.

Extracting the Cosmic Microwave Background

from multi-frequency observations.

J.-F. Cardoso, C.N.R.S.

Institut d'Astrophysique de Paris

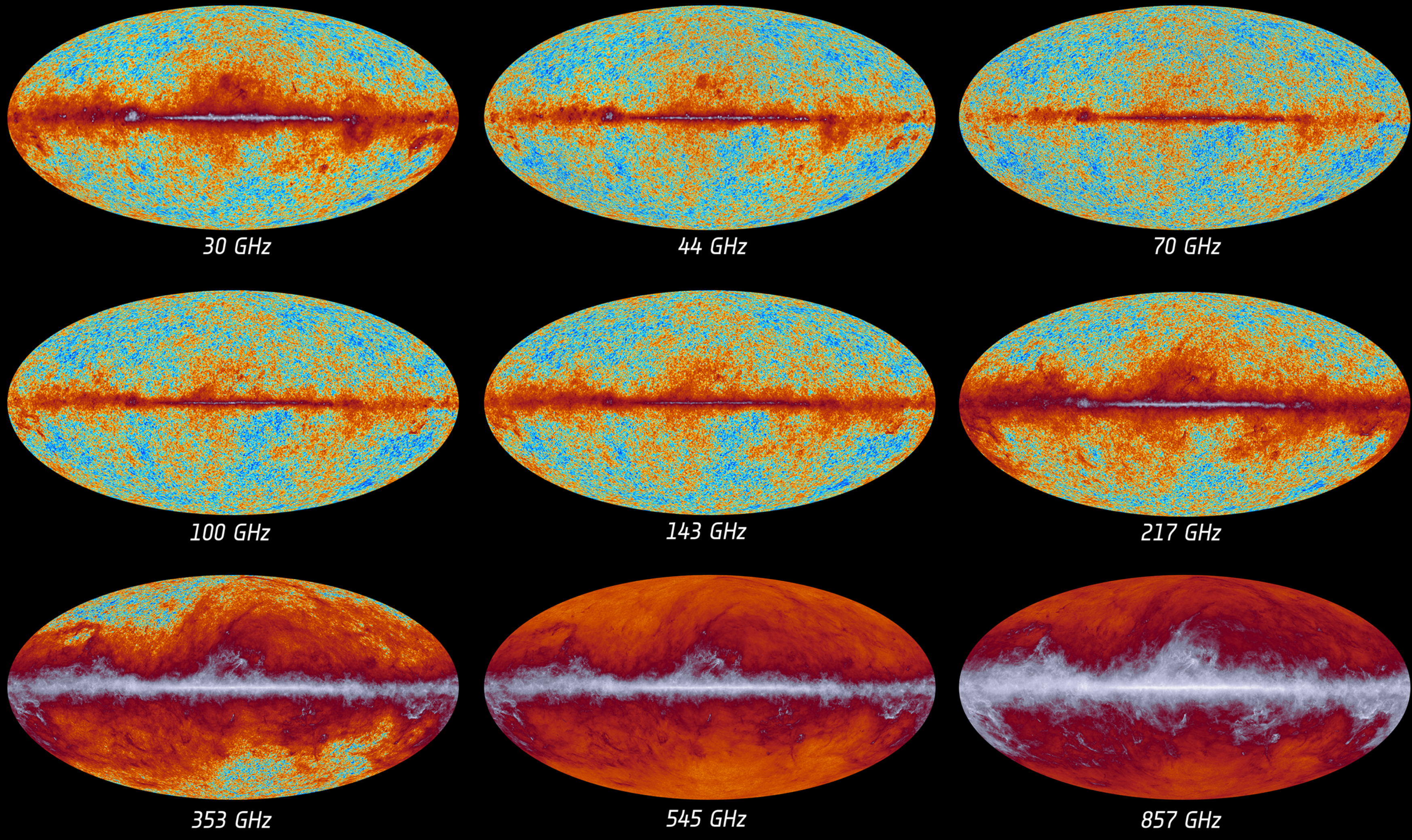
XIIIth School of Cosmology. 12-18/11/17, IESC, Cargese: The CMB from A to Z.

Context

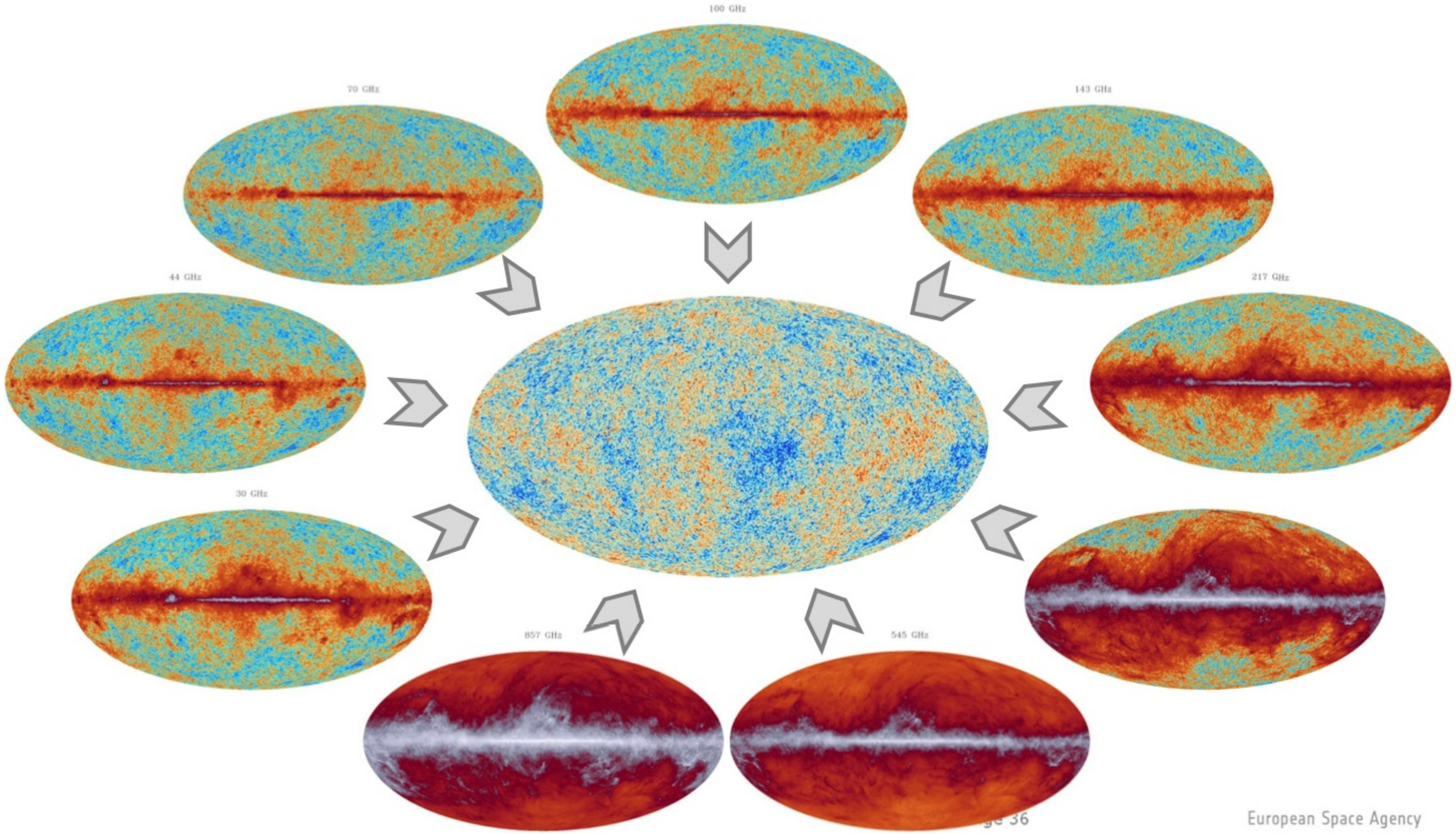
planck

The sky as seen by Planck

esa



Combining the 9 Planck channels into a map of the Cosmic background

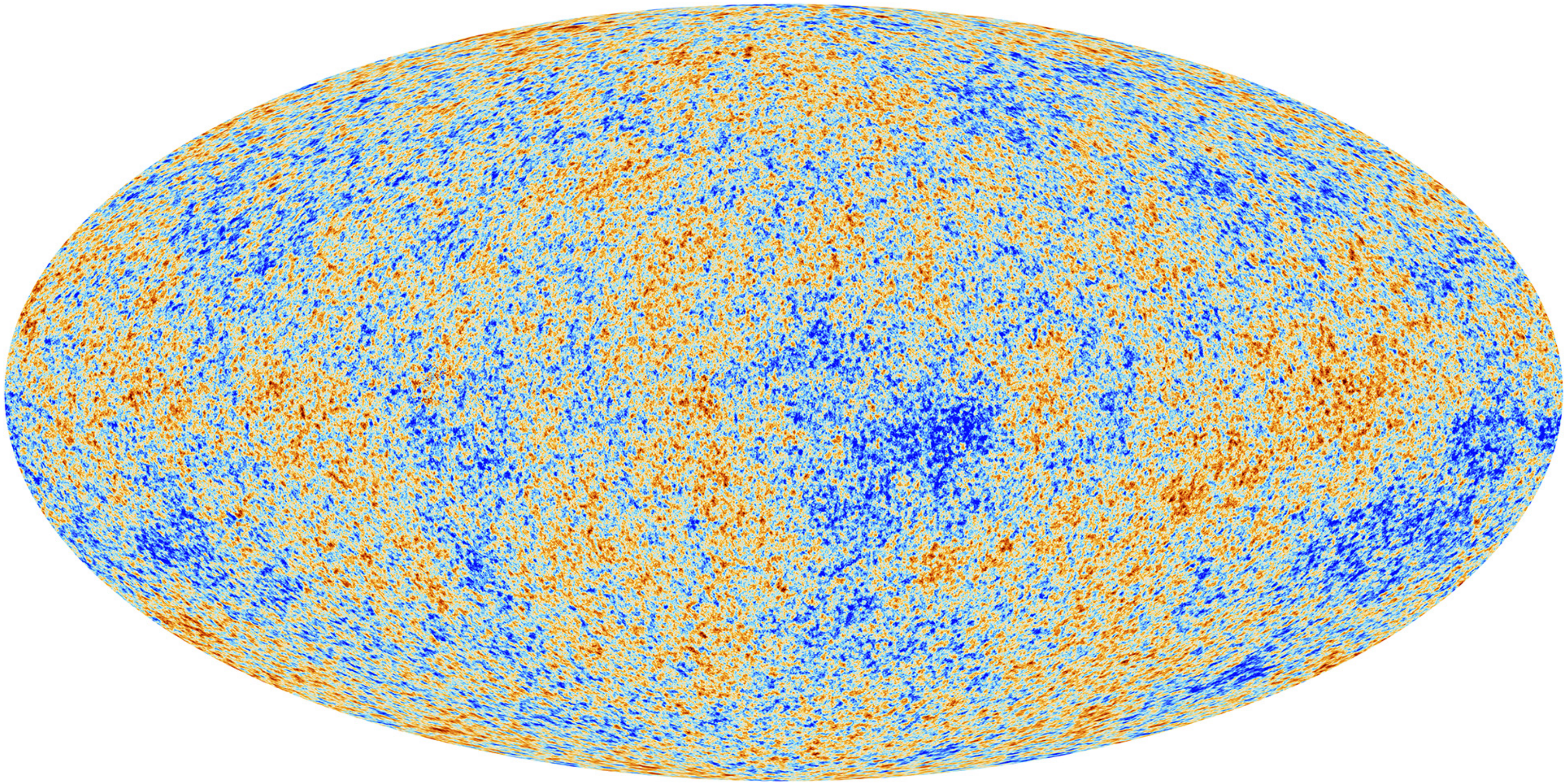


Getting rid of all (?) foregrounds: the noble art of component separation.

Or is it just CMB cleaning ?

Credit: ESA, FRB

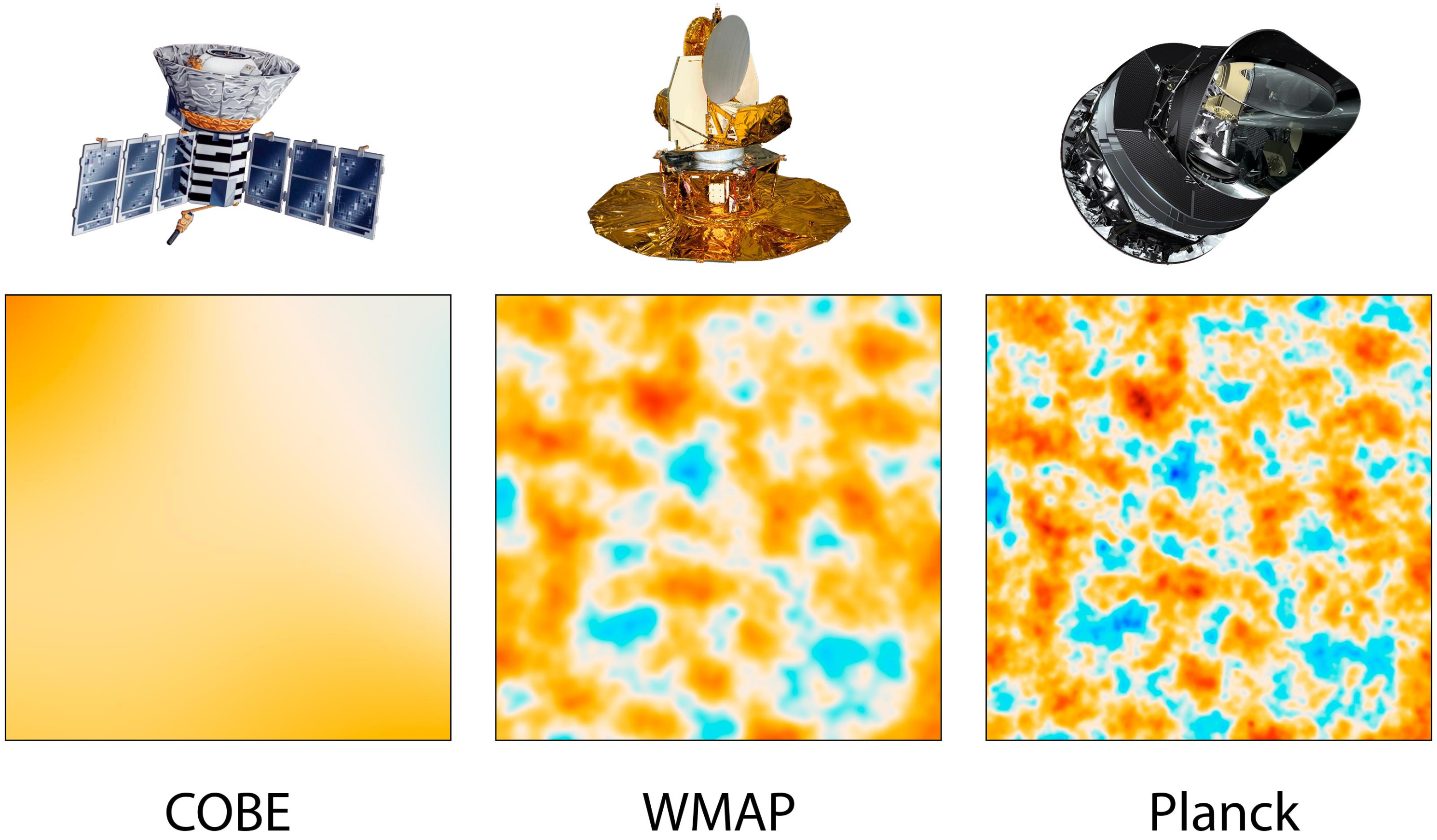
The oldest image in the world, by Planck



Color scale : $\pm$ 300 μK

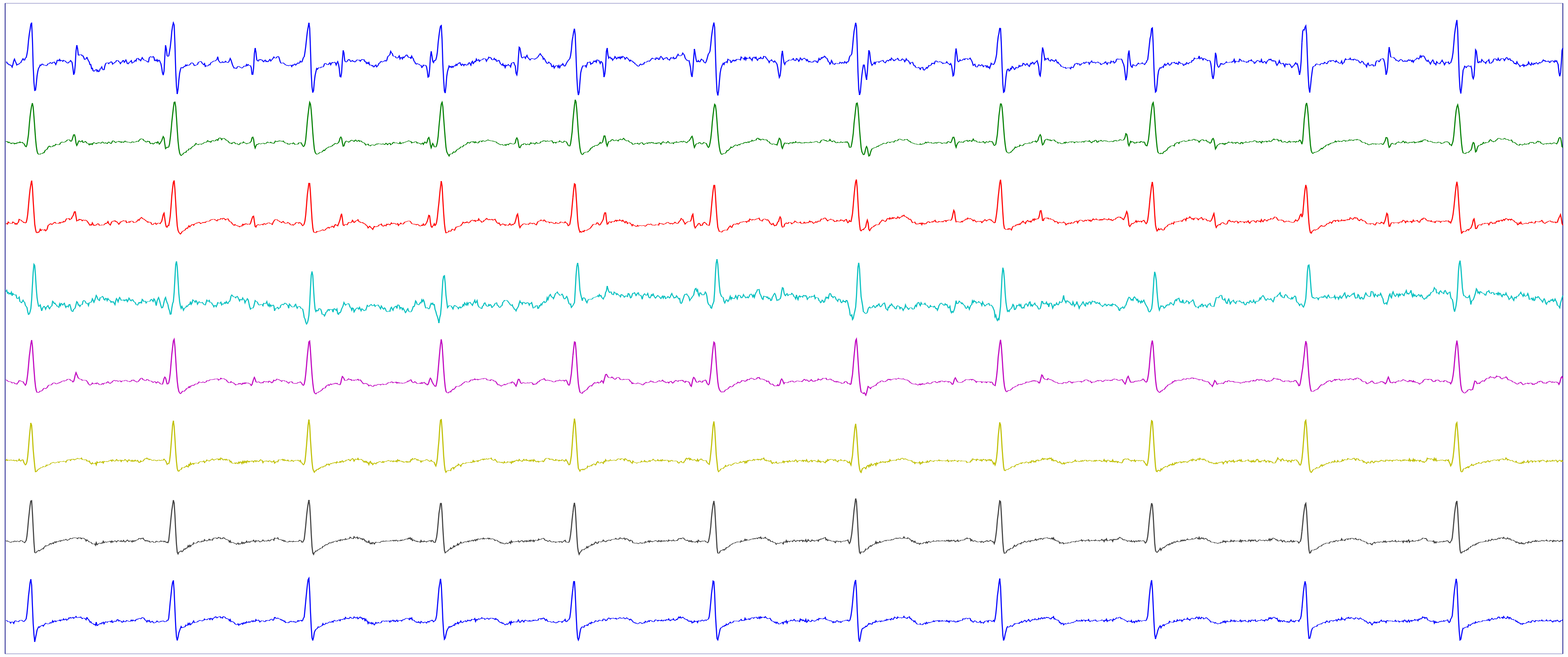
But some parts of it are more recent: inpainting...

