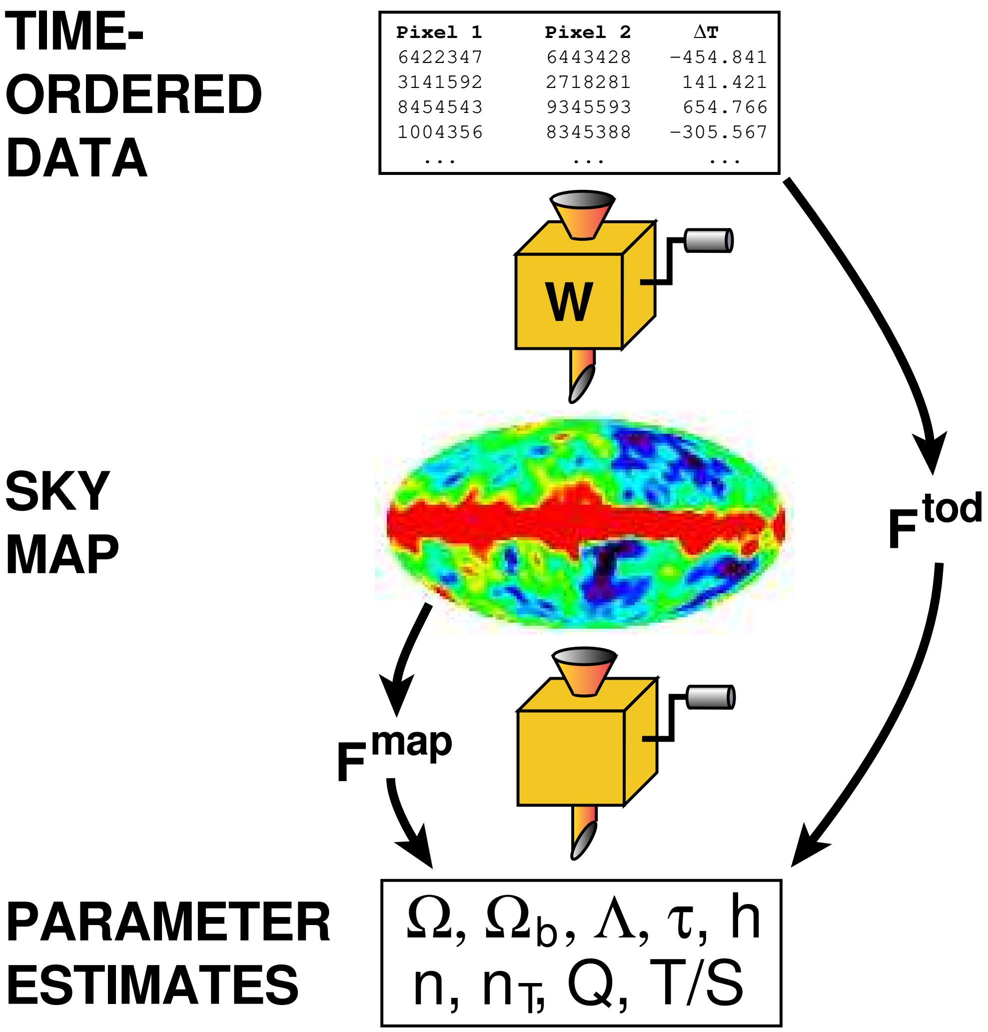

Figure 1: Map-making as an intermediate step in measuring cosmological parameters. If $\mathbf{F}^{\rm map}=\mathbf{F}^{\rm tod}$, then the map-making method $\mathbf{W}$ is lossless, which means that parameter estimation based on the map gives just as small error bars as using all the time-ordered data.