Looking for components in the Universe's oldest data set

Jean-François Cardoso

Centre National de la Recherche Scientifique
École Nationale Supérieure des Télécommunications
http://tsi.enst.fr/~cardoso/stuff.html

Abstract

Modern astronomical experiments offer many exciting challenges in the field of statistical signal processing. This paper gives an introduction to the problem of imaging the cosmic microwave background (CMB). The CMB is a "fossil" electromagnetic radiation ubiquitous in the universe since early times; its measurement and analysis currently are very active research topics in cosmology.

The first part of this paper is a basic introduction to the physics behind the CMB and to its place in Big Bang theory: we describe the origin of the CMB and outline its relevance to cosmological models and the current experimental efforts.

Accurately measuring and analyzing the CMB is a daunting task. In particular, CMB measurements are contaminated by other components from various origins. These components —some of them also of cosmological interest— can be separated by combining the sky maps made at several frequencies. This is the *component separation* problem which is a classic problem of statistical signal processing. The second part of this paper reviews specific aspects of the component separation problem in the context of CMB analysis.

1 Introduction

Astronomy has always offered challenging signal processing problems. This is even more true today with new astronomical experiments, able to collect enormous data sets, and with the implementation on powerful computers of sophisticated algorithms based on advanced statistical methods.

This paper briefly describes the problem of extracting 'components' from multi-spectral measurements. This is motivated by the exploitation of data to be collected by the Planck space mission which will offer measurements of the cosmic microwave background in 10 spectral channels with unprecedented sensitivity and spatial resolution.

The first part of this paper contains a brief description of the cosmological and experimental context. The second gives an overview of the signal processing aspects. Although my background is in physics, I am not by any means an expert in cosmology: the paper should be taken as an invitation to get to the widely available introductory literature on the fascinating topic of cosmology.

2 Some cosmology

The aim of cosmology is to provide a coherent model for the formation of the Universe. The 'Big Bang' currently is the widely accepted conceptual framework for modeling the Universe's history: there are many debates going on in the cosmology community, but the most controversial parts deal with the first *instants* of the Universe. The part of the story which concerns us is the formation of the cosmic microwave background (CMB) which reflects events which occurred when the Universe was about 300.000 years old. This is an early time compared to the 12 billions years which have elapsed since then but this is still pretty late with respect to the most tumultuous events of the early Universe. Actually, CMB formation is governed by non controversial physics, involving familiar particles (electrons, protons, photons...) at familiar energy levels (3000K).

This section gives a very brief overview of some CMB issues: its origin, significance, measurement and statistical features.

2.1 The Big Bang

The 'Big Bang' scenario is that of an expanding universe which starts very dense and very hot and cools down as it expands. As temperature decreases over time, increasingly stable particles become the dominant species of the universe. In the standard scenario, after about a few hundred thousand years, the temperature has reached an all-time low of about 3000K and the universe appears essentially as a gas of familiar particles: protons, neutrons —mostly bound in helium nuclei— electrons and photons (as well as weakly interacting neutrinos).

As long as the temperature remains high enough, electrically neutral atoms cannot form because they would be shattered by thermal agitation. Thus, the photons, carriers of the electromagnetic interaction, interact easily with the electrically charged electrons and nuclei. The situation is that of a thermal equilibrium between all these strongly interacting particles. As a result, the Universe is opaque: light cannot travel without quickly interacting with charged particles.

However, an important milestone is reached when the temperature further decreases due to the continued expansion: a low temperature corresponds to photons with low energy (on the average). When photon energy becomes smaller than the binding energy between electrons and nuclei (this happens at a temperature of about 3000K), photons are no longer energetic enough to break apart the atoms formed by the combination of electrons and nuclei. Thus, below this temperature, the stable form of matter is that of a gas of electrically neutral atoms rather than an ionized gas.

This 'atomic recombination'[1] has a dramatic impact on the subsequent course of events. Because atoms are electrically neutral entities, the interaction between matter (the atoms) and radiation (the photons) becomes very weak after atomic recombination: the photons become free to travel through the cosmos; the universe has become transparent to photons. From then on, the photon gas and the matter are essentially uncoupled and start evolving independently.

[1] I do not know why this is traditionally called 'recombination' since the particles were not combined into atoms before. The term 'atomic combination' would seem more appropriate.

Matter, on one hand, in the form of clouds of (mostly) hydrogen and helium, will continue expanding after the atomic recombination, starting from a highly homogeneous state. However, small deviations from the average matter density are the seeds for large scale structures in the universe: local over-densities are gravitationally instable, tending to collapse on themselves (in competition with the global cosmic expansion), forming the precursors of galaxies or galaxy clusters.

Radiation, on the other hand, has a much more peaceful future after atomic recombination: in the absence of any significant possibility of interaction, it will essentially 'stay there'. Today, most photons in the Universe are CMB photons, *i.e.* photons which were liberated by the atomic recombination (the photons which were later produced by stars actually are a weak minority with respect to CMB photons). These CMB photons have had virtually no interaction since the time of atomic recombination. For this reason, they are a 'fossil light' in a very strong sense; they form the 'cosmic microwave background' (CMB) that we still observe today.

Fossil radiation, however, has undergone one major modification since atomic recombination: due to cosmic expansion, the wavelengths of all photons have stretched by the same factor as the Universe itself. This factor is close to 1000 since atomic recombination. As the energy of a photon is proportional to its frequency, the average CMB photon now appears much colder, by a factor of about 1000. This is the reason why the fossil radiation is to be found today in the microwave domain. Its discovery in 1965 by Penzias and Wilson (1965) won them a Nobel prize. Its existence is considered as one of the most highly supportive facts in favor of the standard cosmological model.

