The inconvenience of a single Universe.

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

Institut d'Astrophysique de Paris

- The photons who come in cold

- A single Universe

Mathematics of Imaging Workshop #2
Institut Henri Poincaré, March 2019

The inconvenience of a single planet

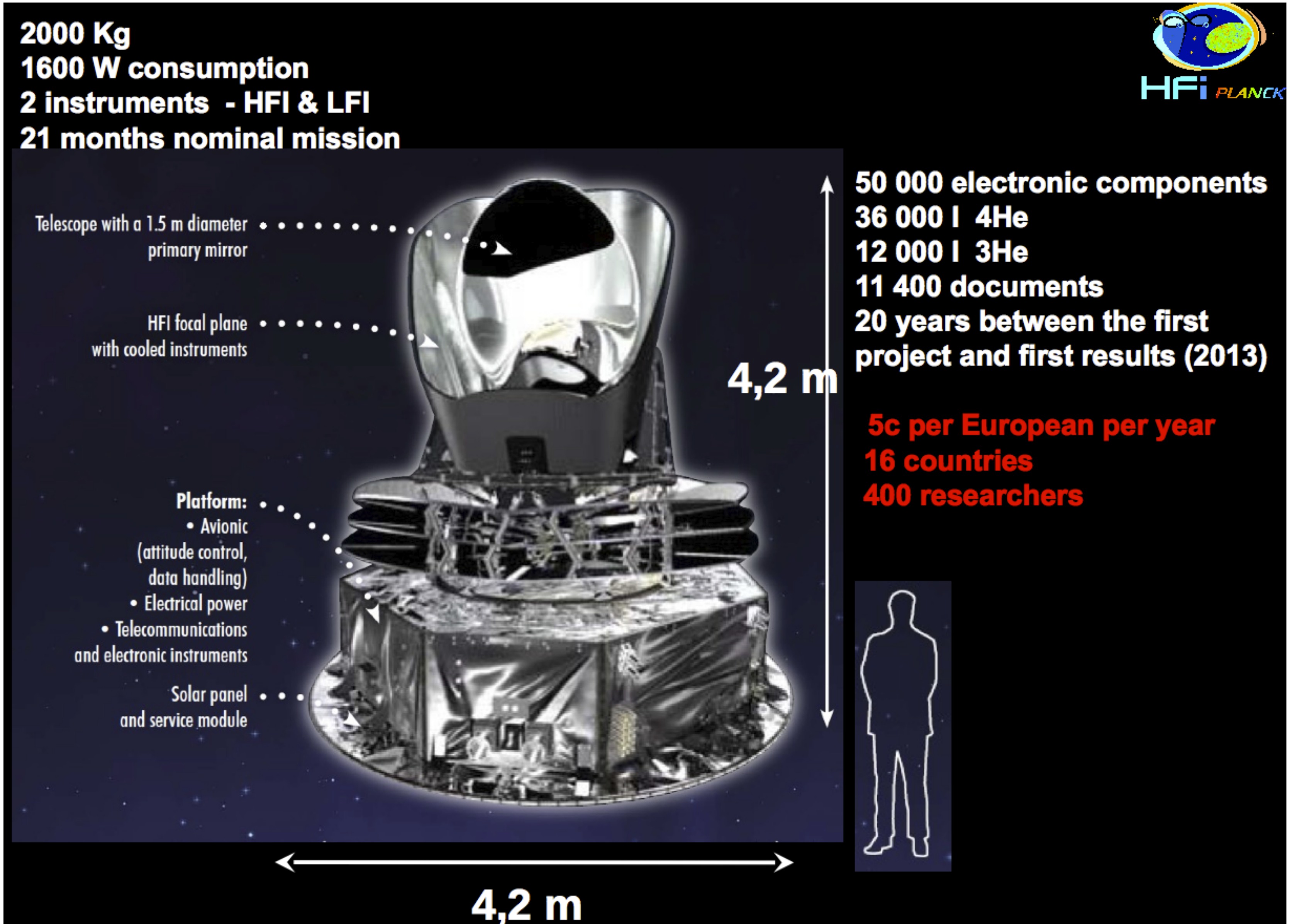
The inconvenience of a single planet

There is no planet B.