COBE 1992, WMAP 2001, Planck 2013



Gains in resolution also made possible by an increased frequency coverage, making foreground cleaning at shorter wavelengths more manageable.

And now for something biomedical. ECG: Electro-Cardiography



8 EEG electrodes located on the thorax and abdomen of a pregnant woman.

Looking for linear decompositions: Data = Mixing matrix × Sources.

Can we extract the heartbeat of the (soon to be) baby ?

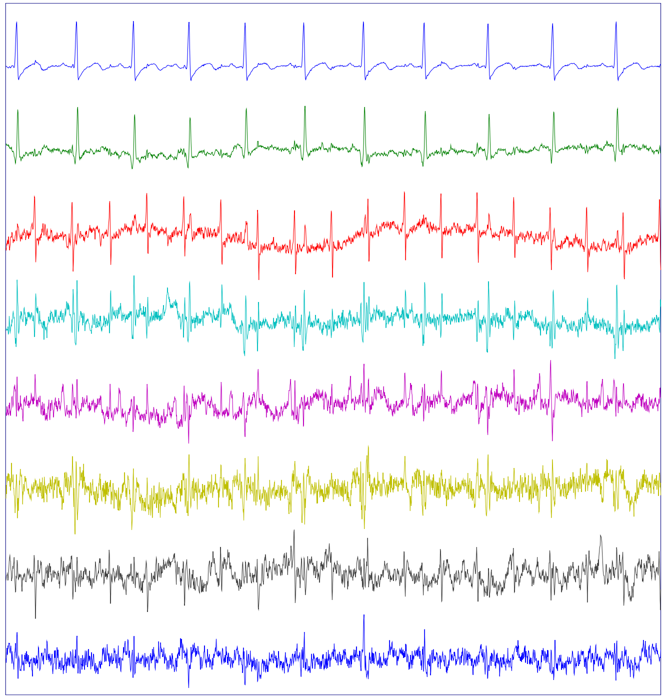
Data Credits: Daisy database

Principal component analysis

=

95	-51	-33	8	2	0	2	2
87	53	16	6	-2	-5	4	4
94	17	14	0	11	6	1	1
-9	105	-28	-5	1	0	0	0
97	26	0	5	0	-1	-9	-9
89	7	1	0	-11	8	0	0
97	-6	1	-4	-1	-4	1	1
92	-32	-3	-15	0	-1	0	0

×

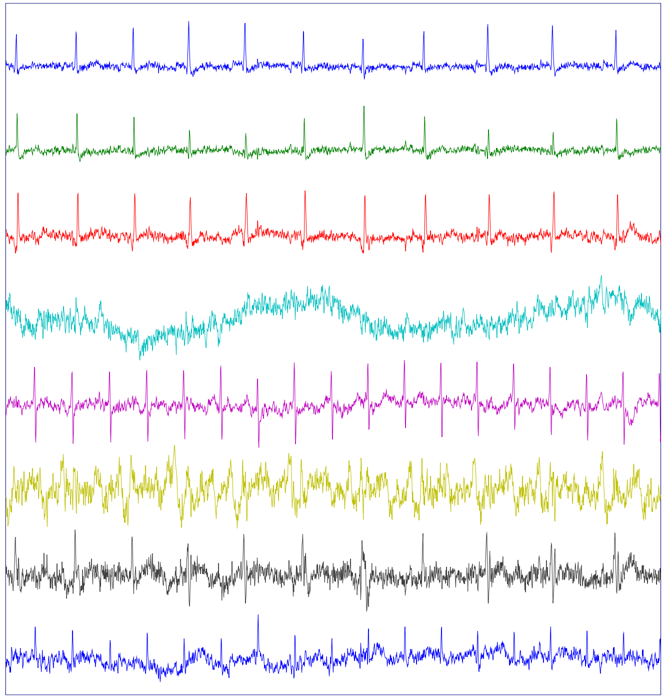


- Orthogonal mixture, uncorrelated components $\frac{1}{T}\sum_t y_i(t) y_j(t) = 0$ for $i \neq j$
- Decorrelation is weak (always posible), orthogonality is implausible.

. . . the baby signal is more visible but not separated yet.

Independent component analysis

$$
= \begin{bmatrix}
98 & 60 & -71 & 5 & -34 & 17 & 3 & 3 \\
93 & 114 & 37 & 0 & 10 & 5 & 6 & 6 \\
101 & 99 & -4 & 0 & 13 & 6 & 6 & 6 \\
-17 & 47 & 79 & 33 & -19 & -13 & -1 & -1 \\
106 & 102 & 7 & 0 & -10 & 4 & 10 & 10 \\
110 & 71 & -8 & 0 & -1 & 1 & 0 & 0 \\
107 & 82 & -22 & 0 & 0 & 7 & 17 & 17 \\
112 & 42 & -43 & 0 & -1 & 11 & 20 & 20
\end{bmatrix} \times
$$

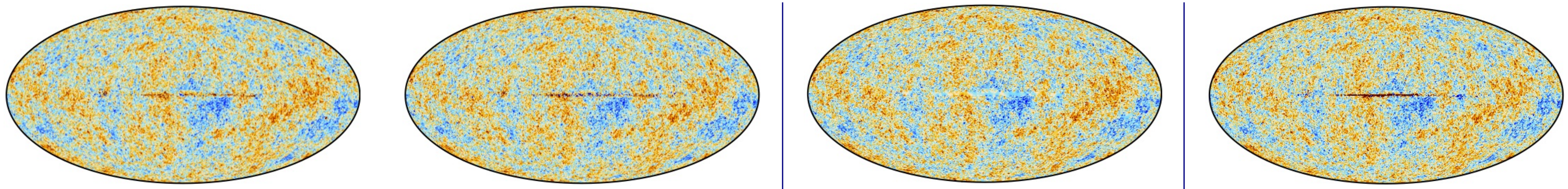


- Linear decomposition into “the most independent sources”
- Blind: only independence is at work but it must go beyond decorrelation.
- Independence is statistically very strong but often physically plausible.
- Weak assumptions $\longrightarrow$ wide applicability.

. . . and **the fetal signal is clearly visible** (it is the fifth extracted source, at about twice the beat rate).

Actually, this works because the signals are non Gaussian, even **sparse**.
Would it work for CMB ? How to do it accurately?

Four CMB anisotropy maps delivered to the Planck Legacy Archive (2013)



NILC	SEVEM	SMICA	C-R
$\ell_{\mathrm{SNR}=1} = 1790$	$\ell_{\mathrm{SNR}=1} = 1790$	$\ell_{\mathrm{SNR}=1} = 1790$	$\ell_{\mathrm{SNR}=1} = 1550$
Wavelet space	Wawelet-like	Harmonic space	Pixel space
non-parametric	non-parametric	semi-parametric	parametric

- Various assumptions about the foregrounds.
- Various filtering schemes (space-dependent, multipole-dependent, or both).

Some requirements for producing a CMB map

- The method should be accurate and high SNR (obviously).

- The method should be linear in the data:

  1. It is critical not to introduce non Gaussianity

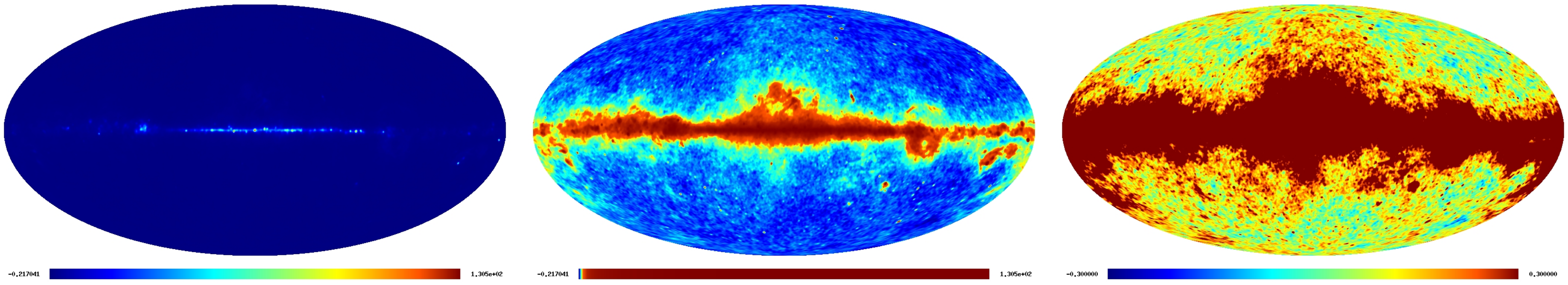
  2. Propagation of simulated individual inputs should be straightforward

- The result should be easily described (e.g. map=beam*sky+noise) with a well defined transfer function (beam control).

- The method should be fast enough for thousands of Monte-Carlo runs.

- The method should be able to support wide dynamical ranges, over the sky, over angular frequencies, across channel frequencies.

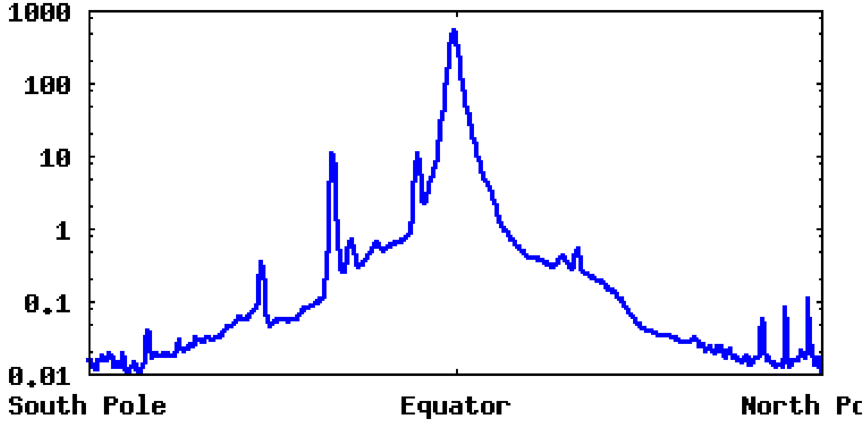
Wide dynamics over the sky



Left: The W-MAP K band. Natural color scale [-200, 130000] μK.

Middle: Same map with an equalized color scale.

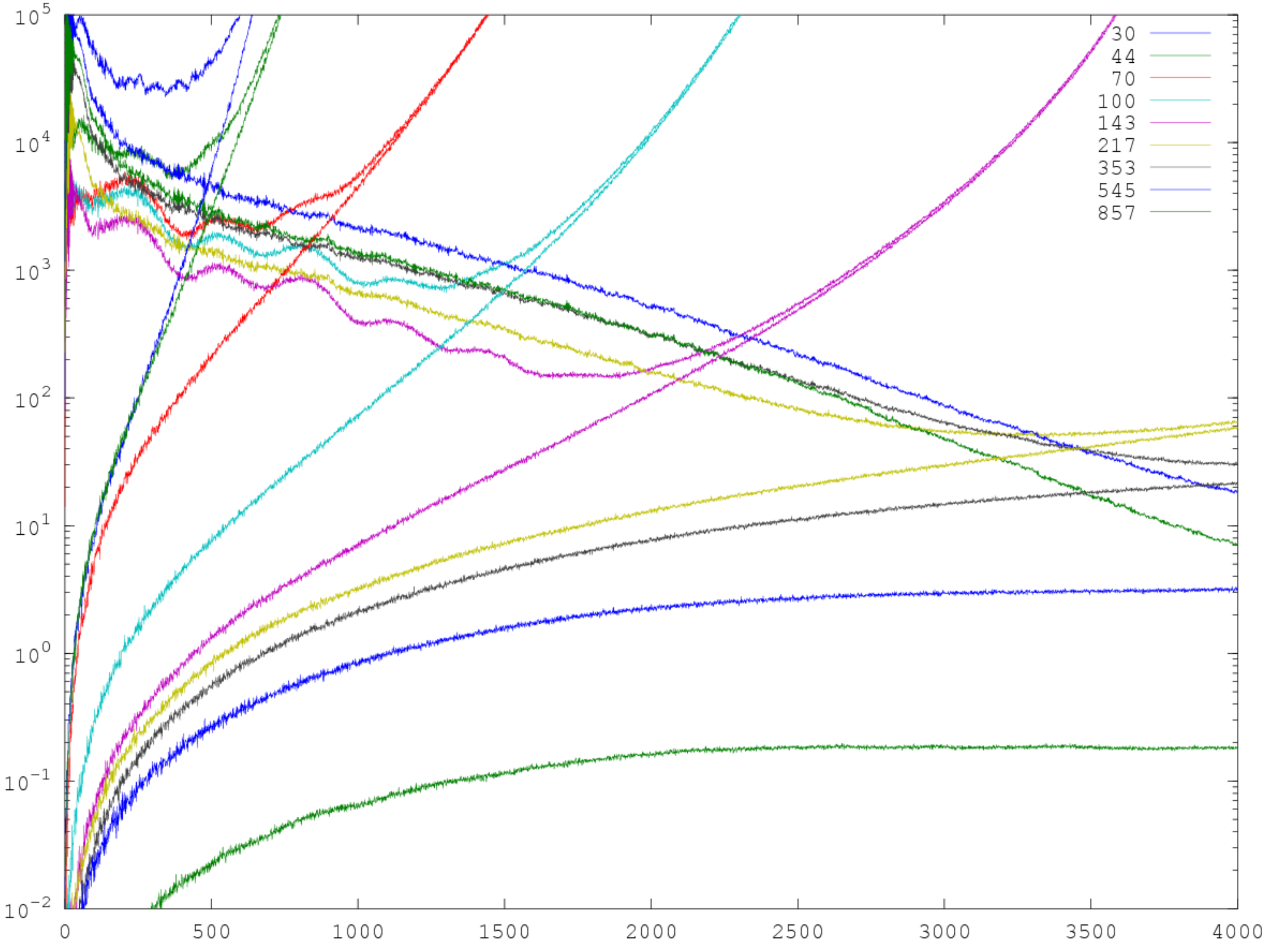
Right: Same map with a color scale adapted to CMB: [-300, 300] μK.



Average power as a function of latitude on a log scale for the same map.

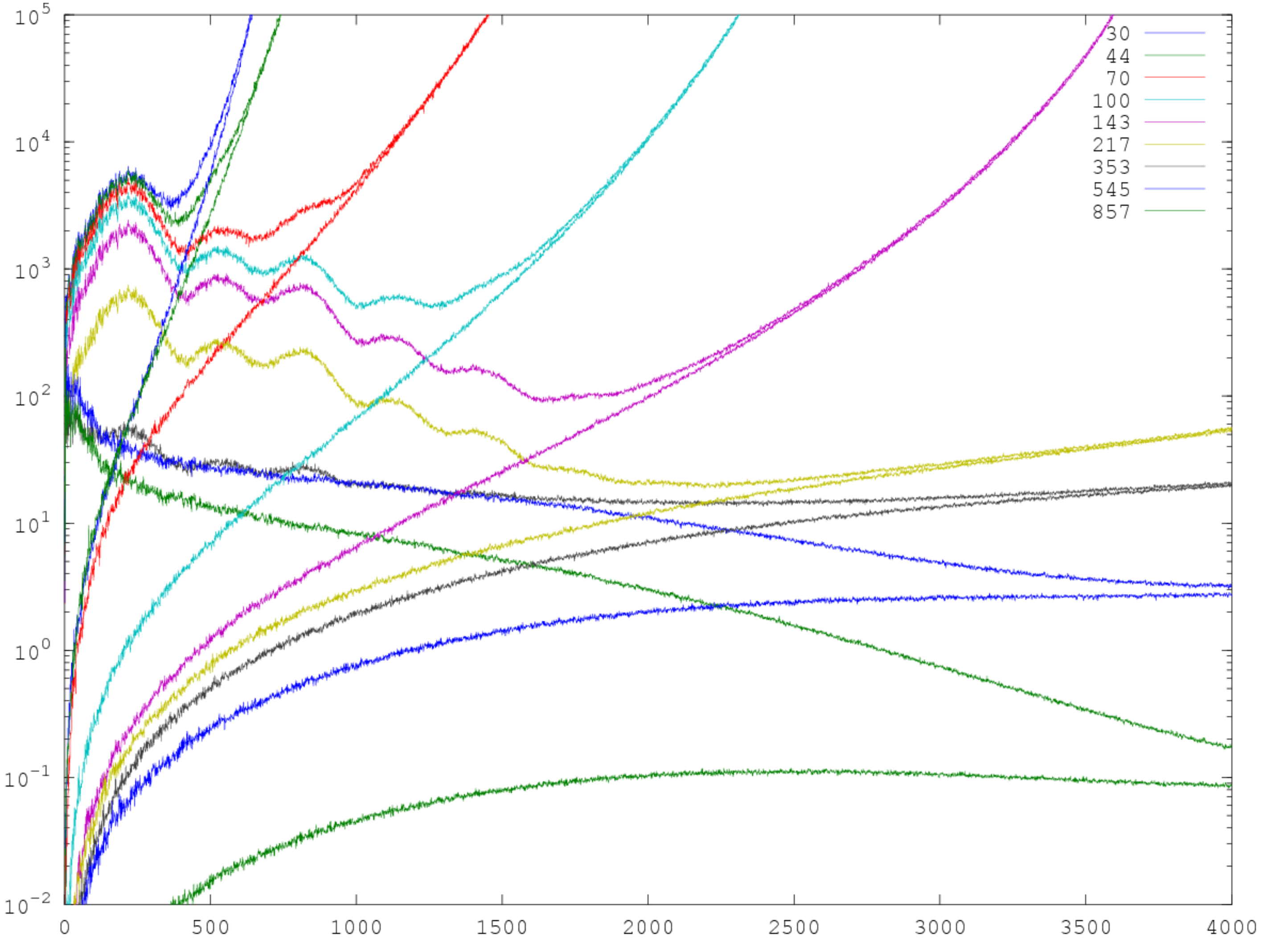
Do we really want to compute two-point correlations over the whole sky?