Very accurate statements can be made about the spectral distribution of the fossil radiation. This is because the energy distribution of light in thermal equilibrium with matter at a given temperature T —the so called 'black body' spectrum— depends only on temperature, as given by the celebrated Planck's formula. This spectrum has a single bump centered around a frequency ν such that $h\nu = kT$ where h is Planck's constant and k is the Boltzmann's constant. Since the time of atomic recombination, the fossil light no longer has anything to be in equilibrium with, but it turns out that stretch-

ing the wavelengths of all photons from a Planck distribution preserves the Planckian shape of the spectrum. This is the reason why the energy of the CMB, as observed today, still is distributed as a black body radiation. Planck's formula predicts with a very high accuracy (fig. 1) the ob-

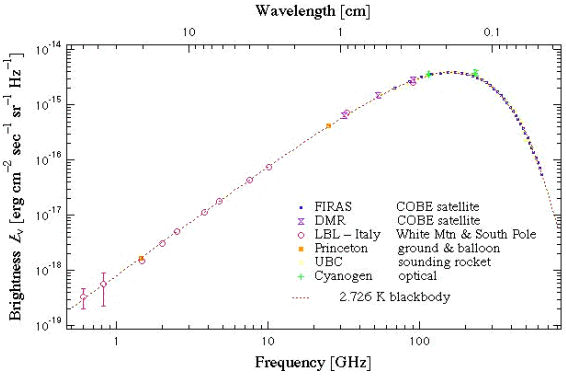

Figure 1: Measurements of the CMB spectrum by various experiments (space-based, airborne and ground based) and their fit as a black body spectrum at temperature T_o=2.726K

served CMB spectrum as corresponding to photons in thermal equilibrium with matter at temperature $T_o = 2.726K$.

The fossil microwave background is highly isotropic: whichever direction in the sky is turned a microwave antenna, it picks up a background microwave radiation at temperature very close to T_o.

2.2 CMB anisotropies

More recent technology, however, allows for better measurements of the 'sky temperature': it is now possible to measure the apparent temperature of the CMB coming from a particular direction in the sky with increased accuracy and improved angular resolution. The Cosmic Background Explorer (NASA's COBE satellite), launched in 1989 has been the first instrument able to detect some intrinsic CMB 'anisotropy', at a level of a part in 100,000. Figure 4 shows the spatial distribution over the sky of these temperature fluctuations. Again, the amplitude of these fluctuations is very small, of the order of a few micro-Kelvins (we note in the passing that producing a sky map like 2 actually involves a huge amount of data processing and model fitting which is not described here).

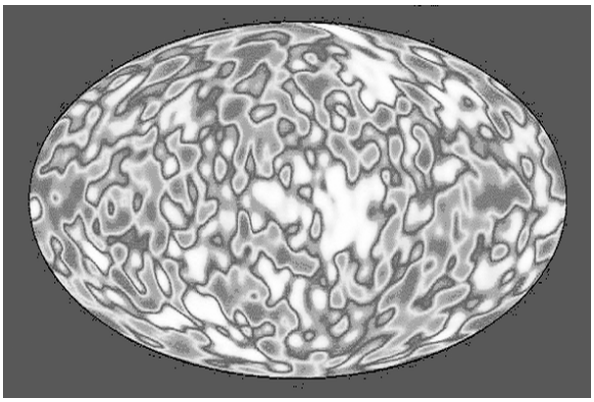

Figure 2: COBE (1989) measurements of the temperature fluctuations over the sky. The amplitude of the fluctuations is of a few micro Kelvins with respect to an average temperature of about 2.7 K.

These tiny variations in the intensity of the CMB over the sky reveal how matter and energy were spatially distributed at the time of atomic recombination. These measurements came after several unsuccessful (because of insufficient sensitivity) attempts and much at the relief of Big Bang theory supporters. If these initial inhomogeneities had not been found (or if they had been found to be of an even smaller amplitude), it would have been difficult to explain how the large scale structures of the universe (galaxies, galaxy clusters) could have formed: CMB anisotropies are understood as the signatures of inhomogeneities in the distribution of matter itself, inhomogeneities which have later evolved and grown (due to gravitational instability) in competition with the Universe's expansion.

Cosmologists are interested in the statistical distribution of the CMB inhomogeneities because the characteristics of their spatial spectrum can be directly related to some important physical quantities.

Harmonic spectrum. Let us be a bit more specific about the data. Measurements are done in several electromagnetic bands $k = 1, \ldots, K$. Processing the k-th band (and skipping lot of important details regarding 'map making') yields a map $\Delta T_k(\theta, \phi)$ of the deviation, in direction (θ, ϕ), of the apparent temperature of the CMB radiation from its mean value T_o. To a good approximation, the temperature fluctuation $\Delta T_k(\theta, \phi)$ seen in the kth band has the same spatial pattern in the other

bands. In other words, one expects the factorization

$$\Delta T_k(\theta,\phi) = A_k \; S(\theta,\phi) \qquad (1)$$

where A_k accounts for the (differential) strength of the CMB emissivity in a given band and $S(\theta,\phi)$ is a spatial pattern common to all bands.

The spatial pattern can be decomposed into spherical harmonics as

$$S(\theta,\phi) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\theta,\phi), \qquad (2)$$

where the doubly indexed set $Y_{lm}(\theta,\phi)$ of spherical harmonic functions is to the sphere what sines and cosines of the discrete Fourier transform are to a one-dimensional interval. A spherical harmonic Y_{lm} accounts for spatial patterns with an angular resolution of $2\pi/l$.