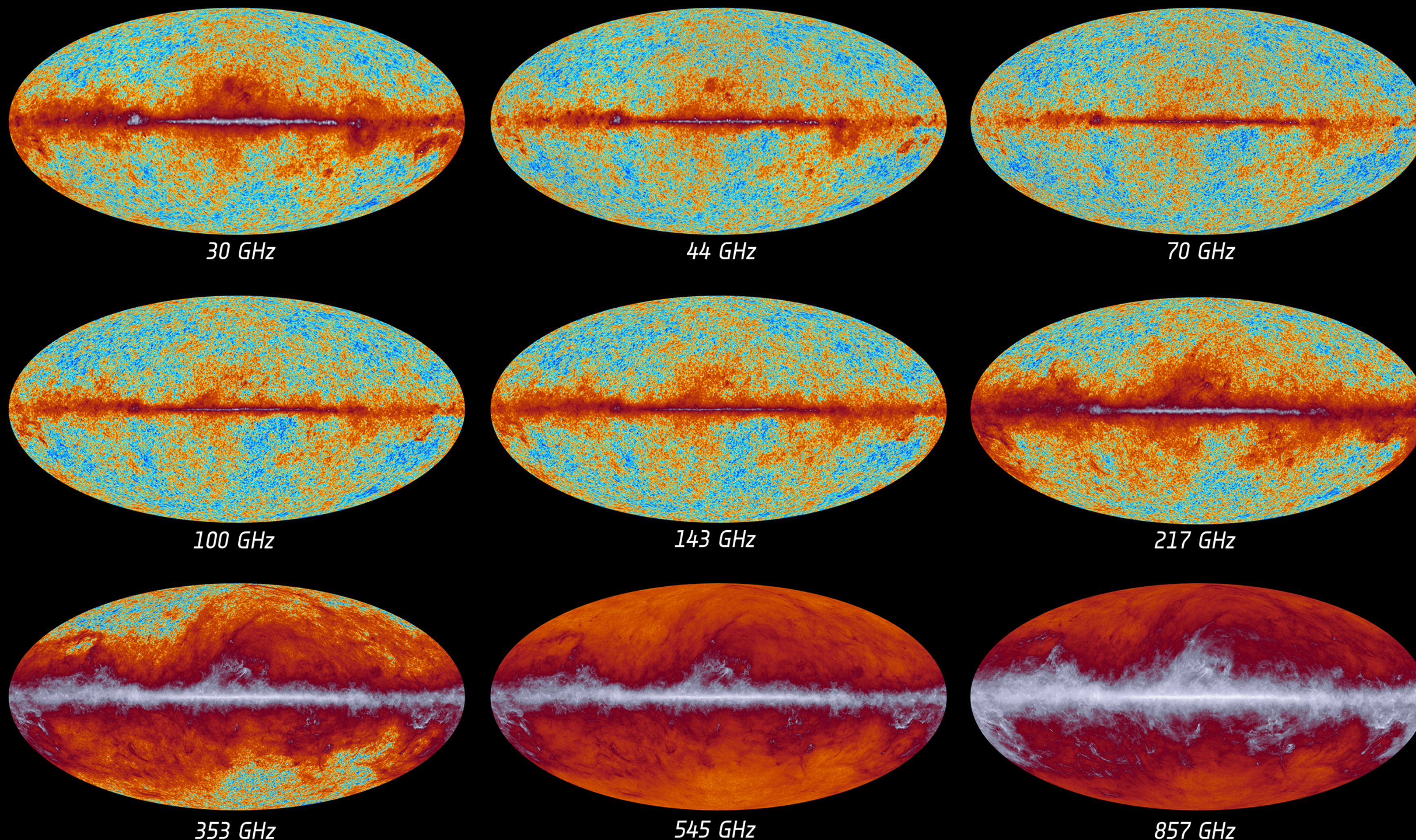
The inconvenience of a single Universe

The photons who come in cold: Big Bang theory