Wide spectral dynamics, SNR variations



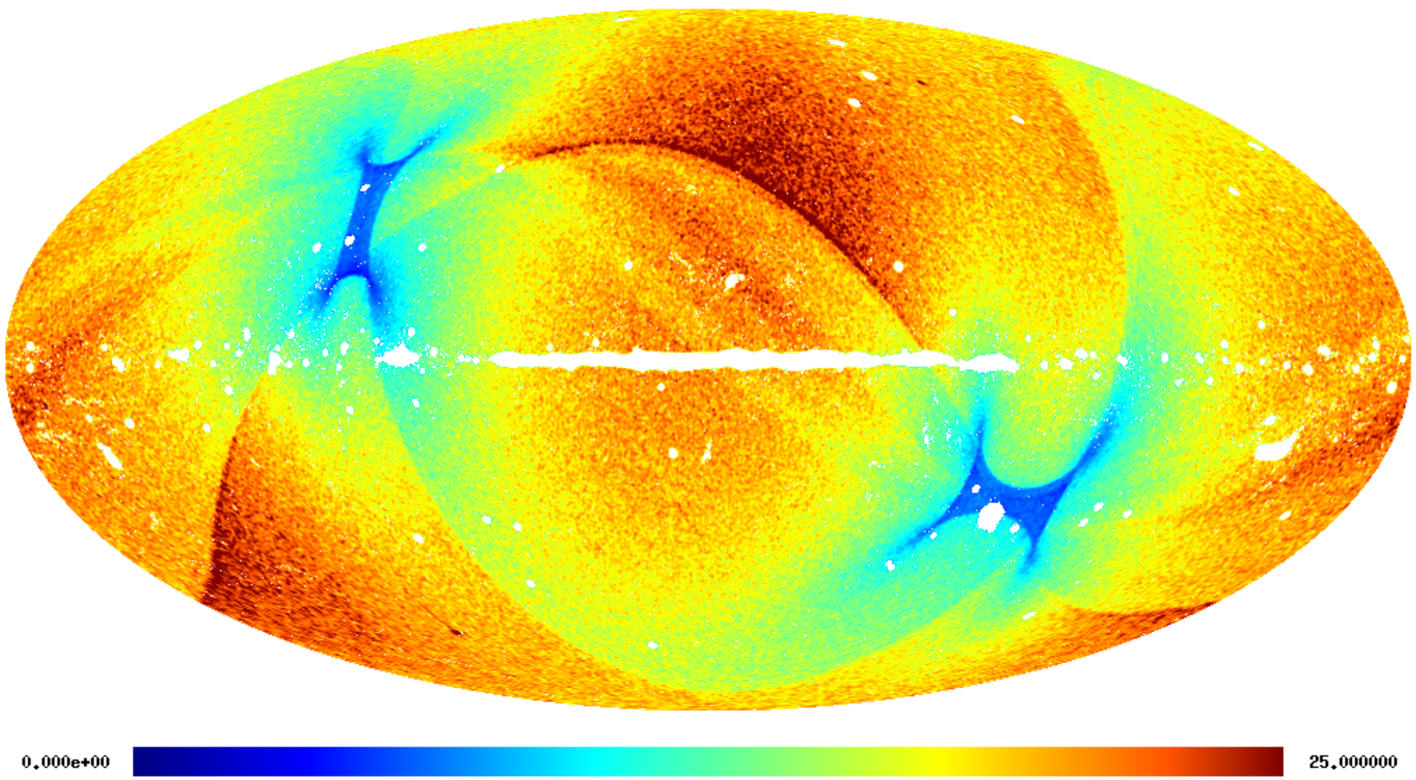
$\widehat{C}(\ell) \cdot \ell(\ell+1)/2\pi$ in $[\mu K_{\mathrm{RJ}}]^2$ for $f_{\mathrm{sky}} = 0.99$.

Wide spectral dynamics, SNR variations



$\widehat{C}(\ell) \cdot \ell(\ell+1)/2\pi$ in $[\mu K_{\mathrm{RJ}}]^2$ for $f_{\mathrm{sky}} = 0.40$.

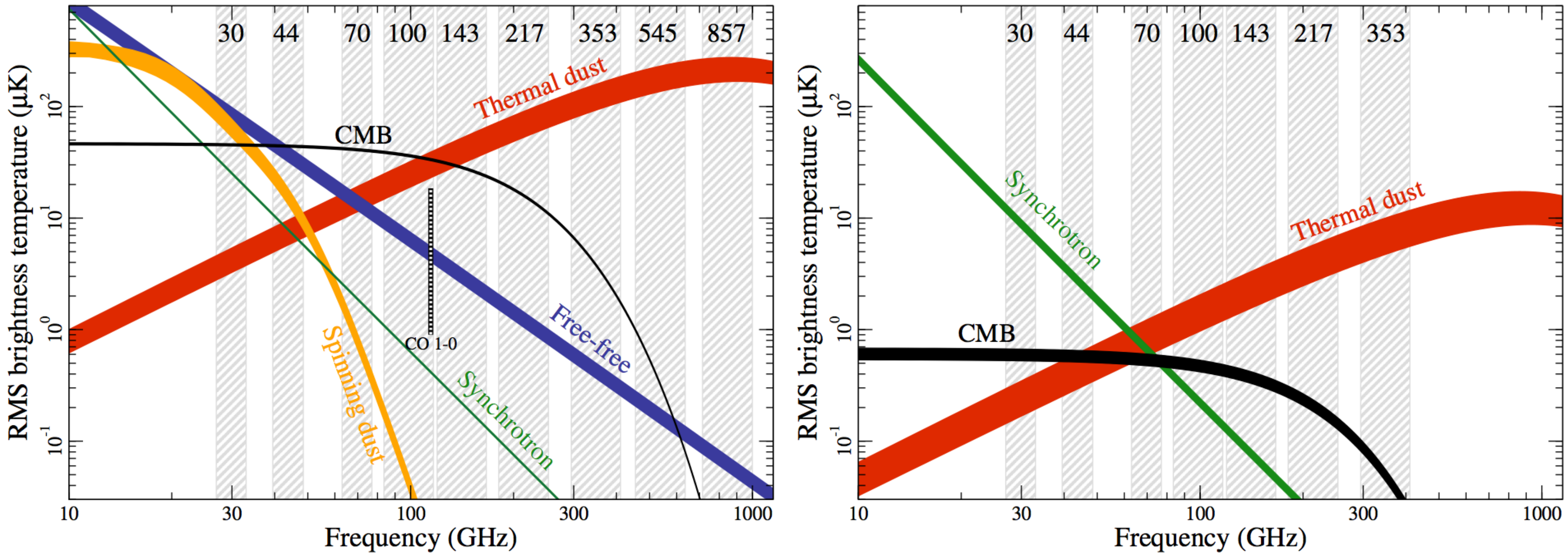
And what about the noise ?



Noise RMS in μK in the SMICA map.

Distribution of Galactic foregrounds. 1) Frequency distribution

Planck Collaboration: Diffuse foregrounds component separation



Brightness temperature RMS as a function of frequency and astrophysical component, for both temperature (left) and polarization (right). For temperature, each component is smoothed to an angular resolution of 1 degree FWHM, and the lower and upper edges of each line are defined by masks covering 81 and 93% of the sky, respectively. For polarization, the corresponding smoothing scale is 40 arcminutes, and the sky fractions are 73 and 93%.

From Planck 2014 foreground paper.

Distribution of Galactic foregrounds. 2) Spatial distribution

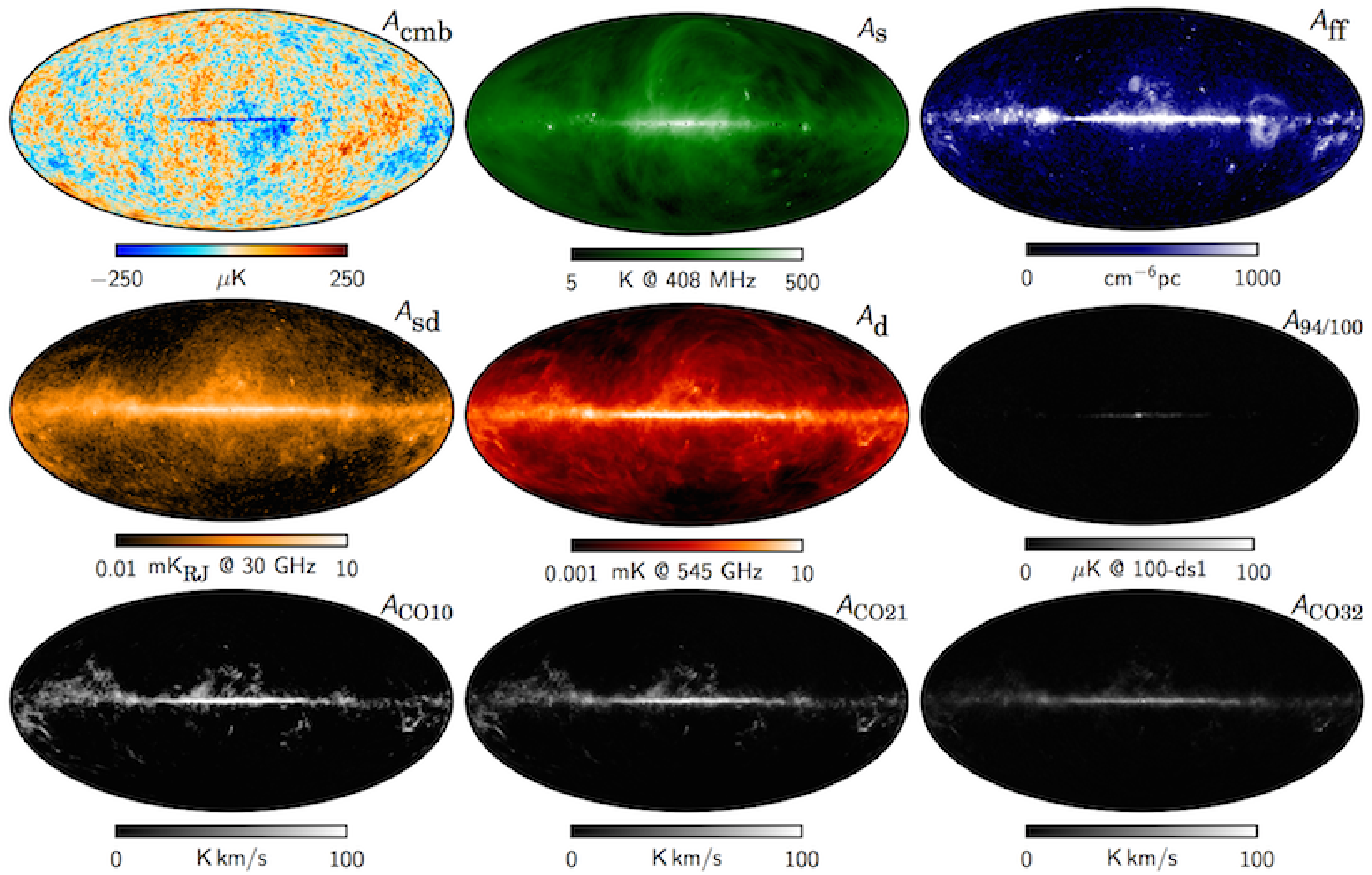


Fig. 5. Maximum posterior amplitude intensity maps derived from the joint baseline analysis of *Planck*, *WMAP* and 408 MHz observations. From left to right and top to bottom, the components are 1) CMB temperature; 2) synchrotron brightness temperature at 408 MHz; 3) free-free emission measure; 4) spinning dust brightness temperature at 30 GHz; 5) thermal dust brightness temperature at 545 GHz; 6) 94/100 GHz line emission, evaluated for the 100-ds1 detector map; and 7–9) CO line emission for J=1→0, J=2→1, J=3→2. Panels 2–5 employ the non-linear HDR color scale, while all other employ linear color scales.

Planck 2015 Galactic foregrounds: (CMB), synchrotron, free-free, spinning dust, thermal dust, CO lines.

Extra Galactic foregrounds

And then, there are extra Galactic foregrounds such as

- The Cosmic Infrared Background,
- Galaxy clusters seen via the SZ effect

And then point sources (non diffuse by definition)

- IR Galaxies
- Radio Galaxies

which can be masked out.

Proverb

It's not how dirty it is.

It's how well you can clean it up.

- **Key factors helping CMB extraction**

i) No occlusion of the CMB by (very?) coherent foregrounds.

ii) Good calibration, *i.e.* the instrument response to the CMB is well determined.

iii) Statistical independence between CMB and foregrounds+noise.

i+ii+iii) enough for simple methods to work.

- **But optimal processing wants a complete statistical model, doesn't it?**

a) CMB very close to a realization of a Gaussian stationary field (good).

b) No realistic statistical model for Galactic foregrounds (bad).

c) And there is only one sky (bad)...

These lectures focuses on the importance (or lack thereof) of statistical modeling.

For doing so, we often simplify as much as possible the data model.

The mixing model for rigid components

The ith map, observed at frequency ν_i, is a noisy superposition of components:

$$X_i(\theta,\phi) = \sum_{c=1}^{C} X_i^c(\theta,\phi) + N_i(\theta,\phi). \qquad c = \text{cmb, dust, SZ}, \ldots$$

If the emission of component c changes with ν_i while keeping the same spatial pattern, then that component is said to 'scale rigidly' and we have

$$X_i^c(\theta,\phi) = A_i^c \; S_c(\theta,\phi)$$

If all components scale rigidly or, i.o.w. are fully coherent, then stacking the sky maps seen at all F observation frequencies:

$$X(\theta,\phi) = \begin{bmatrix} X_1(\theta,\phi) \\ \vdots \\ X_F(\theta,\phi) \end{bmatrix} = \mathbf{A}S(\theta,\phi) + N(\theta,\phi) \qquad A: \text{the } F \times C \text{ mixing matrix.}$$

In these lectures, we focus on that simple model and consider the statistial aspects of component separation, that is, the best recovery of S given X and various amounts of prior information.

This is a simplified setting, complications may appear later. . .

The component separation problem may not be what you think

- If the beams have been perfectly corrected and
- if there are no more foreground emissions than channels and
- if each foreground is fully coherent so that an accurate model is

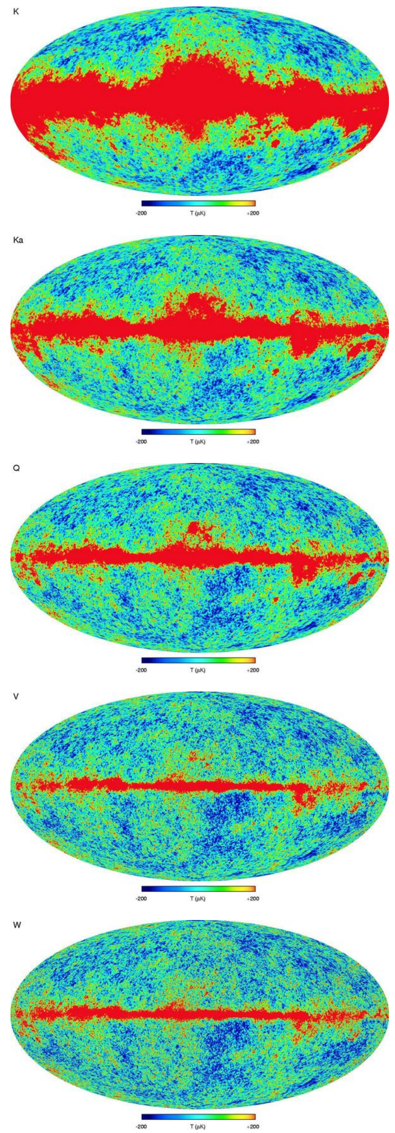
$$X(\theta,\phi) = \begin{bmatrix} X_1(\theta,\phi) \\ \vdots \\ X_d(\theta,\phi) \end{bmatrix} = \mathbf{A}S(\theta,\phi) + N(\theta,\phi)$$

- if the mixing matrix A is known perfectly,
- if all fields are Gaussian stationary processes with known spectra,

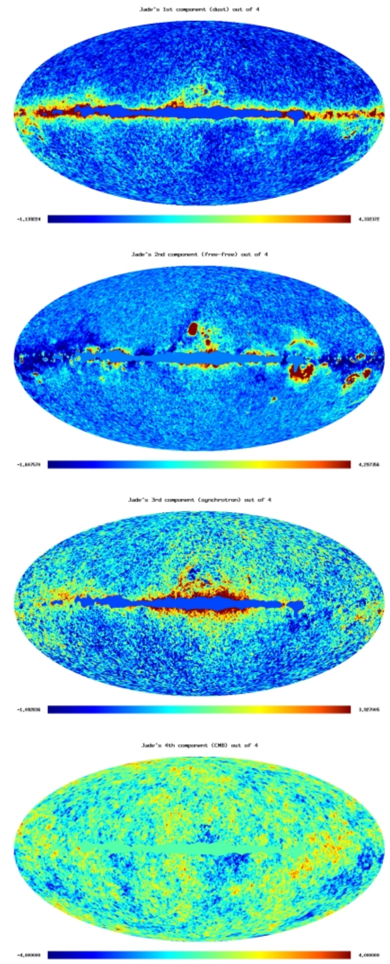
then, optimal solution(s) can be easily derived (see below).

The main issue is dealing with the uncertainties
and the assumptions/approximations in the above statements.