For a given multi-pole l, there are $2l+1$ spherical harmonics whose coefficients a_{lm} are uncorrelated with variance independent of m when $S(\theta,\phi)$ is the realization of an isotropic process on the sphere. The so-called C_l spectrum then is $C_l = E\{|a_{lm}|^2\}$ and its empirical estimate is

$$\hat{C}_l = \frac{1}{2l+1} \sum_{m} |a_{lm}|^2. \qquad (3)$$

Let us stress at this point that the CMB sky stands still (on the human time scale): there is only one Universe and one realization of $S(\theta,\phi)$ available to us. If the CMB pattern is modeled as the realization of a Gaussian isotropic process on the sphere, the empirical spectrum $\{\hat{C}_l\}_{l\geq 0}$ is an exhaustive statistic.

An example of the predicted CMB spectrum is given at figure 3 which is an (appropriately rescaled) plot of the C_l for two values of a density parameter in the standard model.

The peaks in the harmonic spectrum are signatures of 'acoustic oscillations', so called because they result from the competition between inertia and elasticity of the medium. This is the reason why the shapes and locations of the peaks directly carry physical information. Cosmologists have come up with various competing models for the harmonic spectrum, all depending on key cosmological parameters like matter density and energy density at recombination time.

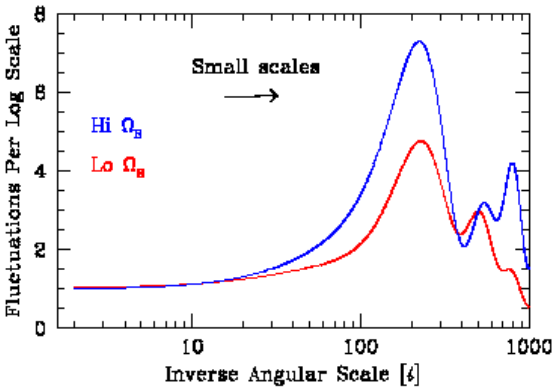

Figure 3: The predicted shape of the harmonic spectrum C_l for two values of the density parameter.

2.3 Really the oldest data set?

Every paper needs a catchy title; claiming the 'oldest data set in the universe' certainly sounds catchy to me. But are the CMB anisotropy sky maps really the oldest images that can be formed today? Recall that before the atomic recombination the universe was opaque to light and that since atomic recombination, CMB photons have traveled through the Universe, virtually unaffected (except by the expansion of the universe itself). Thus, CMB sky maps certainly are the oldest images that can be formed today from electromagnetic waves. What other signal carriers could be available that would have originated before atomic recombination? One can think of gravitational waves and of neutrinos. Current technology, however, does not offer the possibility to form images from these signals (gravitational waves still are to be detected!). Thus it seems safe to claim, that for some time at least, CMB sky maps are the oldest images one can dream of contemplating.

2.4 The Planck mission

We have already mentioned the first detection (and imaging) of CMB anisotropies by the COBE satellite in 1989. Since then, several experiments have improved on these ground breaking results, either ground based, space based or balloon borne. In 2007, ESA (the European Spatial Agency) plans to launch the ultimate CMB experiment with un-

precedented sensitivity and angular resolution. An intuitive view of the gain in resolution is offered by figure 4. It illustrates (in simulated sky maps) how

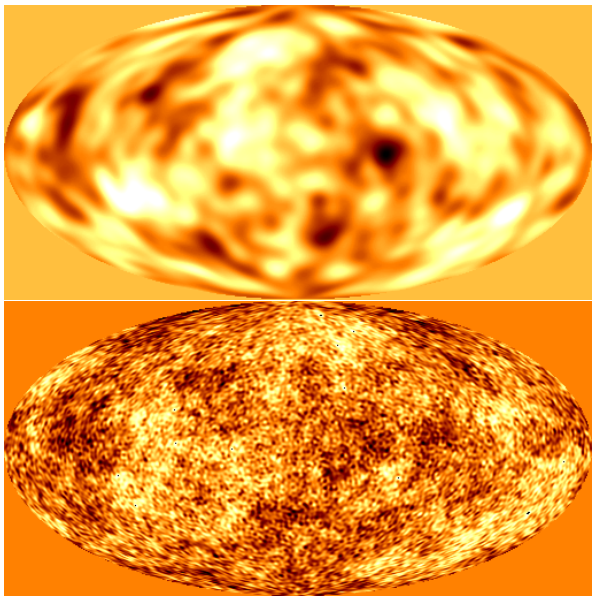

Figure 4: Simulation of CMB anisotropies at COBE's resolution (top) and at Planck's resolution (bottom)

the COBE resolution compares to the expected resolution of Planck.

High angular resolution is of significance to CMB studies for the following reason. Low resolution instruments (like those carried by COBE) give details in the sky corresponding to regions so large that they cannot have been causally connected in the past. Therefore they cannot inform us about the physics of the *interactions* at the time of recombination. A striking example is the first acoustic peak seen in figure 3. It is centered around $l = 200$, way above COBE's resolution. This first peak has been unambiguously detected by more recent experiments but the Planck mission is expected to yield even better estimates of the harmonic spectrum, confirming (or not) the presence of the following peaks and quantifying characteristics with increased accuracy. It will place fundamental constraints on models of the birth and evolution of the Universe.