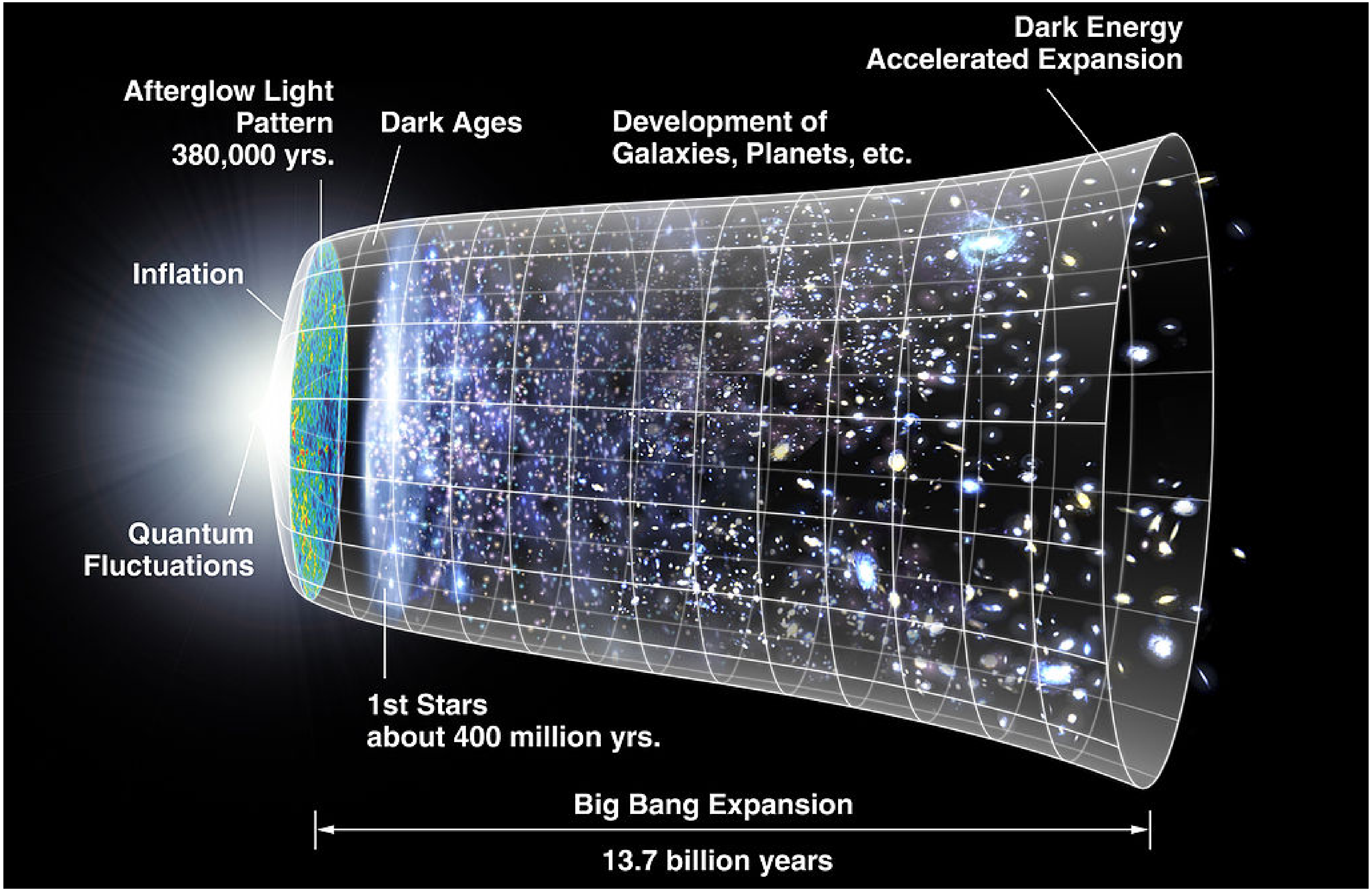
The Planck mission from the European Spatial Agency