Maximal uncertainty: Blind component separation (a.k.a. ICA)



$$= \begin{bmatrix} A_{11} & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & A_{54} \end{bmatrix} \times$$

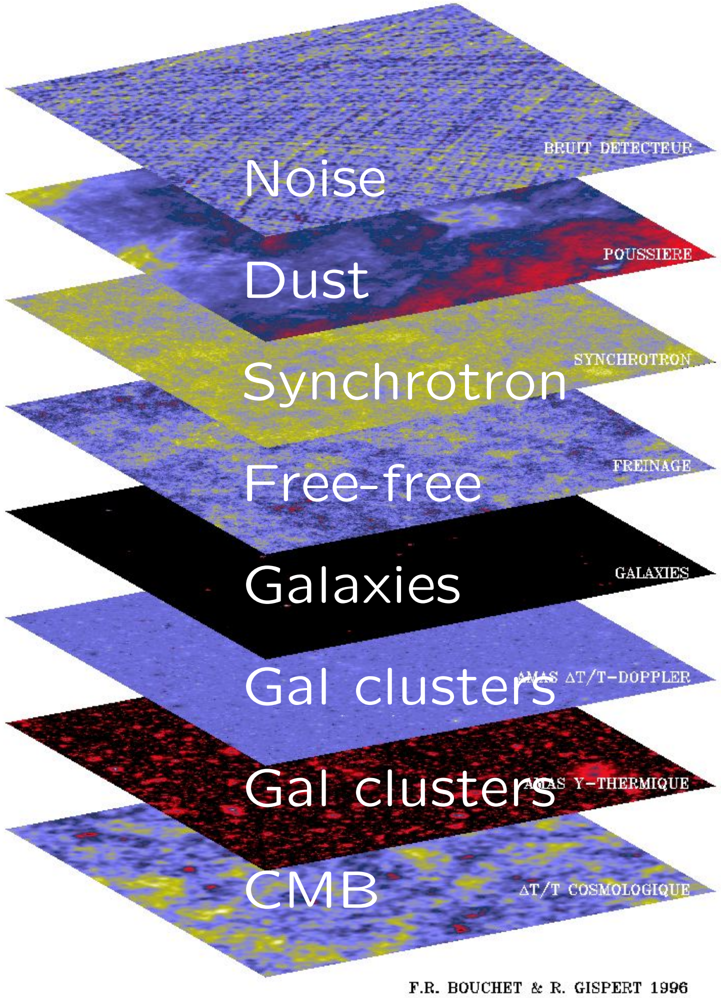


?
?
?
CMB!

$$+ \begin{bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \end{bmatrix}$$

An ICA method (JADE) finds uncorrelated, maximally non Gaussian components.
Here, results on 5 W-MAP channels degraded to common resolution.

Foregrounds and how to get rid of them ?



Various **foreground** emissions (both galactic and extra-galactic) pile up in front of the CMB.

But they do so additively !

Even better, most scale rigidly with frequency: each frequency channel sees a different mixture of each astrophysical emission:

$$\mathbf{d} = \mathbf{A}\mathbf{s} + \mathbf{n} \quad \text{(data = mixture x sources + noise)}$$

Such a linear mixture can be inverted . . . if the mixing matrix $\mathbf{A}$ is known. How to find it or do without it ?

1 Trust astrophysics and use parametric models, or

2 Trust your data and the power of statistics.

Diffuse foregrounds vs point sources

Not the same !

You mask resolved point sources.

You do compsep for diffuse foregrounds.

Best linear combination(s!)

The best MSE predictor

Try to predict a vector X based on the observation of a vector Y.

Assume a probablistic relation between X and Y, represented by their joint probability distribution $p(X, Y)$.

Problem: What the best predictor in the MSE, that is, what is the function $f(Y)$ giving the minimum mean squared error:

$$\min_f \mathbb{E}\|X - f(Y)\|^2$$

The solution is the conditional expectation of X given Y

$$f^\star(Y) = \mathbb{E}(X|Y).$$

Often called 'the Wiener filter'.

Proof: $\mathbb{E} = \int p(X, Y) = \int p(X|Y)p(Y)$

The best linear filter

Best (in the MSE sense) linear predictor W of X given Y:

$$\min_W \mathbb{E}\|X - WY\|^2$$

Depends on $R_{xx} = \mathrm{Cov}(X)$ and on $R_{xy} = \mathrm{Cov}(X, Y)$, and only on that:

$$W^\star = R_{xy} R_{yy}^{-1}$$

regardless of the distribution of (X, Y) (finite variance)

For (jointly!) Gaussian vectors X and Y, the Wiener filter boils down to:

$$\mathbb{E}(X|Y) = R_{xy} R_{yy}^{-1} Y$$

This is linear in Y !

Statistical efficiency versus simplicity

For non Gaussian observations, the best processor (in terms of mean squared error) is non linear. **BUT**,

1. In order to implement the best non linear processor on non Gaussian variables, one needs to know or to estimate the non Gaussian part of their distribution.
2. The best non linear filtering may be significantly (or immensely) more difficult to implement.
3. Non-linear filtering may induce non Gaussianities !
4. The characterization and propagation of errors is much harder for non linear processing.
5. The CMB is Gaussian-distributed in a first very good approximation.

A quick look at the Gaussian scalar Wiener filter

Scalar Gaussian signal in uncorrelated Gaussian noise:

$$y = x + n$$

Then $R_{xy} = \sigma_x^2$ and $R_{yy} = \sigma_x^2 + \sigma_n^2$ and we find a simple downweighting:

$$\hat{x} = W_\star y = R_{xy} R_{yy}^{-1} y = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_n^2}\, y = \frac{1}{1 + \mathrm{SNR}^{-1}}\, y \qquad \mathrm{SNR} = \frac{\sigma_x^2}{\sigma_n^2}$$

The relative reconstruction error

$$\frac{\mathbb{E}(\hat{x} - x)^2}{\sigma_x^2} = \cdots = \frac{1}{1 + \mathrm{SNR}} \leq 1$$

If you're smart, you never make more than 100% error. ;-)

1. 'Better safe than sorry' or 'If SNR is bad, don't even try'.
2. Not 'unbiased' (what a poor choice of words!)
3. No information gain (or loss, for that matter).
4. The story becomes interesting only for vector processing.

Wiener filter for stationary processes

Consider a noisy pixelized CMB map: $x_p = s_p + n_p$
where $\mathbb{E}s_p^2 = \sigma^2_{\text{cmb}}$ and $\mathbb{E}s_p^2 = \sigma_n^2$ is the variance of the noise in each pixel.

A pixel-wise Wiener filter produces an estimated CMB: $\widehat{s}_p = x_p \frac{\sigma^2_{CMB}}{\sigma^2_{CMB}+\sigma^2_n}$.

That is excessively boring and useless. Cannot we use the inter-pixel correlation of the CMB which is ignored in the pixel-wise processor? Maybe some kind of local averaging ?

Yes! Do it in harmonic space where the model becomes $x_{\ell m} = s_{\ell m} + n_{\ell m}$ with

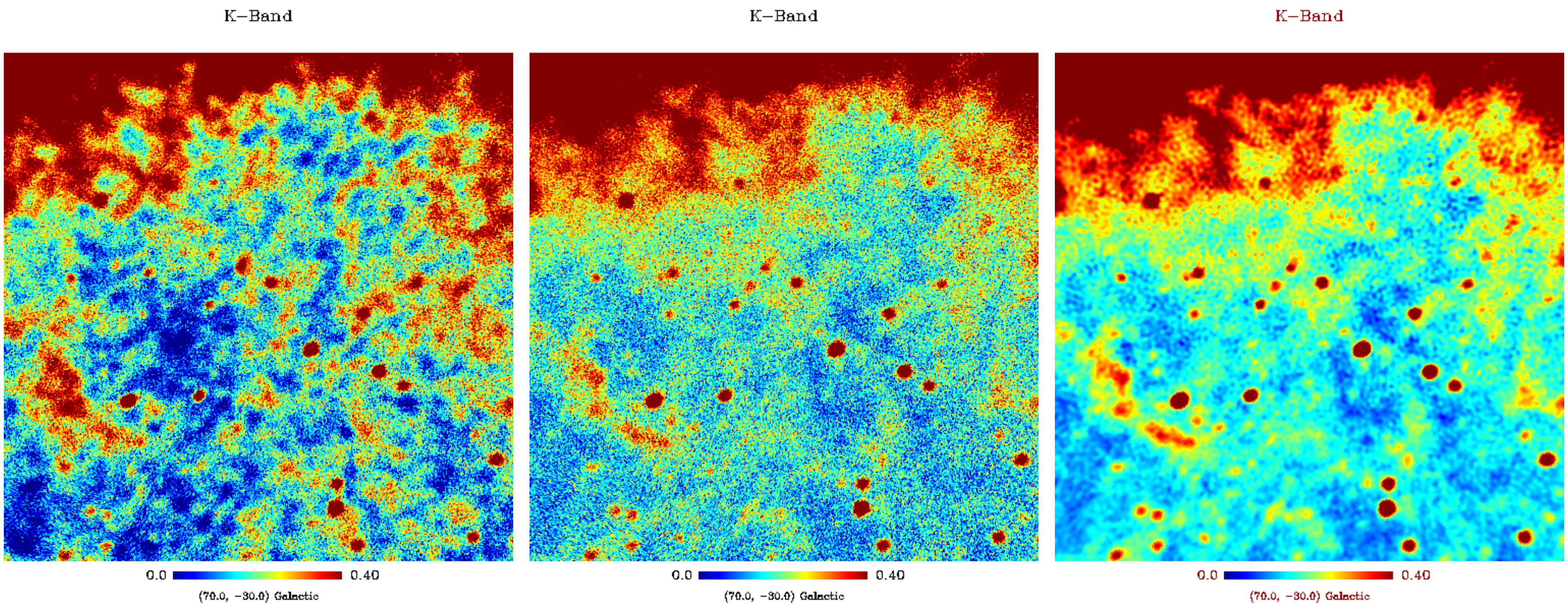
$$\mathbb{E}(s_{\ell m} s_{\ell' m'}) = C(\ell)\ \delta_{\ell\ell'}\,\delta_{mm'} \quad \mathbb{E}(n_{\ell m} n_{\ell' m'}) \approx \Omega\,\sigma_n^2\ \delta_{\ell\ell'}\,\delta_{mm'} \quad \Omega = \frac{4\pi}{N_{\text{pix}}}$$

It exposes the SNR contrast and justifies mode-wise processing, namely:

$$\widehat{s}_{\ell m} = x_{\ell m}\ \frac{C(\ell)}{C(\ell) + \Omega\,\sigma_n^2} \qquad \text{The 'Wiener beam'}$$

That does correspond to smoothing (the MSE-optimal one).

A (double) example from Gosh *et al.*



1) Total emission in WMAP K band
2) After subtracting a Wiener filtered version of (an estimated) CMB map.
3) After applying the Wiener to the previous result.

Does Planck deliver Wiener filtered CMB maps ?
No, we deliver maps with an effective Gaussian beam of 5 arc-minutes FWHM which is simple and similar to it. But feel free to taylor the beam to your needs. . .

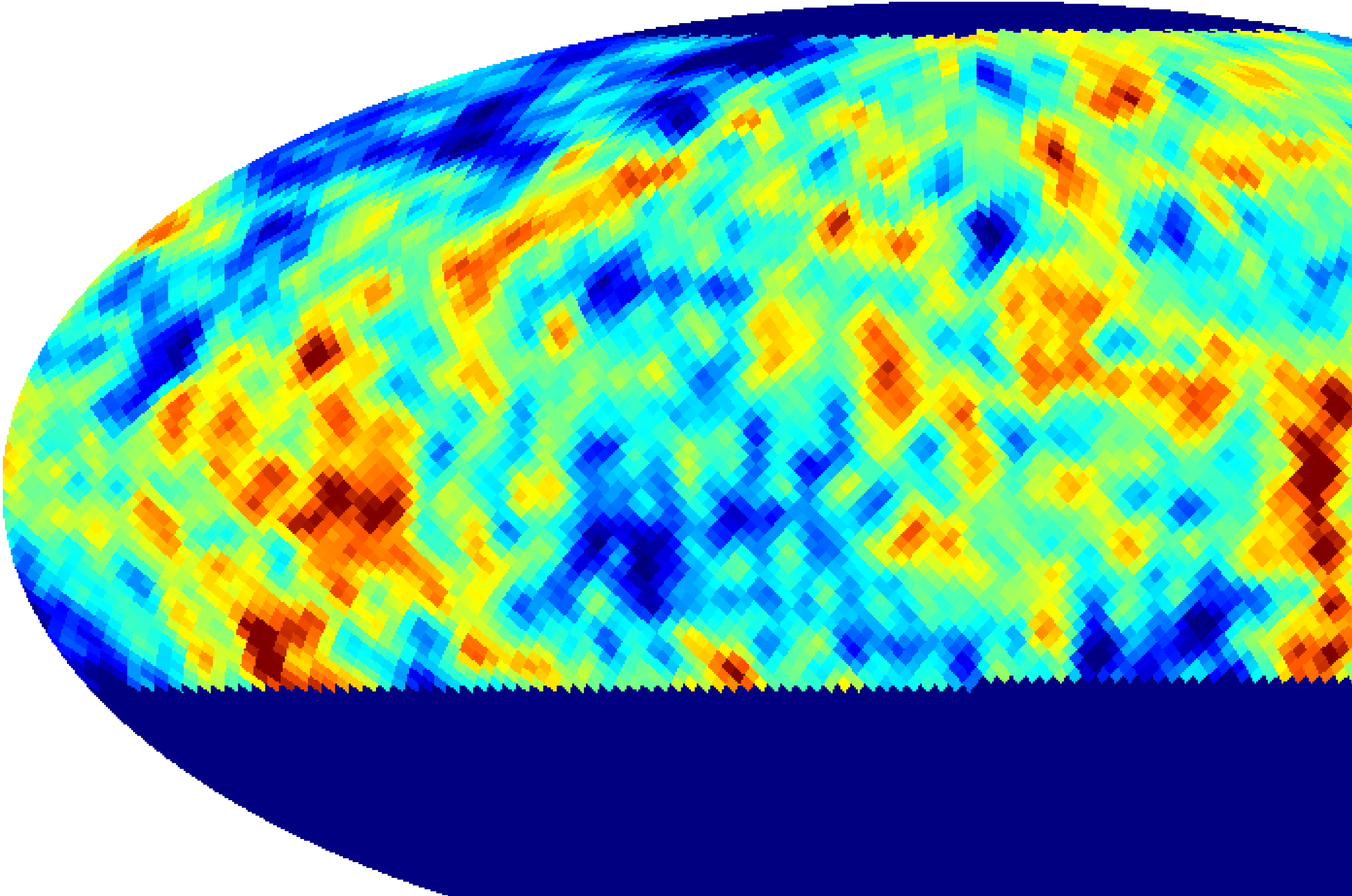
Wiener filter exploting the inter-pixel correlation.

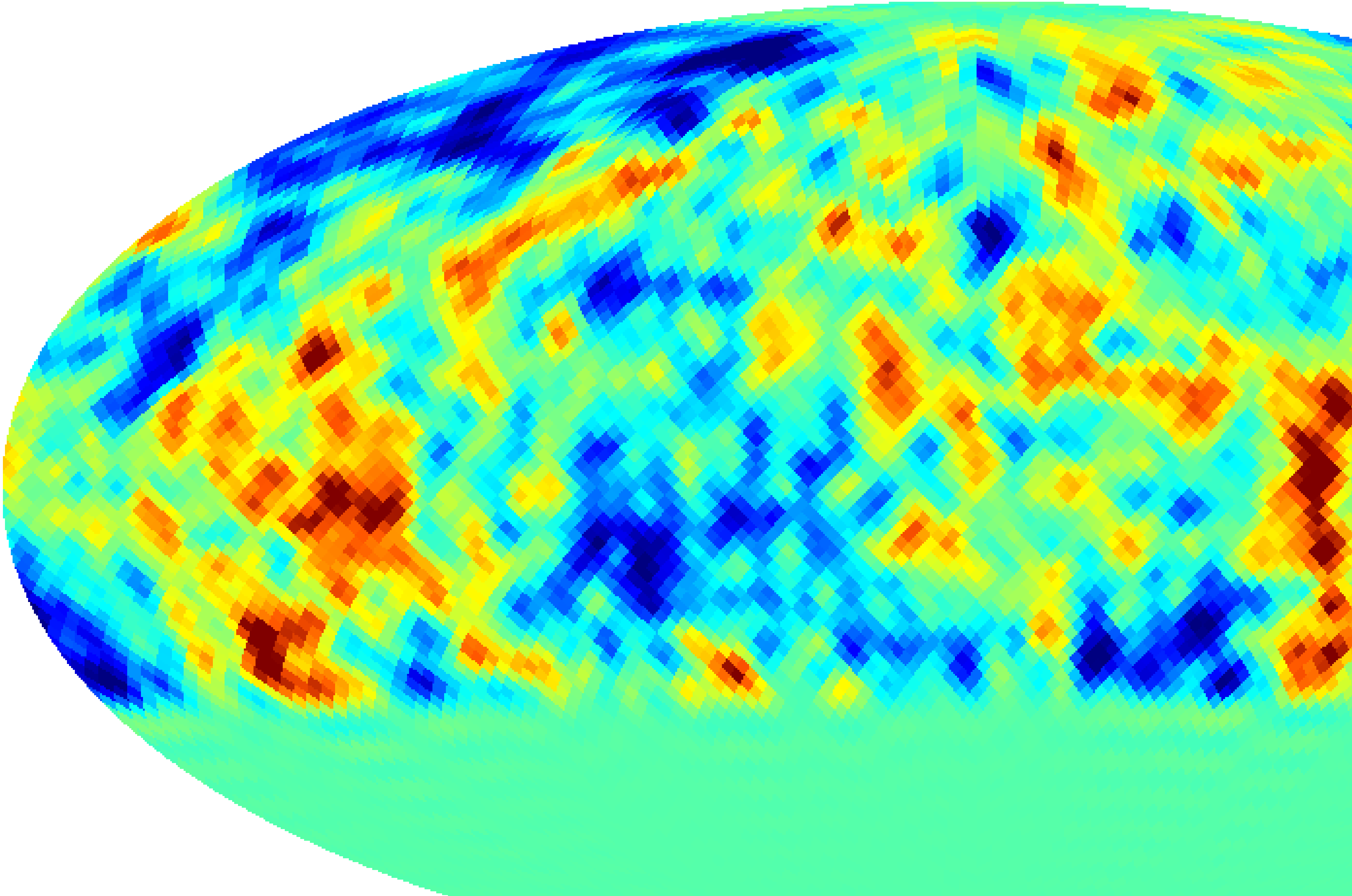
Toward inpainting, that is the computation of a constrained realization.

Consider a spherical map,
modeled as a realization of a stationary Gaussian process,
hence characterized by its angular spectrum,
which encodes the two-point angular correlation.

