Available online at www.sciencedirect.com
SCIENCE DIRECT®
C. R. Physique 4 (2003) 871–879
COMPTES RENDUS PHYSIQUE

The Cosmic Microwave Background/Le rayonnement fossile à 3K

CMB map-making and power spectrum estimation

Jean-Christophe Hamilton [a,b,c]

[a] *LPNHE (CNRS-IN2P3), Paris VI & VII, 4, place Jussieu, 75252 Paris cedex 05, France*
[b] *LPSC (CNRS-IN2P3), 53, avenue des Martyrs, 38026 Grenoble cedex, France*
[c] *PCC (CNRS-IN2P3), 11, place Marcelon Berthelot, 75231 Paris cedex 05, France*

Presented by Guy Laval

Abstract

CMB data analysis is in general performed through two main steps: map-making of the time data streams and power spectrum extraction from the maps. The latter basically consists in the separation between the variance of the CMB and that of the noise in the map. Noise must therefore be deeply understood so that the estimation of CMB variance (the power spectrum) is unbiased. General techniques to make maps from time streams and to extract the power spectrum from them are presented in this article. We will see that exact, maximum likelihood solutions are in general too slow and hard to deal with to be used in modern experiments such as Archeops and should be replaced by approximate, iterative or Monte Carlo approaches that lead to similar precision. ***To cite this article: J.-Ch. Hamilton, C. R. Physique 4 (2003).***

Résumé

Fabrication de cartes de CMB et estimation du spectre de puissance. L'analyse de données CMB consiste en général en deux étapes : fabrication de cartes à partir des données temporelles et extraction du spectre de puissance à partir des cartes. La dernière étape consiste essentiellement en une séparation entre les variance du CMB et du bruit sur la carte. Le bruit doit par conséquent être parfaitement compris afin que l'estimation de la variance du CMB (le spectre de puissance) soit non biaisée. Je présente dans cet article des techniques générales pour fabriquer des cartes à partir de données temporelles et en extraire le spectre de puissance. Nous verrons que les méthodes exactes, maximisant la vraisemblance, sont en général trop lentes et difficiles à manipuler pour être utilisables dans les expériences récentes comme Archeops. On doit donc les remplacer par des méthodes approximatives, itératives ou reposant sur des simulations Monte Carlo qui conduisent à une précision comparable. ***Pour citer cet article : J.-Ch. Hamilton, C. R. Physique 4 (2003).***

Keywords: Cosmology; Cosmic Microwave Background anisotropies; Data analysis; Early Universe

Mots-clés : Cosmologie ; Anisotropies du fond de rayonnement cosmique ; Traitement des donnés ; Univers primordial

1. Introduction

The cosmological information contained in the Cosmic Microwave Background (CMB) anisotropies is encoded in the angular size distribution of the anisotropies, hence in the angular power spectrum and noted C_ℓ. It is of great importance to be able to compute the C_ℓ spectrum in an unbiased way. The simplest procedure to obtain the power spectrum is to first construct a map of the CMBA from the data timelines giving the measured temperature in one direction of the sky following a given scanning strategy on the sky; this is known as the map-making process. The C_ℓ from this map can be extracted; this is the

E-mail address: hamilton@in2p3.fr (J.-Ch. Hamilton).

1631-0705/$ – see front matter
doi:10.1016/S1631-0705(03)00108-7

power spectrum extraction. Various effects usually present in the CMB data make these two operations nontrivial. The major effect is related to the unavoidable presence of instrumental and photon noise. Noise in the timelines is correlated and appears as low frequency drifts that are still present in the map. A good map-making process minimizes these drifts, but in most cases, they are still present in the map. They have to be accounted for in the power spectrum estimation, as the signal power spectrum is nothing but an excess variance in the map at certain angular scales compared to the variance expected from the noise. The CMBA power spectrum will therefore be unbiased only if the noise properties are known precisely.

This article presents the usual techniques that allow an unbiased determination of both the CMBA maps and power spectrum. In Section 2 we will describe the data model and the data statistical properties required for the techniques presented here to be valid. Sections 3 and 4 deal with map-making and power spectrum estimation techniques, respectively.

