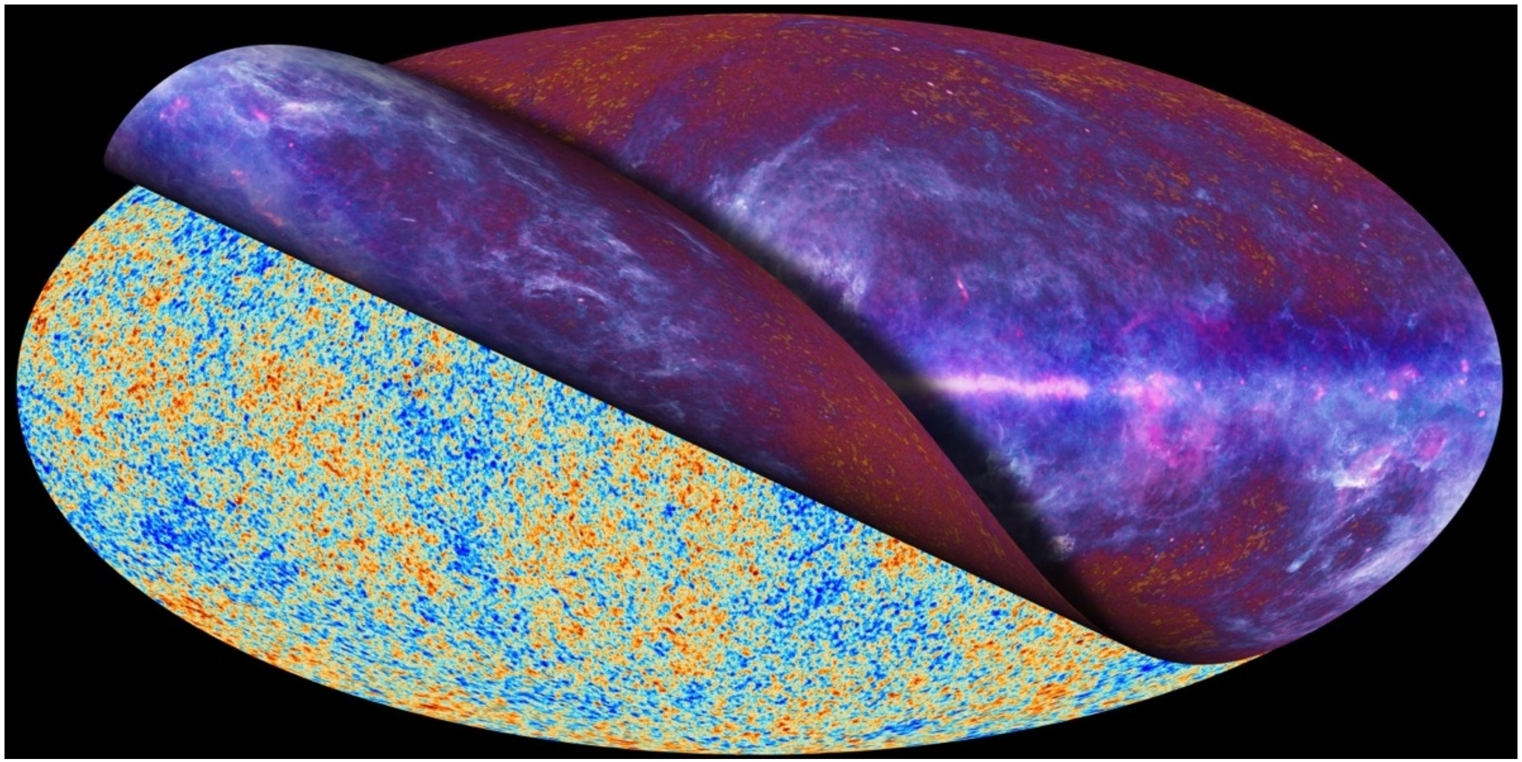
.

# CMB and ICA

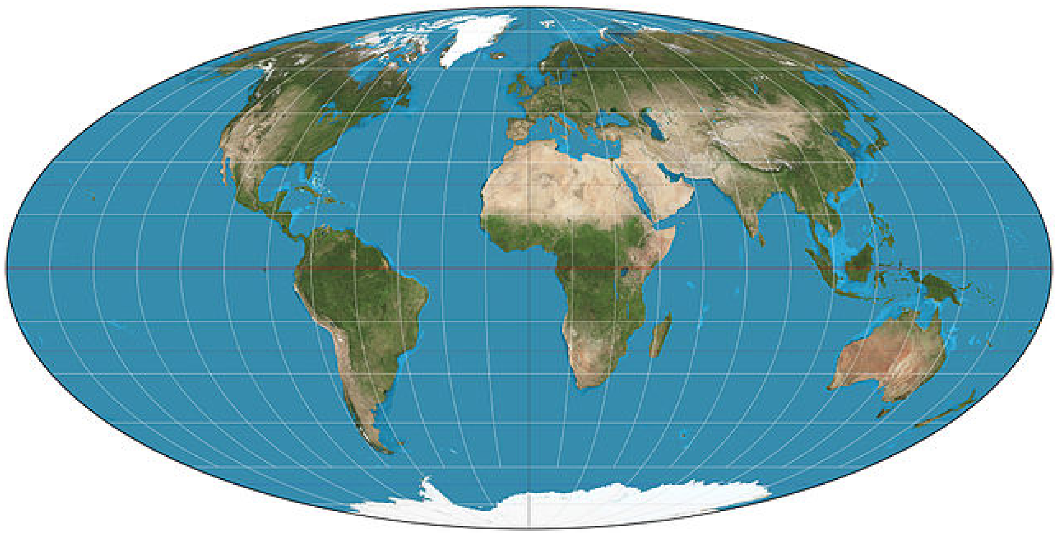
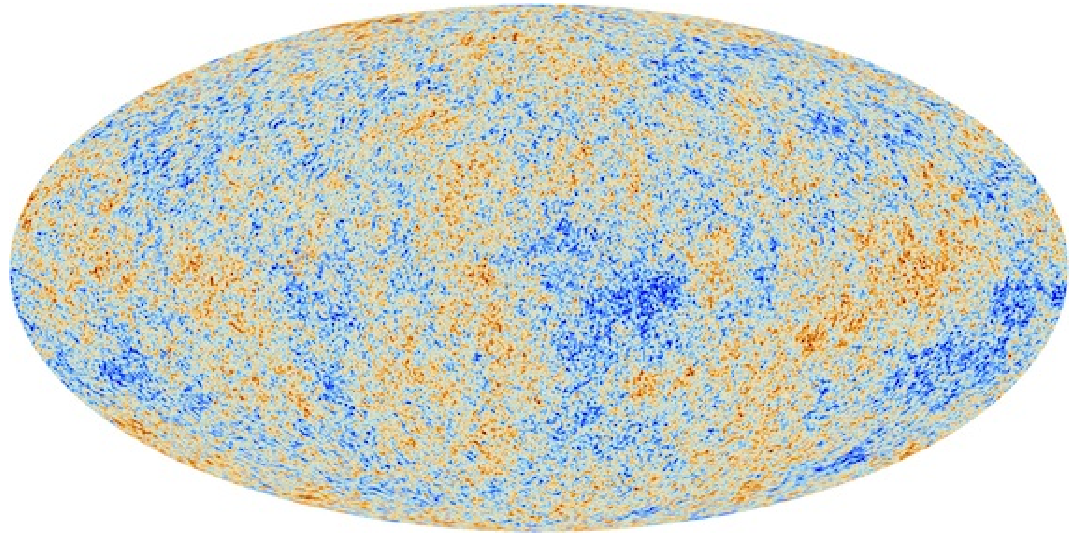


pub

J.-F. Cardoso, CNRS (Télécom-ParisTech, IAP, APC) with the the Planck collaboration

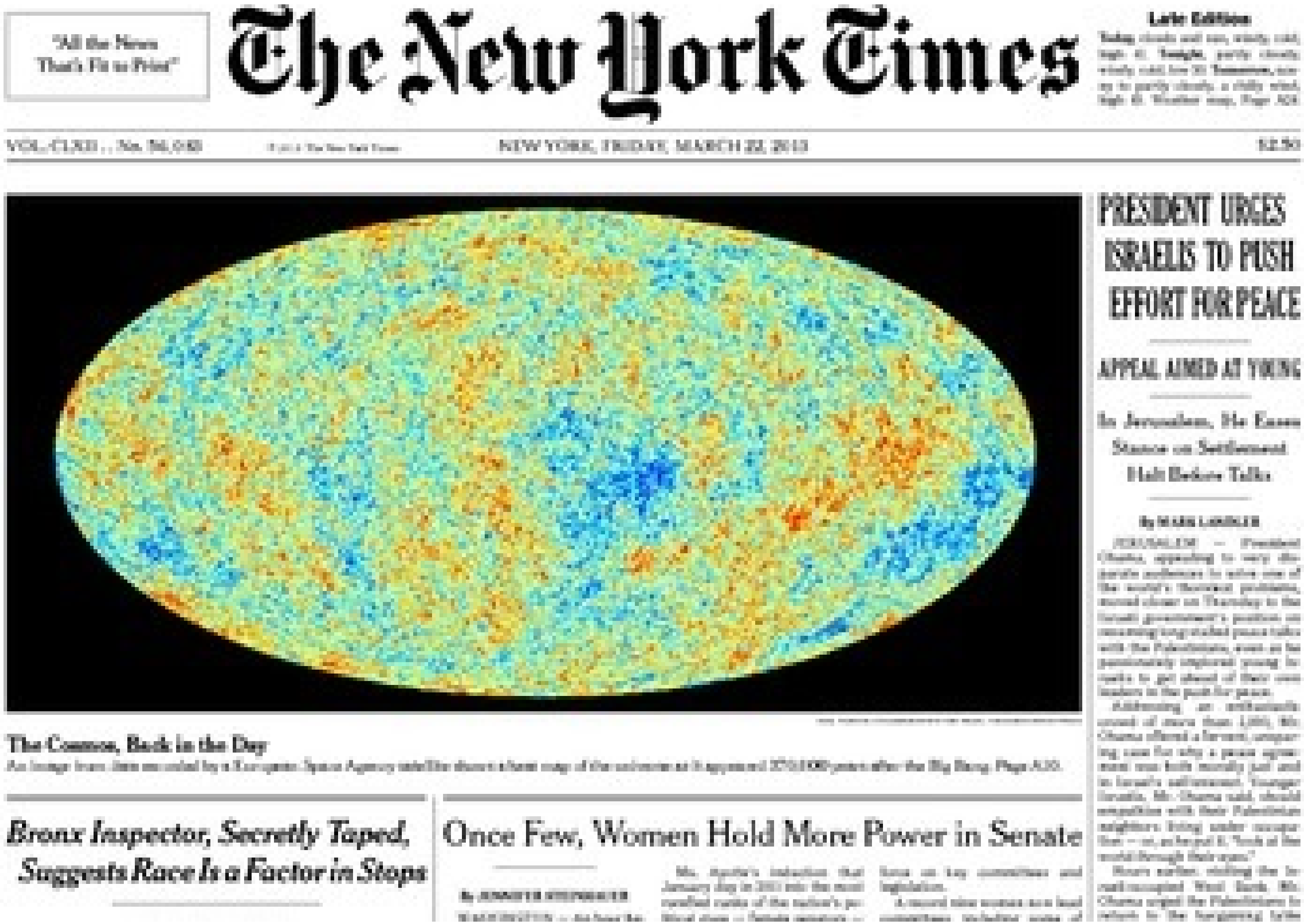
École d'Été BASMATI. Porquerolles, juin 2015.

**Remarque préliminaire:**

Tous nos ovales sont des "projections de Mollweide" de la surface d'une sphère sur le plan.

# "The cosmos, back in the day". Big Science, big news.



"All the News That's Fit to Print"

The New York Times

Late Edition

NEW YORK, FRIDAY, MARCH 22, 2013

$2.50

**The Cosmos, Back in the Day**

PRESIDENT URGES ISRAELIS TO PUSH EFFORT FOR PEACE

APPEAL AIMED AT YOUNG

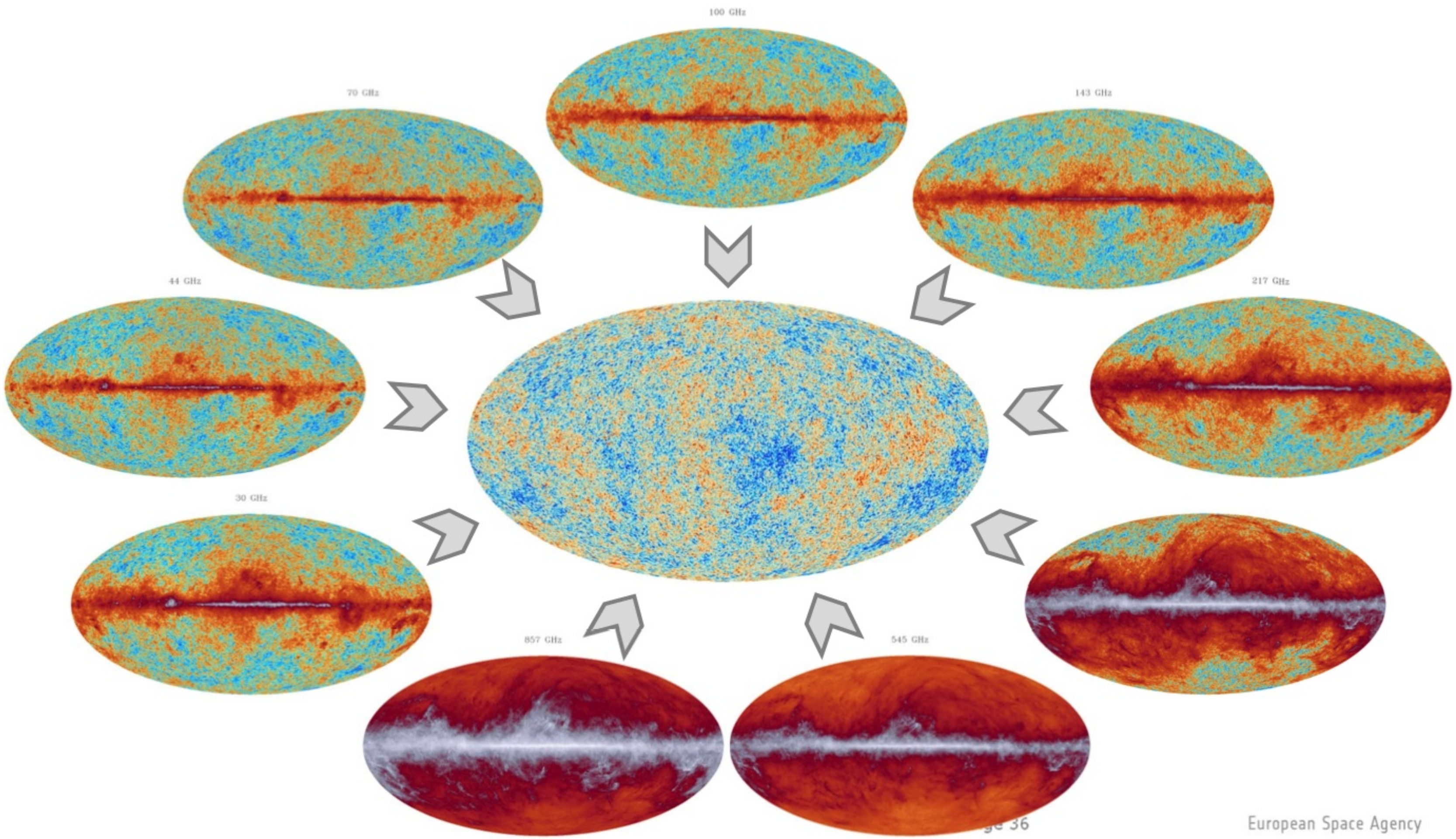
In Jerusalem, He Eases Stance on Settlement Halt Before Talks

*Bronx Inspector, Secretly Taped, Suggests Race Is a Factor in Stops*

Once Few, Women Hold More Power in Senate

## The 9 Planck frequency channels...

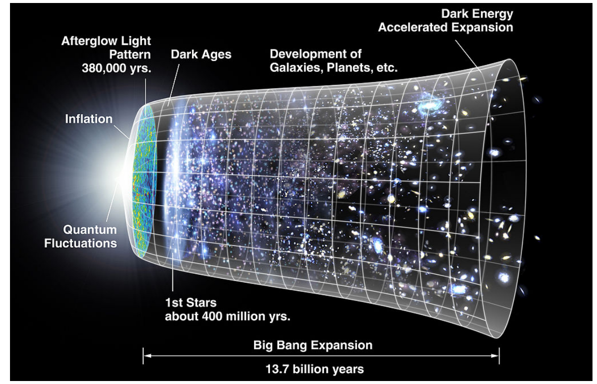
...and the extraction of the CMB by an ICA method.



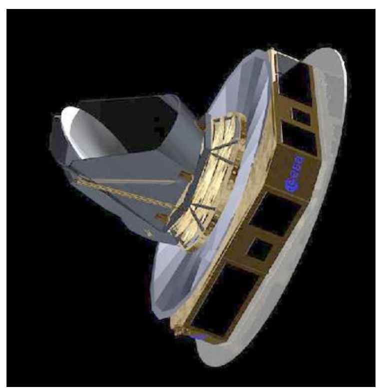

Color scale: hundreds of micro-Kelvins. Credits: ESA, FRB.

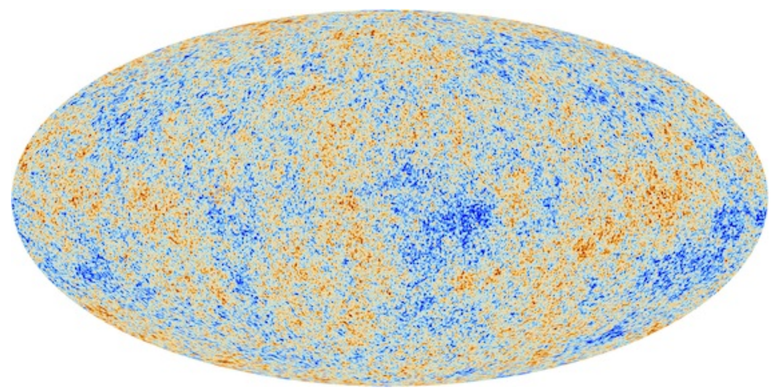
# Notre plan

1. **La théorie (le Big Bang)**: un peu de cosmologie, l'histoire de l'Univers, à grands traits.




2. **Les observations**: le satellite Planck; ce qu'il a vu, et comment.



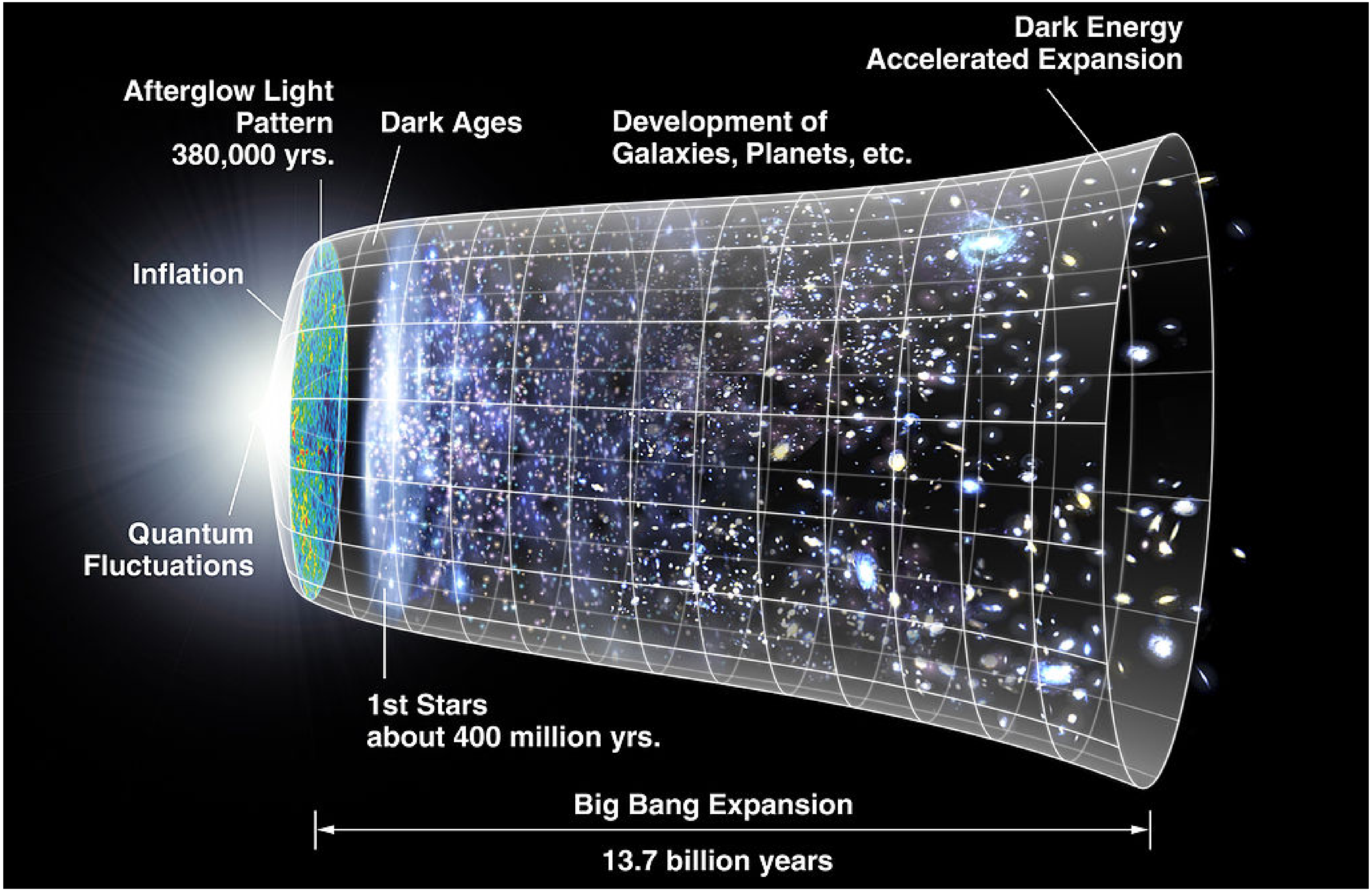


3. **Leur recontre**: extraire la Science des observations.