Figure 5 shows the satellite with its tilted primary mirror and a large shield at its basis. The

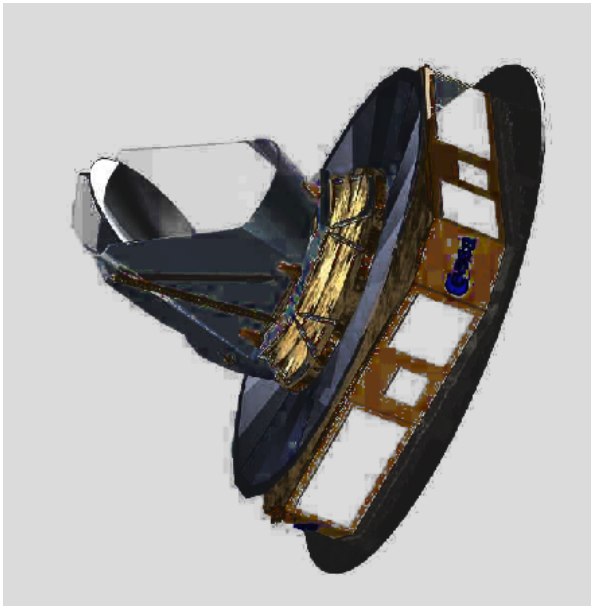

Figure 5: The Planck satellite, standing up

shield is designed to protect the instruments from the Sun. This is crucial because the high frequency instruments will use *bolometers* to measure the intensity of the microwave radiation in the direction pointed to by the primary mirror. Each bolometer essentially is a piece of silicon: when exposed to microwave radiation, it should heat up, resulting in a slight change of conductivity which can then be electrically measured. Detecting variations in intensity corresponding to micro-Kelvins from an average radiation temperature of $2.7K$ requires that the silicon itself should be kept at an even lower temperature. Actually, Planck also is an exercise in high-tech cryogenics: it uses liquid helium to reach and then maintain an operating temperature close to $0.1K$ for the bolometers. The satellite will only live as long as its helium supply. This also is a major reason to place the satellite in orbit at the 'L2' Lagrange point: this is a very special point where on the Sun-Earth line where a satellite can be stabilized and be always shielded from the Sun by the Earth.

Another key aspect of the mission is its ability to make *multi-spectral images*: Planck carries two instruments: the LFI (low frequency instrument) has four channels between 30 and 100 GHz; the HFI (high frequency instrument) has six channels

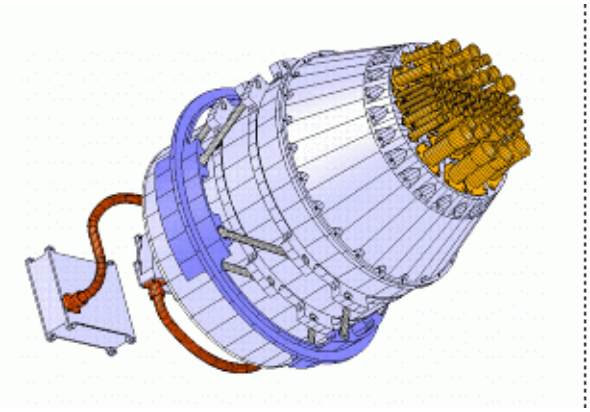

Figure 6: The Planck satellite: the six arrays of bolometers at the heart of the instrument

centered around 100, 143, 217, 353, 545 and 857 GHz (only the HFI uses bolometers, the LFI using radiometers). Figure 6 shows the six arrays of bolometers, at the focal center of the secondary mirror. More precisely it shows the filters and the horns, conducting and filtering the EM waves to the bolometers themselves.

Since a total of 10 spectral bands will be available for imaging, an important issue is the combination of the corresponding sky maps in a coherent fashion in order to get the best out of the instruments. In particular, the diversity offered by all the bands can be used to extract from the data several 'components' of physically distinct origins, including the microwave background but also other potentially interesting 'foregrounds'. This is the topic of the next section.

3 Looking for components

The previous section outlined a simple model (1) of the CMB data. Among the many issues which have been ignored is the problem of contamination: the CMB radiation is not the only source of temperature anisotropy which is picked up by the detectors. Actually several physical phenomena have already been identified as contributing to the microwave signals picked up by detectors. These are briefly described below at section 3.1 together with some indications of their statistical properties. Some of these components, besides the CMB, are of interest to cosmology. Thanks to the fact that sky maps can be acquired in several frequency bands, there is a possibility to *separate* these components using techniques from the array processing tool-box. We review some of the basics in section 3.2 and then briefly discuss the use of *blind source separation* techniques at section 3.3.

3.1 Components

When forming sky maps in the frequency range of interest (millimeter and sub-millimeter wavelength), there are several 'foregrounds' which are superimposed to the cosmic microwave background. In this paper, we only list the two dominant contaminations.

At lower frequencies, an important contaminating foreground is due to the thermal Sunyaev-Zel'dovich (SZ) emission. It results from the scattering of CMB photons on hot electrons gas in the Universe. This means that the hot electron gas 'warms up' the CMB photons. Such hot electron clouds are mostly located within galaxy clusters and these clusters are barely resolved by the instruments (about 1 arc-minute). Therefore, the spatial pattern associated with the SZ emission is close to that of a point process. Successfully separating the SZ emission from the data yields one of the important by-products expected from CMB imaging: it would allow for the collection of data about the statistical distribution of the galaxy clusters.

At higher electromagnetic frequencies, another important contaminating foreground is 'dust emission'. This is due to the presence of tiny dust grains organized in clouds *within our own galaxy*. Heated at temperatures around 17K, they do contribute significantly in the band of interest. Their spatial pattern is that of... clouds.