2. Data model

The initial data are time ordered information (TOI) taken along the scanning strategy pattern of the experiment. The detector measures the temperature of the sky in a given direction through an instrumental beam. This is equivalent to saying that the underlying sky is convolved with this instrumental beam and that the instrument measures the temperature in a single direction of a N_p pixellised convolved sky noted $\mathbf{T}$. The N_t elements TOI noted $\mathbf{d}$ may therefore be modelled as:

$$\mathbf{d} = A \cdot \mathbf{T} + \mathbf{n}. \tag{1}$$

The pointing matrix A relates each time sample to the corresponding pixel in the sky. A is a $N_t \times N_p$ matrix that contains a single 1 in each line as each time sample is sensitive to only one pixel is the convolved sky.[1] The noise TOI $\mathbf{n}$ in general has a nondiagonal covariance matrix N given by:[2]

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

The most important property of the noise, that will be used widely later, is that it *has to be* Gaussian and piece-wise stationary. Both assumptions are crucial as they allow major simplifications of the map-making and power spectrum estimation problems, namely Gaussianity means that all the statistical information on the noise is contained in its covariance matrix and stationarity means that all information is also contained in its Fourier power spectrum, leading to major simplifications of the covariance matrix: the noise depends only on the time difference between two samples and N is therefore a Toeplitz matrix $N_{ij} = N_{|i-j|}$ completely defined by its first line and is very close to being circulant.[3] Such a matrix is diagonal in Fourier space. Its first line is given by the autocorrelation function of the noise, that is, the inverse Fourier transform of its Fourier power spectrum ($\star$ is the convolution operator):[4]

$$N_{i0} = \langle \mathbf{n} \star \mathbf{n} \rangle \equiv \left\langle \mathcal{F}^{-1}\left[\left| \mathcal{F}(\mathbf{n}) \right|^2 \right] \right\rangle. \tag{3}$$

3. Map-making techniques

The map-making problem is that of finding the best estimate $\widehat{\mathbf{T}}$ of $\mathbf{T}$ from Eq. (1) given $\mathbf{d}$ and A. The noise $\mathbf{n}$ is, of course, unknown. We will address the two main approaches to this problem, the first being the simplest one and the second one being the optimal one. An excellent detailed review on map-making techniques for the experts is [1].

3.1. Simplest map-making: coaddition

The simplest map-making that one can think about is to neglect the effects of the correlation of the noise. One can just average the data falling into each pixel without weighting them. This procedure is optimal (it maximises the likelihood) if the noise in each data sample is independent, that is, if the noise is white. In a matrix notation, this simple map-making can be written:

$$\widehat{\mathbf{T}} = \left[A^t \cdot A \right]^{-1} \cdot A^t \cdot \mathbf{d}, \tag{4}$$

[1] Different forms for A can however be used in case of differential measurements or more complex scanning strategies.

[2] the symbols $\langle\ \rangle$ mean that we take the ensemble average over an infinite number of realisations.

[3] Saying that the matrix is circulant is an additionnal hypothesis, but a very good approximation for large matrices.

[4] $\mathcal{F}$ denotes the Fourier transform (in practice, a FFT algorithm is used).

where the operator A^t just projects the data into the correct pixel and $[A^t \cdot A]$ counts the sample falling into each pixel. This simple map-making has the great advantage of the simplicity. It is fast ($\propto N_t$) and robust.

However, in the case of realistic correlated noise, the low frequency drifts in the timelines induce stripes in the maps along the scans of the experiment. These stripes are often much larger than the CMBA signal that is searched for and therefore should be avoided. Various destriping techniques have been proposed to avoid these stripes. A method exploiting the redundancies of the Planck mission[5] scanning strategy has been proposed by [2] and extended to polarisation by [3]. This kind of method aims at suppressing the low frequency signal by requiring that all measurements made in the same direction at different instant coincide to a same temperature signal. Another method has recently been proposed for the Archeops[6] data analysis and estimates the low frequency drifts by minimizing the cross-scan variations in the map due to the drifts [4]. The simplest method for removing the low frequency drifts before applying simple map-making is certainly to filter the timelines so that the resulting timeline has almost white noise. The filtering can consist in prewhitening the noise or directly setting to zero contaminated frequencies. The computing time (CPU) scaling of the filtering + coaddition process is modest and dominated by filtering ($\propto N_t \log N_t$). This method, however, also removes part of the signal on the sky and induces ringing around bright sources which has to be accounted for in later processes.

3.2. Optimal map-making

The most general solution to the map-making problem is obtained by maximizing the likelihood of the data given a noise model [5,6]. As the noise is Gaussian, its probability distribution is given by the N_t dimensional Gaussian:

$$P(\mathbf{n}) = \frac{1}{|(2\pi)^{N_t} N|^{1/2}} \exp\left[-\frac{1}{2}\mathbf{n}^t \cdot N^{-1} \cdot \mathbf{n}\right]. \tag{5}$$

Assuming no prior knowledge on the sky temperature, one gets from Eq. (1) the probability of the sky given the data:

$$P(\mathbf{T}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{T}) \propto \frac{1}{|(2\pi)^{N_t} N|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{d} - A \cdot \mathbf{T})^t \cdot N^{-1} \cdot (\mathbf{d} - A \cdot \mathbf{T})\right]. \tag{6}$$