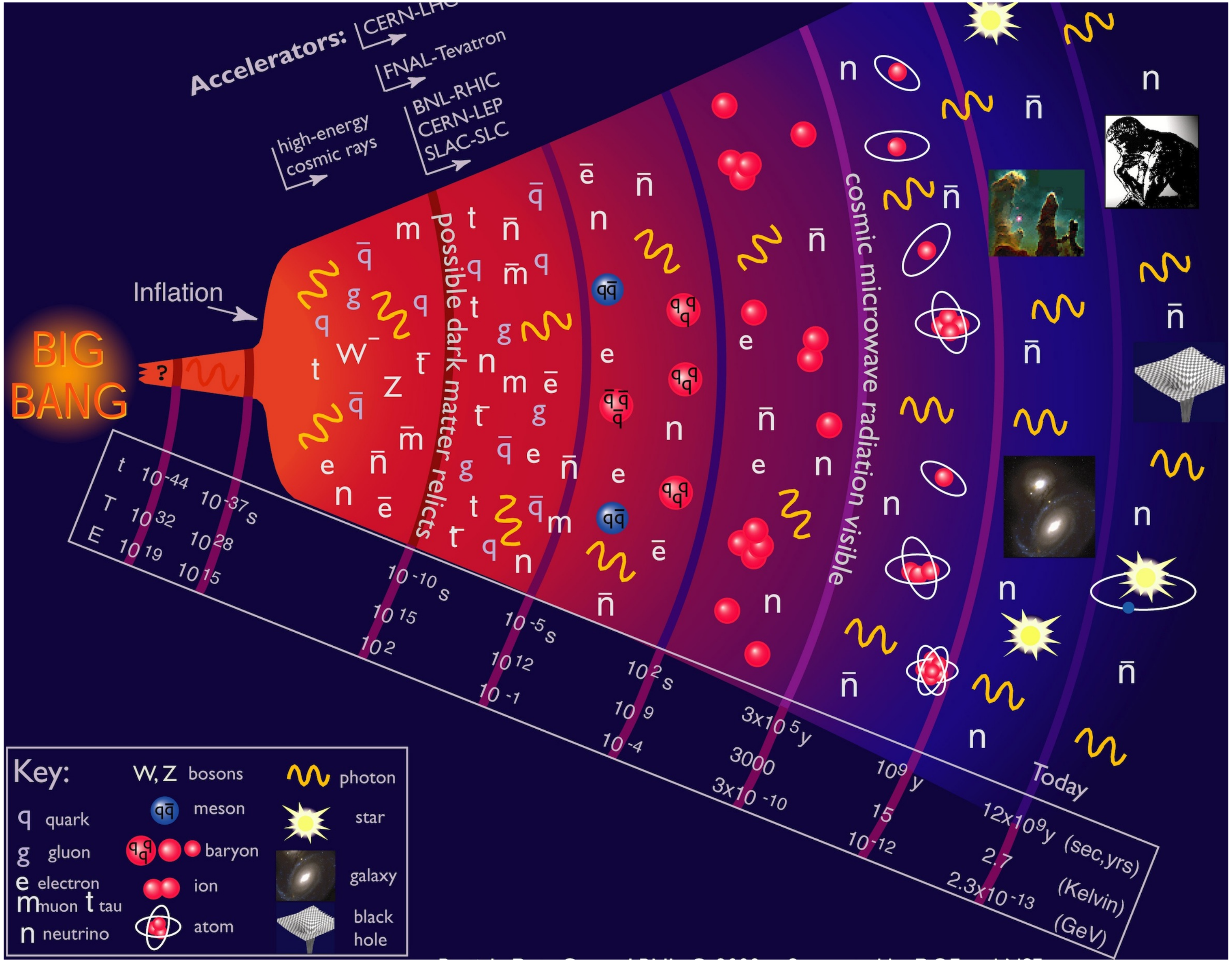
L'âge de l'Univers est . . .

# Big Bang

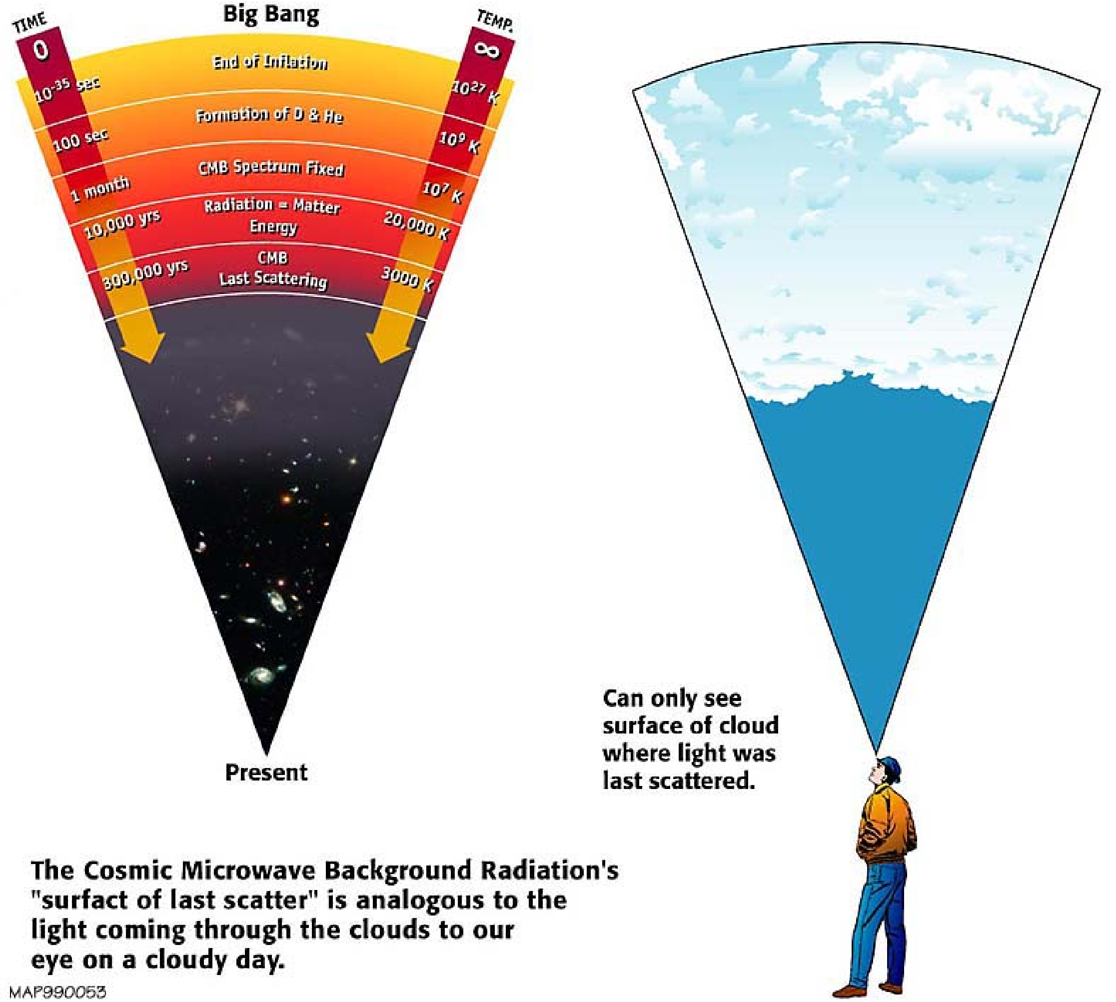
## L'histoire de l'Univers en une image



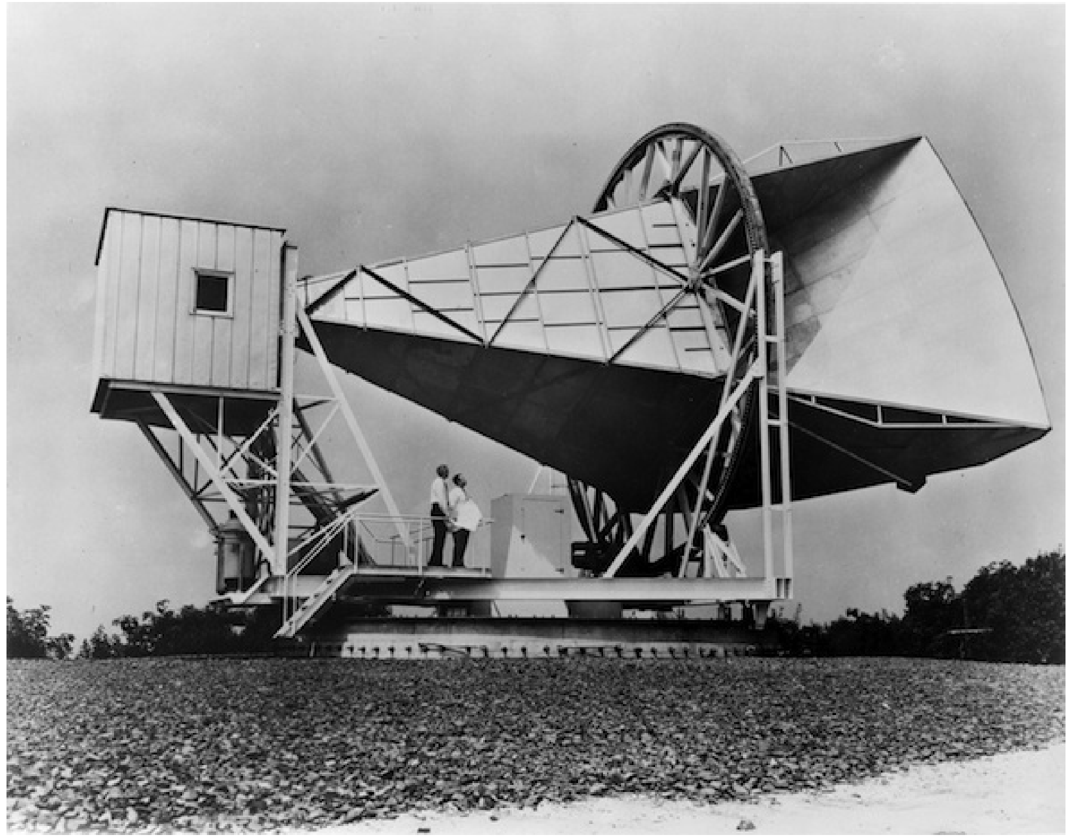
Quand l'Univers devient **transparent**, la lumière se **fossilise**.



# La plus vieille image du monde



## Peut-on réellement percevoir une lumière si lointaine?



Deux gars l'ont vue, sans faire exprès, en 1965. Nobel pour Penzias et Wilson!

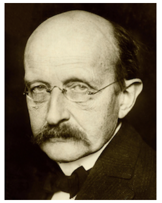
Et ils l'ont trouvée **uniforme** et **froide**: à peu près 3 degrés Kelvin. **C'est-à-dire?**

.

# Lumière, matière et température

## La théorie (Planck)

La répartition d'énergie en function de la fréquence $\nu$ d'un rayonnement de température $T$ est théorisée par la loi de Planck

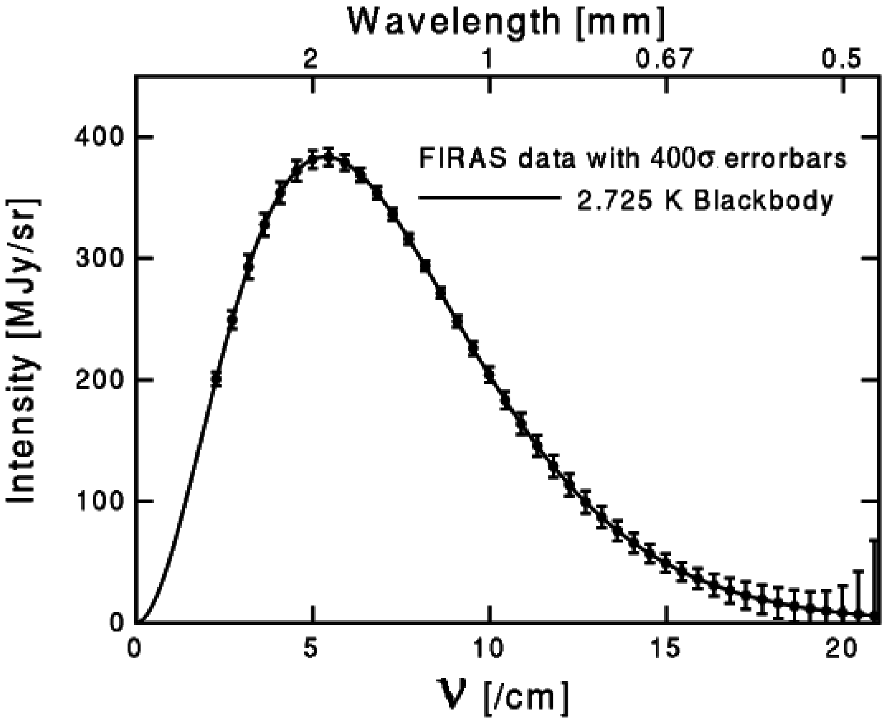


$$I(\nu) = \frac{2h\nu^3}{c^2} \frac{1}{e^{h\nu/kT} - 1}$$

avec les honorables constantes:

- $c$: vitesse de la lumière
- $h$: constante de Planck
- $k$: constante de Boltzmann

## Les mesures (COBE 1992)

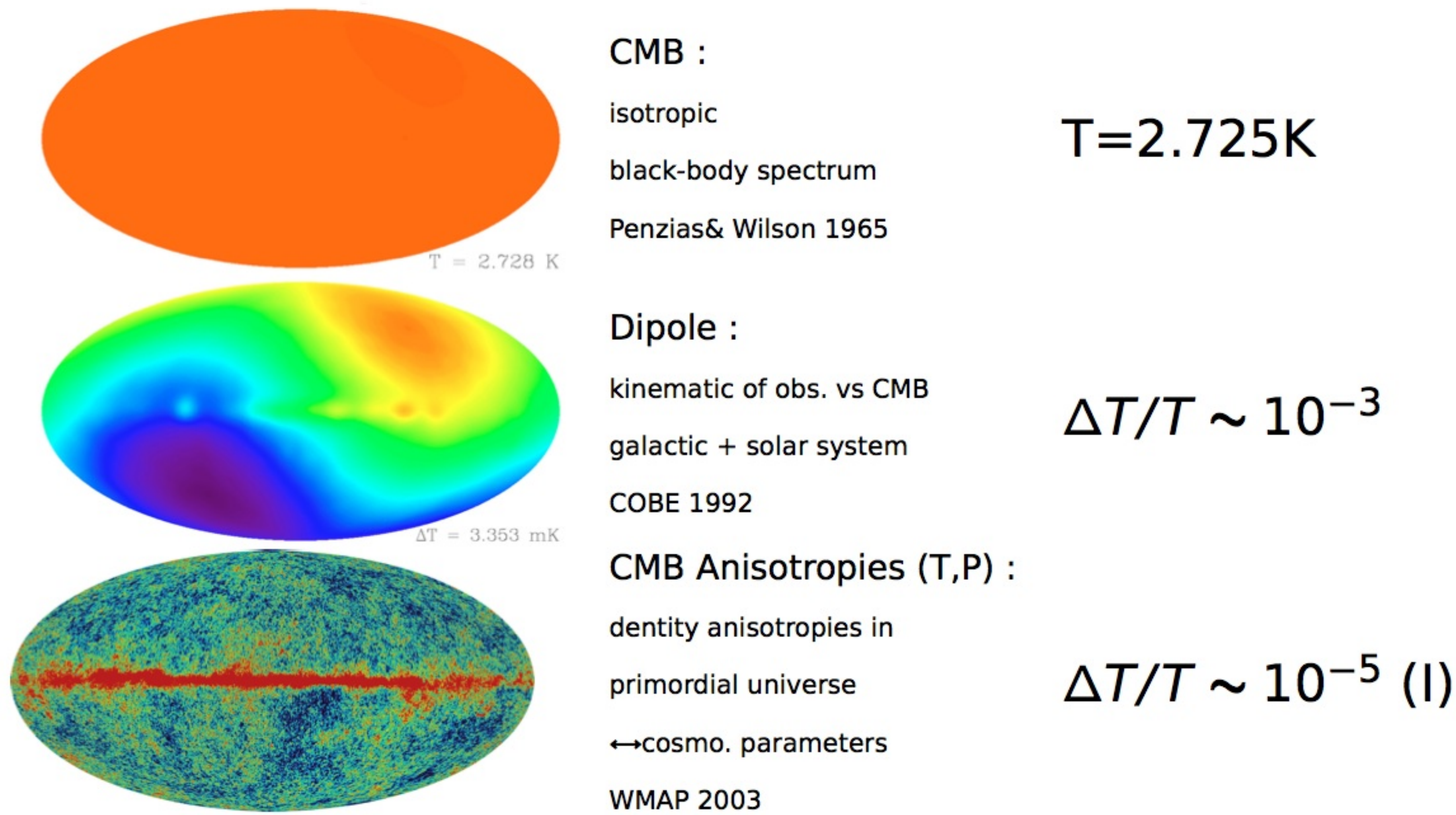



**Éblouissant accord** entre la mesure et la forme théorique!

L'Univers est rempli de vieux photons froids à 2.725 degrés Kelvin.

**Et donc**, il s'est dilaté 1000 fois depuis la recombinaison: $z_{\rm rec} \approx 1000$.

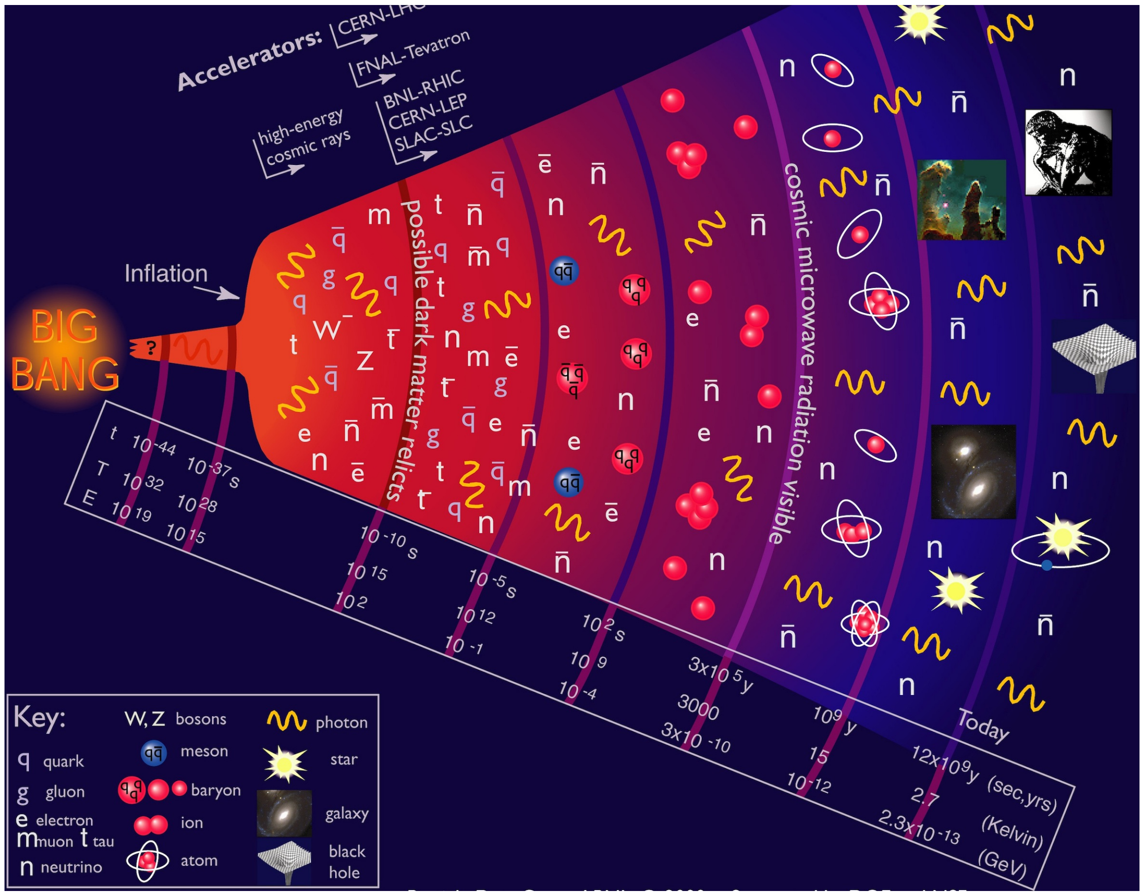
# Le rayonnement fossile est isotrope, mais pas trop

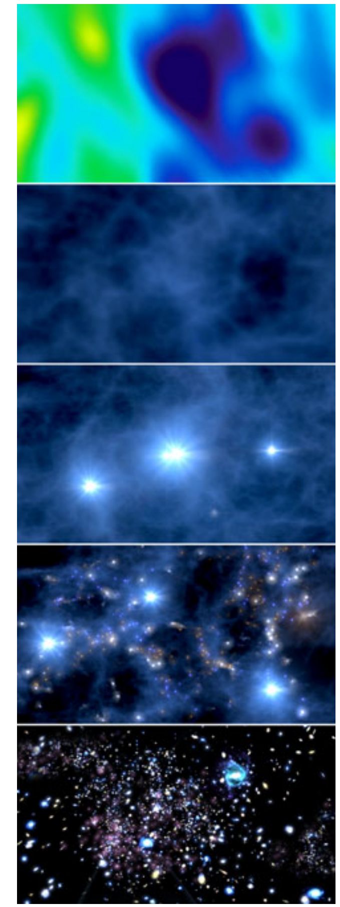



Des anisotropies de 0,0001 degrés (Kelvin) et de 1 degré (centigrade).