Task: If only a part of the map is visible (call it Y),
predict the masked part (call it X by its conditional expectation given the visible part.

That is, compute $\mathbb{E}X|Y$.





Wiener filter for noisy mixtures

Things get more interesting with vector observations.
Assume F noisy mixtures of C components (in pixel space, harmonic space,...):

$$x = As + n \qquad \text{with an } F \times C \text{ mixing matrix } A$$

and with $\text{Cov}(s) = \mathbf{S}$ and $\text{Cov}(n) = \mathbf{N}$.

Often (but not always), we consider redundant observations: 'tall' matrix $F \geq C$.

The Gaussian Wiener estimate is

$$\widehat{s} = W_\star x \quad \text{with} \quad W_\star = R_{sx} R_{xx}^{-1}.$$

Now, $R_{sx} = \mathbf{S}A^\dagger$ and $R_{xx} = \text{Cov}(As + n) = A\mathbf{S}A^\dagger + \mathbf{N}$ so

$$W_\star = \mathbf{S}A^\dagger (A\mathbf{S}A^\dagger + \mathbf{N})^{-1}$$

So the best reconstruction of the observations is

$$A\widehat{s} = (A\mathbf{S}A^\dagger)(A\mathbf{S}A^\dagger + \mathbf{N})^{-1} x = \text{Cov(signal)}\ \ \text{Cov(signal + noise)}^{-1}\ x$$

Compare to the scalar case. But what happens at high SNR ?

The alternate form of Wiener and the high SNR limit

Pepole with really tall matrices love the second form of the Wiener filter:

$$W_\star = \mathbf{S}A^\dagger(A\mathbf{S}A^\dagger + \mathbf{N})^{-1} = (A^\dagger\mathbf{N}^{-1}A + \mathbf{S}^{-1})^{-1}A^\dagger\mathbf{N}^{-1}.$$

making it clear that, in the high SNR limit, $A^\dagger\mathbf{N}^{-1}A \ll \mathbf{S}^{-1}$,
the Wiener filter becomes

$$W_\star \to W_\infty = (A^\dagger\mathbf{N}^{-1}A)^{-1}A^\dagger\mathbf{N}^{-1}$$

The global reconstruction of AS is by the filter AW_∞

$$AW_\infty = A(A^\dagger\mathbf{N}^{-1}A)^{-1}A^\dagger\mathbf{N}^{-1}$$

1) AW_∞ does not depend on the signal covariance $\mathbf{S}$ and
2) AW_∞ depends on A only via $\mathsf{Span}(A)$, *i.e.* is invariant under $A \to AT$.
3) It also reads

$$AW_\infty = \mathbf{N}^{\frac{1}{2}}\,\Pi\,\mathbf{N}^{-\frac{1}{2}}$$

where Π is the orthogonal projector onto $\mathsf{Span}(\mathbf{N}^{-\frac{1}{2}}A)$.

Geometric interpretation: see AW_∞ as an oblique projector.
Statistical interpretation: leaves out uncorrelated noise.

What do we get out of the BLUE?

Best linear unbiased estimate (BLUE):

If $x = As + n$, then find matrix W_u such that
$\mathbb{E}\|Wx - s\|^2$ is minimum under the 'unbiasedness' constraint, that is, $WA = I$.

That is a pure, no compromise, noise-fighting device.

One easily finds:

$$W_\mathsf{u} = (A^\dagger \mathbf{X}^{-1} A)^{-1} A^\dagger \mathbf{X}^{-1} \qquad \text{with } \mathbf{X} = \mathrm{Cov}(x).$$

Notes:

1) W_u needs only $\mathbf{X}$ which can be replaced by a plain sample estimate!

2) AW_u is an oblique projector, just as $AW_\infty = A(A^\dagger \mathbf{N}^{-1} A)^{-1} A^\dagger \mathbf{N}^{-1}$.

High SNR Wiener and the BLUE

For $x = As + n$, with $\mathbf{X} = \mathrm{Cov}(x) = A\mathbf{S}A^\dagger + \mathbf{N}$, two forms of Wiener

$$W_\star = \mathbf{S}A^\dagger\mathbf{X}^{-1} = (A^\dagger\mathbf{N}^{-1}A + \mathbf{S}^{-1})^{-1}A^\dagger\mathbf{N}^{-1}. \tag{1}$$

and two limits: BLUE W_u (enforces unbiasedness) and high SNR Wiener W_∞:

$$W_\mathsf{u} = (A^\dagger\mathbf{X}^{-1}A)^{-1}A^\dagger\mathbf{X}^{-1} \qquad W_\infty = (A^\dagger\mathbf{N}^{-1}A)^{-1}A^\dagger\mathbf{N}^{-1}.$$

Both clearly are left inverses of A since $W_\mathsf{u}A = W_\infty A = I_C$.

Because of eq. (1), AW_u and AW_∞ must be identical projectors. See why?

Therefore

$$W_\mathsf{u} = W_\infty$$

that is, the Wiener filter converges to the BLUE at high SNR.

Signal subspace and oblique projection

We saw $P = AW_\infty = AW_{\mathsf{u}}$ and $P^2 = P = \mathbf{N}^{\frac{1}{2}}\,\Pi\,\mathbf{N}^{-\frac{1}{2}}$

- Whiten the noise
- Project orthogonally onto the signal subspace
- Unwhiten

Final take:
Among all projectors onto $\mathsf{Span}(A)$, we look for the one yielding an uncorrelated decomposition

$$x = x_s + x_n = Px + (I - P)x \qquad \mathsf{Cov}(x_s, x_n) = 0$$

Now

$$\mathsf{Cov}(x_s, x_n) = (I-P)\mathsf{Cov}(x)P^\dagger = (I-P)\mathbf{X}P^\dagger = (I-P)(A\mathbf{S}A^\dagger + \mathbf{N})P^\dagger = (I-P)\mathbf{N}P^\dagger$$

so P is the projector onto $\mathsf{Span}(A)$ orthogonally wrt to the norm defined by $\mathbf{N}$ or $\mathbf{X}$.

Wiener and the BLUE

For $x = As + n$, we can connect the 'true Wiener' and the BLUE:

$$W_\star = \mathbf{S}A^\dagger(A\mathbf{S}A^\dagger + \mathbf{N})^{-1} = \mathbf{S}A^\dagger\mathbf{X}^{-1} = (A^\dagger\mathbf{N}^{-1}A + \mathbf{S}^{-1})^{-1}A^\dagger\mathbf{N}^{-1}$$
$$W_\mathsf{u} = W_\infty = (A^\dagger\mathbf{N}^{-1}A)^{-1}A^\dagger\mathbf{N}^{-1} = (A^\dagger\mathbf{X}^{-1}A)^{-1}A^\dagger\mathbf{X}^{-1}$$

because they share the same row space $\mathsf{Span}(\mathbf{N}^{-1}A) = \mathsf{Span}(\mathbf{X}^{-1}A)$.

Then, let's rephrase in terms of BLUE output. Define

$$s_\mathsf{u} = W_\mathsf{u}x = s + n_\mathsf{u} \quad \text{with} \quad \mathbf{N}_u = \mathsf{Cov}(n_\mathsf{u}) = (A^\dagger\mathbf{N}^{-1}A)^{-1}.$$

We find the $C \times F$ Wiener W is the concatenation of
1) $C \times F$ compression by W_u without 'bias' or information loss followed by
2) $C \times C$ reversible reshaping (biasing) by $1/(1 + \mathsf{SNR}^{-1})$:

$$W_\star = \underbrace{(I + \mathbf{N}_\mathsf{u}\mathbf{S}^{-1})^{-1}}_{\text{reshape}} \underbrace{W_\mathsf{u}}_{\text{project}}$$

The ILC (in excruciating detail)

Combining all 9 Planck channels, non parametrically: the ILC

1/ Linear combination:

Stack the 9 Planck channels into a data 9×1 vector $\mathbf{x} = [x_{30}, x_{44}, \ldots, x_{545}, x_{857}]^\dagger$ and estimate the CMB signal $s(p)$ in pixel p by weighting the inputs:

$$\widehat{s}(p) = \mathbf{w}^\dagger \mathbf{x}(p) \qquad p = 1, \ldots, N_{\text{pix}}$$

2/ Known calibration: At frequency ν, the CMB signal $s(p)$ has amplitude a_ν and is independently contaminated by $f_\nu(p)$:

$$x_\nu(p) = a_\nu\, s(p) \ + \ f_\nu(p) \qquad \text{or} \qquad \mathbf{x}(p) = \mathbf{a}\, s(p) + \mathbf{f}(p)$$

Then, the best ($\min_w \langle (s - \mathbf{w}^\dagger \mathbf{d})^2 \rangle_p$) unbiased ($\mathbf{w}^\dagger \mathbf{a} = 1$) estimator is:

$$\mathbf{w} = \frac{\widehat{\mathbf{C}}^{-1}\, \mathbf{a}}{\mathbf{a}^\dagger \widehat{\mathbf{C}}^{-1}\, \mathbf{a}} \qquad \text{with} \qquad \widehat{\mathbf{C}} = \langle \mathbf{x}\mathbf{x}^\dagger \rangle_p, \quad \text{the \underline{sample} covariance matrix.}$$

That is known as ILC (Internal Linear Combination) in CMB circles, as MVBF (Minimum Variance Beam Former) in array processing, the BLUE elsewhere. . .

The ILC and its Wiener version

1. For $x = As + n$, recall the BLUE estimator:

$$\widehat{s} = W_{\mathsf{u}} x = (A^\dagger \mathbf{X}^{-1} A)^{-1} A^\dagger \mathbf{X}^{-1} x \qquad \text{with } \mathbf{X} = \mathrm{Cov}(x).$$

2. Assume we look for a single component: the CMB. Matrix A reduces to a single column vector $A = [\mathbf{a}]$ (and $\mathbf{a} = \mathbf{1}$ in CMB units).
   The BLUE in any domain reduces to

$$\widehat{s} = W_{\mathsf{u}} x = \frac{\mathbf{a}^\dagger \mathbf{X}^{-1} x}{\mathbf{a}^\dagger \mathbf{X}^{-1} \mathbf{a}} \qquad \text{'Internal' linear combination.}$$

3. Optionnally Wienerize the ILC map *i.e.* impose the Wiener beam:

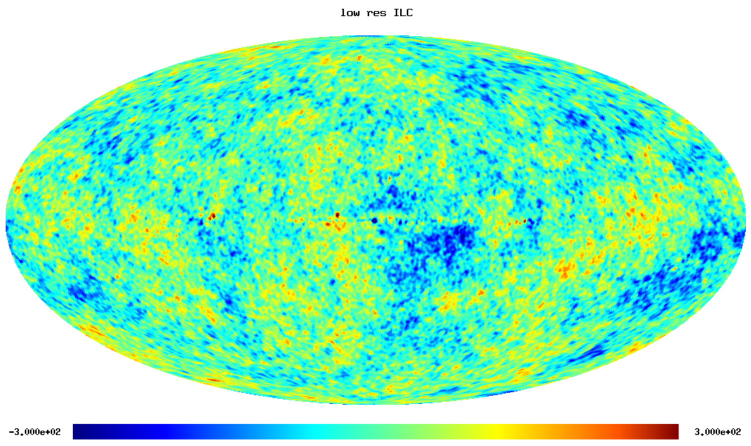
$$\widehat{\widehat{s}}_{\ell m} = \widehat{s}_{\ell m} \frac{C_\ell}{C_\ell + N_\ell} \qquad \text{Reversible smoothing minimizing overall MSE}$$

4. Estimation of missing quantities.
   - BLUE: $\mathbf{X}$ estimated by a sample average $\widehat{\mathbf{X}}$ in the appropriate domain.
   - Wiener: What about $C_\ell, \mathbf{N}_\ell$? Estimation using Planck jackknives.

Is pixel-based ILC good enough for Planck data ?

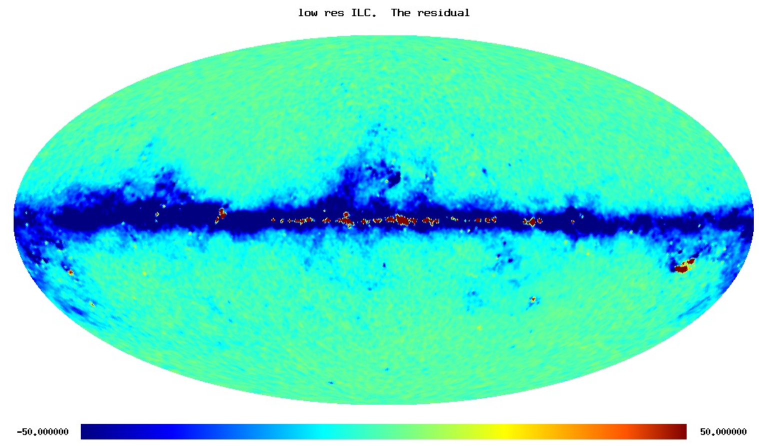
ILC looks good: linear, unbiased, min. MSE, very blind, very few assumptions: knowing $\mathbf{a}$ (calibration) and the CMB uncorrelated from the rest (very true).

However, a simulation result shows poor quality:



$\longleftarrow$ ILC map on a $\pm 300\mu K$ color scale

Error on a $\pm 50\mu K$ color scale $\longrightarrow$



Two things, at least, need fixing:

- localization (in real space? harmonic space? both ?) and
- chance correlations.

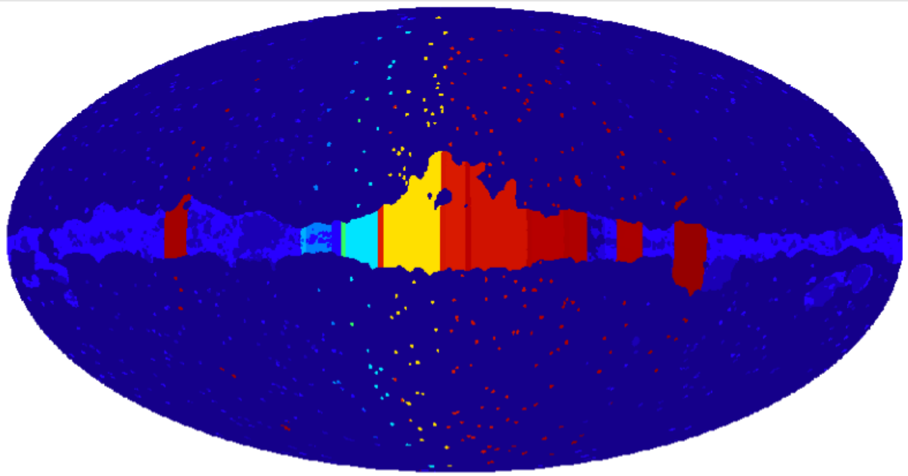
Spatial localization

- Pixel-based ILC

$$\widehat{s}(\theta,\phi) = \frac{\mathbf{a}^\dagger \widehat{\mathbf{X}}^{-1} x(\theta,\phi)}{\mathbf{a}^\dagger \widehat{\mathbf{X}}^{-1} \mathbf{a}} \qquad \widehat{\mathbf{X}} = \frac{1}{N_{\text{pix}}} \sum_{\text{pix}} x(\theta_p,\phi_p) x(\theta_p,\phi_p)^\dagger$$

with the data covariance matrix estimated globally over the sky.

- Localize by working on sky domains and stitching the results, a la W-MAP

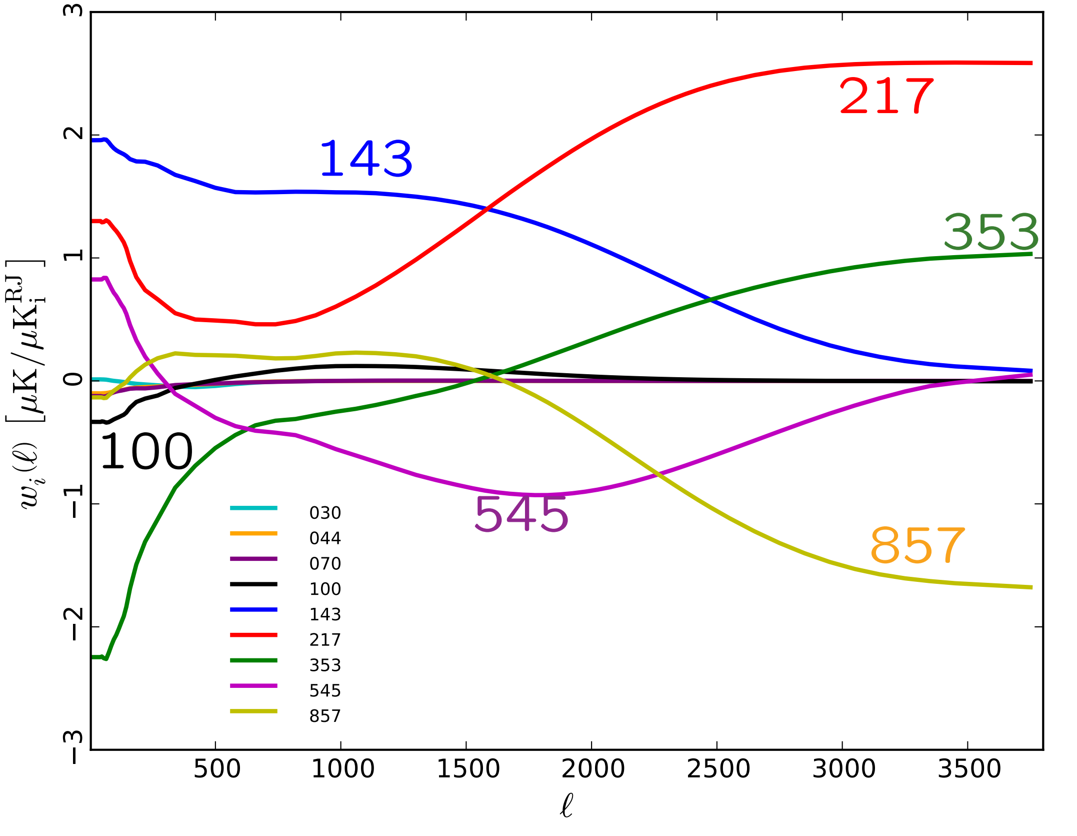


Multipole localization : Harmonic ILC

- Localize in harmonic space (Tegmark)

$$\widehat{s}_{\ell m} = \mathbf{w}_\ell \, \mathbf{x}_{\ell m} \qquad \mathbf{w}_\ell = (\mathbf{a}^\dagger \widehat{\mathbf{X}}_\ell^{-1} \mathbf{a})^{-1} \, \mathbf{a}^\dagger \widehat{\mathbf{X}}_\ell^{-1} \qquad \widehat{\mathbf{X}}_\ell = \text{Smooth}\left[\frac{1}{2\ell+1} \sum_m \mathbf{x}_{\ell m} \mathbf{x}_{\ell m}^\dagger\right]$$

- The SMICA CMB map is synthesized from spherical harmonic coefficients $\widehat{s}_{\ell m}$ obtained by harmonic ILC.
- At high ℓ, spectral covariance matrices well estimated by (smoothed) $\widehat{\mathbf{X}}_\ell$.
- At lower ℓ, chance correlation must be fought using a model.