Maximizing this probability with respect to the map leads to solving the linear equation:

$$A^t \cdot N^{-1} \cdot A \cdot \mathbf{T} = A^t \cdot N^{-1} \cdot \mathbf{d} \tag{7}$$

with solution:[7]

$$\widehat{\mathbf{T}} = \left(A^t \cdot N^{-1} \cdot A\right)^{-1} \cdot A^t \cdot N^{-1} \cdot \mathbf{d}. \tag{8}$$

One therefore just has to apply this linear operator to the data timeline to get the best estimator of $\mathbf{T}$; note that $\widehat{\mathbf{T}}$ is also the minimum variance estimate of the map. The covariance matrix of the map is:

$$\mathcal{N} = \left(A^t \cdot N^{-1} \cdot A\right)^{-1}. \tag{9}$$

Problems arise when trying to implement this simple procedure, the timeline data and the maps are in general very large: the typical dimensions of the problem are $N_t \simeq 6 \times 10^7$ and $N_p \simeq 10^5$ for Archeops.

The maximum likelihood solution requires both N^{-1} and $(A^t \cdot N^{-1} \cdot A)^{-1}$ which are not easy to determine. Two approaches can be used at this point: one can try to make a brute force inversion of the problem, relying on huge parallel computers, or one can try to iteratively approach the solution, hoping that convergence can be reached within a reasonable time.

3.3. Brute force inversion

The brute force optimal map-making parallel implementation is freely available as the MADCAP [7] package. It is a general software designed to produce an optimal map for any experiment by solving directly Eq. (7). The use of this package requires access to large parallel computers.

The only assumption that is made in MADCAP map-making is that the inverse time-time noise covariance matrix can be obtained directly without inversion from the noise Fourier power spectrum:

$$N_{i0}^{-1} \simeq \left\langle \mathcal{F}^{-1}\left[\frac{1}{|\mathcal{F}(\mathbf{n})|^2}\right]\right\rangle. \tag{10}$$

[5] http://astro.estec.esa.nl/Planck/

[6] http://www.archeops.org/

[7] One can remark here that simple map-making is equivalent to optimal map-making if the noise covariance matrix is diagonal, which is consistent with what was said before.

This assumption is not perfectly correct on the edges of the matrix but leads to a good estimate of the inverse time covariance matrix for the sizes we deal with. This allows this step to scale as $N_t \log N_t$ operations rather than the N_t^2 required by a Toeplitz matrix inversion. In most cases, the time correlation N_τ length is less than the whole timestream N_t so that N is band-diagonal. For Archeops, we have $N_\tau \simeq 10^4$.

The next step is to compute the inverse pixel noise covariance matrix $\mathcal{N}^{-1} = (A^t \cdot N^{-1} \cdot A)$ and the noise weighted map $A^t \cdot N^{-1} \cdot \mathbf{d}$; both operations scale as $N_t \times N_\tau$ when exploiting the structure of A and N. The last step is to invert $\mathcal{N}^{-1}$ and multiply it by $A^t \cdot N^{-1} \cdot \mathbf{d}$ to get the optimal map. Unfortunately, $\mathcal{N}^{-1}$ has no particular structure that can be exploited and this last step scales as a usual matrix inversion $\propto N_p^3$ and largely dominates the CPU required by MADCAP for the usual large datasets (e.g., Archeops).

We can remark here that MADCAP provides the map covariance matrix $\mathcal{N}$ for free, as a byproduct. This matrix is crucial for estimating the power spectrum, as will be seen in Section 4.

3.4. Iterative solutions

The other possibility is to solve Eq. (7) through an iterative process such as the Jacobi iterator, or more efficiently, a conjugate-gradient [8]. Both converge to the maximum likelihood solution.

The use of the Jacobi iterator for solving for the maximum likelihood map in CMB analysis was first proposed by [9]. The basic algorithm is the following. We have to solve the following linear system (see Eq. (7)):

$$\Gamma \cdot \mathbf{x} = \mathbf{y}. \tag{11}$$

The Jacobi iterator starts with an approximation Λ_0 of Γ^{-1} and iterates to improve the residuals R:

$$\Lambda_0 \cdot \Gamma = I - R. \tag{12}$$

In order to converge, the algorithm requires the first approximation to be good enough so that the eigenvalues of R are all smaller than 1 (a good estimate in general is $\Lambda_0 = [\operatorname{diag} \Gamma]^{-1}$). We can therefore expand:

$$\Gamma^{-1} = (I - R)^{-1} \cdot \Lambda_0 = \left(I + R + R^2 + \cdots\right) \cdot \Lambda_0. \tag{13}$$