# L'Univers sauvé par les anisotropies

A peine âgé de 380,000 ans, l'Univers est dans une situation délicate:

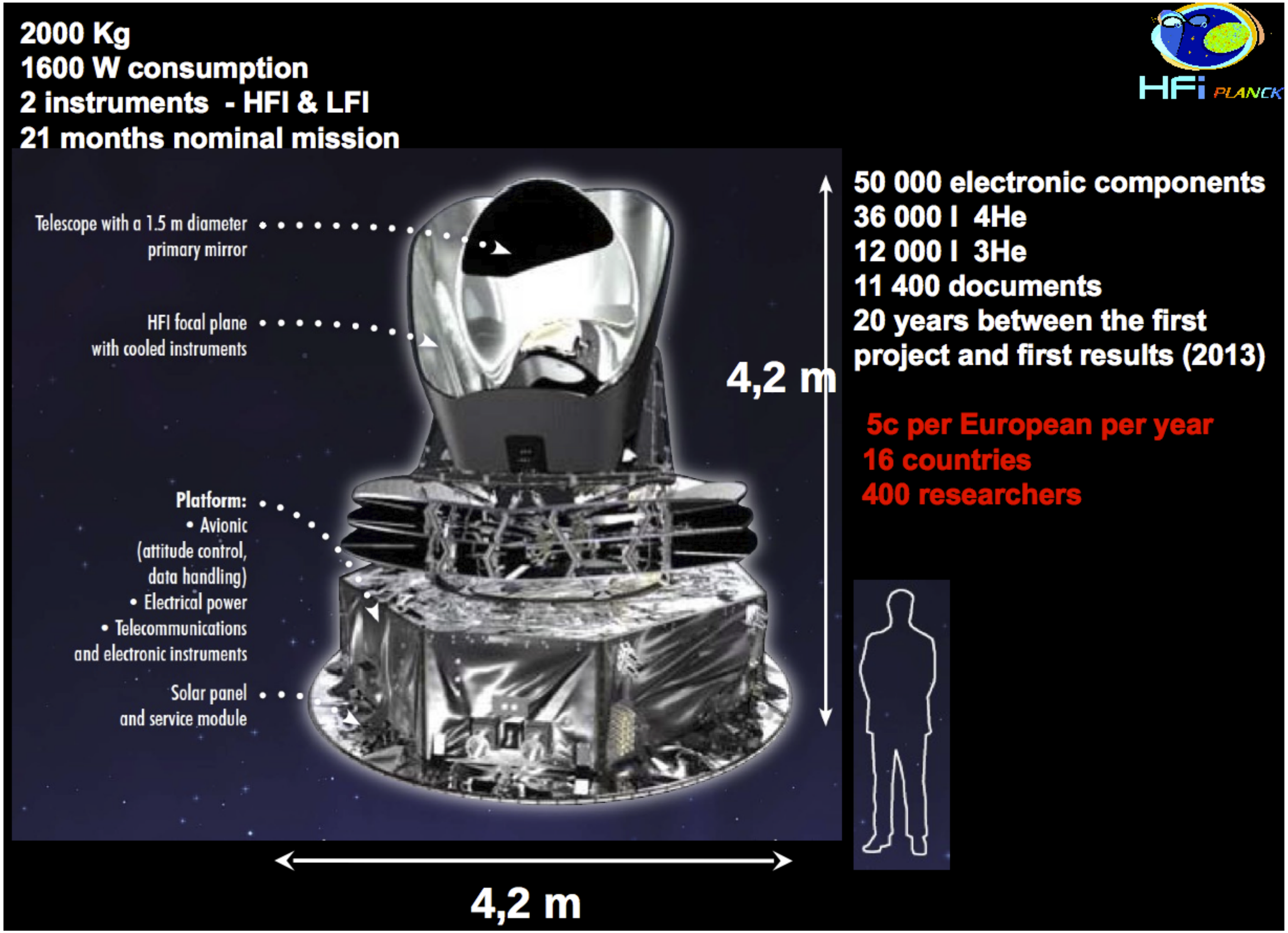


Beaucoup reste donc à faire: les étoiles, les galaxies, les planètes, la vie. . .
Tout va-t-il donc partir à vau-l'eau? Il faut initier les grandes structures!

# Planck

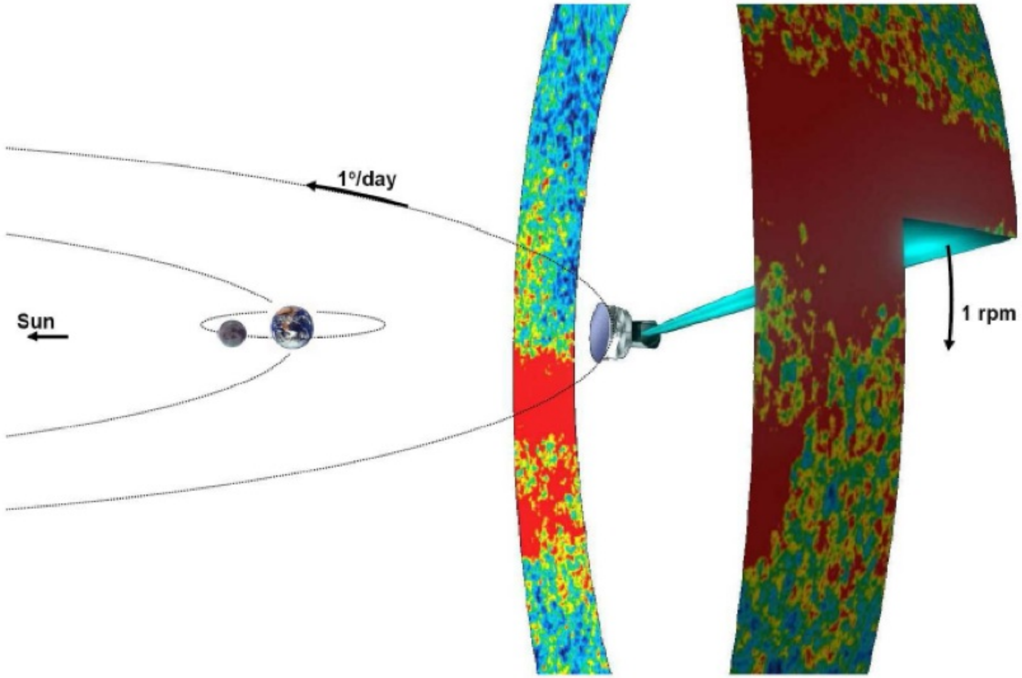
# Pour mieux y voir: la mission Planck

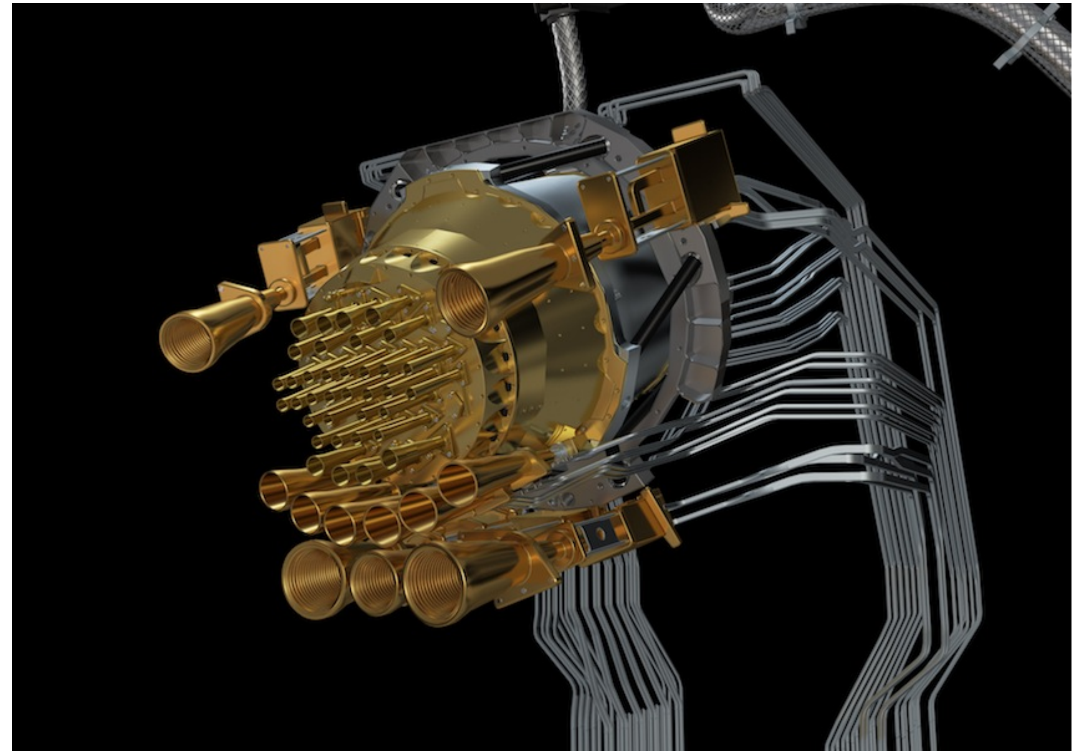


## En route pour le deuxième point de Lagrange!!

# Les cosmologistes adorent le 2ème point de Lagrange Terre-Soleil

## Scanning the sky

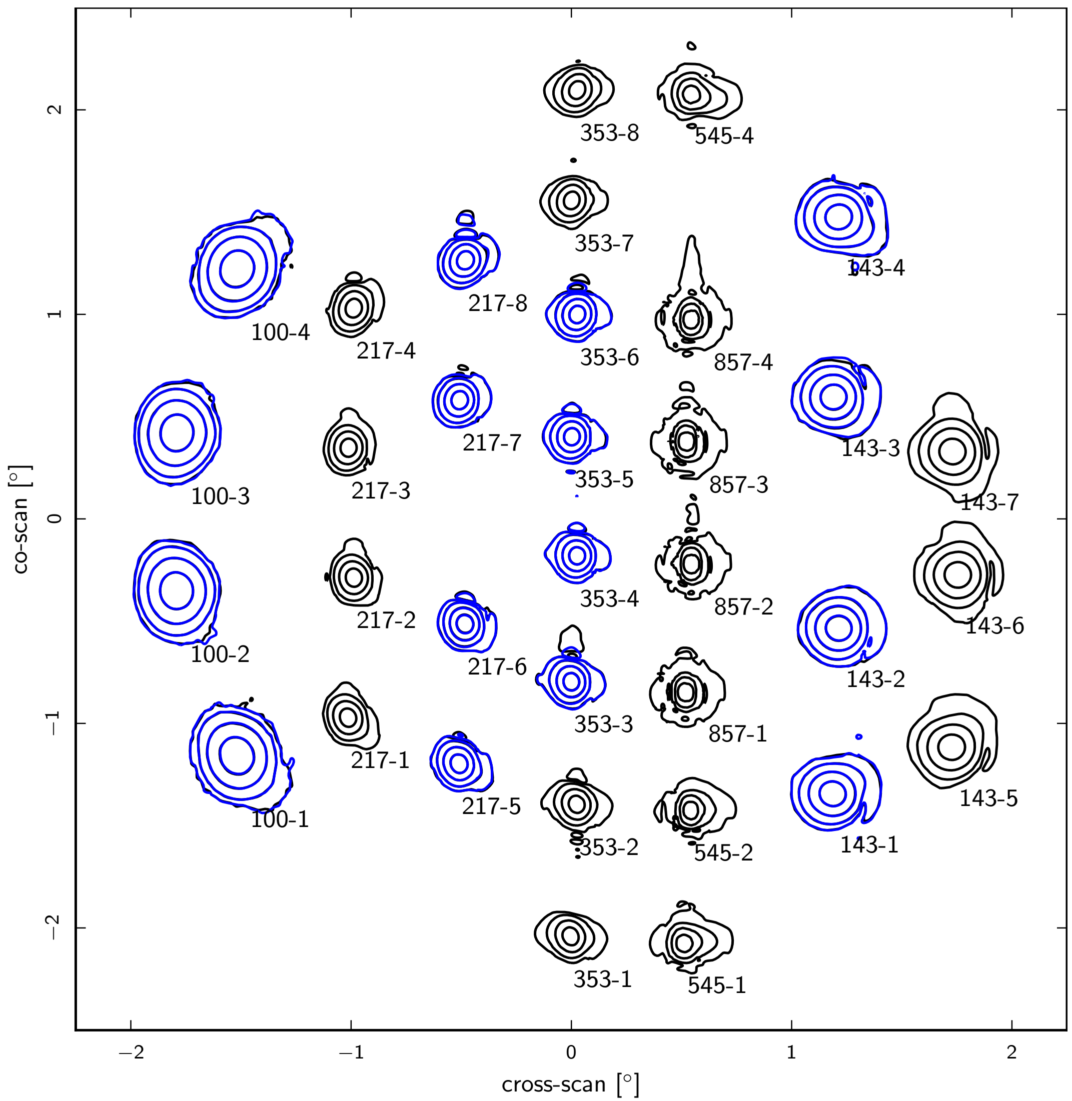


Le téléscope focalise la lumière vers 52 cornets (HFI) qui la filtrent et la guident vers les bolomètres refroidis à 0.1 degrés Kelvin.

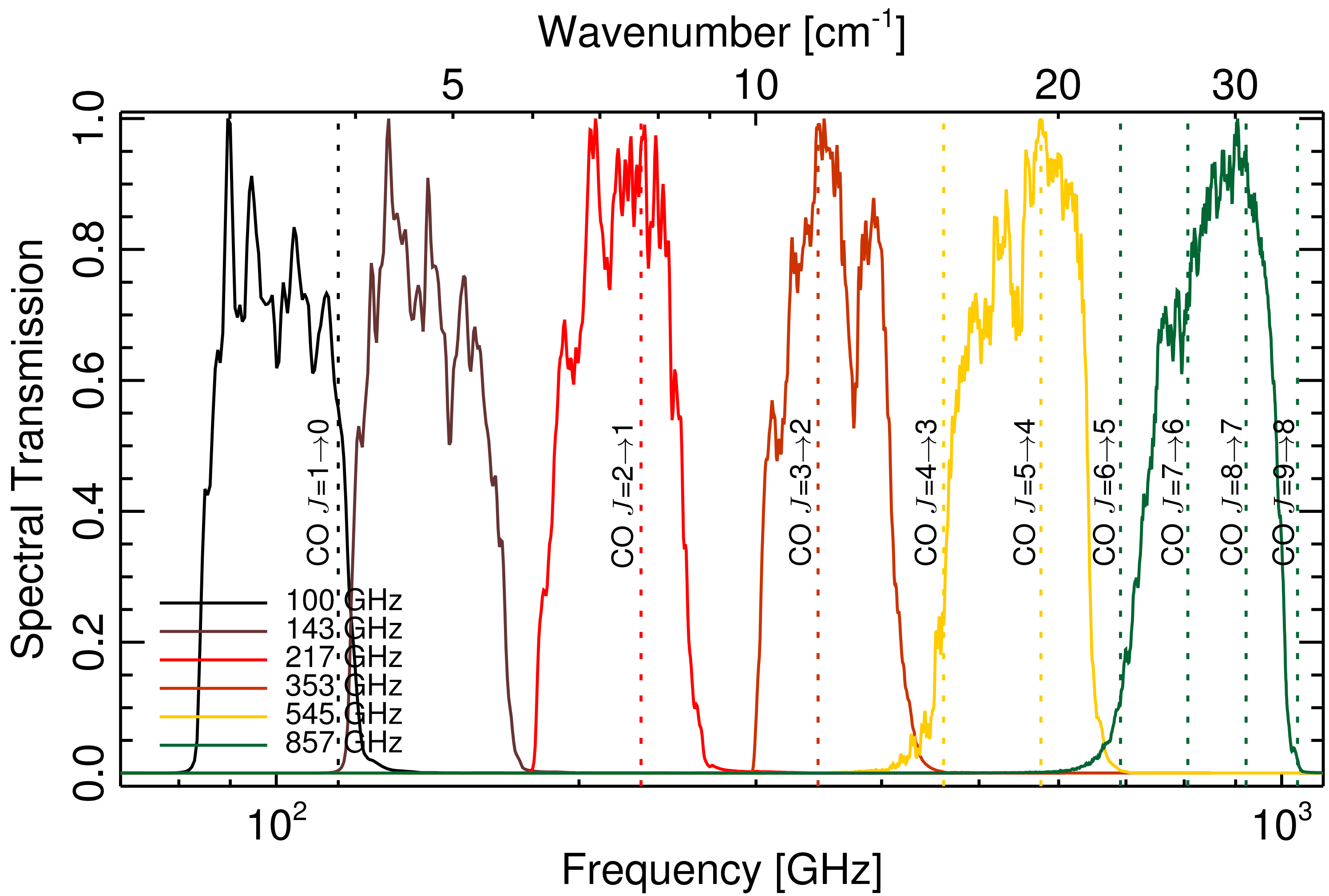
Notez les tuyaux: Planck est aussi un exploit frigorifique.

# Le plan focal de Planck HFI

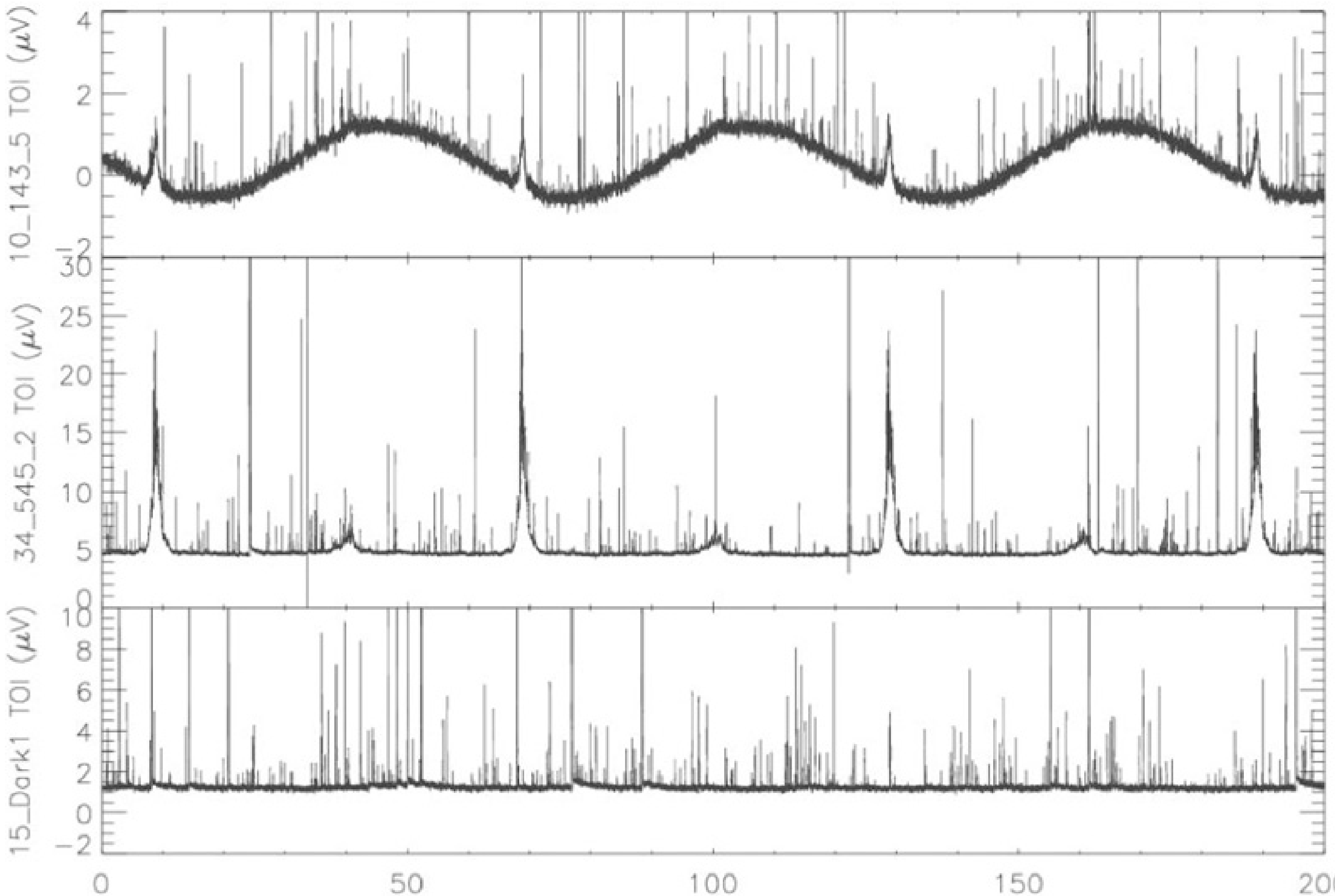



5 arc-minute resolution in best channels → $50\,10^6$ pixels over the sky.