Other contributions may be less important and are not described here. See fig 7 for typical realizations of the components listed so far: CMB, galactic dust and SZ emission. For most of these components, it is expected that, to some good approximation, the spatial pattern of each component remains constant over all detectors (frequency channels) and that only its *amplitude* (emissivity) will vary. In other words, we expect to form, for each frequency channel $k = 1, K$, a sky map of the temperature anisotropy $\Delta T_k(\theta, \phi)$ which can be mod-

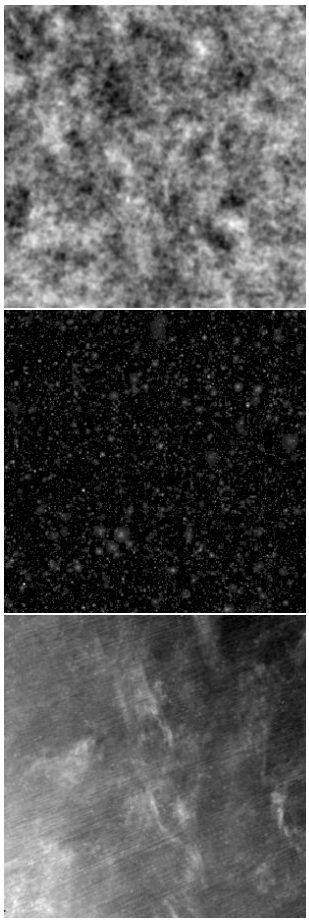

Figure 7: Typical samples of three major components: CMB, extra-galactic SZ effect, intra-galactic dust

eled to a reasonable approximation as:

$$\Delta T_k(\theta,\phi) = \sum_{j=1}^{J} A_{kj}\ S_j(\theta,\phi) + V_k(\theta,\phi) \qquad (4)$$

Model (4) is just the linear superposition of factors like (1) with some noise added. Indeed another important problem besides contamination of the CMB by various foregrounds is observation noise, present at a significant level because of the minuteness of the effect to be detected. It is not very informative to give an overall value of the expected signal to noise ratio because it depends very much of the spatial frequency (or harmonic) since most of the components are dominated by the low-frequency part of their spectrum. See figure 8 and the accompanying text below.

We should add that eq. (4) ignores diffraction effects, or more precisely it ignores differential diffraction effects. Indeed, the point spread function of the instrument varies very significantly across detectors because the resolution of the detector is mostly determined by diffraction. From 100 to 850 GHz, the wavelength does change a lot (this is why high-frequency detectors are packed around the focal point of the secondary mirror, giving more room to the low-frequency, less resolving detectors at the periphery.) In this paper, however, we do not discuss how to compensate for the beam and, for simplicity, we shall ignore this effect.

Sampling. A complication of processing sky maps is that one must deal with data points sampled on the sphere. A specific sampling scheme [7], adapted to stochastic processes living on the sphere and to the computation of the harmonic coefficients (2) has to be devised. A scheme, known as 'HEALPix' (Hierarchical Equal Area isoLatitude Pixelisation of the sphere) will be used for the processing of Planck data. It has the specific property that all pixels are located on lines of constant latitude, allowing for a 'semi-fast' computation of the harmonic coefficients.

In the following, for the sake of exposition and to keep in line with our early stage of experimentation, we only present simulations over a small region of the sky: our sky patch has an extension of a $12.5^\circ \times 12.5^\circ$ and is sampled over a 300×300 pixel grid; this is small enough to ignore curvature effects so that Euclidean coordinates can be used rather than spherical coordinates.

Spectral SNR. In order to give an idea of the order of magnitude of various quantities, we report here on preliminary tests on simulated observations for the $K = 6$ channels of the HFI (high frequency instrument) which are centered around the microwave frequencies 100, 143, 217, 353, 545 and 857 GHz. Our simulations only include the three astrophysical components listed above: CMB, galactic dust, and SZ emission from galaxy clusters. We have also added spatially white noise at the level currently expected.

The simulated temperature anisotropies $\Delta T_k(\theta,\phi)$ for the $K = 6$ channels are stacked

in a $K \times 1$ vector

$$X(\vec{r}) = [\Delta T_1(\theta,\phi), \ldots, \Delta T_6(\theta,\phi)]^\dagger$$

where $\vec{r} = \vec{r}(\theta,\phi)$ is a vector in the tangent plane to the sky patch. With similar notations, the noisy mixture model (4) is rewritten as

$$X(\vec{r}) = A\ S(\vec{r})\ +\ V(\vec{r}) \tag{5}$$

with a 6×3 (in our experiments) 'mixing matrix' A whose components describe the emissivity of each component in each band.

Figure 8 shows the 'spectral SNR', that is, the expected power at detector k of the j-th component at spatial frequency ν. This is estimated by averaging the squared modulus of the two-dimensional Fourier transform of each map $S_j(\vec{r})$ over a thin 'spectral ring' where the modulus of the spatial frequency is close to ν and multiplying the result by A_{kj}^2. Figure 8 also shows a flat noise contribution: a flat spectrum is probably not very realistic but its level at low frequency is. The vertical scale on fig. 8 is in dB and the horizontal scale is in terms of the sampling frequency of our maps (300 pixels for 12.5°). The plot is limited to $\nu < 0.3$ rather than $\nu < 0.5$ because our model for CMB synthesis is not accurate (yet) above this value.