Let us define $\Lambda_n = (I + R + R^2 + \cdots + R^n) \cdot \Lambda_0$ so that $\Gamma^{-1} = \lim_{n\to\infty} \Lambda_n$. We have the relationship $\Lambda_{j+1} = \Lambda_0 + R \cdot \Lambda_j$. If we define $\mathbf{x}_j = \Lambda_j \cdot \mathbf{y}$, it is straightforward to show that:

$$\mathbf{x}_{j+1} - \mathbf{x}_j = \Lambda_0 \cdot (\mathbf{y} - \Gamma \cdot \mathbf{x}_j) \tag{14}$$

which defines the Jacobi iterator. When going back to the usual CMB notation for maps and timelines, one gets:

$$\mathbf{T}_{j+1} - \mathbf{T}_j = \left[\operatorname{diag}\left(A^t \cdot N^{-1} \cdot A\right)\right]^{-1} \cdot A^t \cdot N^{-1} \cdot (\mathbf{d} - A \cdot \mathbf{T}_j) \tag{15}$$

which looks rather complicated but is in fact very simple to implement: the operation $A \cdot \mathbf{T}_j$ just consists in reading the map at iteration j with the scanning strategy ($\propto N_t$), and the matrix $\operatorname{diag}(A^t \cdot N^{-1} \cdot A)$ is just the white noise level variance divided by the number of hits in each pixel. It is diagonal and therefore does not require proper inversion. The only tricky part here is the multiplication $N^{-1} \cdot (\mathbf{d} - A \cdot \mathbf{T}_j)$, given the fact that N^{-1} is unknown. As the noise is stationary, N is Toeplitz and circulant[8], the multiplication by N^{-1} can be done in Fourier space directly through:

$$N^{-1} \cdot \mathbf{x} \simeq \mathcal{F}^{-1}\left[\frac{\mathcal{F}(\mathbf{x})}{|\mathcal{F}(\mathbf{x})|^2}\right] \tag{16}$$

which requires $N_t \log N_t$ operations. Finally, each iteration is largely dominated by the latter so that the final CPU time scales like $N_{it} \times N_t \log N_t$ where N_{it} is the number of iterations.

Unfortunately the convergence of such an iterator is very slow and makes it rather inefficient as it is. A significant improvement was proposed by [10] in the publicly available software MAPCUMBA. They noted that the convergence was actually very fast on small scales (compared to the pixel) but that the larger scales were converging slowly. They proposed a multigrid method where the pixel size changes at each iteration so that the global convergence is greatly accelerated (see Fig. 7 of [10]), making this iterative map-making really efficient. A conjugate gradient solver instead of the Jacobi iterator is implemented in the software Mirage [11] and accelerates again the convergence significantly. A new version of MAPCUMBA also uses a conjugate gradient solver, as well as MADmap [12].

[8] again, it is not exactly circulant but it is an excellent approximation as the matrix is large

If obtaining an optimal map is now quite an easy task using an iterative implementation (the presence of strong sources, such as the galactic signal, however, complicates this simple picture), they do not provide the map noise covariance matrix $\mathcal{N} = (A^t \cdot N^{-1} \cdot A)^{-1}$ which is of great importance when computing the CMB power spectrum in the map in order to be able to make the difference between noise fluctuations and real signal fluctuations. The only way to obtain this covariance matrix using these iterative methods is through a large Monte Carlo simulation that would reduce the advantage of iterative map-making compared to brute-force map-making.

3.5. Map-making comparisons

The precision of the MADCAP, MAPCUMBA and Mirage implementations are shown in Fig. 1 with the same CMB and noise simulation based on Archeops realistic conditions. The three resulting maps were kindly provided by [13]. The six maps on the left are respectively from top left to bottom right: initial CMB fluctuation, coaddition of the timeline without filtering, coaddition of the timeline with white noise only (i.e., the true optimal map that has to be reconstructed), MADCAP residual map (difference between MADCAP reconstructed map and the white noise map), MAPCUMBA residual map and Mirage residual map. All maps are shown with the same color scale. The first remark that can be made is that the stripes are indeed a real problem and that straight coaddition is not to be performed. The three different optimal map-making codes give very similar results, especially MADCAP and MAPCUMBA. In all cases, as can be also seen in the right panel of Fig. 1, the residuals are much smaller than the CMB fluctuations that are searched for. The three map-making implementations can therefore be considered are unbiased.[9]

Finally one can summarize the comparison as following: iterative and brute-force optimal map-making give very similar results as far the optimal map is concerned. The brute force inversion provides the map noise covariance matrix for free, which is a major point as will be seen in next section. The computer requirements are, however, much larger than for iterative map-making. The latter should therefore be used when the power spectrum estimation can be carried out without the knowledge of the map noise covariance matrix, in general using a Monte Carlo technique (see next section). In this case, one should seriously consider the filtering + coaddition map-making that is by far the fastest, but removes part of the signal. This is, however, accounted for (see Section 4.2) also using a Monte Carlo technique.