# Les canaux spectraux moyens de Planck HFI

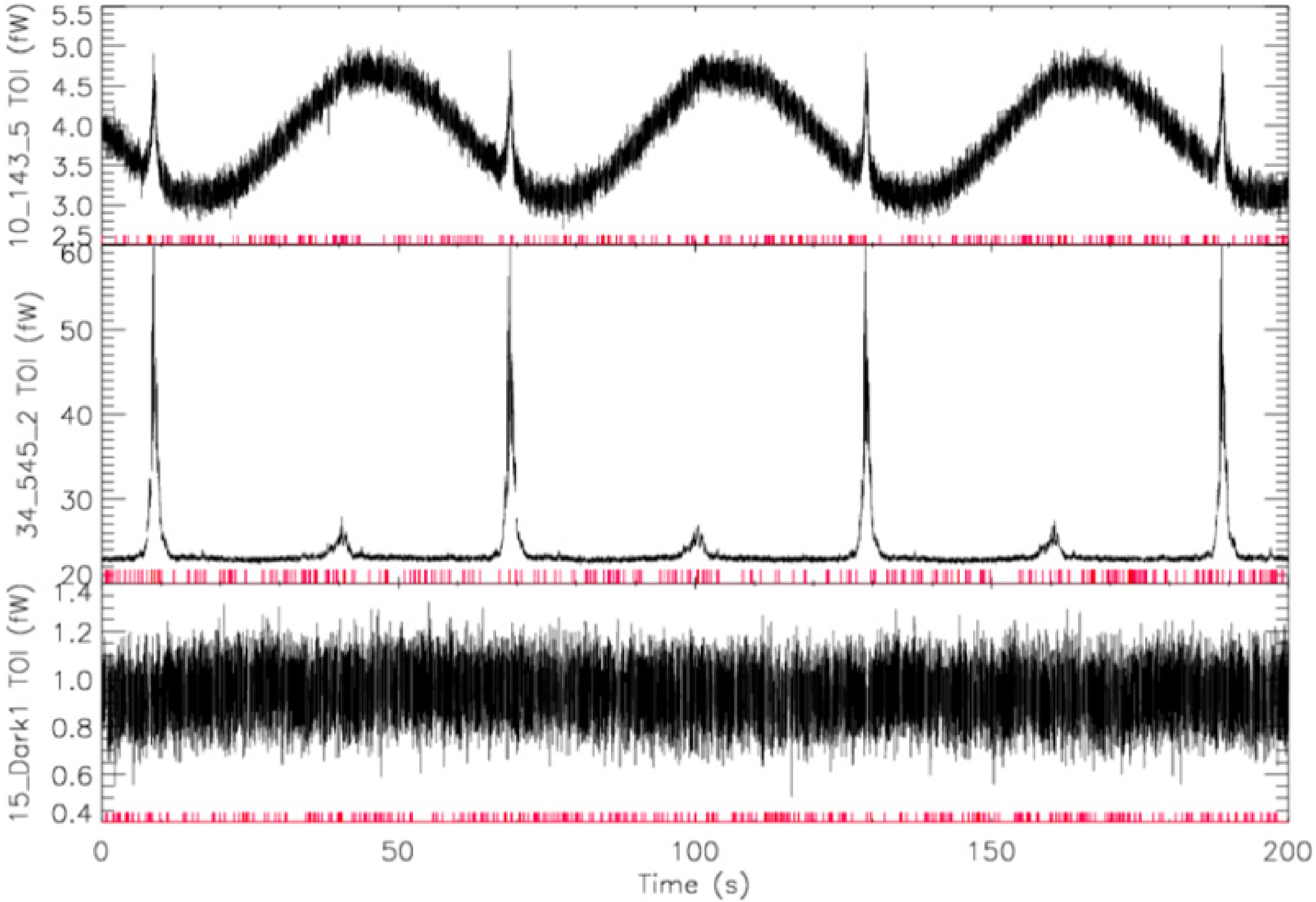


## Signals from a 143 GHz, a 545 GHz, a dark bolometer



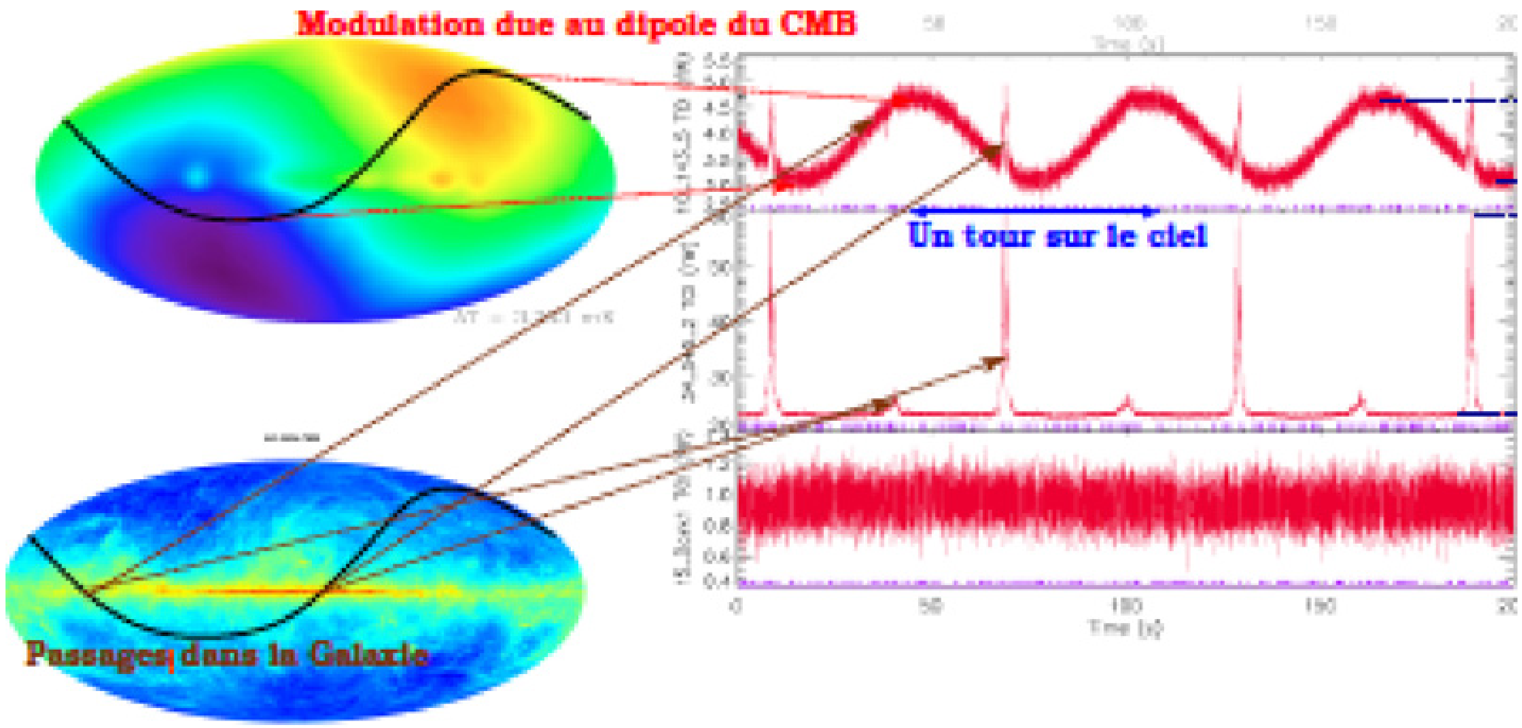

A serious case of glitchopathy.

# Les mêmes, après deglitching




After some glitchotherapy (thanks to recent advances in glitchology).

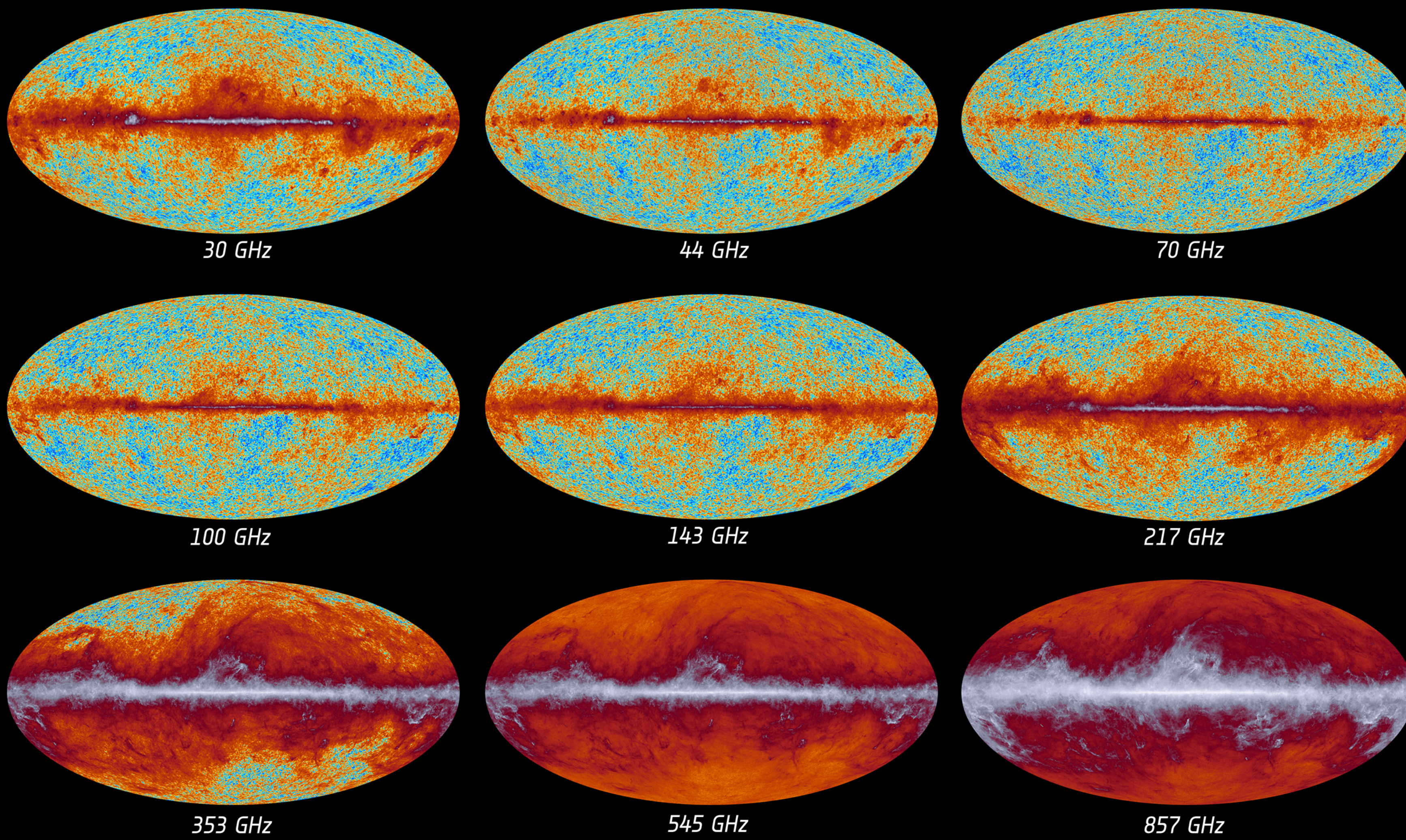
# Map Making needed



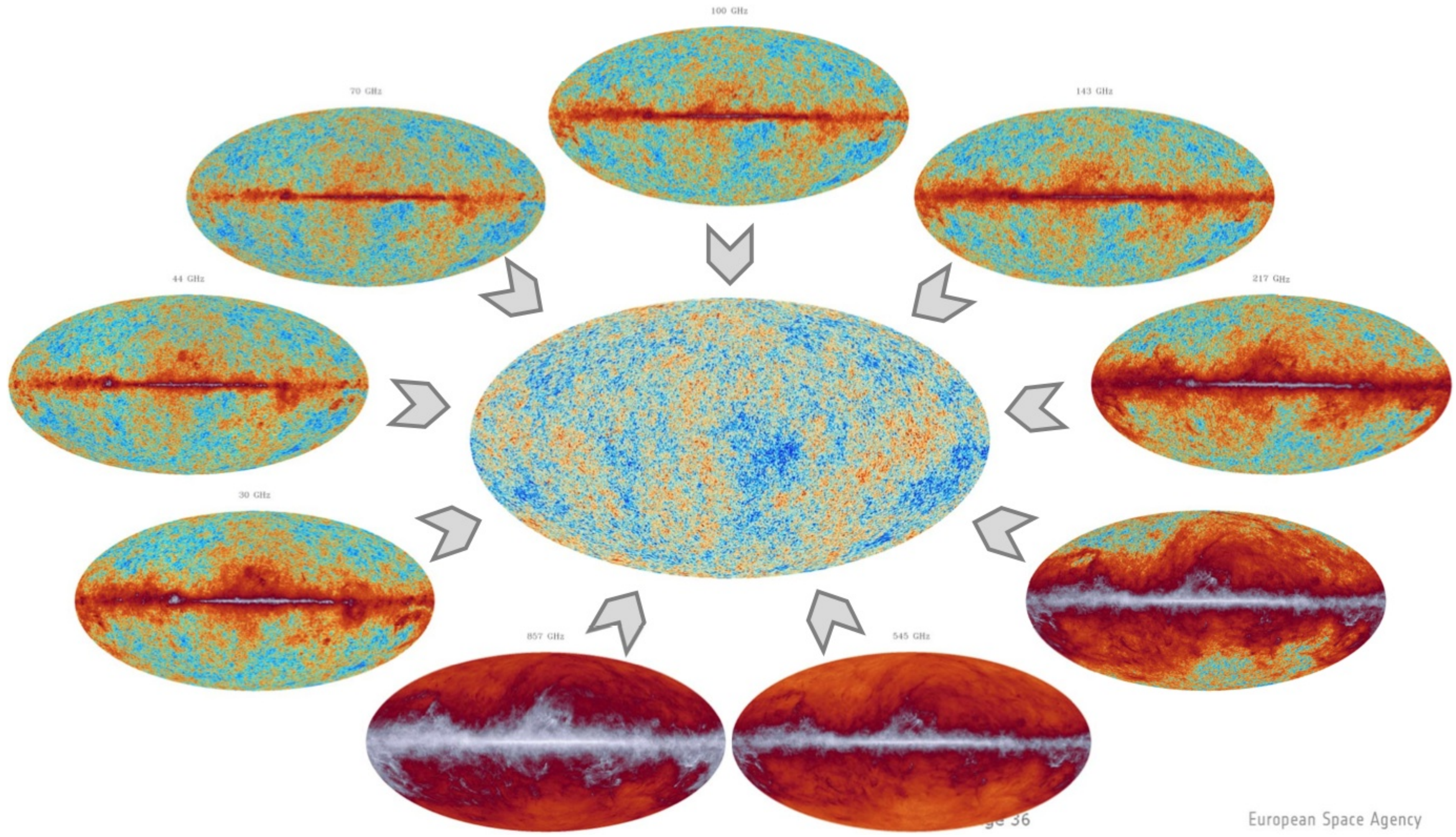

The cosmological signal is still under the noise.

How to go from noisy time lines to less noisy spherical maps: another adventure in big data.

# The sky as seen by Planck

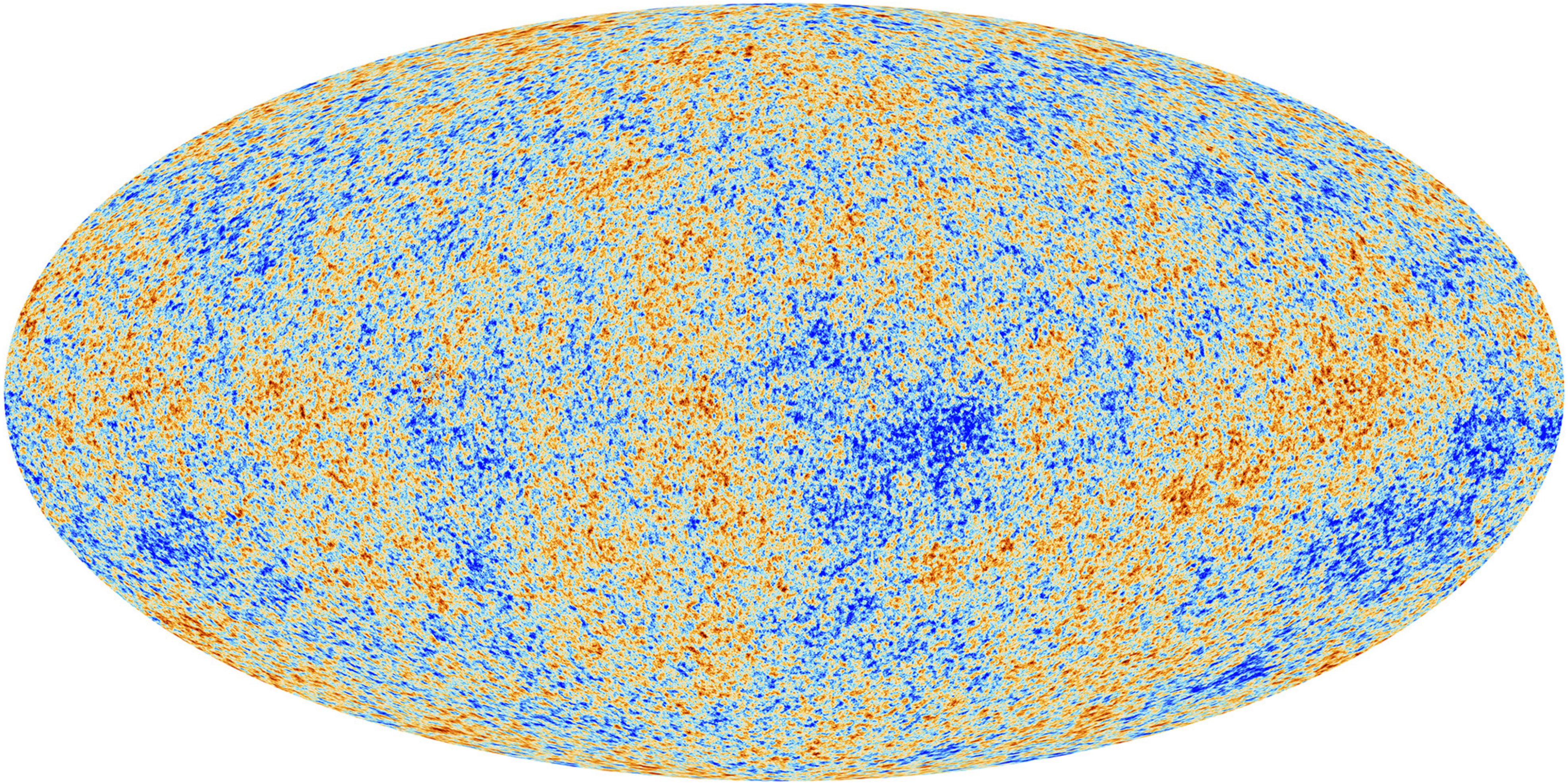


# Combinaison des 9 canaux Planck pour extraire le rayonnement fossile



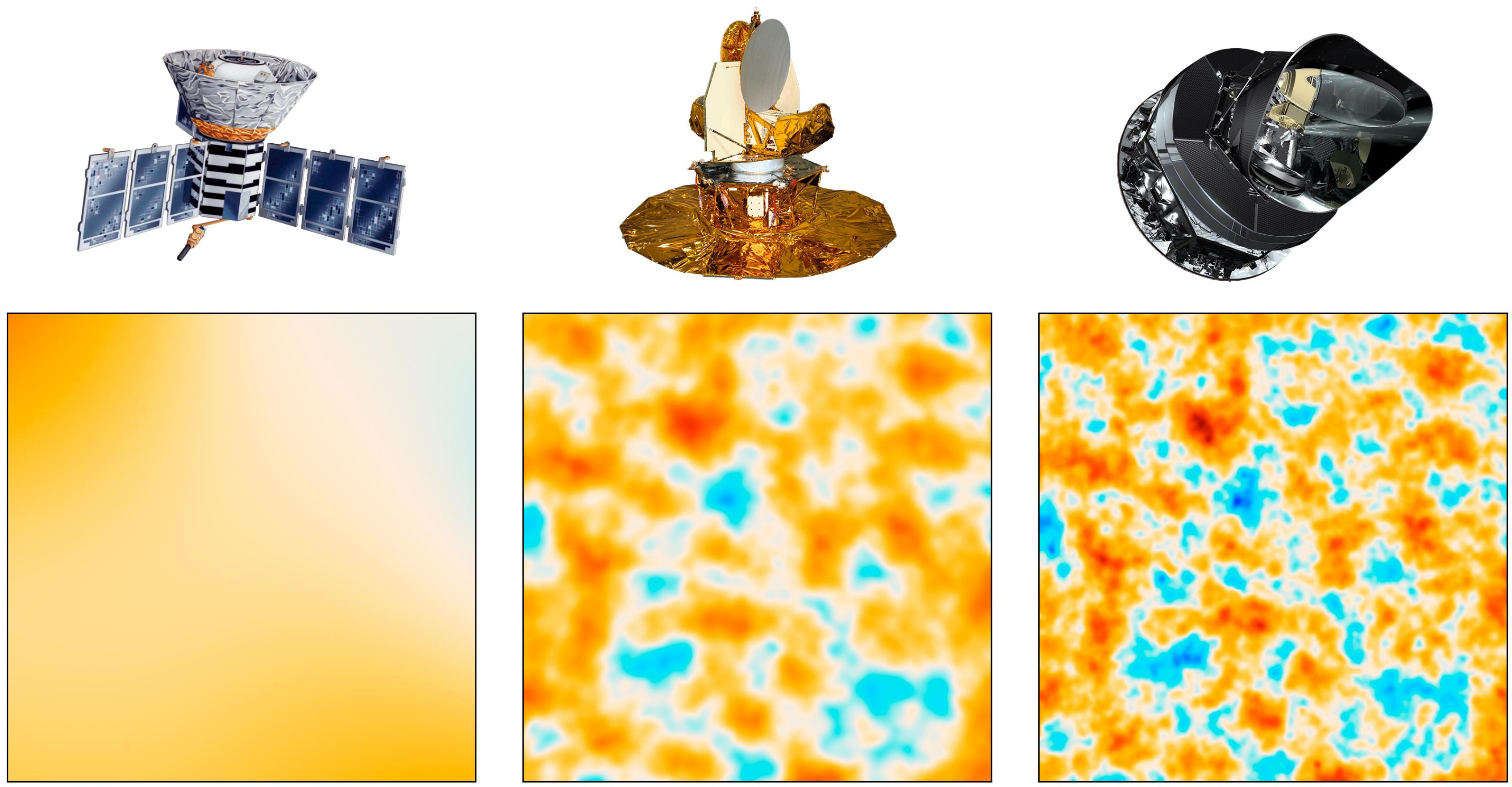

Credit: ESA, FRB

## La plus vieille image du monde, par Planck



Echelle de couleur: ± 300 millionièmes de degré.

# COBE 1992, WMAP 2001, Planck 2013



COBE WMAP Planck

Fine, but how does one do Science with such a boring image ?

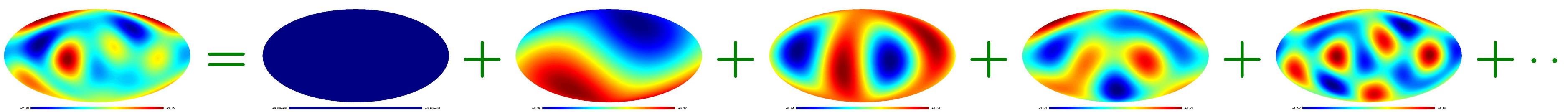
# Cosmological inference on the sphere

## Multipole decomposition and angular frequencies

A spherical field $X(\theta,\phi)$ can decomposed into 'harmonic' components:

$$X(\theta,\phi) = \sum_{\ell\geq 0} X^{(\ell)}(\theta,\phi) \qquad [\theta,\phi] = [\text{(co)latitude, longitude}]$$

called monopole, dipole, quadrupole, octopole, . . . , multipole, indexed by the (discrete) angular frequency, $\ell = 0, 1, 2, \ldots$, thusly:



$$X(\theta,\phi) = X^{(0)}(\theta,\phi) + X^{(1)}(\theta,\phi) + X^{(2)}(\theta,\phi) + X^{(3)}(\theta,\phi) + \cdots$$

- Compare $R^n$ where $\Delta e^{i\vec{k}\cdot\vec{r}} = -\|\vec{k}\|^2 \, e^{i\vec{k}\cdot\vec{r}}$ to $S^2$ where $\Delta X^{(\ell)} = -\ell(\ell+1)\, X^{(\ell)}$ .