This figure shows a few important points in terms of imaging processing. First we note that all the components are dominated by the low frequency part of their spectrum. The power spectrum decreases steadily with spatial frequency, (very) roughly as $1/\nu$. We also note that the component of interest (CMB) is stronger than the noise only on the first three detectors. The relative noise level is significant especially in the high frequency part of the spectrum, which is the most interesting since it represents the correlations at the smallest angular scales.

3.2 Separation of noisy mixtures

In this section, we briefly discuss the reconstruction of components from noisy linear mixtures as modeled at eq. (4). First we note that because the SNR condition is varying wildly through the frequency range, it seems natural to process the data in the frequency domain. Fourier transforming the data does not change the basic structure (4) which is the starting point for component separation. Since we are considering processes whose distribution can be assumed to be spatially homogeneous, we can further exploit the fact that the Fourier coefficients of stationary processes are (at least asymptotically) uncorrelated.

Let us the recall some basics of the classic problem of reconstructing a vector S from noisy observations assumed to follow a noisy instantaneous mixture model:

$$X = AS + V. \tag{6}$$

After Fourier transforming the data, we have one such equation at each Fourier frequency; matrix A is constant through the frequency domain while the statistics of S and V are not assumed to be frequency independent. At a given frequency, the linear reconstruction of S from X via a matrix B reads:

$$\hat{S} = BX. \tag{7}$$

for some matrix B. The Wiener filter is the matrix B_W which minimizes the covariance of the reconstruction error $B_W = \arg\min_B \mathrm{Cov}(\hat{S} - S)$. It is easily found to be

$$B_W = (A^\dagger R_v^{-1} A + R_s^{-1})^{-1} A^\dagger R_v^{-1} \tag{8}$$

where $R_x = \mathrm{Cov}(X)$, $R_s = \mathrm{Cov}(S)$, etc.

Two remarks are in order about the Wiener solution. First, with this solution, the components are not reconstructed with unit gain. Indeed, it is not difficult to see that $[B_W A]_{jj} \leq 1$ for all j (with equality in the noise free case). Therefore, we will rather consider a reconstruction by

$$B_N = \mathrm{diag}([B_W A]_{11}, \ldots, [B_W A]_{nn})^{-1} B_w \tag{9}$$

where the diagonal prefactor ensures that $[B_N A]_{jj} = 1$ for all j. This solution seems more appropriate for building estimates of the component spectra while the Wiener filter should probably be preferred for building images for visual inspection [6]. We note that the diagonal prefactor in (9) does *not* change the signal to noise ratio; it just provides unbiased estimates of the variance of the components.

Second, while it may seem that knowledge of all parameters is needed to compute the Wiener filter (8) or its normalized version (9), this is actually not the case when one is interested in a *single* component. Indeed, denoting b_j the j-th column of

$B_N^\dagger$, the j-th component is recovered as $b_j^\dagger X$. Using the alternate expression $B_W = R_s A^\dagger R_x^{-1}$ for the Wiener filter, it is found that, for uncorrelated components,

$$b_j = (a_j^\dagger R_x^{-1} a_j)^{-1} R_x^{-1} a_j \qquad (10)$$

where a_j is the jth column of A. This expression shows that the knowledge of R_x and a_j is all that is needed to compute the normalized Wiener filter b_j for the j-th source. This is of particular relevance in our application because the emissivity of the CMB component can be accurately predicted from the Planck spectrum. Thus the implementation of the Wiener filter (normalized or not) for the CMB component only requires the additional knowledge of the covariance matrix R_x. This quantity however can easily be estimated from the data.

Figure 9 shows the improvement offered by spatial filtering with the normalized Wiener filter (9). Each panel of the figure shows the reconstruction of a given component. The vertical axis of the j-th panel shows, on a dB scale, the power of each component in the estimate $\hat{S}_j = b_j^\dagger X$ as function of the normalized frequency. If the noise level was negligible at all frequencies, the jth panel would show a domination of the j-th component. Because there is a significant amount of noise at all frequencies, each panel shows the best compromise between rejecting each other component while keeping down the noise.

Because figure 8 assumes that the noisy factored model (4) holds, the *shape* of the spectrum of each component does not change from detector to detector; only the level changes. These particular spectral shapes are found again undistorted in figure 9 because we have used the normalized form (9) of the Wiener filter. Thus comparing the two figures allows to gauge the improvement in SNR brought by the Wiener filter in each frequency band. For instance, comparing the CMB component to the noise level, the best detector is the second one which gives $SNR > 1$ for $\nu < 0.2$. After Wiener filtering, the CMB component remains over the noise level for $\nu < 0.28$. The same can be approximately said when comparing the level of CMB to the SZ component. The fact that, in all panels of fig. 9, the Wiener filter does not massively reject the contaminations is a consequence of the difficult SNR conditions: for each component, a better rejection of the other components would entail an even more severe increase in the noise level.

So far, we have discussed *linear* Wiener filtering. For Gaussian signals, the minimization of a quadratic criterion like $\mathrm{Cov}(\hat{S} - S)$ is achieved by a linear filter, indeed. However, as can be seen from our sample sky maps, the SZ and the dust components are non Gaussian. Therefore, one can expect better results by using non Gaussian models and non-linear filters. Such an attempt is described in [4]. The authors compute the expected value of the components, given the observations and a non Gaussian prior on the coefficients of the Fourier transform. This is an intriguing result since a Fourier transform tends to Gaussianize signals. Given the (roughly) point-like structure of the SZ components, one should expect even larger improvement by expressing the non Gaussianity in the original spatial domain rather than in the transform domain.