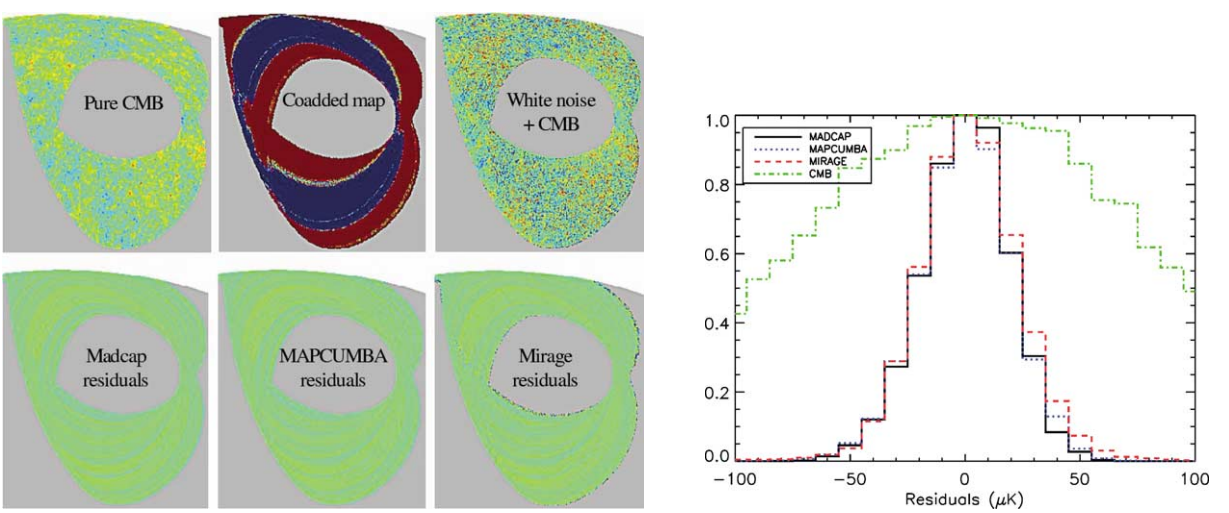

Fig. 1. The six maps on the right show a comparison of results from different map-making implementations on the same simulation (typical of Archeops data). All maps are in Healpix pixellisation [14] and have the same color scale ranging from −500 μK to 500 μK from dark blue to dark red, green corresponding to zero. The histograms of the residuals is shown on the right and are more than three times smaller than the actual CMB fluctuation.

[9] Let us note that the noise model that was used for MADCAP is the true one, not an estimation. This makes, however, little difference.

4. Power spectrum estimation techniques

We know want to compute the power spectrum C_ℓ of the map $\mathbf{T}$ whose noise covariance matrix $\mathcal{N}$ might be known or not, depending on the method that was used before to produce the map. The map is composed of noise and signal (from now on $\mathbf{n}$ is the noise on the map pixels):

$$\mathbf{T} = \mathbf{s} + \mathbf{n}. \tag{17}$$

The signal in pixel p can be expanded on the $Y_{\ell m}(\theta_p, \phi_p)$ spherical harmonics basis:

$$s_p = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m} B_\ell Y_{\ell m}(\theta_p, \phi_p), \tag{18}$$

where B_ℓ stands for the beam[10]. If the CMBA are Gaussian, the variance of the $a_{\ell m}$, called the angular power spectrum and denoted C_ℓ contains all the cosmological information:

$$\langle a_{\ell m} a^\star_{\ell' m'} \rangle = C_\ell \delta_{\ell \ell'} \delta_{m m'}. \tag{19}$$

The map covariance matrix (assuming no correlation between signal and noise) is:

$$M = \langle \mathbf{T} \cdot \mathbf{T}^t \rangle = \langle \mathbf{s} \cdot \mathbf{s}^t \rangle + \langle \mathbf{n} \cdot \mathbf{n}^t \rangle \tag{20}$$

$$= S + N \tag{21}$$

and the signal part is related to the C_ℓ:

$$S_{pp'} = \langle s_p s_{p'} \rangle = \sum_\ell \frac{2\ell+1}{4\pi} C_\ell B_\ell^2 P_\ell(\chi_{pp'}), \tag{22}$$

where $\chi_{pp'} = \cos(\mathbf{u}_p \cdot \mathbf{u}_{p'})$, $\mathbf{u}_p$ being the unit vector towards pixel p and P_ℓ are the Legendre polynomials.