- The (empirical) angular spectrum :

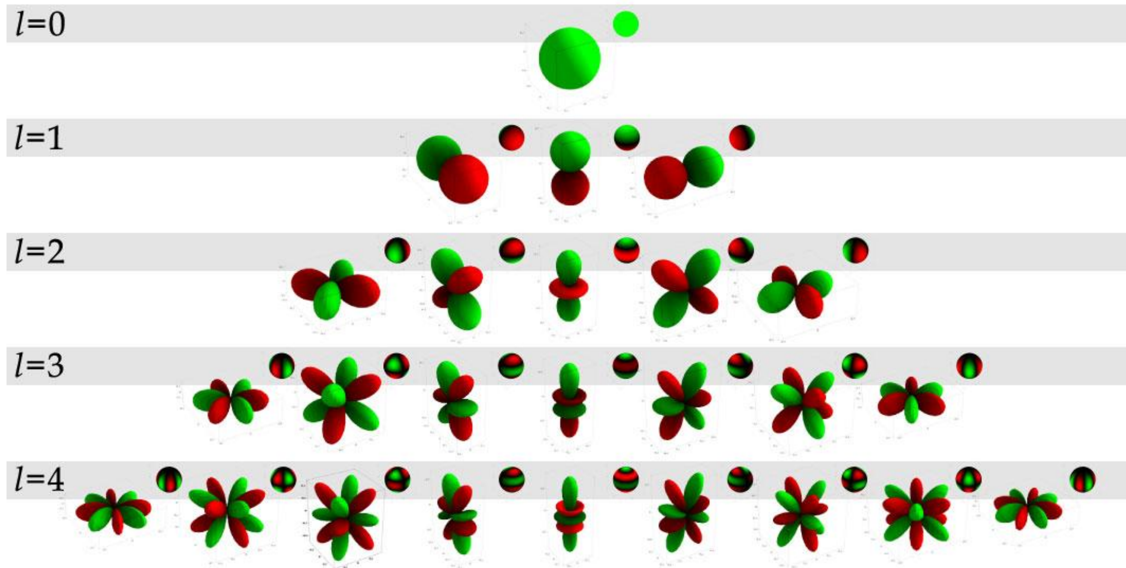
$$\widehat{C}_\ell \stackrel{\text{def}}{=} \int\int_{S^2} \left(X^{(\ell)}(\theta,\phi)\right)^2 / (2\ell+1)$$

. . . quantifies how power is distributed across (angular) scales.

# Fourier on the sphere: Spherical harmonic decomposition

- An ortho-basis for spherical fields: the spherical harmonics $Y_{\ell,m}(\theta,\phi)$:

$$X(\theta,\phi) = \sum_{\ell\geq 0} \sum_{m=\ell}^{m=-\ell} a_{\ell,m}\ Y_{\ell,m}(\theta,\phi) \quad \longleftrightarrow \quad a_{\ell,m} = \int_\theta \int_\phi Y_{\ell,m}(\theta,\phi)\ X(\theta,\phi)$$

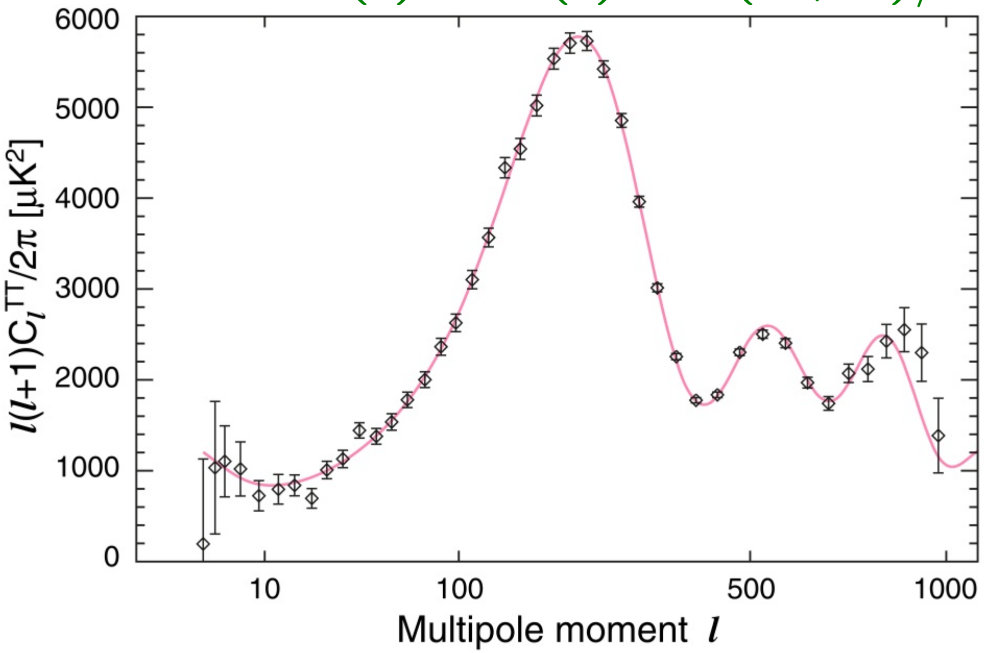



- The decomposition into multipoles and the angular spectrum:

$$X^{(\ell)}(\theta,\phi) = \sum_{m=\ell}^{m=-\ell} a_{\ell,m}\ Y_{\ell,m}(\theta,\phi) \qquad \widehat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=\ell}^{m=-\ell} a_{\ell,m}^2$$

# The angular spectrum of the CMB, measured by W-MAP.

$$D(\ell) = C(\ell) \times \ell(\ell+1)/2\pi$$




Black ink: empirical spectrum (W-MAP 5yr).

Red ink: best-fit theory spectrum.

Three acoustic peaks: congrats, W-MAP!

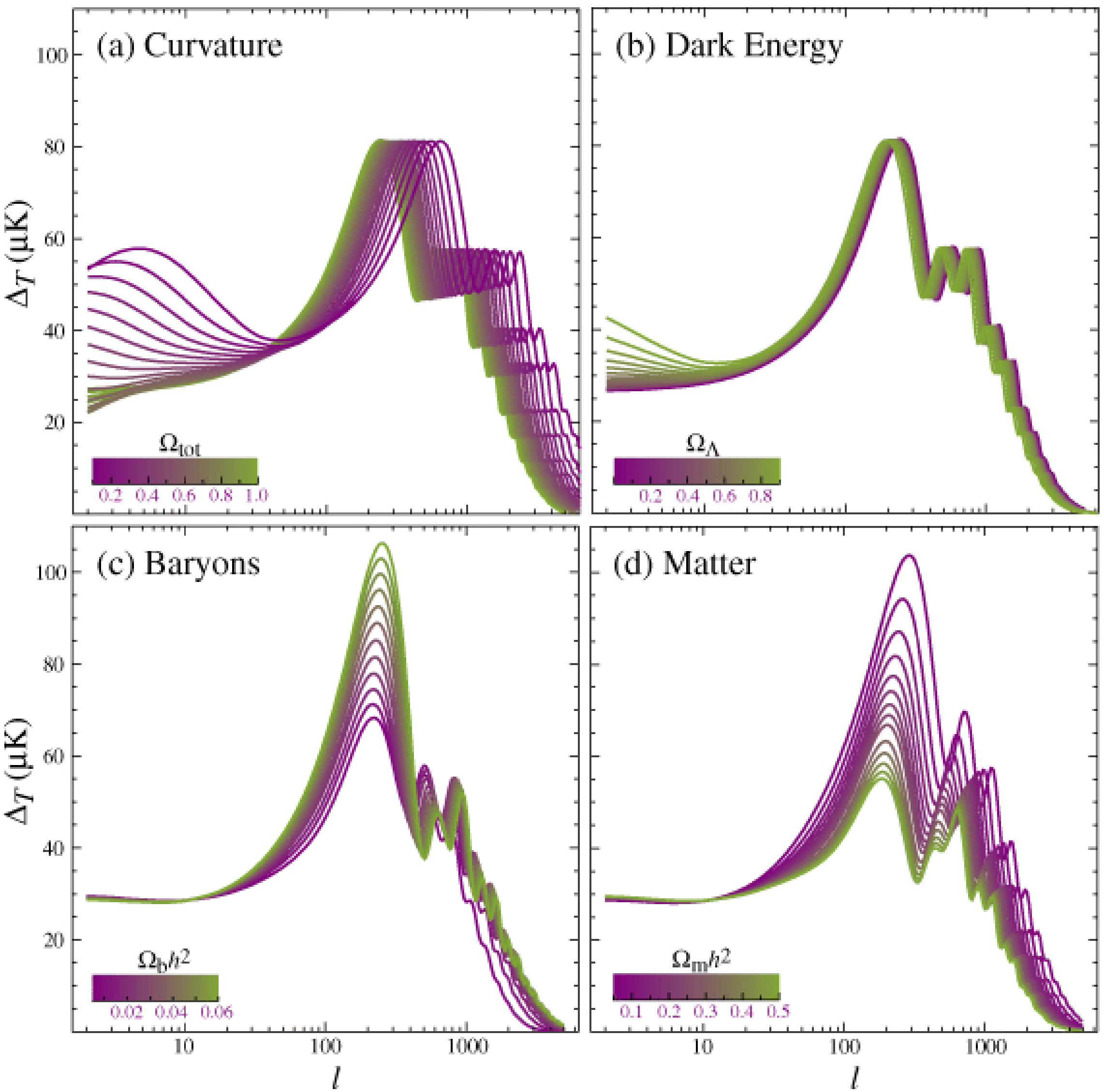
- CMB is fixed on the sky, so only $2\ell+1$ coefficients for estimating $C\ell)$:

$$\text{Var}\Big(\widehat{C}_\ell / \mathbb{E}\widehat{C}_\ell\Big) = \frac{2}{2\ell+1}$$ 'Cosmic variance' on a fixed full sky

- For W-MAP 5: Cosmic variance dominates for $\ell \leq 540$

- Instrumental noise dominates at higher multipoles.

# Theoretical angular spectrum of the CMB




A cosmological model has to predict the angular spectrum of the CMB as a function of "cosmological parameters".

Left: examples of the dependance of the spectrum on some parameters of the $\Lambda - \text{CDM}$ model.

Important note : we plot $\ell^2 C_\ell$.
Large scales dominate the power.

## Angular spectrum and likelihood

- The spherical harmonic coefficients $a_{\ell,m}$ of a stationary random field are uncorrelated with variance $C_\ell$, the angular power spectrum:

$$\mathsf{E}\Big(a_{\ell,m}\ a_{\ell' m'}\Big) = C_\ell\ \delta_{\ell\ell'}\ \delta_{mm'}$$

- Thus, for a stationary Gaussian field, the empirical spectrum

$$\widehat{C}_\ell = \frac{1}{2\ell+1} \sum_{m=\ell}^{m=-\ell} a_{\ell,m}^2$$

is a sufficient statistic since the likelihood then reads:

$$-2 \log P(X|\{C_\ell\}) \ = \ \sum_{\ell \geq 0} (2\ell+1) \left( \frac{\widehat{C}_\ell}{C_\ell} + \log C_\ell \right) + \mathsf{cst}$$

**Super compression!**

- It is also a spectral mismatch:

$$-2 \log P(X|\{C_\ell\}) \ = \ \sum_{\ell \geq 0} (2\ell+1)\ k(\widehat{C}_\ell / C_\ell) + \mathsf{cst'} \qquad k(u) \stackrel{\mathsf{def}}{=} u - \log u - 1$$

## The likelihood of our Universe, in an ideal nutshell

• Cosmologists build physical models predicting a Gaussian stationary CMB sky with an angular spectrum depending on fundamental cosmologic parameters:

$$C_\ell = C_\ell(\alpha) \qquad \alpha = (\Omega_\Lambda, \Omega_m, H_0, \ldots)$$

• Instrumentalists painfully measure the angular spectrum $\widehat{C}_\ell$ of the CMB sky.

• Statisticians know how to adjust theory to data :

$$-2 \log p(\mathrm{CMB}|\alpha) = \sum_{\ell \geq 0} (2\ell + 1) \left( \frac{\widehat{C}_\ell}{C_\ell(\alpha)} + \log C_\ell(\alpha) \right) + \mathrm{cst.}$$

• In real life, things (the likelihood, the estimate $\widehat{C}_\ell$) are much more complicated, but we still match a model spectrum to an empirical spectrum.

## Angular correlation and angular spectrum.

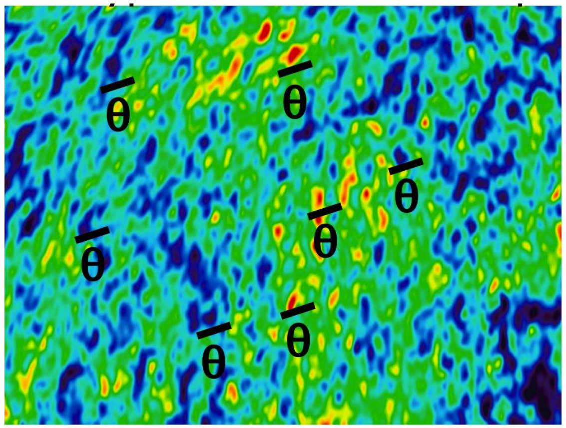
Average correlation over all directions separated by an angle $\theta$

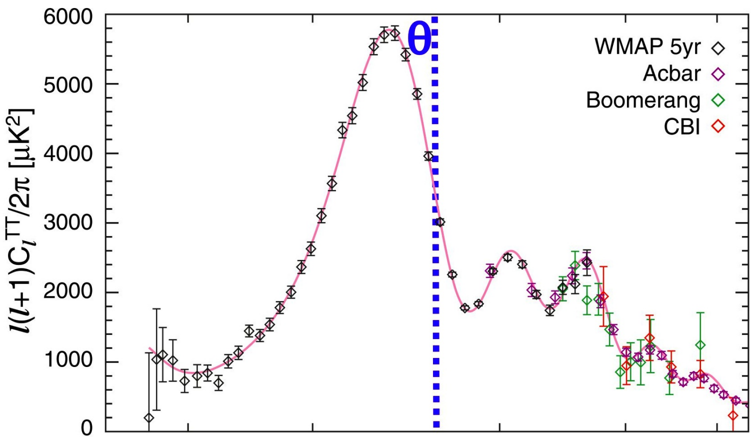
$$\rho(\theta) \overset{\text{def}}{=} \langle X(\vec{n}) X(\vec{n}') \rangle_{\vec{n}\cdot\vec{n}'=\cos\theta}$$ Angular correlation $\rho(\theta)$

Wiener-Kinchin on the sphere, connecting correlation and spectrum:

$$\rho(\theta) = \sum_{\ell \geq 0} \frac{2\ell+1}{4\pi} C(\ell)\, P_\ell(\theta)$$ Angular spectrum $C(\ell)$

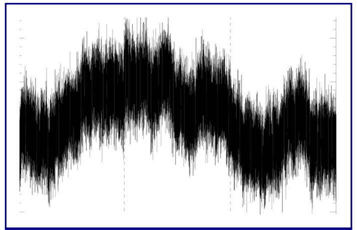
$C(\ell) \times \ell(\ell+1)/2/\pi$



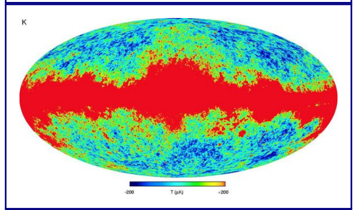




The angular spectrum shows "acoustic peaks", signatures of a vibrating plasma.

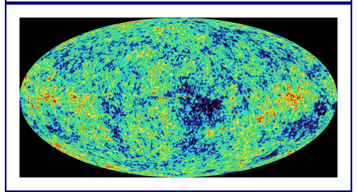
# The big pipeline picture, from time series to cosmology, ideally



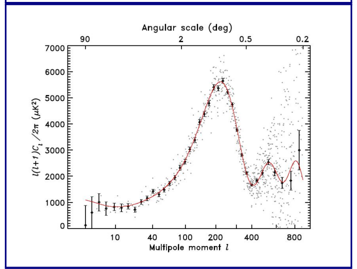
← SCANNING THE SKY.
Make sure you capture those $\mu K$s in your time lines.
Deglitch, flag, deconvolve, calibrate. . .



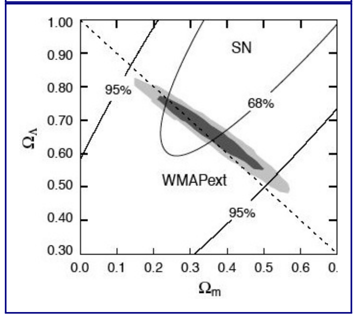
← MAP MAKING: from time lines to spherical maps.
Here, the microwave sky at 23 GHz seen by W-MAP.



← COMPONENT SEPARATION: from several frequency channels maps to a component map. Here, pure (?) CMB from WMAP.




← SPECTRAL ESTIMATION: a bumpy ‘angular spectrum’.

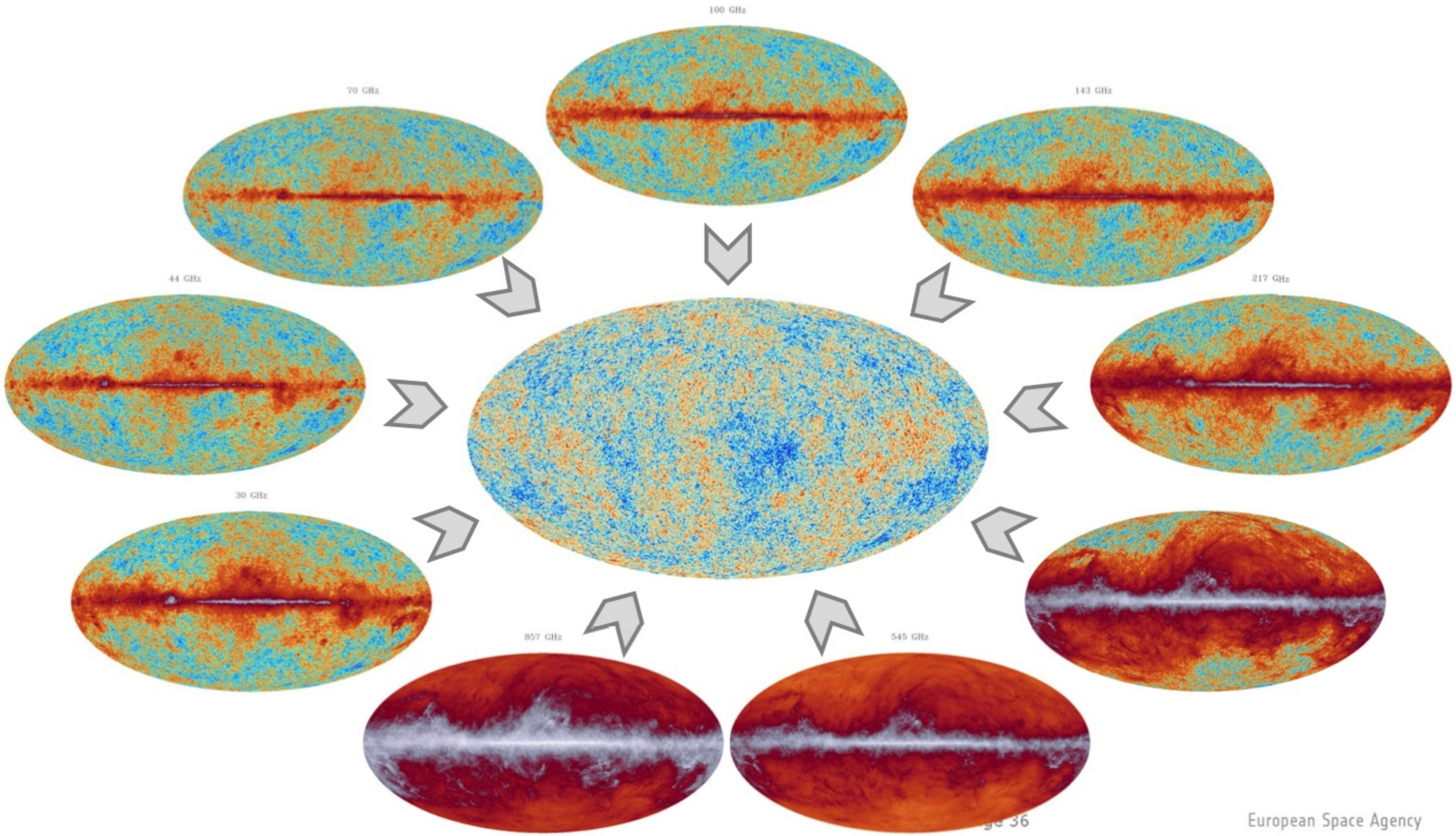



← LIKELIHOOD ANALYSIS. Here, likelihood of matter $\Omega_m$ and vacuum $\Omega$ energy densities in front of CMB data (and supernovae).

→ “Thus”, the Universe is flat and $13.7 \pm 0.2$ billions years old (says WMAP). . .

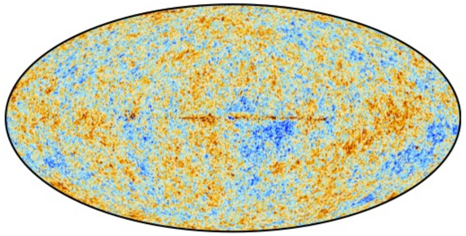
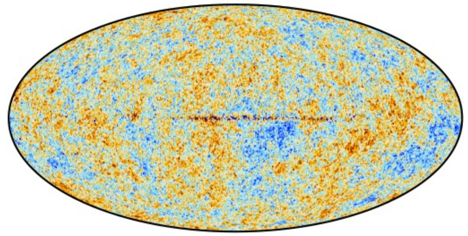
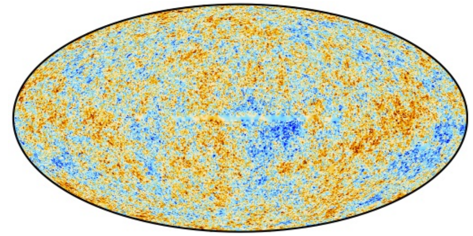
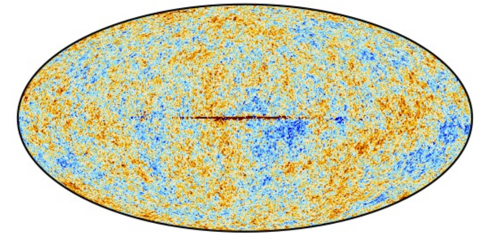
# Component separation

# Extracting the CMB from the 9 Planck frequency channels




Color scale: hundreds of micro-Kelvins. Credits: ESA, FRB.

# Four CMB anisotropy maps delivered to the Planck Legacy Archive

|  |  |  |  |
|---|---|---|---|
| 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).
- The SMICA (Spectral Mismatch ICA) method selected for the 'Main product' for the Planck CMB map.