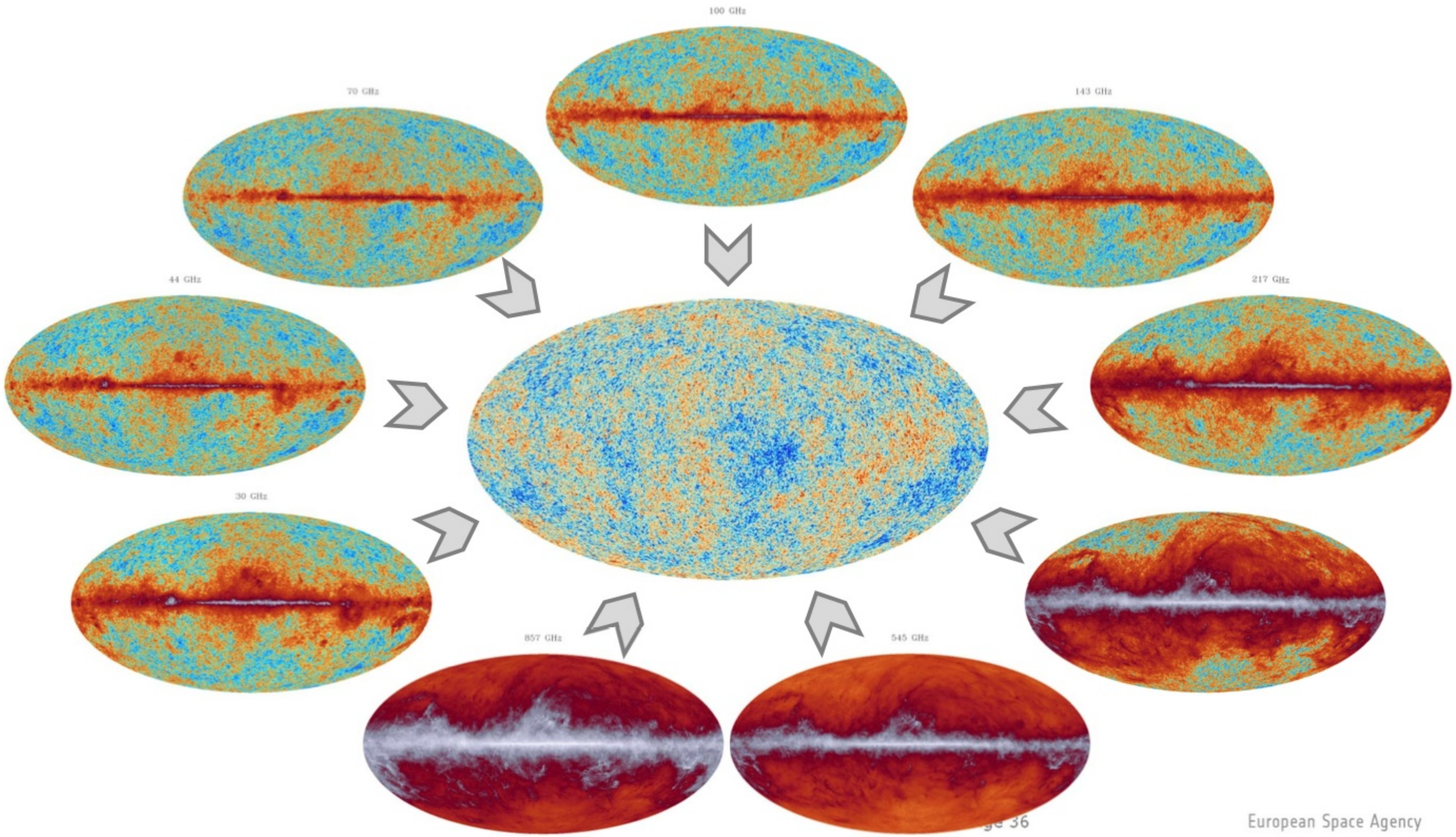
The sky as seen by Planck



Cosmic background (and pesky foregrounds)



Extracting the CMB from Planck frequency channels



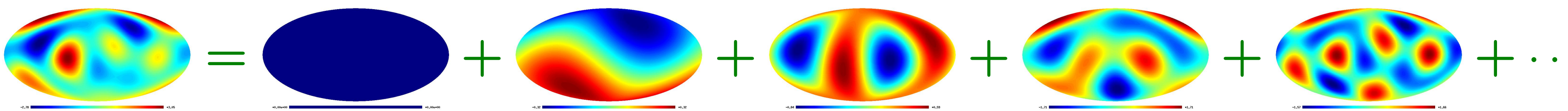
Color scale: hundreds of micro-Kelvins. Credits: ESA, FRB.

Multipole decomposition and angular frequencies

- A spherical field $X(\theta,\phi)$ can be 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:



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

- Projection onto the $(2\ell + 1)$-dimensional spherically invariant subspaces.