One therefore has a direct relation between the map and noise covariance matrices and the angular power spectrum:

$$M = N + \sum_\ell \frac{2\ell+1}{4\pi} C_\ell B_\ell^2 P_\ell(\chi_{pp'}). \tag{23}$$

The power spectrum estimation consists in estimating C_ℓ from $\mathbf{T}$ and N (that can be unknown) using this relation.

4.1. Maximum likelihood solution

Full details concerning this can be found in [15,16]. As for the map-making problem, the maximum likelihood solution proceeds by writing the probability for the map, given its covariance matrix assuming Gaussian statistics:[11]

$$P(C_\ell|\mathbf{T}) \propto P(\mathbf{T}|C_\ell) = (2\pi)^{-N_p/2} \exp\left[-\frac{1}{2}\left[\left(\mathbf{T}^t \cdot M^{-1} \cdot \mathbf{T} \right) + \mathrm{Tr}(\ln M) \right] \right] \tag{24}$$

and we therefore want to maximize the likelihood function through $\partial L / \partial C_\ell = 0$:

$$L(C_\ell) = -\frac{1}{2}\left[\left(\mathbf{T}^t \cdot M^{-1} \cdot \mathbf{T} \right) + \mathrm{Tr}(\ln M) \right]. \tag{25}$$

Tedious calculations lead to the solution:

$$C_\ell = \sum_{\ell'} F^{-1}_{\ell \ell'} \times \mathrm{Tr}\left[\left(\mathbf{T} \cdot \mathbf{T}^t - N \right) \cdot M^{-1} \cdot \frac{\partial S}{\partial C_\ell} \cdot M^{-1} \right], \tag{26}$$

where F is the Fisher matrix:

$$F_{\ell \ell'} = \mathrm{Tr}\left[\frac{\partial S}{\partial C_\ell} \cdot M^{-1} \cdot \frac{\partial S}{\partial C_{\ell'}} \cdot M^{-1} \right]. \tag{27}$$

[10] It is the Legendre transform of the instrumental beam under the assumption that it is symmetric.

[11] The trace appears from $|M|^{-1} = \exp[-\mathrm{Tr}(\ln M)]$ as the trace is invariant.

Eq. (26) let C_ℓ appear in both sides (in M) in an uncomfortable way and therefore cannot be solved simply. The method usually used [15,7] is the Newton–Raphson iterative scheme: One starts from an initial guess for the binned power spectrum[12] $\mathbf{C}^{(0)}$ and iterates until convergence following:

$$\mathbf{C}^{(i+1)} = \mathbf{C}^{(i)} + \delta\mathbf{C} \tag{28}$$

with:

$$\delta\mathbf{C} = -\left[\frac{\partial^2 L}{\partial \mathbf{C}^2}\bigg|_{\mathbf{C}=\mathbf{C}_i}\right]^{-1} \cdot \frac{\partial L}{\partial \mathbf{C}}\bigg|_{\mathbf{C}=\mathbf{C}_i} \tag{29}$$

the likelihood L being that of Eq. (25). Convergence is usually reached after a few iterations. The explicit form of the derivatives of Eq. (29) is:

$$\frac{\partial L}{\partial C_b} = \frac{1}{2}\left(\mathbf{m}^{\mathrm{T}} \cdot M^{-1} \cdot \frac{\partial S}{\partial C_b} \cdot M^{-1} \cdot \mathbf{m} - \mathrm{Tr}\left[M^{-1} \cdot \frac{\partial S}{\partial C_b}\right]\right), \tag{30}$$

$$\frac{\partial^2 L}{\partial C_b \partial C_{b'}} = -\mathbf{m}^{\mathrm{T}} \cdot M^{-1} \cdot \frac{\partial S}{\partial C_b} \cdot M^{-1} \cdot \frac{\partial S}{\partial C_{b'}} \cdot M^{-1} \cdot \mathbf{m} + \frac{1}{2}\,\mathrm{Tr}\left[M^{-1} \cdot \frac{\partial S}{\partial C_b} \cdot M^{-1} \cdot \frac{\partial S}{\partial C_{b'}}\right], \tag{31}$$

where the index b denotes the bin number.

Each iteration will then require a large number of large matrix operations forcing such an algorithm to be implemented on large memory parallel supercomputers. MADCAP [7] is the common implementation of this algorithm and scales as $2(N_b + \frac{2}{3})N_p^3$ operations per iteration. The CPU/RAM/Disk problem is therefore even cruder for the power spectrum than for the map-making. This algorithm leads to the optimal solution accounting correctly for the noise covariance matrix and additionally provides the likelihood shape for each bin through the various iterations, allowing a direct estimate of the error bars.