## 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

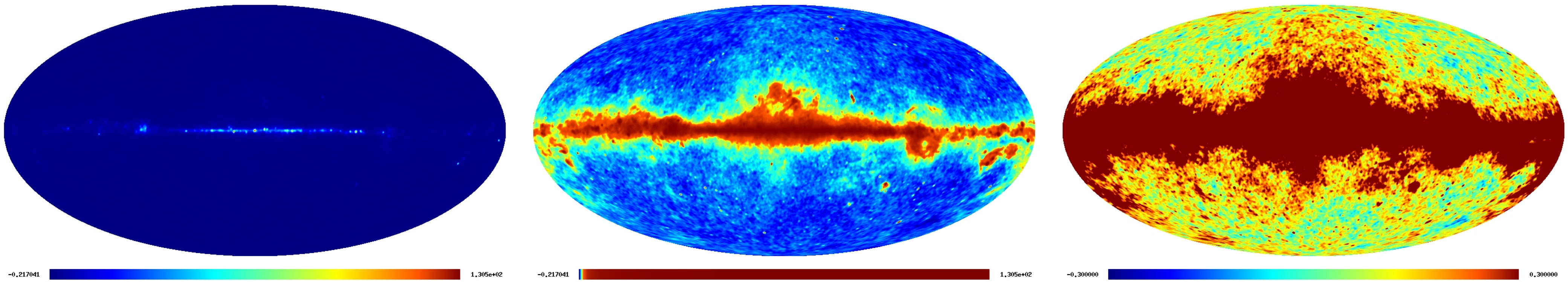
  2. Propagation of simulated individual inputs should be straightforward

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

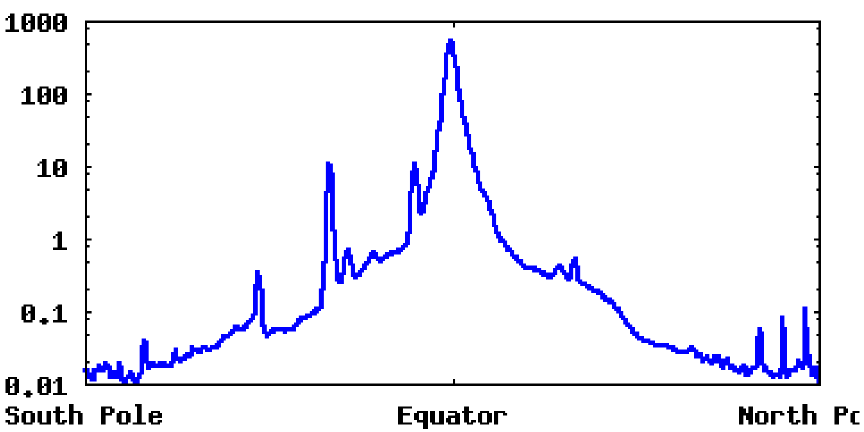
- The method should be fast enough for thousands of Monte-Carlo runs.

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

# Wide dynamics over the sky



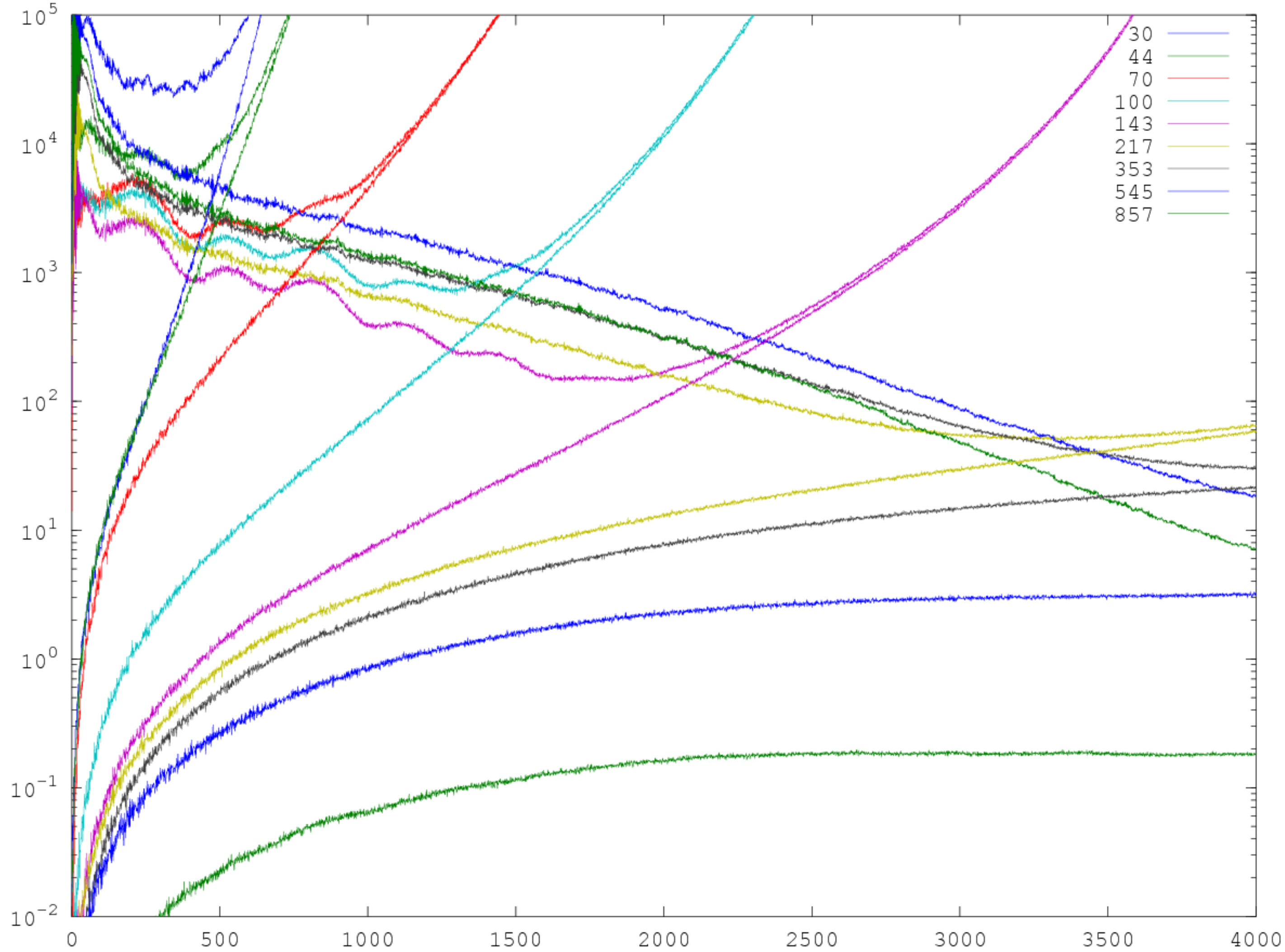

Left: The W-MAP K band. Natural color scale [-200, 130000] $\mu K$.
Middle: Same map with an equalized color scale.
Right: Same map with a color scale adapted to CMB: [-300, 300] $\mu K$.




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

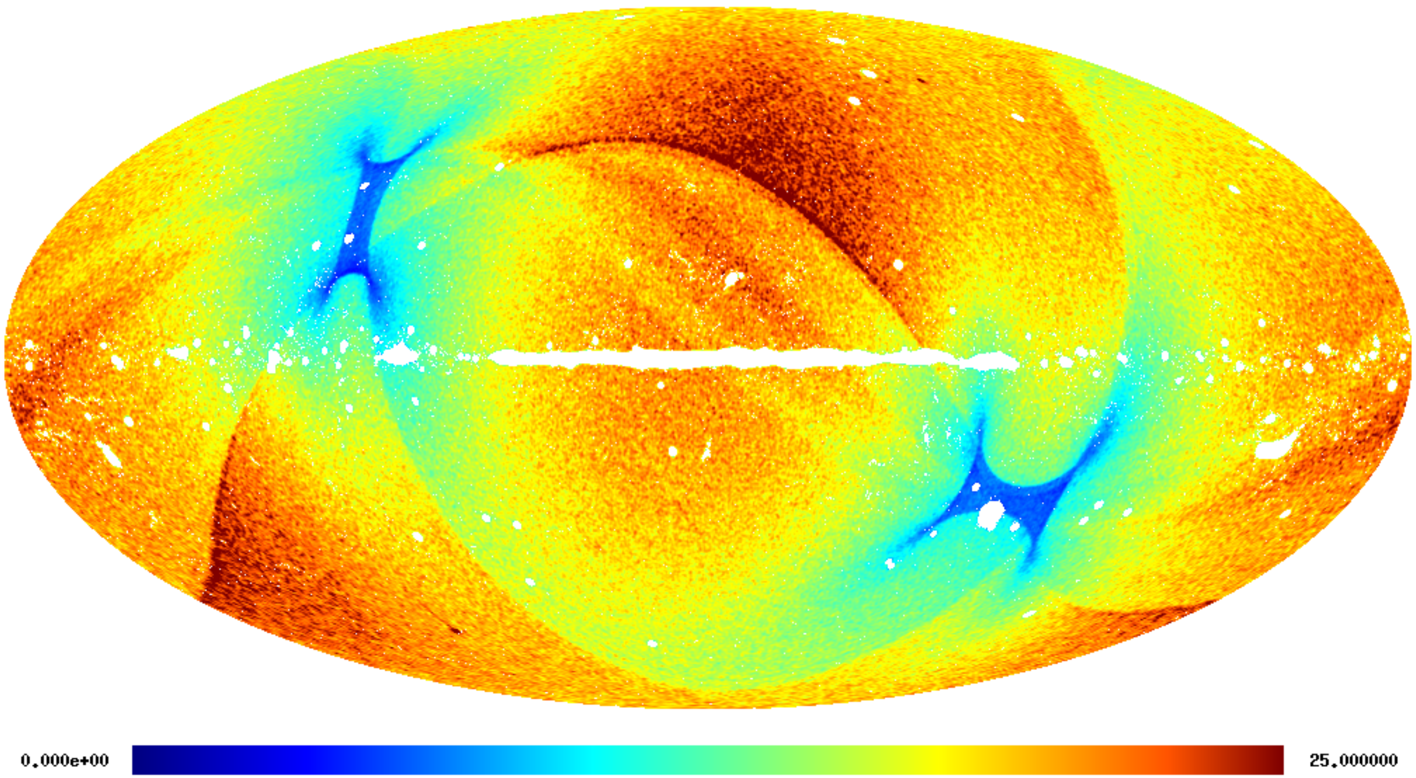
Do we really want to estimate covariance matrices over the whole sky?

# Wide spectral dynamics, SNR variations



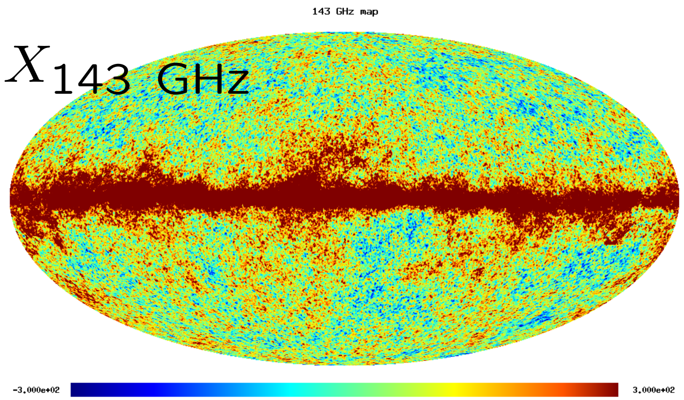

$\widehat{C}(\ell) \cdot \ell(\ell + 1)/2\pi$ in $[\mu K_{\mathrm{RJ}}]^2$ for $f_{\mathrm{sky}} = 0.99$.

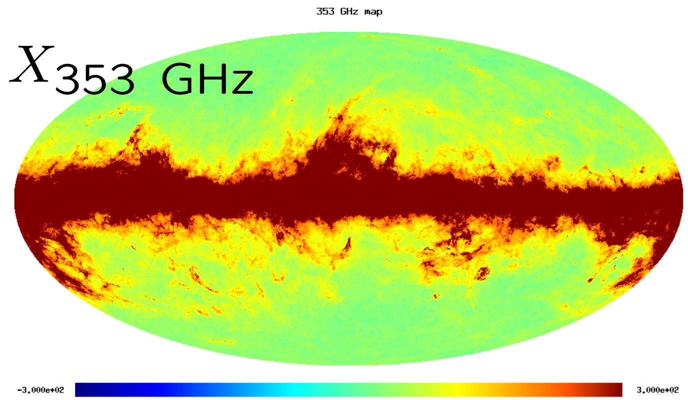
# And what about the noise ?

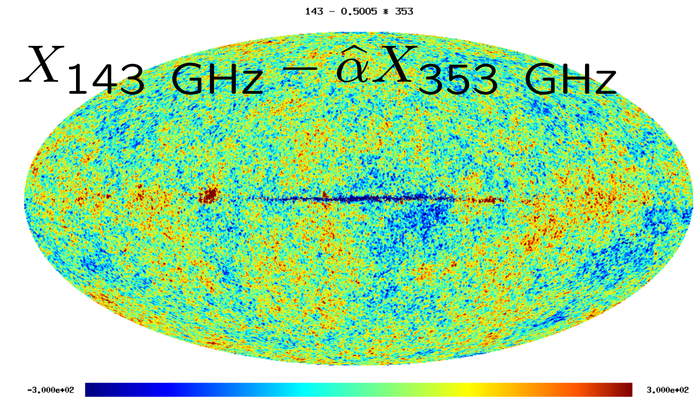



Noise RMS in $\mu K$ in the SMICA map.

# 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:

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

Find $\alpha$ by correlation: $\hat{\alpha} = \langle X_{143} X_{353} \rangle / \langle X_{353} X_{353} \rangle$ and a clean (?) CMB map as

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

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

## A closer look at a toy contamination model

First channel $d_1$ contains the signal of interest $s_1$ with an independent contamination by $s_2$, perfectly observed in the second channel $d_2$:

$$d_1 = s_1 + \alpha s_2$$
$$d_2 = s_2$$

We estimate $\alpha$ by cross-correlation, to get

$$\widehat{s}_1 = d_1 - \widehat{\alpha}\, d_2 = d_1 - \frac{\langle d_1 | d_2 \rangle}{\langle d_2 | d_2 \rangle} d_2 = s_1 - \frac{\langle s_1 | s_2 \rangle}{\langle s_2 | s_2 \rangle} s_2$$

There are 'chance correlations' between $s_1$ and $s_2$ —by which $\langle s_1 | s_2 \rangle \neq 0$— because we cannot average over Universes...

Empirical correlation between two maps, expressed in harmonic space:

$$\langle s_1 | s_2 \rangle = \iint_{S^2} s_1(\vec{n})\, s_2(\vec{n})\, d\vec{n} \; = \; \sum_\ell \sum_m \; s^{(1)}_{\ell,m} s^{(2)}_{\ell,m}$$

Because our random fields have $1/\ell^2$ (say) spectra,
the harmonic sum is dominated by a few large scale coefficients.

## Processing the toy model optimally

First we move losslessly to harmonic space

$$d^{(1)}_{\ell,m} = s^{(1)}_{\ell,m} + \alpha\, s^{(2)}_{\ell,m}$$
$$d^{(2)}_{\ell,m} = s^{(2)}_{\ell,m}$$

The CMB is Gaussian stationnary: $s^{(1)}_{\ell,m} \sim \mathcal{N}(0, C_\ell)$ and, even more importantly, the spherical harmonic coefficients are uncorrelated: the likelihood reads

$$\log p(d^{(1)}|\alpha) = -\frac{1}{2}\sum_\ell \sum_m \left(d^{(1)}_{\ell,m} \;-\; \alpha\, d^{(2)}_{\ell,m}\right)^2 / C_\ell$$

Likelihood maximization wrt $\alpha$ yields a chance correlation :

$$\langle s_1 | s_2 \rangle \;=\; \sum_\ell \sum_m \; s^{(1)}_{\ell,m} s^{(2)}_{\ell,m} / C_\ell$$

**1)** the amplitude of each term in the sum is equalized by the $1/C_\ell$ factors.

**2)** No need to model the contaminant statistically: it is observed deterministically.

## Combining all 9 Planck channels, non parametrically: the ILC

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

$$\widehat{s}(p) = \mathbf{w}^\dagger \mathbf{d}(p) \qquad p = 1, \dots, N_{\mathsf{pix}}$$

At frequency $\nu$, the CMB signal $s(p)$ has amplitude $a_\nu$ and contaminated by $f_\nu(p)$

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

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

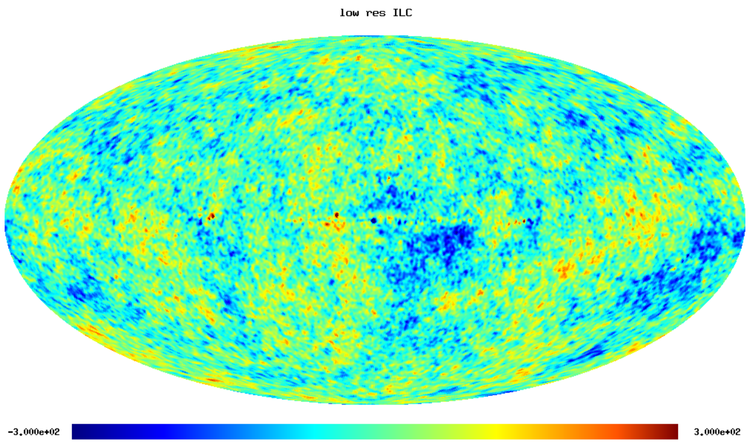
$$\mathbf{w} = \frac{\widehat{\mathbf{C}}^{-1}\,\mathbf{a}}{\mathbf{a}^\dagger \widehat{\mathbf{C}}^{-1}\,\mathbf{a}} \qquad \text{with} \qquad \widehat{\mathbf{C}} = \langle \mathbf{d}\mathbf{d}^\dagger \rangle_p, \quad \text{the sample covariance matrix}$$

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

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

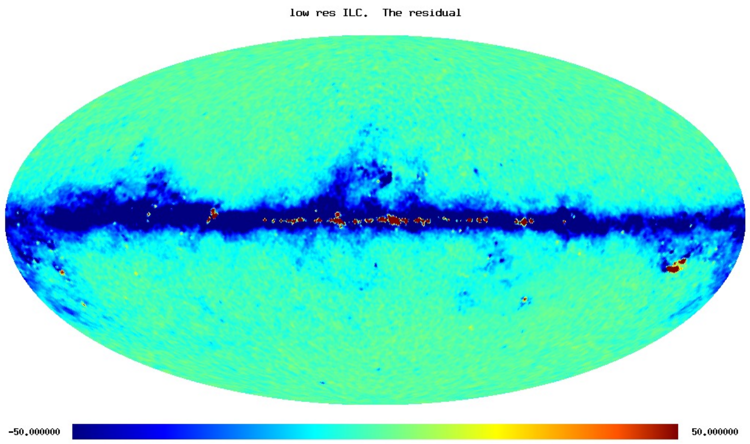
## Is the ILC good enough for Planck data ?

A simulation result:




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

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




ILC looked promising, but something went wrong.

Actually two things, at least, need fixing:

- harmonic dependence and
- chance correlations.

# SMICA: Linear filtering in harmonic space

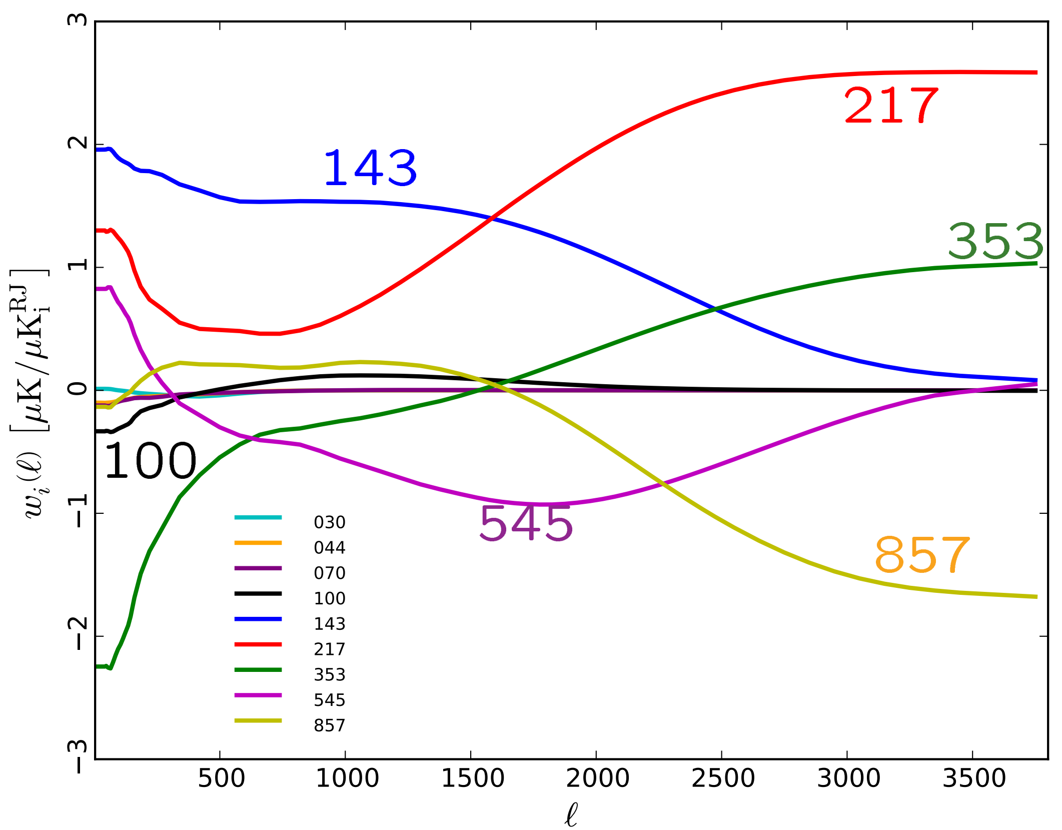
Since resolution, noise and foregrounds vary (wildly) in power over channels and angular frequency, the combining weights should depend on $\ell$.

The SMICA CMB map is synthesized from spherical harmonic coefficients $\hat{s}_{\ell,m}$, obtained as linear combinations:

$$\hat{s}_{\ell,m} = \mathbf{w}_\ell^\dagger \, \mathbf{d}_{\ell,m} \quad \text{with, again,}$$

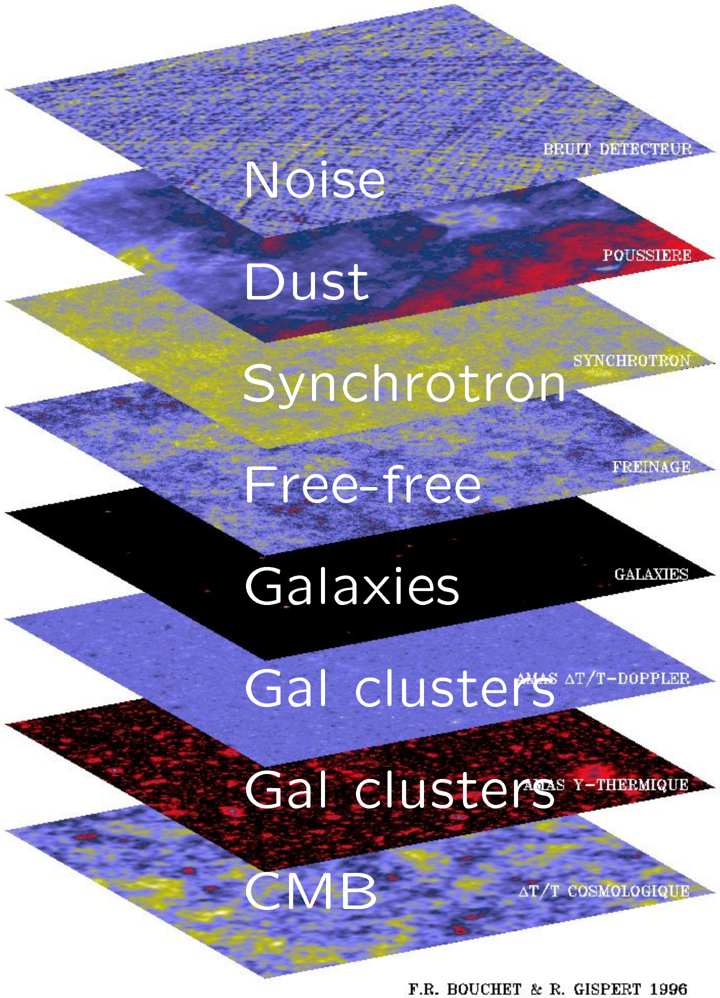
$$\mathbf{w}_\ell = \frac{\mathbf{C}_\ell^{-1}\mathbf{a}}{\mathbf{a}^\dagger \mathbf{C}_\ell^{-1}\mathbf{a}} \qquad \mathbf{C}_\ell = \mathrm{Cov}(\mathbf{d}_{\ell,m})$$

- At high $\ell$, the (spectral) covariance matrices $\mathbf{C}_\ell$ are well estimated by their sample counterparts
- At lower $\ell$, we need to get smarter.