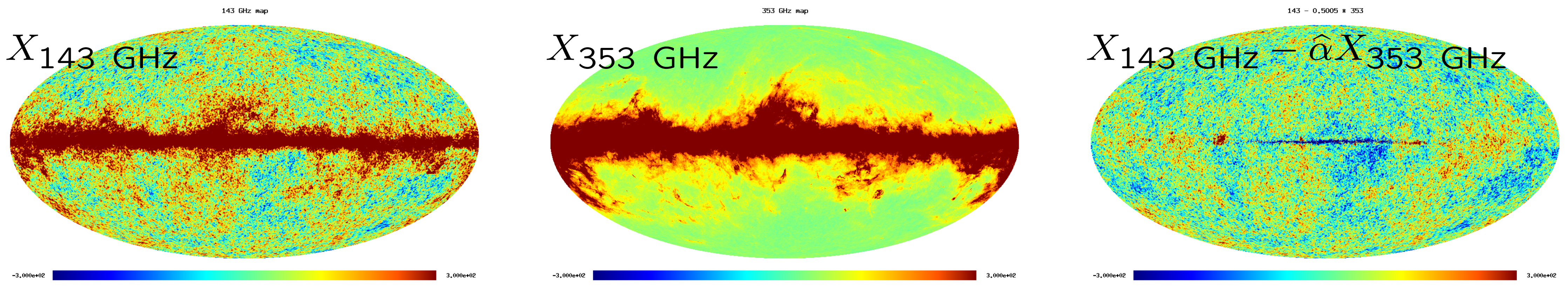
- Bonus: trivial control of an effective beam.

Double direction-multipole localization : Needlet ILC

- Localize in both space and multipole using wavelets.

$\rightarrow$ Needlet ILC 'optimally' exposes the local SNR condition.

Simple CMB cleaning by "template removal"



Assume that the 353 GHz channel sees only dust emission
and that the 143 GHz channel sees CMB plus a rescaled dust pattern:

$$X_{143} = \text{CMB} + \alpha\ X_{353}$$

Find α by cross-correlation and get a clean (?) CMB map as

$$\widehat{\text{CMB}} = X_{143} - \frac{\langle X_{143} X_{353} \rangle}{\langle X_{353} X_{353} \rangle}\ X_{353} \quad \text{where } \langle \cdot \rangle \text{ denotes a pixel average}$$

The result (top right) does not look bad, but it is !

Note: By construction $\langle \widehat{\text{CMB}}\ X_{353} \rangle = 0$.

Template removal and the ILC

Single template removal works if one channel contains the signal of interest plus unspecified contamination while the second channel does not carry at all the signal of interest.

So it's like

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \mathbf{a} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Now

If $\widehat{\mathbf{X}} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$ then $\hat{s} = \dfrac{\mathbf{a}^\dagger \widehat{\mathbf{X}}^{-1} \mathbf{x}}{\mathbf{a}^\dagger \widehat{\mathbf{X}}^{-1} \mathbf{a}} = \cdots = x_1 - \dfrac{r_{12}}{r_{22}}\, x_2$

so, yes, template fitting is a special case of BLUE/ILC.

Nulling out components with known emissivity

Want to get unbiased CMB but block a component
with a known emissivity $\mathbf{b}$, like SZ emission ?

The ILC is easily generalized to minimize the mean squared error
under both constraints $\mathbf{w}^\dagger\mathbf{a} = 1$ and $\mathbf{w}^\dagger\mathbf{b} = 0$:
just take $\mathbf{w}$ as the first column of

$$\mathbf{C}^{-1}[\mathbf{a}\ \mathbf{b}]\,([\mathbf{a}\ \mathbf{b}]^\dagger\mathbf{C}^{-1}[\mathbf{a}\ \mathbf{b}])^{-1}$$

Or apply the BLUE with $A = [\mathbf{a}\ \mathbf{b}]$ to get the same result (remember $W_{\mathsf{u}}A = I$).

Fighting chance correlations

Beware notations $x \to d$ and $w \to w^\dagger$

Another derivation of the BLUE/ILC, with sample statistics.

Estimate the CMB as a linear combination of the data $\widehat{s}(\vec{\eta}) = \mathbf{w}^\dagger \mathbf{d}(\vec{\eta})$.

Assume $\mathbf{d}(\vec{\eta}) = \mathbf{a}\ s(\vec{\eta}) + \mathbf{c}(\vec{\eta})$ and $\mathbf{w}^\dagger \mathbf{a} = 1$. Then $\mathbf{w}^\dagger \mathbf{d}(\vec{\eta}) = \widehat{s} = s + \mathbf{w}^\dagger \mathbf{c}(\vec{\eta})$,

and the pixel-averaged output power is

$$\langle \widehat{s}^2 \rangle_P = \langle (\mathbf{w}^\dagger \mathbf{d})^2 \rangle_P \ = \ \langle s^2 \rangle_P + 2\, \mathbf{w}^\dagger \langle \mathbf{c}\, s \rangle_P + \langle (\mathbf{w}^\dagger \mathbf{c})^2 \rangle_P$$

where $\langle \cdot \rangle_P$ defines a pixel-based average over the sky:

$$\langle f \rangle_P \ \stackrel{\text{def}}{=}\ \frac{1}{N_{\text{pix}}} \sum_p f(\vec{\eta}_p) \ \propto \ \iint_{S^2} f(\vec{\eta}).$$

Thus, if $\langle \mathbf{c}\, s \rangle_P = 0$, then $\arg\min \langle (\mathbf{w}^\dagger \mathbf{d})^2 \rangle_P = \arg\min \langle (\mathbf{w}^\dagger \mathbf{c})^2 \rangle_P$

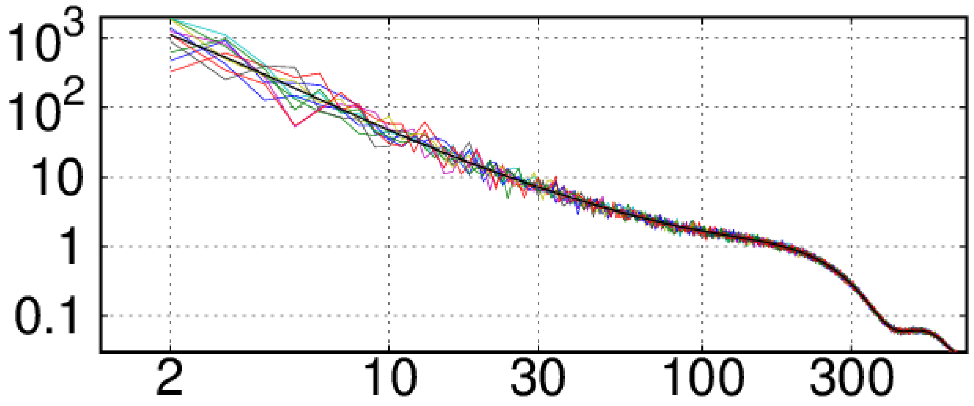
$$\widehat{\mathbf{w}}_P = (\mathbf{a}^\dagger \widehat{\mathbf{C}}_P^{-1} \mathbf{a})^{-1}\ \widehat{\mathbf{C}}_P^{-1} \mathbf{a} \text{ where } \widehat{\mathbf{C}}_P \stackrel{\text{def}}{=} \langle \mathbf{d}(\vec{\eta}) \mathbf{d}(\vec{\eta})^\dagger \rangle_P$$

so ILC, BLUE, do the right thing.

But $\boxed{\langle \mathbf{c}\, s \rangle_P \neq \mathsf{E}(\mathbf{c} s) = 0 \text{ *i.e.* 'chance correlation' between CMB and contaminants}}$.

Having 50 million pixels to average over, we should not worry. Or should we ?

Correlations



Empirical angular spectra in units of $(\mu K)^2$ of 10 simulated CMB maps drawn independently with a typical angular spectrum $C(\ell)$.

Sharp decrease with $\ell \rightarrow$ the CMB map is dominated by its large angular scales.

We do not have 50 millions independent pixels$\rightarrow$ Bad averaging in pixel space !

But we do have 50 millions independent harmonic coefficients (well, not really. . .).

We can look again at the chance correlation term:

$$\langle s(\vec{\eta})\mathbf{c}(\vec{\eta})\rangle_P = \iint_{S^2} s(\vec{\eta})\mathbf{c}(\vec{\eta}) = \sum_\ell \sum_m s_{\ell m}\mathbf{c}_{\ell m}$$

and see that only a few independent terms are contributing to the correlation.

Cleaning coherent contamination.

- Maximal coherent contamination:

$$\mathbf{d}(\vec{\eta}) = \mathbf{A}\mathbf{s}(\vec{\eta}) = [\mathbf{a}\ \mathbf{H}] \begin{bmatrix} s(\vec{\eta}) \\ \mathbf{f}(\vec{\eta}) \end{bmatrix} \qquad \text{or} \qquad \mathbf{D} = \mathbf{A}\mathbf{S} = [\mathbf{a}\ \mathbf{H}] \begin{bmatrix} S \\ \mathbf{F} \end{bmatrix} \quad \text{(matrix form)}$$

where matrix $\mathbf{A}$ is $n \times n$ and has vector $\mathbf{a}$ in its first column, where $\mathbf{H}$ is an (unknown) set of $n-1$ columns and where $\mathbf{f}(\vec{\eta})$ is a vector of $n-1$ 'foregrounds'

- The contamination is fully coherent here *i.e.* it can be completely nulled out : this is a noise-free model, only representative at large scales.

- Note: if chance correlations were magically suppressed, that is, if

$$\widehat{\mathbf{C}}_P = \langle \mathbf{d}\mathbf{d}^\dagger \rangle_P \qquad \xrightarrow{\text{magic}} \qquad \widehat{\mathbf{C}}_{\text{fake}} = \mathbf{a}\mathbf{a}^\dagger \langle s^2 \rangle_P + \mathbf{H} \langle \mathbf{f}\mathbf{f}^\dagger \rangle_P \mathbf{H}^\dagger,$$

the ILC filter $\widehat{\mathbf{w}}_{\text{fake}} = (\mathbf{a}^\dagger \widehat{\mathbf{C}}_{\text{fake}}^{-1} \mathbf{a})^{-1} \widehat{\mathbf{C}}_{\text{fake}}^{-1} \mathbf{a}$ would give perfect foreground rejection.

Preprocessing or re-parameterization.

In the mixing matrix $\mathbf{A} = [\mathbf{a}\ \mathbf{H}]$, matrix $\mathbf{H}$ is unknown.
But replacing $\mathbf{H}$ with a fixed arbitrary matrix $\mathbf{T}$ and inverting yields

$$\left[\mathbf{a}\,\mathbf{T}\right]^{-1}\mathbf{D} \stackrel{\text{def}}{=} \begin{bmatrix} Y \\ \mathbf{G} \end{bmatrix} \quad \text{and} \quad \left[\mathbf{a}\,\mathbf{T}\right]^{-1}\mathbf{A} \stackrel{\text{def}}{=} \begin{bmatrix} 1 & \mathbf{v}^\dagger\mathbf{K} \\ \mathbf{0}_{(n-1)\times 1} & \mathbf{K} \end{bmatrix}$$

for unknown $\mathbf{K}$ and $\mathbf{v}$. The model $\mathbf{D} = \mathbf{AS} = [\mathbf{a}\ \mathbf{H}] \begin{bmatrix} S \\ \mathbf{F} \end{bmatrix}$ is turned into

$$\begin{bmatrix} Y \\ \mathbf{G} \end{bmatrix} = \begin{bmatrix} S + \mathbf{v}^\dagger\mathbf{KF} \\ \mathbf{KF} \end{bmatrix}$$

In other words, pre-processing* the data $\mathbf{D}$ yields Y and $\mathbf{G}$ such that

$$Y = S + \mathbf{v}^\dagger\mathbf{G}$$

meaning that the contamination of S in Y by $\mathbf{G}$ is available deterministically.

*That pre-processing is only introduced here as a 'mathematical device' but a similar idea is actually implemented in the SEVEM algorithm for CMB extraction.

A likelihood $p_{\mathbf{D}}(\mathbf{D}|\mathbf{A})$.

Two ingredients for the likelihood of $\mathbf{D} = \mathbf{A}\begin{bmatrix} S \\ \mathbf{F} \end{bmatrix}$:

$$\underbrace{p_{\mathbf{S}}(\mathbf{S}) = P_S(S) \cdot P_{\mathbf{F}}(\mathbf{F})}_{\text{1) Statistical independence}} \quad \text{and} \quad \underbrace{\begin{bmatrix} \mathbf{a}\ \mathbf{T} \end{bmatrix}^{-1} \mathbf{D} \stackrel{\text{def}}{=} \begin{bmatrix} Y \\ \mathbf{G} \end{bmatrix} = \begin{bmatrix} S + \mathbf{v}^\dagger \mathbf{K}\mathbf{F} \\ \mathbf{K}\mathbf{F} \end{bmatrix}}_{\text{2) (Pre-processed) mixing model}}.$$

After just a little bit of work:

$$p_{\mathbf{D}}(\mathbf{D}|\mathbf{A}) = \underbrace{p_S(Y - \mathbf{v}^\dagger \mathbf{G})}_{\text{CMB}} \cdot \underbrace{p_{\mathbf{F}}(\mathbf{K}^{-1}\mathbf{G}) |\det(\mathbf{K})|^{-N}}_{\text{foregrounds}} \cdot \underbrace{|\det \begin{bmatrix} \mathbf{a}\ \mathbf{T} \end{bmatrix}|^{-N}}_{\text{Constant}}.$$

Thus the maximum likelihood solution for the signal of interest is

$$\widehat{S}^{\mathsf{ML}} = Y - \hat{\mathbf{v}}^\dagger \mathbf{G} \qquad \hat{\mathbf{v}} = \arg\max_{\mathbf{v}}\ p_S(Y - \mathbf{v}^\dagger \mathbf{G})$$

and this value depends neither on $\mathbf{K}$ nor on the contamination model $p_{\mathbf{F}}(\cdot)$.

Maximum likelihood solutions for some models p_S.

The ML solution $\hat{\mathbf{v}} = \arg\max_{\mathbf{v}} p_S(Y - \mathbf{v}^\dagger \mathbf{G})$ can be found in some simple models.

- If S is modelled as N i.i.d. Gaussian samples, $\log P_S(S) = -\frac{1}{2}\sum_p s_p^2/\sigma^2 + \text{cst}$, and then $\widehat{S}^{\text{ML}}$ exactly as produced by pixel-based ILC.

- If S is taken non Gaussian i.id.: $\log P_S(S) = \sum_p \log p(s_p)$, then

$$\sum_p \psi(\hat{s}_p^{\text{ML}})\ \mathbf{g}_p = 0 \text{ with } \psi(s) \stackrel{\text{def}}{=} -p'(s)/p(s) \qquad \text{(compare to regular ICA)}$$

- If S is a Gaussian stationary spherical field with spectrum $C(\ell)$, then

$$\log P_S(S) = -\frac{1}{2}\sum_\ell \sum_m s_{\ell m}^2 / C(\ell) + \text{cst}$$

takes again the form of an ILC filter except that the covariance matrix $\widehat{\mathbf{C}}_H$ is computed in the harmonic domain with an inverse-variance weighting.

$$\widehat{\mathbf{w}}_H = \frac{\widehat{\mathbf{C}}_H^{-1}\mathbf{a}}{\mathbf{a}^\dagger \widehat{\mathbf{C}}_H^{-1}\mathbf{a}} \qquad \text{with} \qquad \widehat{\mathbf{C}}_H = \sum_{\ell \le \ell_{\text{max}}} \sum_{m=-\ell}^{m=+\ell} \mathbf{d}_{\ell m}\mathbf{d}_{\ell m}^\dagger / C_\ell.$$

Pixel space versus harmonic space.

Compare the two versions of the ILC filter obtained in pixel space $\widehat{\mathbf{w}}_P$ and in harmonic space $\widehat{\mathbf{w}}_H$ by comparing the empirical covariance matrices $\widehat{\mathbf{C}}_P$ and $\widehat{\mathbf{C}}_H$.

$$\widehat{\mathbf{C}}_P \stackrel{\text{def}}{=} \langle \mathbf{d}\mathbf{d}^\dagger \rangle_P \propto \sum_\ell \sum_{m=-\ell}^{m=+\ell} \mathbf{d}_{\ell m} \mathbf{d}_{\ell m}^\dagger \qquad \widehat{\mathbf{C}}_H \stackrel{\text{def}}{=} \sum_\ell \sum_{m=-\ell}^{m=+\ell} \mathbf{d}_{\ell m} \mathbf{d}_{\ell m}^\dagger / C_\ell.$$

For a flat spectrum $C(\ell) = \text{cst}$, one has $\widehat{\mathbf{w}}_P = \widehat{\mathbf{w}}_H$ but flat spectrum means uncorrelated pixels, so (again) pixel-based ILC = Gaussian i.i.d. Max. Likelihood.

Otherwise, for a correlated field, the two solutions $\widehat{\mathbf{w}}_P$ and $\widehat{\mathbf{w}}_H$ are different.

For a rapidly decreasing $C(\ell)$ like the CMB, they are very different because only a few terms then dominate the sum $\widehat{\mathbf{C}}_P$. The maximum likelihood principle, in its asymptotic wisdom, shows clearly the cure: equalise the summands, the $1/C(\ell)$ factor making the CMB power of each term in $\mathbf{C}_H$ identical at all angular scales.