3.3 Blind component separation

An interesting property of model (5) is that, under some assumptions, it allows for a 'blind separation' of the components. This means that it is possible to identify the mixing matrix A using only the available data and the assumptions that the components are *statistically independent*. This idea currently is the subject of many researches in the signal processing community. It has already been tested on astronomical images by Nuzillard *et al* [5], although in a context where the physical significance of the 'components' is not as clear as in the CMB imaging problem.

There is an obvious benefit to the 'blind approach' for astronomical component separation: the mixing matrix is not perfectly known so that processing the data (using the Wiener filter, for instance) with wrong assumptions may lead to poor separation (or no separation at all). We have mentioned that there are little doubts about the emissivity pattern of the CMB component but the emissivity patterns of the other components may be more difficult to predict. It is also an exciting perspective to use the blind separation approach possibly to *discover* underlying components.

One drawback of a purely blind approach is also obvious: using an estimate of the mixing matrix A

in place of the true one (if it exists and is known) can only lead to performance degradation. Thus, even though the blind approach appears attractive, it is important to make sure that the gain in versatility is not outweighed by a significant loss of accuracy.

Exploiting the statistical assumption of component independence is not straightforward because there are many different possibilities to express it in ways which are amenable to sensible algorithms.

The first attempt at a blind separation of components in the CMB context is due to Baccigalupi *et al* [1]. They use a technique which follows the 'classic' approach of blind source separation: the exploitation of *non Gaussianity* in the components [3]. It is indeed possible to blindly identify a mixture of independent components provided that *at most* one of them is Gaussian (we refer here to the marginal distribution of the components, their time or space dependence being ignored in this approach). This seems valid in the CMB context since, while the CMB component itself has a Gaussian distribution, the other components are expected to be non Gaussian (possibly very strongly so, as the SZ component, for instance). The paper by Baccigalupi reports promising results, showing that the statistical distribution of the components at least allows for blind identifiability. However, their experiments are performed without noise. Taking noise into account will probably require a whole new strategy.

In these proceedings, we report on a different approach [2] which uses spectral statistics in the Fourier domain. This approach does not rely on the non Gaussianity of the components but on their *spectral diversity*. An important benefit of relying on the spectral structure for blind identification is that we can content ourselves with a *Gaussian model* for the noisy data. In this model, the likelihood can be easily optimized with respect to all the parameters of interest. One drawback of the method is that it does require that the spectra of components to be separated have a different shape: two components with proportional spectra cannot be blindly separated. Exact proportionality of spectra is not likely to occur but the quality of separation between any two given components is expected to be poor if they have similar spectral shape.

In the future, we hope to see successful contributions of the blind source separation techniques to the astronomical component separation problem. These solutions should be noise-resistant and able to exploit the statistical structure of the components in a manner adapted to the components: Gaussian models for the CMB, non Gaussian models for components due to the SZ effect, etc.

4 Other aspects

We have only given a glimpse to one of the many issues raised by CMB imaging. Two other important tasks are map making and spectrum estimation.

Map making is the process of building two-dimensional sky maps starting from the one dimensional scan lines acquired by the rotation of the satellite. Some issues are: dealing with the large number of pixels at the Planck resolution, making good use of the redundancy of the scans (many scan lines will pass close to a given pixel of the final map), interpolation to the sampling grid, taking into account beam effects (deconvolution), properly handling the measurement noise and its non stationarity,...

Obtaining a high accuracy estimate of the CMB harmonic spectrum is a major objective of the mission. This is made difficult in particular by the 'galactic cut'. In effect, the equatorial zone of the sky is occupied by the Milky Way, our galaxy. It is not possible to obtain reliable measurements of the CMB temperature from this part of the sky. Thus, one has to 'cut off the galaxy', leaving the statistician with a large gap of missing data along the equator. Other issues regarding spectrum estimation are, again, dealing with the large volume of data and assessing the accuracy of the estimates, a crucial point since the results are ultimately destined to constrain cosmological parameters.

Credits and resources

I would like to thank Luigi Bedini from CNR Pisa, Italy, for introducing me to the component separation problem in the context of CMB estimation. Thanks to Jacques Delabrouille for reading a draft of this paper and checking my physics facts (even though later edition and additions may have ruined

his work!). The simulated CMB data also are due to his group at Collège de France.

Many wonderful sites are available on the Internet to get started with the big bang, the Planck mission and everything cosmologic. Among those, I have particularly appreciated Ned Wright's cosmology tutorial [10] (with its section 'News of the Universe') and the extensive material maintained by Wayne Hu on his home page [9] at the University of Chicago.

The images in this paper are taken from a poster made by the Planck science team. It is available together with additional tutorial material on their web site [8].

References

[1] C. Baccigalupi, L. Bedini, C. Burigana, G. De Zotti, A. Farusi, D. Maino, M. Maris, F. Perrotta, E. Salerno, L. Toffolatti, and A. Tonazzini. Neural networks and separation of cosmic microwave background and astrophysical signals in sky maps. *Monthly Notices of the Royal Astronomical Society*, 318:769–780, November 2000.

[2] J.-F. Cardoso, H. Snoussi, J. Delabrouille, and G. Patanchon. Blind separation of noisy Gaussian stationary sources. Application to cosmic microwave background imaging. In *Proc. EUSIPCO*, 2002.

[3] Jean-François Cardoso. Blind signal separation: statistical principles. *Proceedings of the IEEE. Special issue on blind identification and estimation*, 9(10):2009–2025, October 1998.

[4] M. P. Hobson, A. W. Jones, A. N. Lasenby, and F. R. Bouchet. Foreground separation methods for satellite observations of the cosmic microwave background. *MNRAS*, 300:1–29, October 1998. astro-ph/9806387.