Note: spectral localization is a must. Spatial localization do not seem critical (See NILC perf.).

# Foregrounds and how to get rid of them (at low $\ell$) ?




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.

# Foregrounds, physical components and the mixing matrix

At low $\ell$ (i.e. large angular scales), there are less Fourier modes available for estimating spectral statistics $\widehat{\mathbf{C}}_\ell$: the variability (chance correlations) must be decreased by fitting a model $\mathbf{C}_\ell(\theta)$ to $\widehat{\mathbf{C}}_\ell$.

- 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 $\mathbf{A} = \mathbf{A}(\theta)$:

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

- Which model, which fitting criterion? See next the SMICA case.

# 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 !

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

- Blind identifiability: can it be done? Maths say: yes!

- Fit by $\min_\theta \sum_\ell (2\ell + 1) \left[ \text{trace } \widehat{\mathbf{C}}_\ell \mathbf{C}_\ell(\theta)^{-1} + \log\det \mathbf{C}_\ell(\theta) \right]$ = Gaussian stationary likelihood.

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

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

## Why the stationary Gaussian likelihood is OK (and sparsity useless)

Consider the noise-free, square case : $\mathbf{d} = [\,\mathbf{a} \mid \mathbf{F}\,] \begin{bmatrix} s \\ \mathbf{f} \end{bmatrix}$ with known $\mathbf{a}$.

For any matrix $\mathbf{G}$, the 'preprocessor' $\mathbf{T} = [\,\mathbf{a} \mid \mathbf{G}\,]^{-1}$ ensures

$$\mathbf{TA} = [\,\mathbf{a} \mid \mathbf{G}\,]^{-1}\ [\,\mathbf{a} \mid \mathbf{F}\,] = \begin{bmatrix} 1 & \alpha^\dagger \\ 0 & \mathbf{Y} \end{bmatrix} \quad \text{so that} \quad \widetilde{\mathbf{d}} = \mathbf{Td} = \begin{bmatrix} s + \alpha^\dagger \mathbf{f} \\ \mathbf{Y}\,\mathbf{f} \end{bmatrix}$$

Hence, the first pre-processed channel $\widetilde{\mathbf{d}}_1$ contains the signal of interest $s$ contaminated by a linear combination of the other observed channels $\widetilde{\mathbf{d}}_2, \widetilde{\mathbf{d}}_3, \cdots$.

• Statistical foreground models **not** needed: they are deterministically observed.

• Hence, a statistical model is needed only for the CMB. Since it is Gaussian and stationary, the likelihood has a simple expression which is readily maximized in the cleaning coefficients. That justifies the use of a Gaussian stationary likelihood for fitting the SMICA model.

## More explicitly

Assuming without loss of generality that $\mathbf{Y} = \mathbf{Id}$ and denoting $d = \tilde{\mathbf{d}}_1$, the model is, in each pixel $p$

$$d(p) \;=\; s(p) + \alpha^\dagger \mathbf{f}(p)$$

with $\mathbf{f}(p)$ observed. For the best cleaning, we need the optimal estimate of $\alpha$.

But the likelihood is trivial in harmonic space, with the CMB as 'noise' (!)

$$-2 \log P(d|\alpha) = \sum_\ell (2\ell + 1)\, \frac{(d_{\ell,m} - \alpha^\dagger \mathbf{f}_{\ell,m})^2}{C_\ell} + \text{cst}$$

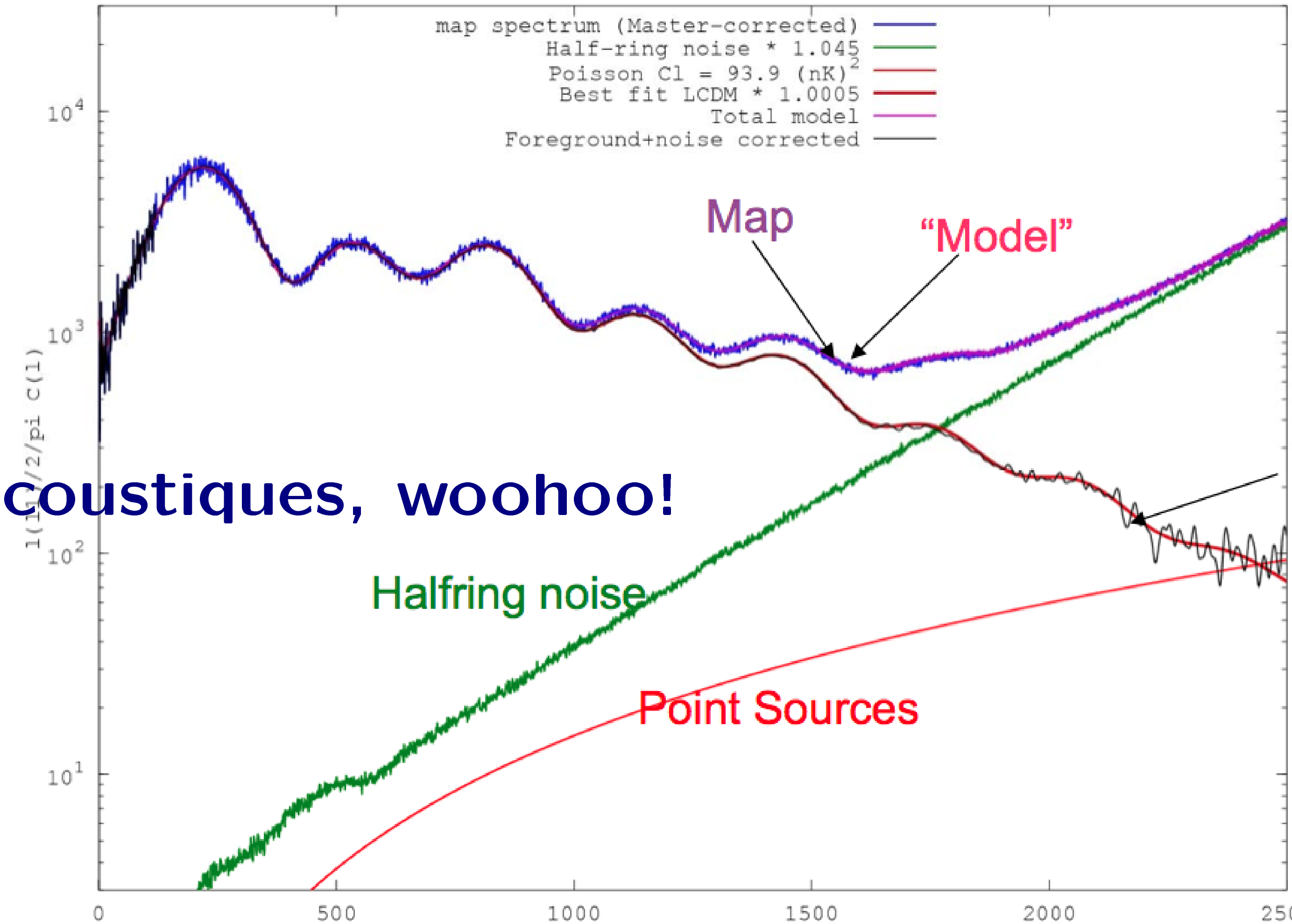
The (trivial) solution corresponds to combining the inputs with weight vector

$$\mathbf{w} = \frac{\widehat{\mathbf{C}}^{-1}\,\mathbf{a}}{\mathbf{a}^\dagger \widehat{\mathbf{C}}^{-1}\,\mathbf{a}} \qquad \text{i.e. an ILC with} \qquad \widehat{\mathbf{C}} = \sum_\ell \sum_m \mathbf{d}_{\ell,m} \mathbf{d}_{\ell,m}^\dagger / C_\ell$$

This is **also** the SMICA solution in the same context:
chance correlation is optimally mitigated in the spectral domain.

# Some results

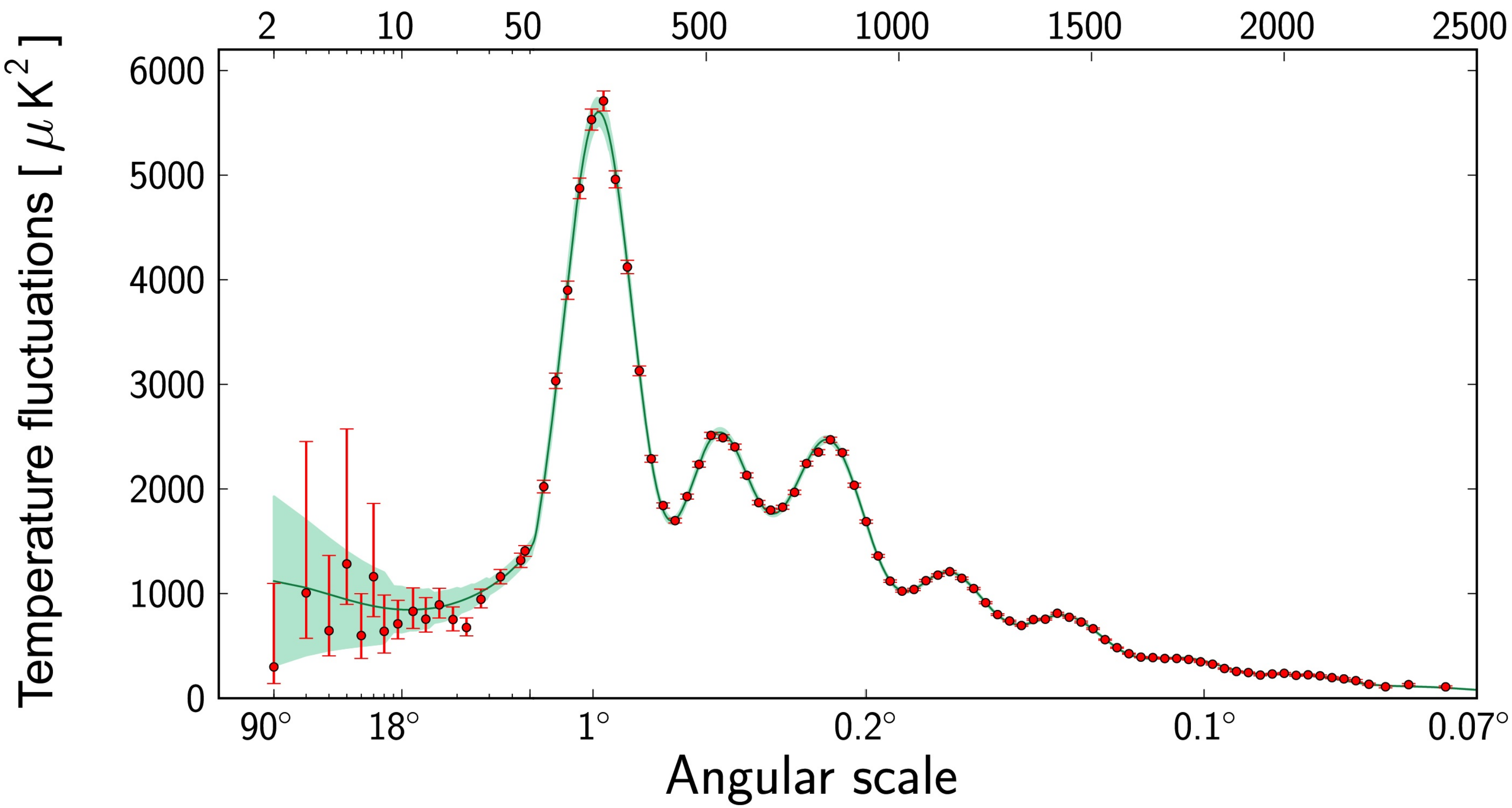
# Spectre angulaire de la carte Planck sur 89% du ciel.




**Sept pics acoustiques, woohoo!**

Les erreurs sont dominées par la variance cosmique jusqu'à $\ell = 1500$ (disons).

## Le spectre angulaire de Planck: contact entre théorie et observations

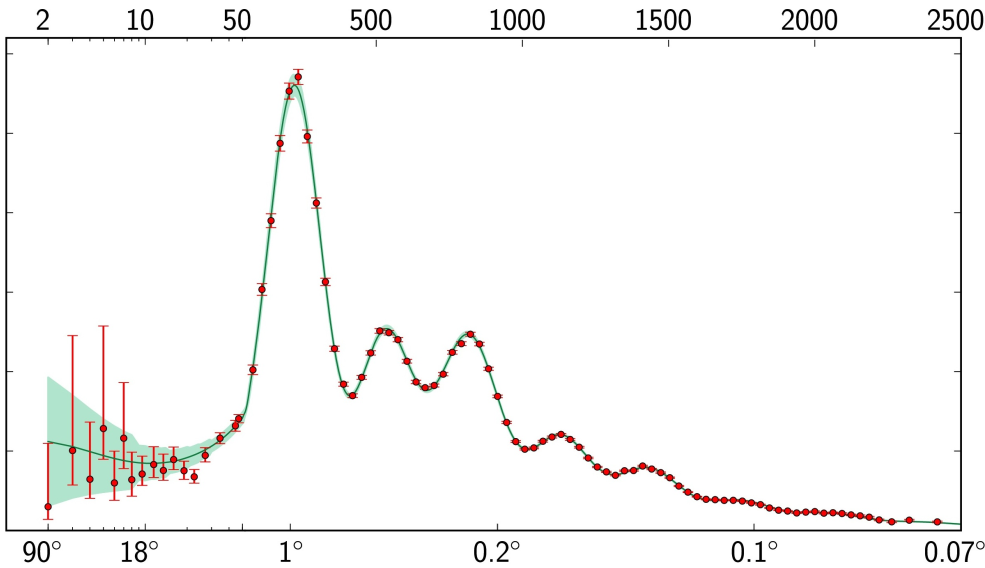



Un superbe ajustement des mesures par les prédictions du plus simple des modèles cosmologiques: le modèle $\Lambda - \mathrm{CDM}$ à 6 paramètres.

# Le modèle standard du Big Bang

Un scénario à 6 paramètres: Le modèle $\Lambda$-CDM.

1. Amplitude $A$ des fluctuations primordiales
2. Leur indice spectral $n_s$
3. Densité de matière noire $\Omega_d$
4. Densité d'énergie noire $\Omega_\Lambda$
5. Taux d'expansion $H_o$ de l'Univers
6. Profondeur optique de ré-ionisation $\tau$

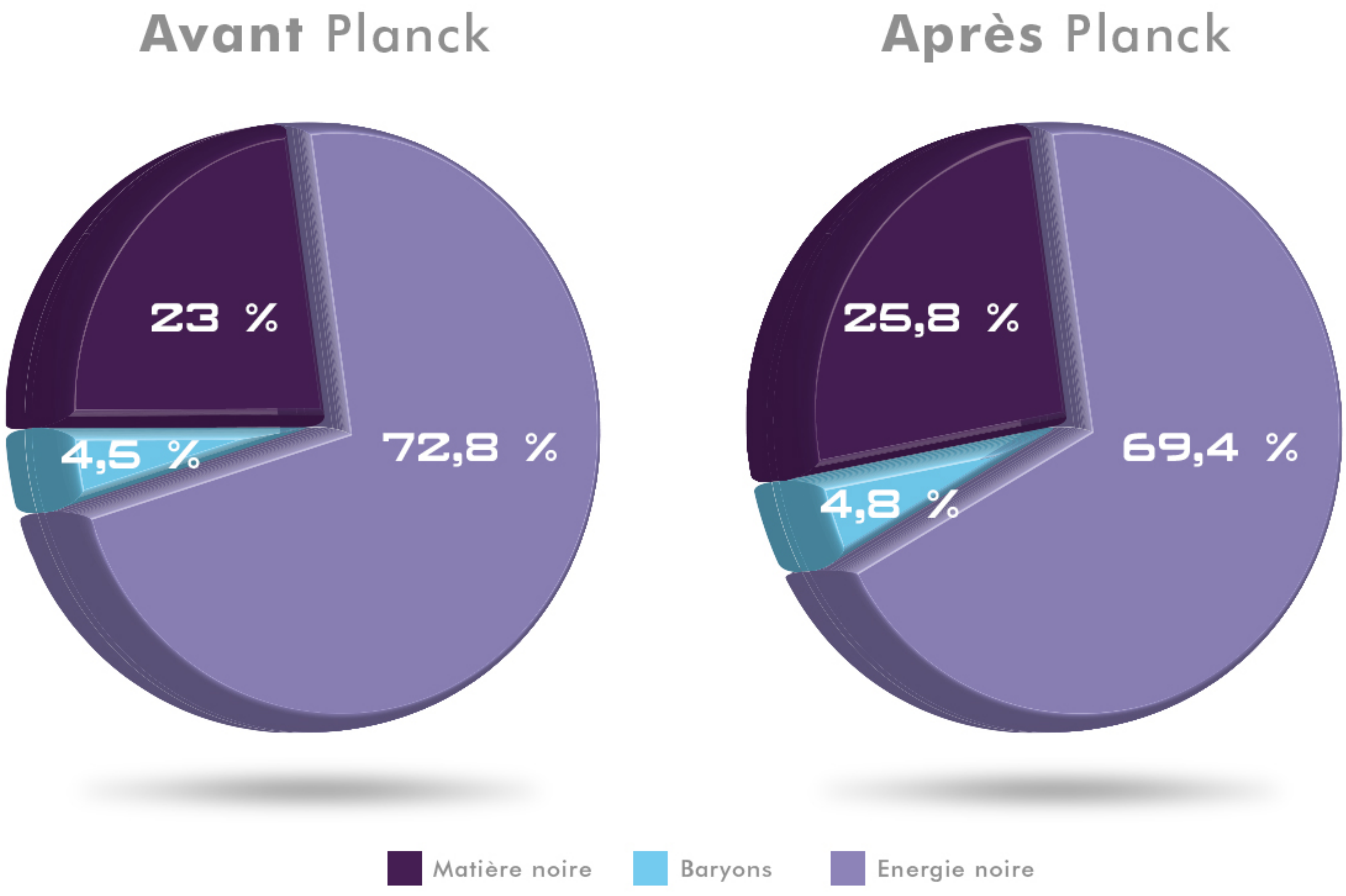



.

Les 2 premiers paramètres $A$ et $n_s$ décrivent le spectre primordial $\rightarrow$ inflation.

Les trois suivants $H_o$, $\Omega_d$, $\Omega_\Lambda$ contrôlent sa "mise en forme".

## Quelques résultats

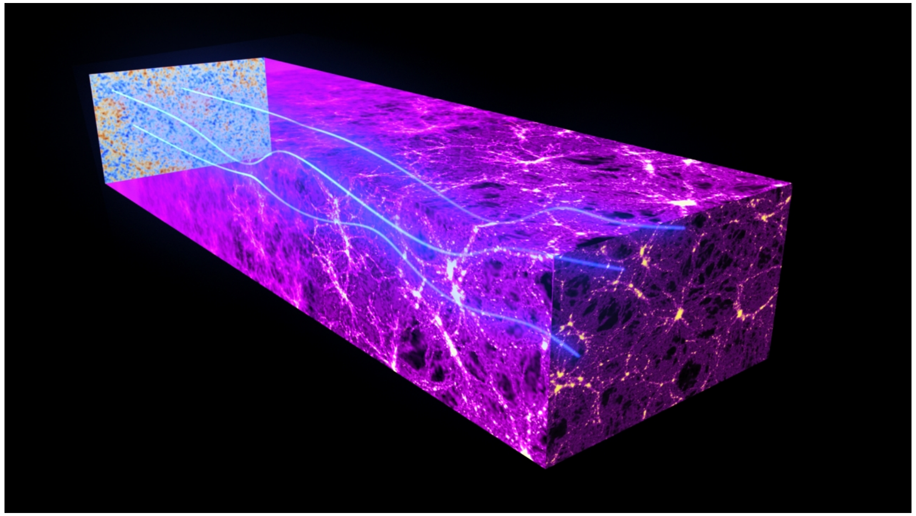


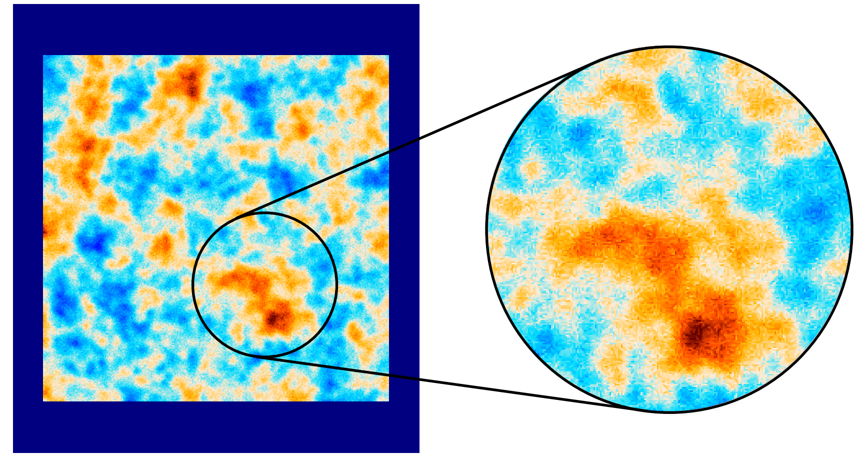

Un taux d'expansion $H_o$ de 67,15 km/s/Mpc et un âge de 13,8 milliards d'années.