Notes:

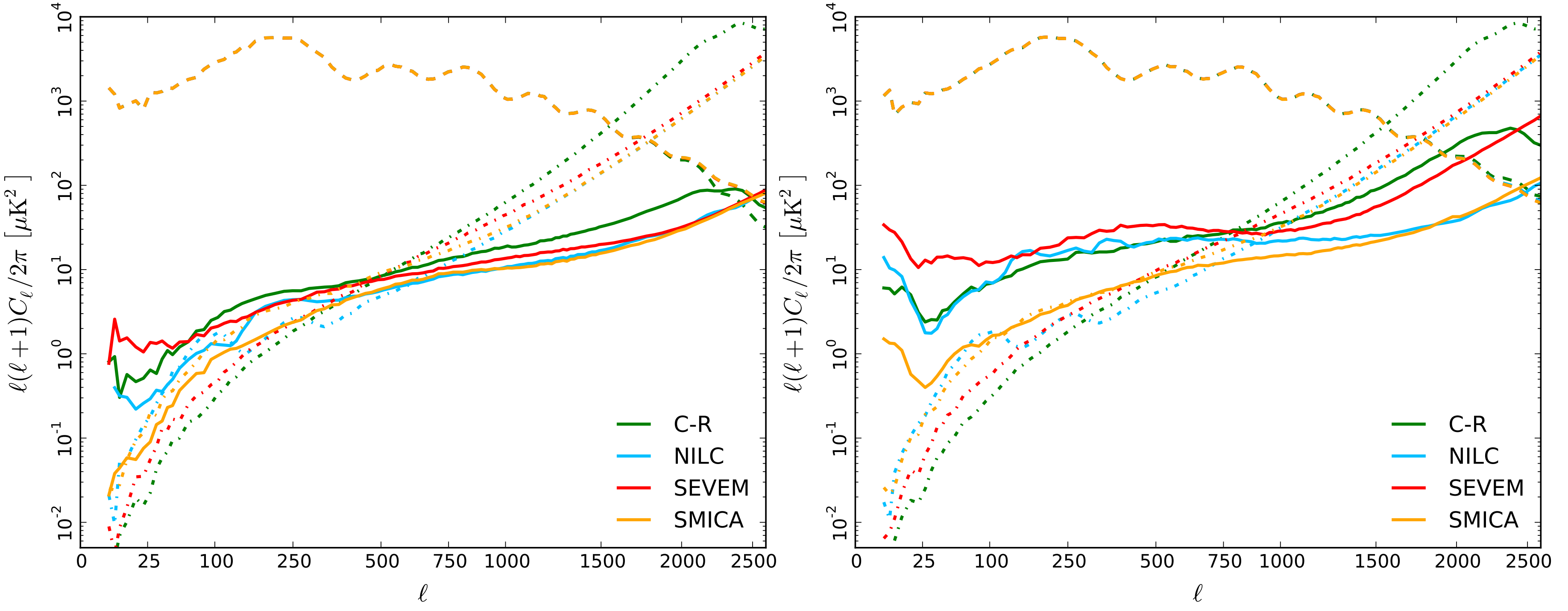
a) Needlet-ILC 2013 uses $\mathbf{w}_H$ in its first scale.

b) It's only a weight: using $w_\ell = \ell^2$ instead of $w_\ell = 1/C(\ell)$ works almost as well.

One more thing

The optimal weight vector $\mathbf{w}$ is the normal vector to the foreground subspace.

How is this is relevant to Planck work ?



Angular power spectra with FFP6 simulations.

CMB (dashed); residual noise (dot-dashed); residual foregrounds (solid).

Left (resp. right) evaluated over a clean (resp. less clean) region of the sky.

Two messages:

- At fine scales: the weights should really be allowed to depend on ℓ.
- At large scales: spatial localisation by wavelet does not seem to help much.

My interpretation: closer to the Galactic plane, the same ennemy, only stronger.

Conclusions for part 2

- Likelihood analysis reveals that for the CMB in the high SNR limit,

\+ ILC suboptimal (also true for any highly correlated signal).

\+ the ‘exact’ MLE is easily implemented in the harmonic domain,

\+ it is an harmonic ILC with a proper weighting of the harmonic coefficients,

\+ it is used in the first scale for the needlet-ILC (NILC) method,

\+ it does not depend <u>at all</u> on the foreground model.

- In my view, it validates using a Gaussian stationary likelihood
and makes sparsity arguments questionable.

- A full solution (accouting for beams, noise,. . .) at all scales
implemented in the SMICA CMB product.

Appendix 1

Number of effective samples

Loss of efficiency in pixel space.

Invariance and chance correlation.

Recall the model after pre-processing: the CMB S contaminated by a linear combination of visible foregrounds $\mathbf{G}$ with unknown weights $\mathbf{w}_\star$:

$$Y = S + \mathbf{v}_\star^\dagger \mathbf{G}$$

- The signal estimated by the ILC is

$$\widehat{S} = Y - \widehat{\mathbf{v}}^\dagger \mathbf{G} \text{ with } \widehat{\mathbf{v}}^\dagger = \widehat{C}_{y\mathbf{g}} \widehat{C}_{\mathbf{gg}}^{-1}$$

and the estimation error is found to be

$$\widehat{S} - S = -\widehat{C}_{s\mathbf{g}} \widehat{C}_{\mathbf{gg}}^{-1} \mathbf{G}.$$

- The error does <u>not</u> depend on $\mathbf{v}_\star$!

We can assume $\mathbf{v}_\star = 0$: no contamination, you always clean too much.

- The error is strictly invariant with respect to any invertible mixing or rescaling of the foregrounds.

- The error strictly vanishes with the chance correlation $\widehat{C}_{s\mathbf{g}}$

Error covariance.

Recall the (invariant) estimation error :

$$\widehat{S} - S = -\widehat{C}_{s\mathbf{g}} \widehat{C}_{\mathbf{gg}}^{-1} \mathbf{G}.$$

We evaluate $\text{Cov}\left(\widehat{C}_{s\mathbf{g}} \widehat{C}_{\mathbf{gg}}^{-1}\right)$ where the Cov is over the CMB for fixed foregrounds.

When statistics are computed in harmonic space with weight w_ℓ, one gets

$$\text{Cov}(\widehat{C}_{s\mathbf{g}}) = \text{Cov}\left(\sum_\ell w_\ell \sum_m s_{\ell m} \mathbf{g}_{\ell m}\right) = \sum_\ell w_\ell^2 \; C_\ell \; \mathbf{\Gamma}_\ell \text{ with } \mathbf{\Gamma}_\ell = \sum_m \mathbf{g}_{\ell m} \mathbf{g}_{\ell m}^\dagger$$

so

$$\text{Cov}(\widehat{C}_{s\mathbf{g}} \widehat{C}_{\mathbf{gg}}^{-1}) = \left[\sum_\ell w_\ell \mathbf{\Gamma}_\ell\right]^{-1} \left[\sum_\ell w_\ell^2 \; C_\ell \; \mathbf{\Gamma}_\ell\right] \left[\sum_\ell w_\ell \mathbf{\Gamma}_\ell\right]^{-1}.$$

Taking $w_\ell = 1$ also yields the error for statistics computed in pixel space.

Optimal weighting

We found

$$\mathrm{Cov}(\hat{C}_{s\mathbf{g}}\hat{C}_{\mathbf{gg}}^{-1}) = \left[\sum_\ell w_\ell \mathbf{\Gamma}_\ell\right]^{-1} \left[\sum_\ell w_\ell^2 \; C_\ell \; \mathbf{\Gamma}_\ell\right] \left[\sum_\ell w_\ell \mathbf{\Gamma}_\ell\right]^{-1} = \mathbf{\Sigma}(w_\ell)$$

Theorem:

$$\mathbf{\Sigma}(w_\ell) \geq \mathbf{\Sigma}(C_\ell^{-1}) = \left[\sum_\ell C_\ell^{-1} \mathbf{\Gamma}_\ell\right]^{-1}$$

The optimal weighting is $w_\ell = C_\ell^{-1}$ and does not depend on the foregrounds.

Relative gain of optimal harmonic weighting

Assume that the spectral inter-frequency correlation of the foregrounds does not change with the angular scale i.e. $\mathbf{\Gamma}_\ell = g_\ell \, \mathbf{\Gamma}$.

Then, the error covariance factors as

$$\mathrm{Cov}(\widehat{C}_{s\mathbf{g}}\widehat{C}_{\mathbf{gg}}^{-1}) = \frac{\sum_\ell w_\ell^2 \; g_\ell \; C_\ell}{(\sum_\ell w_\ell \; g_\ell \;)^2} \cdot \mathbf{\Gamma}^{-1}$$

Now, we can make a simple statement regarding the gain brought in by optimal harmonic weighting $w_\ell = C_\ell^{-1}$ with respect to flat pixel averaging $w_\ell = 1$.

It is like have α times more effective samples since the pixel-based variance is larger by a factor α

$$\alpha = \alpha(C_\ell, g_\ell) = \frac{(\sum_\ell \; g_\ell \; C_\ell)\,(\sum_\ell \; g_\ell / C_\ell)}{(\sum_\ell g_\ell \;)^2} \geq 1$$

That loss also is the ratio between of the (g_ℓ-weighted) arithmetic mean to the harmonic mean of the CMB spectrum over the multipole range.

Some orders of magnitude

- Multipole range $2 \leq \ell \leq 25$

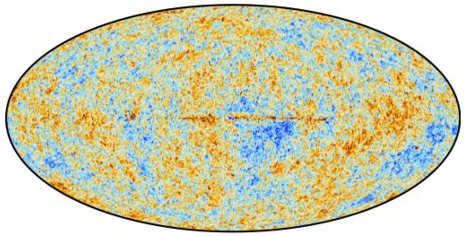
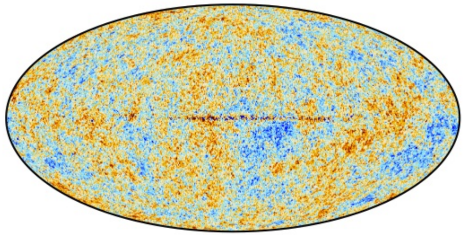
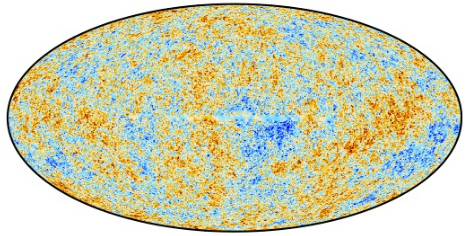
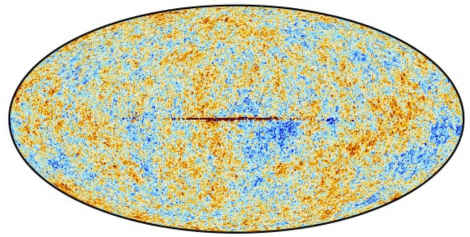
- Galactic foregrounds with $g_\ell = (2\ell + 1)\,\ell^{-2.4}$

Variance decreases by a factor 6.60 with respect to pixel average if optimal weighting $w_\ell = 1/C_\ell$ is used.

Variance decreases by a factor 6.55 with respect to pixel average if suboptimal weighting $w_\ell = \ell^2$ is used.

Filtering

Four CMB cleaning methods used in Planck 2013

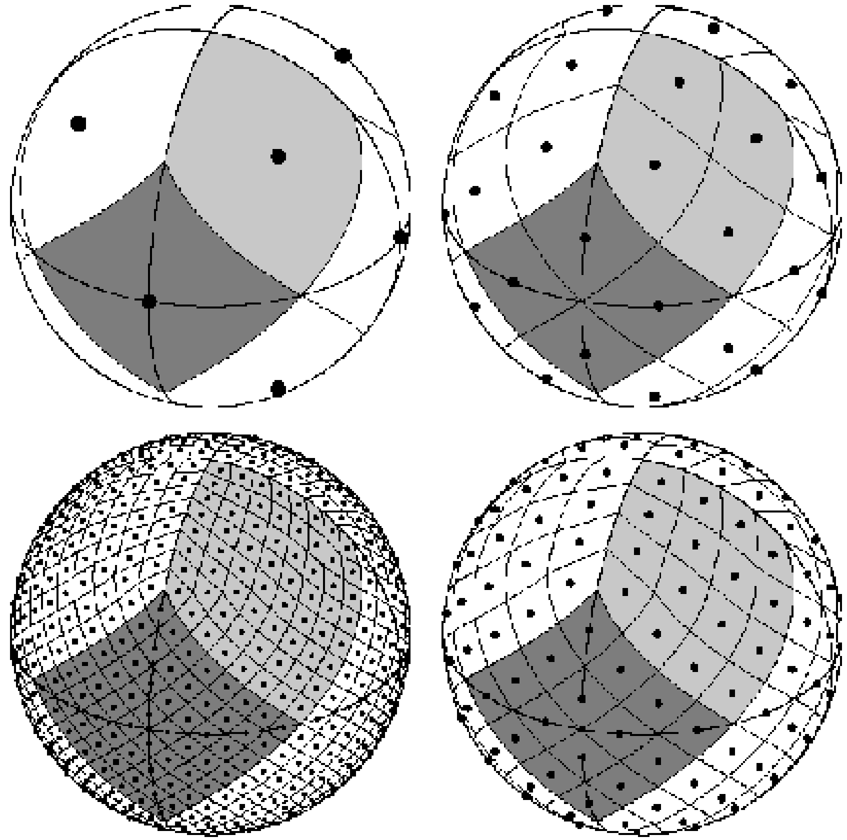
			
NILC	SEVEM	SMICA	C-R
$\ell_{\mathrm{SNR}=1} = 1790$	$\ell_{\mathrm{SNR}=1} = 1790$	$\ell_{\mathrm{SNR}=1} = 1790$	$\ell_{\mathrm{SNR}=1} = 1550$
Wavelet space	Pixel then harmonic	Harmonic space	Pixel space
non-parametric	non-parametric	semi-parametric	parametric

They differ, in particular, by their filtering schemes.

But they all use HEALPix...

HEALPix (Gorski, Hivon *et al.*)

1. Hierarchical structure.
Essential for large data bases, neighborhood search, multi-resolution analysis,. . .
2. Equal pixel area.
Preserves white noise, among other things.
3. Iso-latitude distribution.
Recall $\mathcal{Y}_{\ell m}(\theta,\phi) = P_{\ell m}(\cos\theta)\ e^{im\phi}$.
θ direction: Associated Legendre functions are evaluated via slow recursions.
ϕ direction: FFT possible.



The HEALPix grid at resolution r has $N_{\text{pix}} = 12N_{\text{side}}^2 = 12 \cdot 2^{2r}$ pixels.
It offers synthesis and (approximate) analysis up to $\ell_{\max} \approx 3 \times N_{\text{side}}$:

$$X(\theta_p,\phi_p) = \sum_{\ell\le\ell_{\max}} \sum_{|m|\le\ell} a_{\ell m}\mathcal{Y}_{\ell m}(\theta_p,\phi_p) \qquad \frac{4\pi}{N_{\text{pix}}}\sum_p X(\theta_p,\phi_p)\mathcal{Y}_{\ell m}(\theta_p,\phi_p) \approx a_{\ell m}$$

Jargon: WMAP delivers at $N_{\text{side}} = 512$, Planck at $N_{\text{side}} = 2048$.

Spherical basis and frames

- Unlocalized **orthogonal** harmonic decompostion. Analysis and synthesis

$$X(\xi_p) = \sum_{\ell m} x_{\ell m} \mathcal{Y}_{\ell m}(\xi_p) \quad \leftrightarrow \quad x_{\ell m} = \frac{4\pi}{N_{\text{pix}}} \sum_p X(\xi_p) \mathcal{Y}_{\ell m}(\xi_p)$$

- A simple but very redundant scheme: undecimated wavelets (JL Starck).

- A tighter frame: needlets (Petrushev).