[5] Daniele Nuzillard and Albert Bijaoui. Blind source separation and analysis of multispectral astronomical images. *Astron. Astrophys. Suppl. Ser.*, 147(1):129–138, 2000.

[6] Max Tegmark and George Efstathiou. A method for subtracting foregrounds from multifrequency cmb sky maps. *Monthly Notices of the Royal Astronomical Society*, 281:1297–1314, 1996. astro-ph/9507009.

[7] Healpix: Hierarchical equal area isolatitude pixelisation of the sphere. http://www.eso.org/science/healpix/.

[8] The home page of the science team of Planck. http://astro.estec.esa.nl/Planck/.

[9] The home page of Wayne Hu at the University of Chicago. http://background.uchicago.edu/∼whu/

[10] Ned Wright's Cosmology Tutorial, http://www.astro.ucla.edu/∼ wright/cosmolog.htm

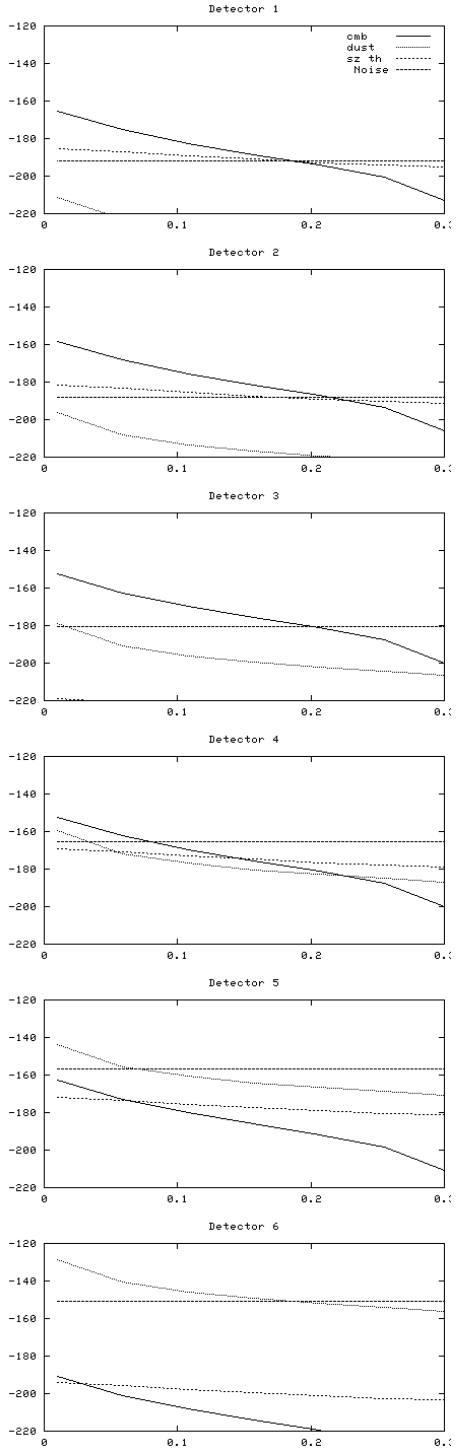

Figure 8: SNRs at the detectors

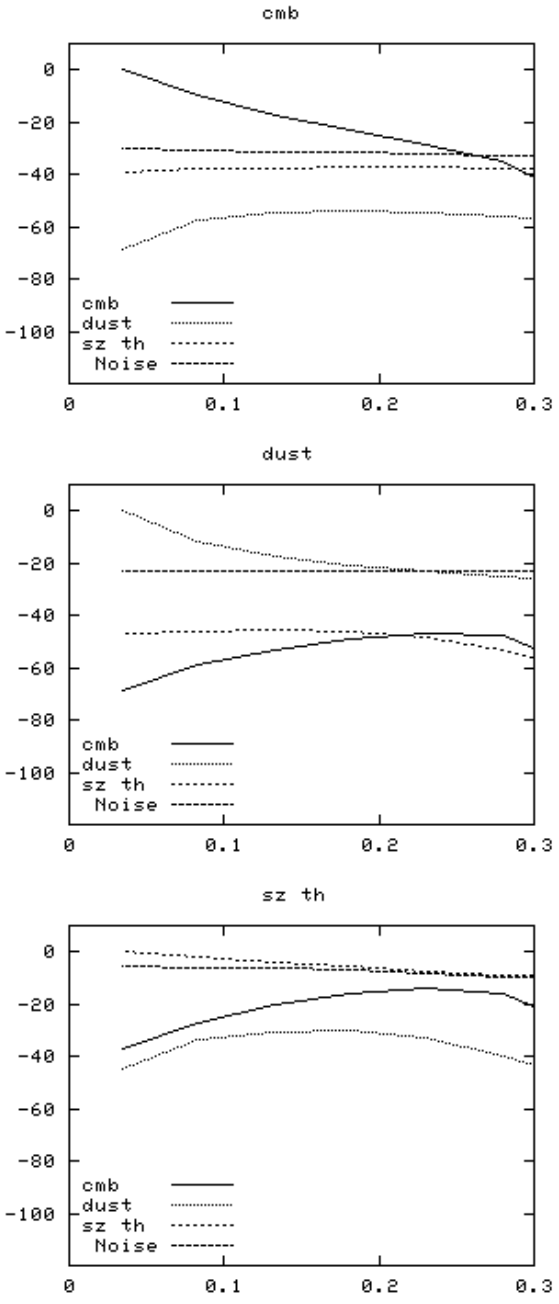

Figure 9: SNRs for each component after normalized Wiener filtering