# Planck results

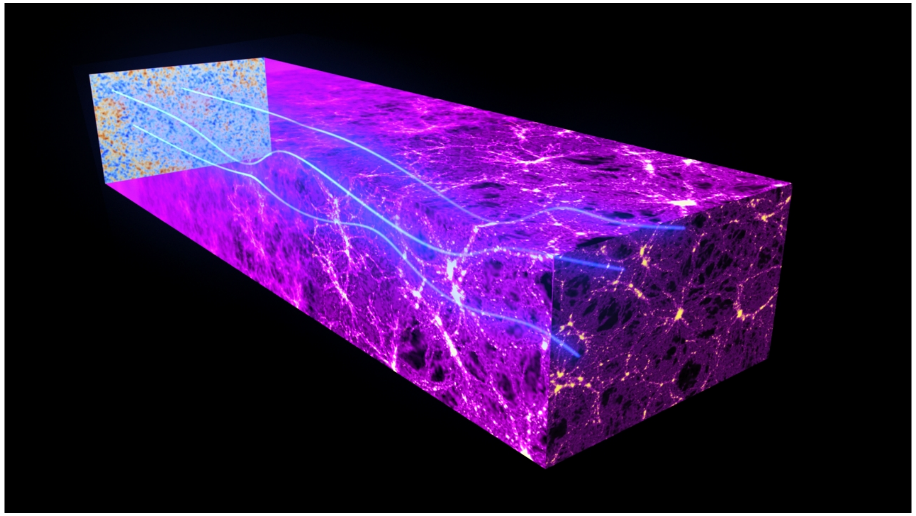
- La plus grosse surprise: pas de grosse surprise.
- Excellente prédiction des observations avec un modèle simple...
  mais un peu bizarre: $\Lambda$-CDM = inconnu + inconnu !
- Sans parler de quelques anomalies marginales...
- Beaucoup d'autres façons d'exploter les données Planck,
  tant en Cosmologie qu'en Astrophysique.
- Magnifique succès scientifique, mais ce n'est pas fini...

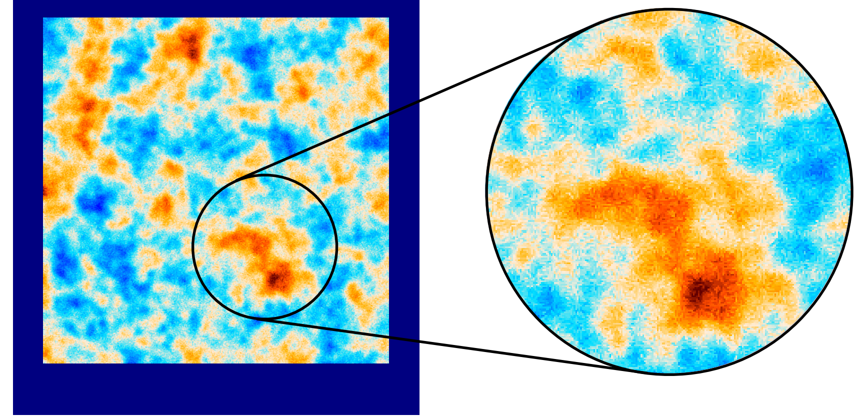
# Plus de science: lentillage gravitationnel





# Plus de science: lentillage gravitationnel





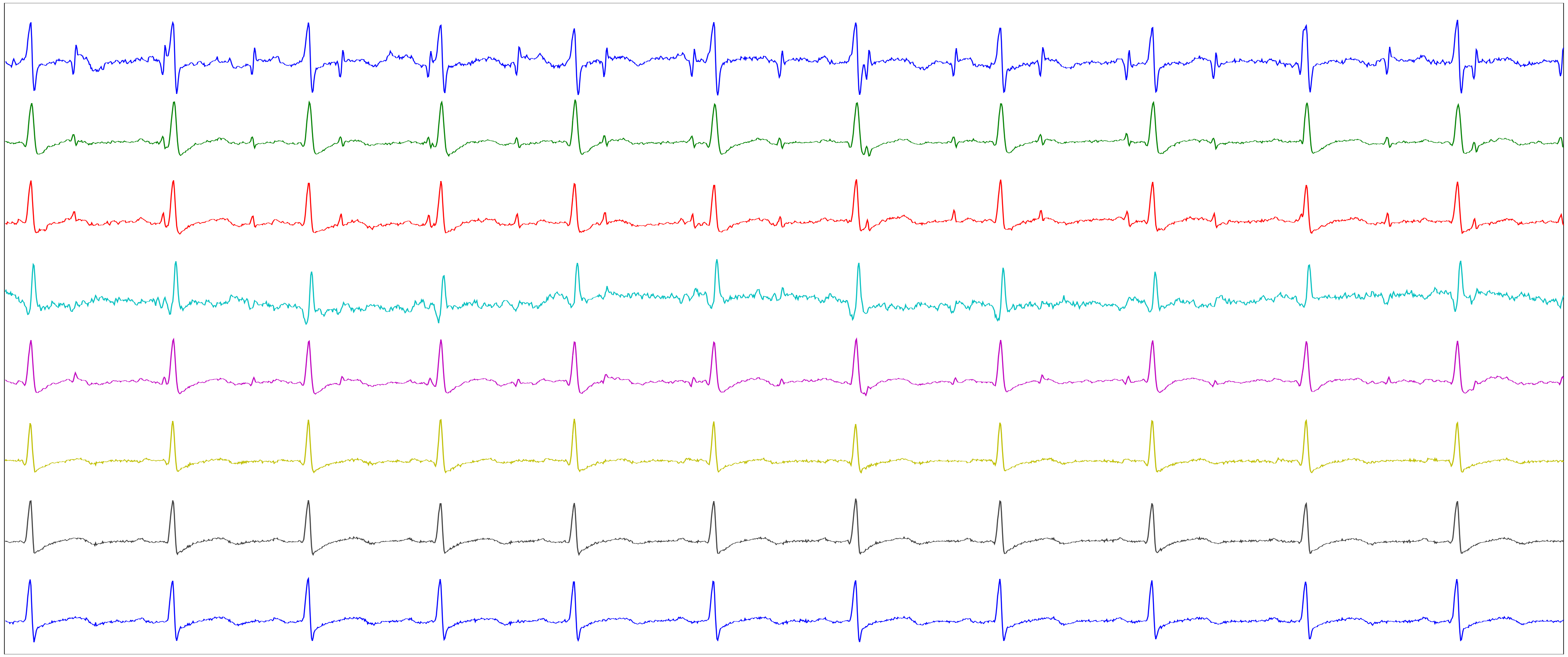
# Thanks and get ready for polarization!

**The scientific results that we present today are a product of the Planck Collaboration, including individuals from more than 100 scientific institutes in Europe, the USA and Canada**

Planck is a project of the European Space Agency, with instruments provided by two scientific Consortia funded by ESA member states (in particular the lead countries: France and Italy) with contributions from NASA (USA), and telescope reflectors provided in a collaboration between ESA and a scientific Consortium led and funded by Denmark.

# A few slides about ICA

# A nice ECG data set



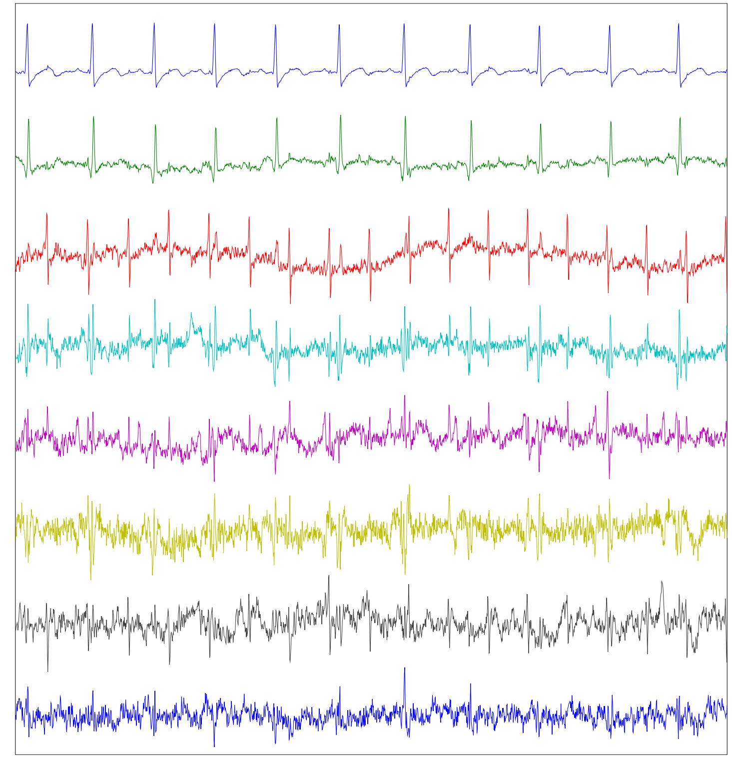
8 electrodes located on the thorax and the abdomen of a pregant woman.

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

Data Credits: Daisy database

# Principal component analysis

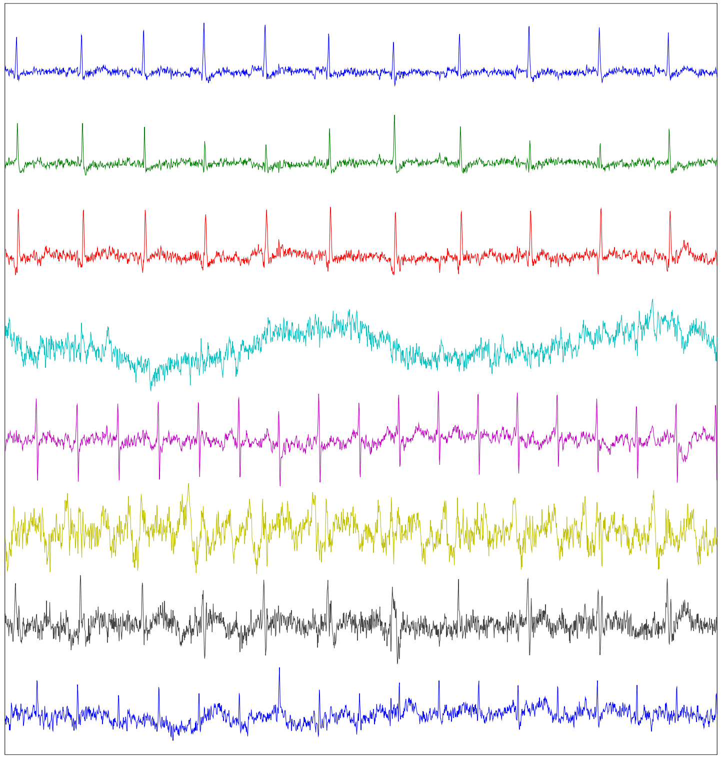
$$= A_{\mathsf{PCA}}^{(8\times 8)} \times$$



- 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.

# Independent component analysis

$$= A_{\mathrm{ICA}}^{(8\times 8)} \times$$



- 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

# The basic ICA model

- An $n \times T$ data set $X = \{x_i(t) \mid 1 \le i \le n,\ 1 \le t \le T\}$ ...

  ... modelled as $X = AS$ with an $n \times T$ source matrix of independent rows.

$$X = A \times \begin{bmatrix} \cdots & S_1 & \cdots \\ & \vdots & \\ \cdots & S_n & \cdots \end{bmatrix}$$

- Two lines of concern for applicability:

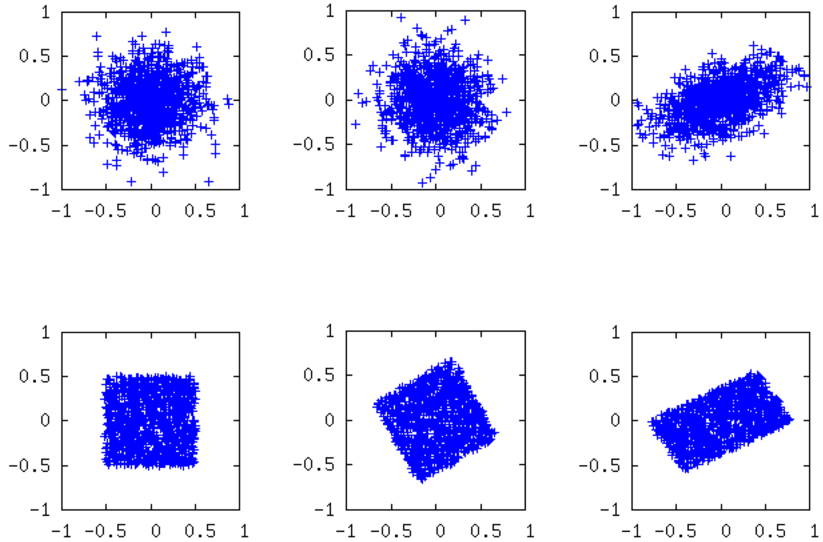
•• Modelling with the component model $X = AS$:
Linear mixing ? Instantaneous mixing ? Square mixing ? Need for noise?

•• Blind identifiability:
Assuming $X = AS$, is the assumption of statistical independence sufficient to recover the sources blindly and efficiently ?

Blindness: why not? If mixing always destroys independence, then recovering independence should unmix, shouldn't it?

# A caricature

Mixing Gaussian (top row) or uniform(bottom) random variables with the identity (left), a rotation (center), a generic transform (right):




Rotation is invisible unless the data **and** the model are non Gaussian.

Mere decorrelation: $\frac{1}{T}\sum_t y_i(t)y_j(t) = 0$ does not cut it.

# How to do it ?

- Ignore time structure but use non Gaussianity . . .
  - Minimize dependence between recovered sources . . .
  - . . . as measured by mutual information, why ? hard !
  - . . . or approximated using high order cumulants, clumsy ?
  - or find the most non Gaussian sources why ? how ?
  - or find sources uncorrelated through non linear functions: *e.g* $\frac{1}{T}\sum_t \phi(y_i(t))\psi(y_j(t)) = 0$ for $i \neq j$. Which functions $\phi$, $\psi$ ?
- . . . or use temporal structure, such as
  - . . . temporal correlations, or spectral diversity,
  - . . . or non stationarity

But how to make sense of all that ? How to do it properly ?

.

# Equivariance and uniform performance

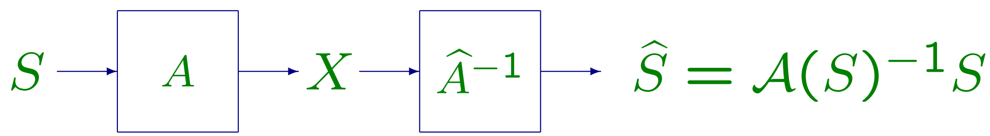
We are contemplating a transformation model: $X = AS$, $A \in \mathrm{GL}(n)$.

An estimator of $A$, *i.e.* a function $X \to \widehat{A} = \mathcal{A}(X)$, is called equivariant if

$$\mathcal{A}(MX) = M\mathcal{A}(X) \quad \text{for any invertible transform } M.$$

Equivariant estimators have uniform performance in the mixing:

$$\widehat{S} = \widehat{A}^{-1}X = \mathcal{A}(AS)^{-1}\, AS = \mathcal{A}(S)^{-1}S \qquad A \text{ is gone!}$$



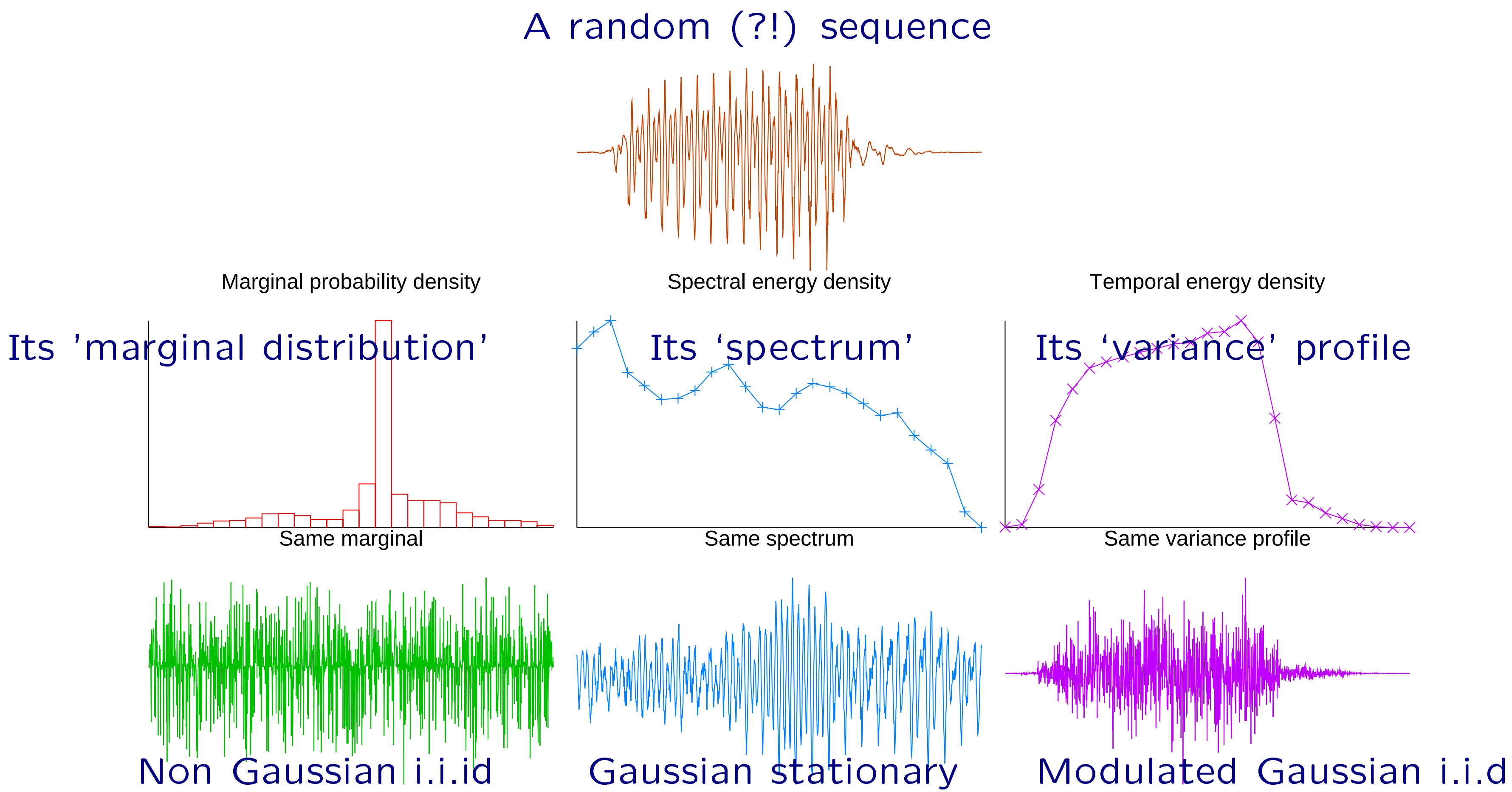

For ICA, any estimator obtained as a function of $Y = A^{-1}X$ only, for instance

$$\text{Maximize} \sum_{i=1}^{n} |\mathrm{kurt}(Y_i)| \text{ subject to } \mathrm{Cov}(Y) = I_n$$

is equivariant.

Uniform performance is fine, but we want it to be (uniformly) good ! MLE ?

# Three points of view on a time series.




- *All models are wrong, but some are useful* — George Box
- *. . . especially for ICA.* — JFC.

# Three simple models away from Gaussian whiteness

Simple statistical models for the $i$-th component $\{S_i(t)\}_{t=1}^T$

0) Gaussian i.i.d. with variance $\sigma_i^2$.

1) Non Gaussian i.i.d. with marginal distribution $p_i(S_i(t))$.

2) Gaussian stationary with power spectrum $C_i(\ell)$.

3) Gaussian independent with variance profile $\sigma_i^2(t)$.

Model 0) has a single parameter and is not strong enough for blind separation: its likelihood leads to measuring independence by global decorrelation.

Models 1), 2) and 3) are parameterized by a one-dimensional density:
a probability density, a spectral density, a temporal density.
They are strong enough for blind separation even with a rough (or very rough) estimation of the parameter function.

# How likelihood+model characterize independence

For a given type of component model, the maximum likelihood estimate $\hat{A}$ of $A$ is characterized by a condition on the recovered components $Y = \hat{A}^{-1}X$:

- for a **Gaussian i.i.d model**

Max. likelihood $\rightarrow \frac{1}{T}\sum_t y_i(t)y_j(t) = 0$ : plain **symmetric** decorrelation = **bad**.

- for **non Gaussian models** with marginal pdf $p_i$ for $s_i(t)$:

Define the (non-linear) score function $\psi_i(u) = -p_i'(u)/p_i(u)$, then

$$\text{Max. likelihood} \longrightarrow \frac{1}{T}\sum_{t=1}^{T}\psi_i(y_i(t))\, y_j(t) = 0 \quad \text{for} \quad i \neq j$$

- for **non stationary Gaussian models** with local variance $\sigma_i^2(t)$ for $s_i(t)$:

$$\text{Max. likelihood} \longrightarrow \frac{1}{T}\sum_{t=1}^{T}\frac{y_i(t)}{\sigma_i^2(t)}\, y_j(t) = 0 \quad \text{for} \quad i \neq j$$

- for **stationary Gaussian models**: Do just as above, but in Fourier space.

The 'local variance' is nothing but the power spectrum.

.

# Gaussian stationary models on the sphere

For $n$ sky maps in channels $\nu_1, \ldots, \nu_n$, modelled as Gaussian stationary:

$$-\log P(\text{data}|\text{model}) = \sum_\ell (2\ell + 1)\, \mathcal{K}\Big(\widehat{\mathbf{C}}(\ell), \mathbf{C}(\ell)\Big) + \text{cst}$$

where the empirical spectral covariance matrix $\widehat{\mathbf{C}}(\ell)$ has entries:

$$[\widehat{\mathbf{C}}(\ell)]_{\nu_1,\nu_2} = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{m=+\ell} \mathbf{d}_{\ell,m}^{\nu_1} \mathbf{d}_{\ell,m}^{\nu_2},$$

and is the natural estimate of the theory spectrum $\mathbf{C}(\ell)$.

The divergence between the data (effectively compressed into $\widehat{\mathbf{C}}(\ell)$) and the model $\mathbf{C}(\ell)$ appears to be measured by

$$\mathcal{K}\Big(\mathbf{C}_1, \mathbf{C}_2\Big) = \text{trace}\Big(\mathbf{C}_1 \mathbf{C}_2^{-1}\Big) - \log\det\Big(\mathbf{C}_1 \mathbf{C}_2^{-1}\Big) - n$$

Back to the SMICA method.

.