4.2. Frequentist approaches

An alternative approach to power spectrum estimation is to compute the so called pseudo power spectrum (harmonic transform of the map, noted $\widetilde{C}_\ell$) and to correct it so that it becomes a real power spectrum. This approach has been proposed and developed in [17,18]. The harmonic transform of the map differs from the true C_ℓ in various ways (we follow the notations from [17]): the observed sky is convolved by the beam and by the transfer function of the experiment so that the observed power spectrum is $B_\ell^2 F_\ell C_\ell$, where B_ℓ characterizes the beam shape in harmonic space and F_ℓ the filtering performed on the data by the analysis process (that may also include electronic filtering by the instrument itself). The observed sky is in general incomplete (at least because of a Galactic cut) leading to the fact that the C_ℓ measured are not independent as they are convolved in harmonic space by the window-function [19]. We therefore have access to $\sum_{\ell'} M_{\ell\ell'} B_{\ell'}^2 F_{\ell'} C_{\ell'}$ where $M_{\ell\ell'}$ is the mode mixing matrix. Finally, the noise in the timelines projects on the sky and adds its contribution $\widetilde{N}_\ell$ to the sky angular power spectrum. At the end, the map angular power spectrum, the *pseudo-*C_ℓ is related to the true C_ℓ via:

$$\widetilde{C}_\ell = \sum_{\ell'} M_{\ell\ell'} B_{\ell'}^2 F_{\ell'} C_{\ell'} + \widetilde{N}_\ell. \tag{32}$$

The frequentist methods propose to invert Eq. (32), making an extensive use of Monte Carlo simulations (details can be found in [17]):

- The pseudo power spectrum of the map $\widetilde{C}_\ell$ is computed by transforming the map into spherical harmonics (generally using Healpix pixellisation and the `anafast` procedure available in the Healpix package [14]).
- The mode mixing matrix is computed analytically through:

$$M_{\ell_1\ell_2} = \frac{2\ell_2 + 1}{4\pi} \sum_{\ell_3} (2\ell_3 + 1)\mathcal{W}_{\ell_3} \begin{pmatrix} \ell_1 & \ell_2 & \ell_3 \\ 0 & 0 & 0 \end{pmatrix}, \tag{33}$$

  where $\mathcal{W}_\ell$ is the power spectrum of the window of the experiment (in the simplest case 1 for the observed pixels and 0 elsewhere, but more complex weighting schemes may be used, as in Archeops [20] or WMAP [21]). In the SpICE approach [18], the $M_{\ell\ell'}$ inversion in harmonic space is replaced by a division in angular space which is mathematically equivalent.

[12] Binned power spectrum means that we do not consider one single mode ℓ but a bin in ℓ as we do not have access in general to all modes due to incomplete sky coverage.

- The beam transfer function is computed from a Gaussian approximation or the Legendre transform of the beam maps or a more complex modelling if the beams are asymetric, such as in [22].
- The filtering transfer function is computed using a signal only Monte Carlo simulation (it should include the pre-processing applied to the time streams). Fake CMB skies are passed through the instrumental and analysis process producing maps and pseudo power spectra. The transfer function is basically computed as the ratio of the input model to the recovered ensemble average. An important point at this step is to check that the transfer function is independent of the model assumed for the simulation. Let us also remark that using a transfer function that depends only on ℓ is a bit daring as the filtering is done in the scan direction, which, in general, corresponds to a particular direction in the sky. This approximation, however, seems to work well and has been successfully applied to Boomerang [23] and Archeops [20].
- The noise power spectrum is computed from noise only simulations passing again through the instrumental and analysis process to produce noise only maps and pseudo power spectra. The noise power spectrum is estimated from the ensemble average of the various realisations.
- Error bars are computed in a frequentist way by producing signal + noise simulations and analysing them as the real data. This allows us to reconstruct the full likelihood shape for each power spectrum bin and the bin-bin covariance matrix.

Such an approach based on simulations has the advantage of being fast: each realisation basically scales as $\propto N_t \log N_t$ for the noise simulation and map-making (if filtering + coaddition is used) and $\propto N_p^{3/2}$ for the CMB sky simulation and pseudo power spectrum computation. An important advantage of such a method is the possibility to include in the simulation systematic effects (beam, pointing, atmosphere, ...) that would not be easily accountable for in a maximum likelihood approach.