Needlet analysis. Bandpass and synthesize (at appropriate resolution):

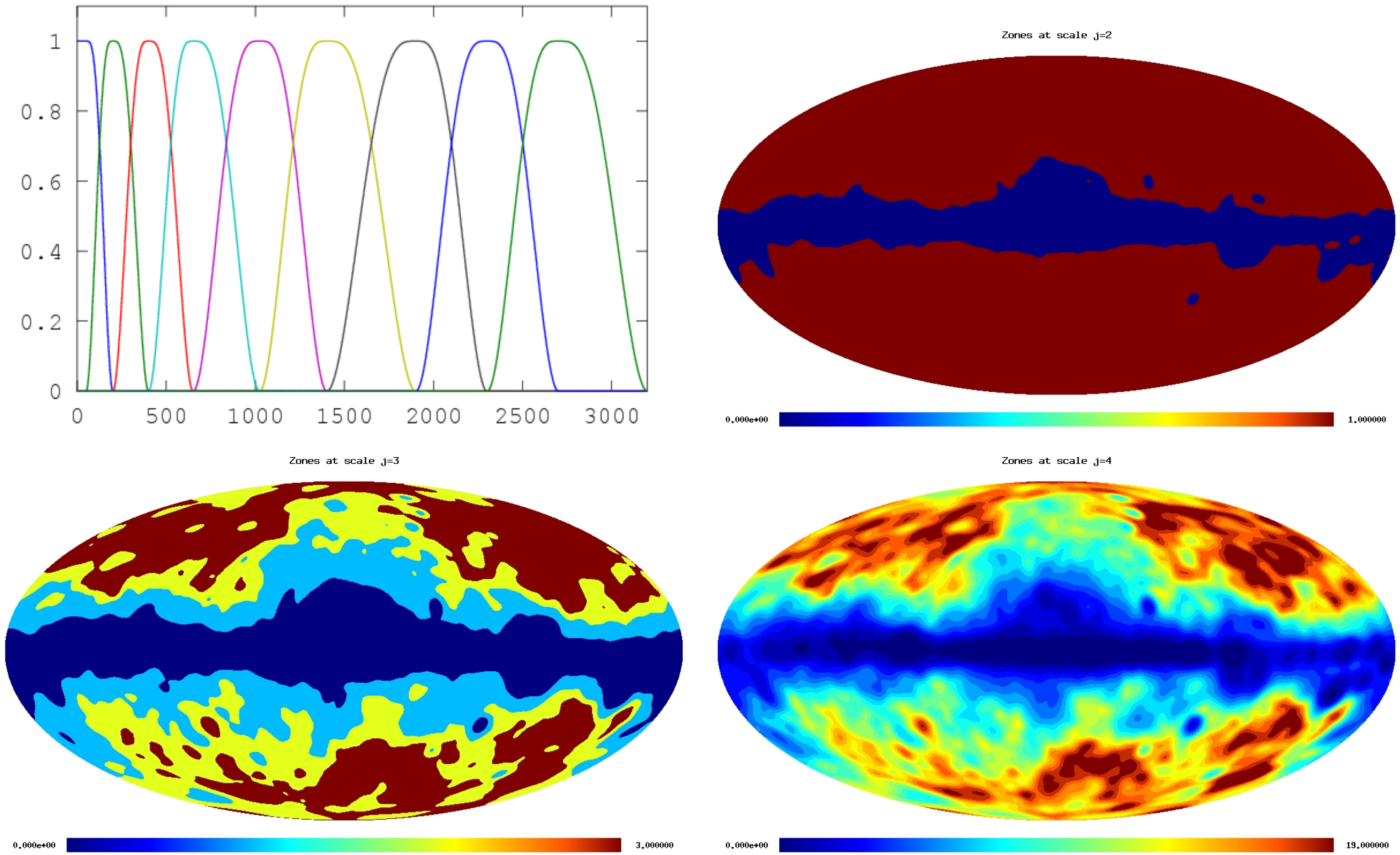
$$x_{j,\ell m} = x_{\ell m} w_j^a(\ell) \qquad x_{jk} = \sum_{\ell m} x_{j,\ell m} \mathcal{Y}_{\ell m}(\xi_k) \qquad \xi_k \in H_j \qquad \text{for } 1 \le j \le J$$

Needlet synthesis: analyze, bandpass, synthesize

$$\tilde{x}_{j,\ell m} = \frac{4\pi}{N_k} \sum_{\xi_k \in H_j} x_{jk} \mathcal{Y}_{\ell m}(\xi_k) \qquad \tilde{x}_{\ell m} = \sum_j w_j^s(\ell) \tilde{x}_{j,\ell m} \qquad \tilde{X}(\xi) = \sum_{\ell m} \tilde{x}_{\ell m} \mathcal{Y}_{\ell m}(\xi)$$

Perfect reconstruction: impose $\sum_j w_j^s(\ell) w_j^a(\ell) = 1$

Needlet ILC in a nutshell



$$\hat{s}_{jk} = \mathbf{w}_{jK}^{\dagger} \mathbf{x}_{jk} \qquad \mathbf{w}_{jK} = \mathrm{ILC}\Big(\langle \mathbf{x}_{jk} \mathbf{x}_{jk}^{\dagger} \rangle_K \Big)$$

Fighting chance correlations with spectral matching

SMICA: Linear filtering in harmonic space

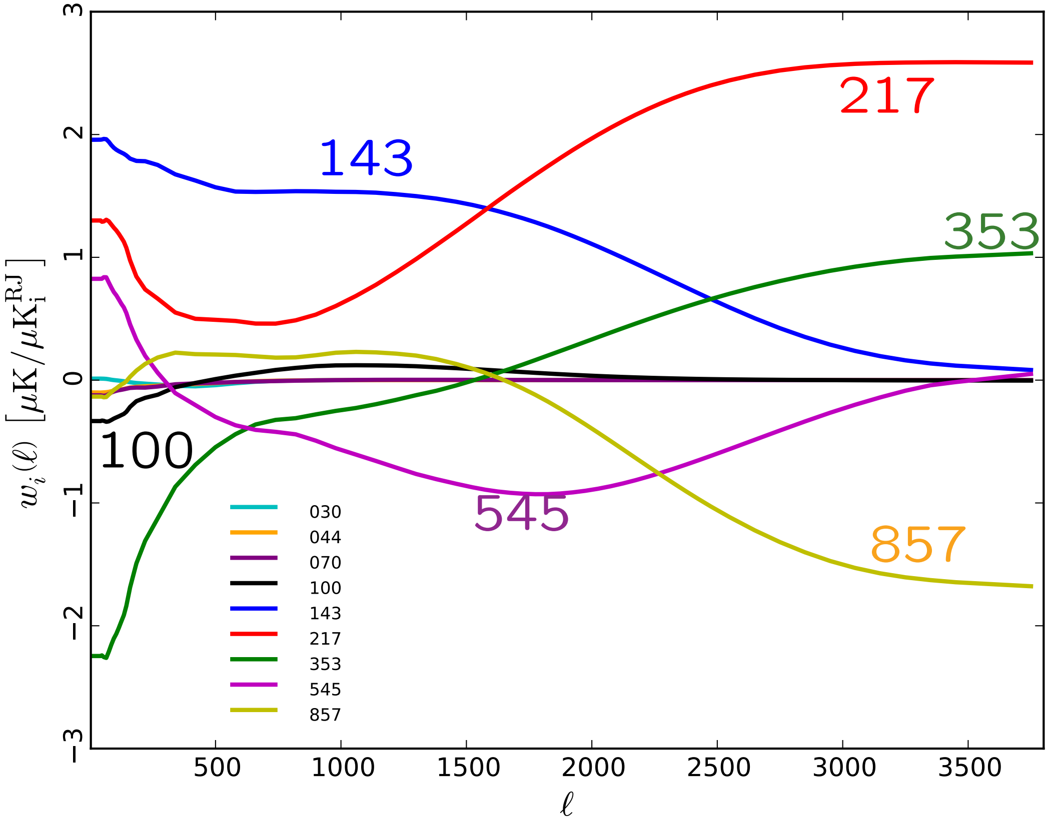
Since resolution, noise and foregrounds vary (wildly) in power over channels and angular frequency, the combining weights should depend on ℓ.

SMICA CMB map synthesized from spherical harmonic coefficients $\widehat{s}_{\ell m}$:

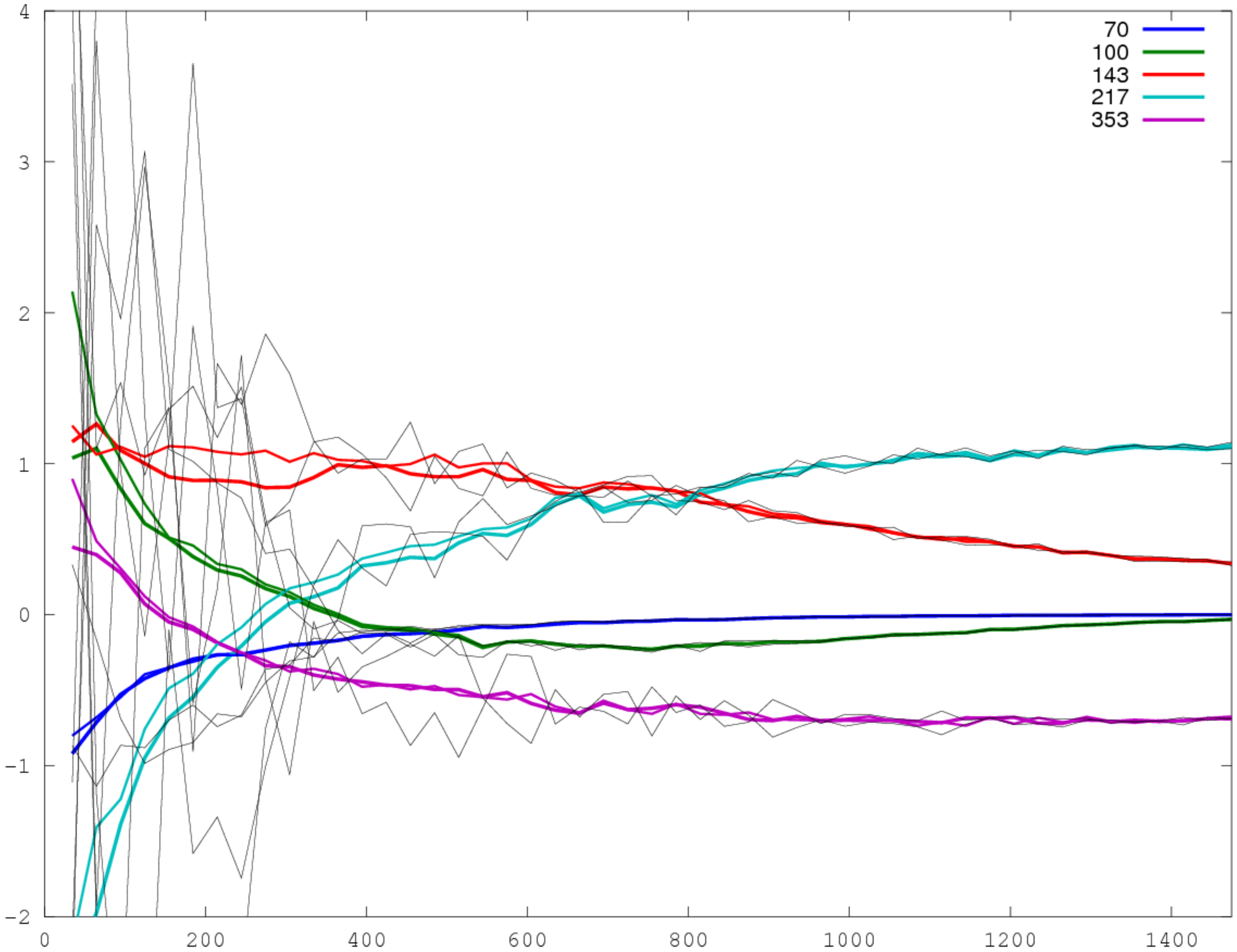
$$\widehat{s}_{\ell m} = (\mathbf{a}^\dagger \tilde{\mathbf{C}}_\ell^{-1} \mathbf{a})^{-1} (\mathbf{a}^\dagger \tilde{\mathbf{C}}_\ell^{-1} \mathbf{d}_{\ell m})$$

with $\tilde{\mathbf{C}}_\ell$ an estimate of $\mathbf{C}_\ell = \mathrm{Cov}(\mathbf{d}_{\ell m})$.

- At high ℓ, spectral covariance matrices well estimated by (smoothed) $\widehat{\mathbf{C}}_\ell$.
- At lower ℓ, chance correlation must be fought using a model.



ILC coefficients: raw (thin lines) and via SMICA modelling (thick lines)



Likelihood for a Gaussian isotropic spherical field.

- The (real) spherical harmonics are a complete orthonormal basis on the sphere:

$$X(\vec{\eta}) = \sum_{\ell \geq 0} \sum_{m=-\ell}^{m=\ell} x_{\ell m} Y_{\ell m}(\vec{\eta}) \qquad \longleftrightarrow \qquad x_{\ell m} = \frac{1}{4\pi} \iint_{S^2} X(\vec{\eta}) \; Y_{\ell m}(\vec{\eta})$$

- The harmonic coefficients $x_{\ell m}$ of an isotropic random field are uncorrelated:

$$\mathsf{E}\left(x_{\ell m} x_{\ell' m'}\right) = C_\ell \; \delta_{ll'} \, \delta_{mm'}$$

+ The angular spectrum $\{C_\ell \stackrel{\text{def}}{=} \mathsf{Var}(x_{\ell m})\}_{\ell \geq 0}$ has a natural sample estimate

$$\widehat{C}_\ell \stackrel{\text{def}}{=} \frac{1}{2\ell + 1} \sum_{m=-\ell}^{m=+\ell} x_{\ell m}^2$$

- . . . which is a sufficient statistic for a Gaussian field since

$$-2 \log p(X) = -2 \log p(X | C_\ell) = \sum_\ell (2\ell + 1) \left[\frac{\widehat{C}_\ell}{C_\ell} + \log C_\ell \right] \; + \; \text{cst}$$

- The likelihood reads like a spectral matching criterion:

$$p(X | C_\ell) \propto \exp -\frac{1}{2} \sum_\ell \; (2\ell + 1) \; k(\widehat{C}_\ell / C_\ell) \qquad k(u) = u - \log(u) - 1$$

The joint likelihood for N Gaussian isotropic maps

If $\mathbf{d}_{\ell m}$ is the $N \times 1$ vector of harmonic coefficients of N spherical maps modelled as Gaussian isotropic, ...

then their joint distribution depends only on spectral $N \times N$ covariance matrices

$$\mathbf{C}_\ell \stackrel{\text{def}}{=} \operatorname{Cov} \mathbf{d}_{\ell m} = \mathbb{E} \mathbf{d}_{\ell m} \mathbf{d}_{\ell m}^\dagger$$

containing angular auto-spectra (on diagonal) and cross-spectra (off diagonal).

A sufficient statistic for the joint likelihood of the N maps is the set

$$\widehat{\mathbf{C}}_\ell \stackrel{\text{def}}{=} \frac{1}{2\ell+1} \sum_m \mathbf{d}_{\ell m} \mathbf{d}_{\ell m}^\dagger \quad \ell = 0, 1, 2, \cdots$$ of sample spectral covariance matrices

and the joint likelihood again is a spectral matching criterion since

$$p(X|C_\ell) \propto \exp -\frac{1}{2} \sum_\ell (2\ell+1)\ \mathcal{K}(\widehat{\mathbf{C}}_\ell, \mathbf{C}_\ell)$$

where $\mathcal{K}(\cdot, \cdot)$ is the Kullbak-Leibler divergence between two positive matrices

$$\mathcal{K}(\mathbf{C}_a, \mathbf{C}_b) \stackrel{\text{def}}{=} \operatorname{trace}(\mathbf{C}_a \mathbf{C}_b^{-1}) - \log\det(\mathbf{C}_a \mathbf{C}_b^{-1}) - N$$

Foregrounds, physical components and the mixing matrix

- Mixing matrix. The 9 Planck channels as noisy linear mixtures of components:

$$\mathbf{d}_{\ell m} = \mathbf{A}(\theta)\ \mathbf{s}_{\ell m}\ +\ \mathbf{n}_{\ell m}$$

- Some models for the mixing matrix $\mathbf{A} = \mathbf{A}(\theta)$:

Model type	Mixing matrix	parameters θ
physical, fixed	$\mathbf{A} = [\mathbf{a}_{\mathsf{cmb}}\ \mathbf{a}_{\mathsf{dust}}\ \mathbf{a}_{\mathsf{CO}}\ \mathbf{a}_{\mathsf{LF}}]$	$\theta = [\ ]$
physical, parametric	$\mathbf{A} = [\mathbf{a}_{\mathsf{cmb}}\ \mathbf{a}_{\mathsf{dust}}(T)\ \mathbf{a}_{\mathsf{CO}}\ \mathbf{a}_{\mathsf{LF}}(\beta)\ ]$	$\theta = (T, \beta)$
non-parametric ($\sim$ ILC)	$\mathbf{A} = [\mathbf{a}_{\mathsf{cmb}}\ \mathbf{F}]$ (a square matrix)	$\theta = \mathbf{F}$
semi-parametric, SMICA	$\mathbf{A} = \mathbf{A}$ (any tall matrix)	$\theta = \mathbf{A}$

SMICA semi-parametric model

- SMICA models the 9 Planck channels as noisy linear mixtures of CMB and 6 "foregrounds":

$$\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ \vdots \\ d_9 \end{bmatrix} = \begin{bmatrix} a_1 & F_{11} & \dots & F_{16} \\ a_2 & F_{21} & \dots & F_{26} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ a_9 & F_{91} & \dots & F_{96} \end{bmatrix} \times \begin{bmatrix} s \\ f_1 \\ \vdots \\ f_6 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ \vdots \\ n_9 \end{bmatrix} \quad \text{or} \quad \mathbf{d}_{\ell m} = [\,\mathbf{a} \mid \mathbf{F}\,] \begin{bmatrix} s_{\ell m} \\ \mathbf{f}_{\ell m} \end{bmatrix} + \mathbf{n}_{\ell m}$$

- SMICA only assumes decorrelation between foregrounds and CMB.

The foregrounds must have 6 (say) dimensions but are otherwise completely unconstrained: they may have any spectrum, any color, any correlation...

So the data model is very blind: all non-zero parameters are free !

$$\mathrm{Cov}(\mathbf{d}_{\ell m}) = [\,\mathbf{a} \mid \mathbf{F}\,] \begin{bmatrix} C_\ell^{\mathrm{cmb}} & 0 \\ 0 & \mathbf{P}_\ell \end{bmatrix} [\,\mathbf{a} \mid \mathbf{F}\,]^\dagger + \begin{bmatrix} \sigma_{1\ell}^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_{9\ell}^2 \end{bmatrix} = \mathbf{C}_\ell(\mathbf{a}, C_\ell^{\mathrm{cmb}}, \mathbf{F}, \mathbf{P}_\ell, \sigma_{i\ell}^2).$$

- Blind identifiability: can it be done? Maths say: yes! (with diversity...)

- Fit by spectral matching $\min_\theta \sum_\ell (2\ell + 1)\ \mathcal{K}\Big(\widehat{\mathbf{C}}_\ell,\ \mathbf{C}_\ell(\theta)\Big)$ = maximum likelihood.

- Only $\mathrm{Span}(\mathbf{F})$, the foreground subspace, is needed to suppress the foregrounds.

It is collectively determined by all the multipoles involved in the fit.

Why spatial localization may not be critical

Two arguments for spatial localization:
a) The (relative) strengths of contaminants
change with angular scale and position on the sky.
b) For some astrophysical components, emmission laws
change (slightly) over the sky, (*e.g.* varying Galactic dust temperature).

Qualitative answers:

a) It's the subspace that counts !

b) Sky-varying emissivity can be accounted for
- locally by letting $\mathbf{A}$ depend on the pixel: $\mathbf{A}(\theta_{\mathsf{pix}})$ (Commander) or
- globally by adding columns to $\mathbf{A}$ (SMICA).

For instance, a sky-varying low-frequency emission $\mathbf{a}_{\mathsf{LF}}(\theta_{\mathsf{pix}})$ could be approximatively represented by two fixed columns over the whole sky: $[\ \mathbf{a}_{\mathsf{LF}}(\langle\theta\rangle),\ d\mathbf{a}_{\mathsf{LF}}/d\theta(\langle\theta\rangle)\ ]$

SMICA vs "basic" ICA

SMICA differs from "basic" ICA in several aspects:

- No use of non Gaussianity.

  Use spectral diversity instead as the source of contrast.

- Explicit handling of the noise

- One big multi-dimensional component to capture all contaminants.

  In other words, non-CMB components can have arbitrary correlation.

  This is no problem so long as one does not care
  about the unmixing of the other astrophysical emissions.

Conclusions