4.3. Cross-power spectra

When several photometric channels are available from the experimental setup, it is possible to compute cross-power spectra between the channel rather than power spectra of an individual channel or of the average of all channels. This has the advantage of suppressing the noise power spectrum (but not its variance, of course) that is not correlated between channels and leaving the sky signal unchanged. The cross-power spectrum of channels i and j is defined as:

$$C_\ell^{i,j} = \frac{1}{2\ell+1} \sum_{m=-\ell}^{\ell} a_{\ell m}^i a_{\ell m}^{j\star}. \tag{34}$$

The cross-power spectrum method can easily be associated with the frequentist approach, simplifying significantly its implementation, since the most difficult part, the noise estimation, is now less crucial as noise disappears and cannot bias the power spectrum estimation. This has been successfully applied in the WMAP analysis [21].

4.4. Which power spectrum estimator should be used?

The maximum likelihood approach is undoubtedly the best method to use if possible, but its CPU/RAM/Disk requirements are such that in practice, with modern experiments, it is very difficult to implement. It should however be considered to check the results on data subsets small enough to make it possible. The frequentist approaches are much faster and provide comparable precision in terms of error bars, and permit the accounting for systematic effects in a simple manner. The tricky part is, however, to estimate the noise statistical properties precisely enough. The same difficulty exists, however, in the maximum likelihood approach where the noise covariance matrix has to be known precisely. It is generally directly computed in the map-making process from the time correlation function, thus displacing the difficulty elsewhere. In any case, the noise model has to be unbiased as the final power spectrum is essentially the subtraction between the pseudo power spectrum of the map and the noise power spectrum. Estimating the noise properties is a complex problem mainly due to signal contamination and pixellisation effects. A general method for estimating the noise in a CMB experiment is proposed in [24] and was successfully applied for the Archeops analysis [20]. When multiple channels are available, the frequentist approach applied on cross-power spectra is certainly the simplest and most powerful power spectrum estimation technique available today as it reduces the importance of the difficult noise estimation process. We can also mention the hierarchical decomposition [25] that achieves an exact power spectrum estimation to submaps at various resolutions, and then optimally combines them.

5. Conclusions

We have shown techniques designed to make maps from CMB data and to extract power spectra from them. In both cases, the brute force, maximum likelihood approach is the most correct, but generally hard to implement in practice. Alternative

approaches, iterative or relying on Monte Carlo simulations, provide similar precision with smaller computer requirements. In all cases, much work has to be done before: firstly by designing the instrument correctly, and afterwards by cleaning the data, flagging bad samples and ending with a dataset that match the minimum requirement of all the methods described in this review: stationarity and Gaussianity.

Acknowledgements

The Author wants to thank the Archeops collaboration for its stimulating atmosphere, A. Amblard and P. Filliatre for reading carefully the manuscript and providing useful inputs.

References

[1] R. Stompor, et al., Phys. Rev. D 65 (2002) 022003.
[2] J. Delabrouille, Astron. Astrophys. Sup. 127 (1998) 555–567.
[3] B. Revenu, et al., Astron. Astrophys. Sup. 142 (2000) 499–509.
[4] A. Bourrachot et al., in preparation.
[5] E.L. Wright, astro-ph/9612006.
[6] M. Tegmark, Phys. Rev. D 56 (1997) 4514.
[7] J. Borrill, in: Proc. of the 5th European SGI/Cray MPP Workshop, 1999, astro-ph/9911389.
[8] W.H. Press, et al., Numerical Recipes in C, Cambridge University Press, 1988.
[9] S. Prunet, et al., astro-ph/0006052.
[10] O. Doré, et al., Astron. Astrophys. 374 (2001) 358D.
[11] D. Yvon et al., in preparation.
[12] C. Cantalupo, http://www.nersc.gov/~cmc/MADmap/doc/index.html.
[13] P. Filliatre, PHD thesis, Université Joseph Fourier Grenoble, France, 2002.
[14] K.M. Gorski, et al., astro-ph/9812350, http://www.tac.dk/~healpix/.
[15] J.R. Bond, A.H. Jaffe, L. Knox, Phys. Rev. D 57 (1998) 2117.
[16] M. Tegmark, Phys. Rev. D 55 (1997) 5895.
[17] E. Hivon, et al., astro-ph/0105302.
[18] I. Szapudi, et al., astro-ph/0107383.
[19] M. White, M. Srednicki, Astrophys. J. 443 (1995) 6.
[20] A. Benoît, et al., Astron. Astrophys. 399 (2003) 19L.
[21] G. Hinshaw, et al., astro-ph/0302217, Astrophys. J., submitted for publication.
[22] M. Tristram, et al., in preparation.
[23] C.B. Netterfield, et al., Astrophys. J. 571 (2002) 604.
[24] A. Amblard, J.-Ch. Hamilton, Astron. Astrophys., submitted for publication.
[25] O. Doré, L. Knox, A. Peel, Phys. Rev. D 64 (2001) 083001.