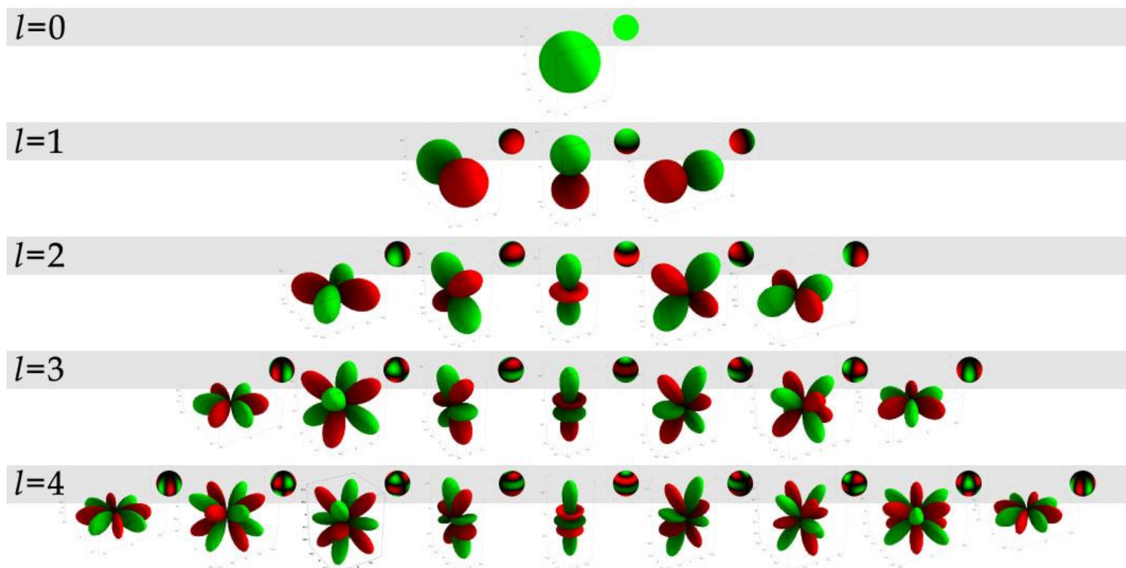
- Distribution of energy across (angular) scales quantified by

  the (empirical) angular spectrum: $\quad \widehat{C}_\ell \stackrel{\text{def}}{=} \dfrac{1}{2\ell+1} \iint_{S^2} X^{(\ell)}(\theta,\phi)^2.$

Fourier on the sphere: Spherical harmonic decomposition

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

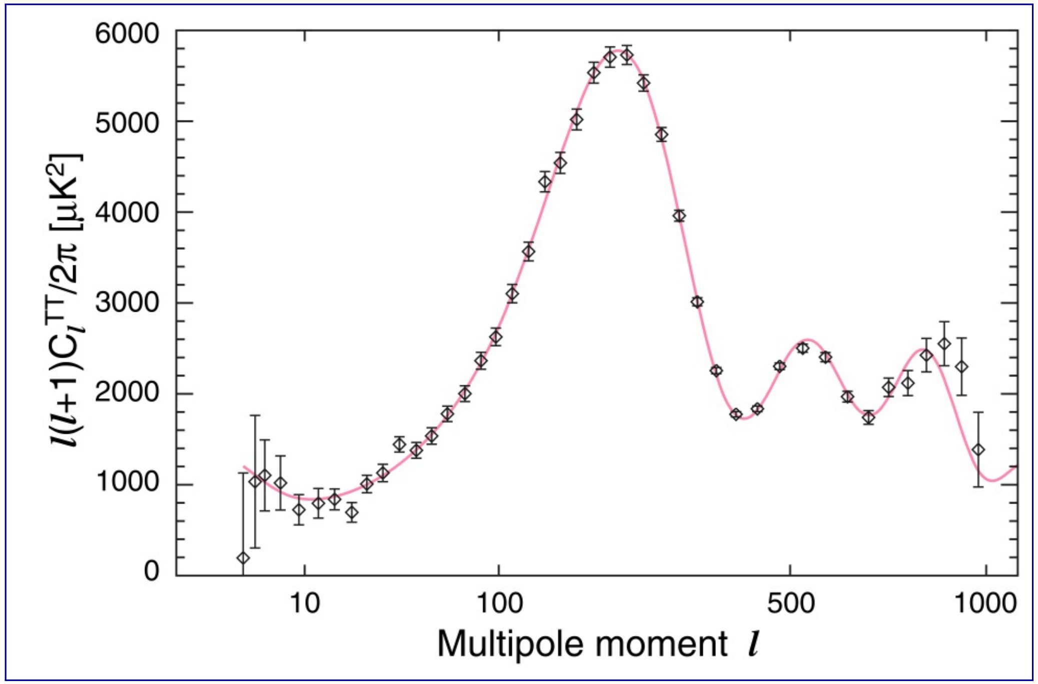
$$X(\theta,\phi) = \sum_{\ell \geq 0} \sum_{-\ell \leq m \leq \ell} a_{\ell m} \; Y_{\ell m}(\theta,\phi) \quad \longleftrightarrow \quad a_{\ell m} \; = \int_\theta \int_\phi Y_{\ell m}(\theta,\phi) \; X(\theta,\phi)$$



- Multipole decomposition and angular spectrum:

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

Angular spectrum of the CMB (as measured/fitted by W-MAP)



- Large scales dominate. We plot:

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

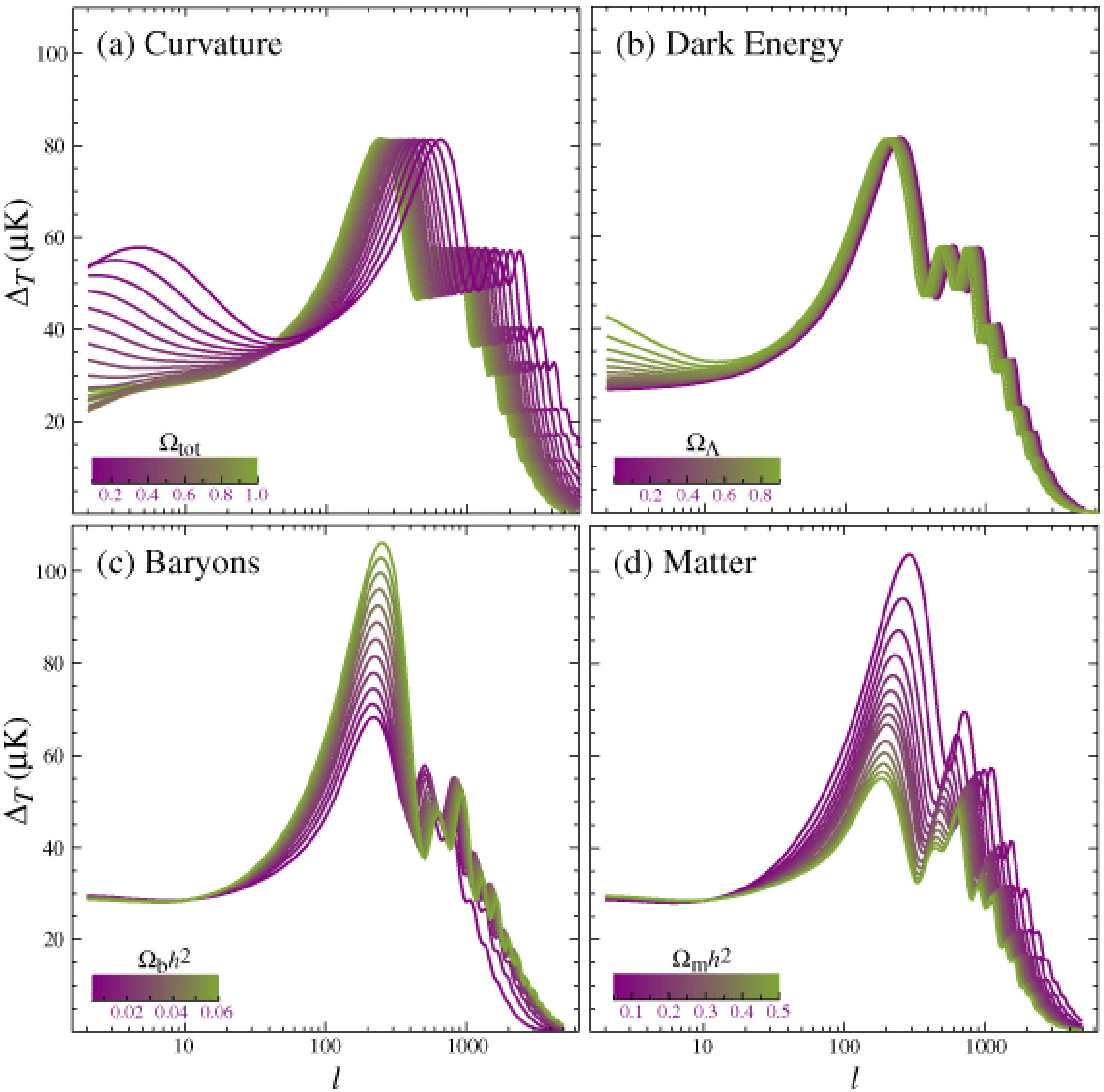
- Acoustic peaks !!!

- One Universe: cosmic variance.

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

then $\mathrm{Var}\left(\hat{C}_\ell / \mathbb{E}\hat{C}_\ell\right) = \frac{2}{2\ell+1}$.

Theoretical angular spectrum of the CMB



A cosmological model predicts the angular spectrum of the CMB as a function of "cosmological parameters".

Examples of the dependence of the spectrum on some parameters of the $\Lambda-$CDM model.

Angular spectrum and likelihood (ideally)

- The spherical harmonic coefficients $a_{\ell m}$ of a stationary random field are uncorrelated with variance C_ℓ, defining 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 <u>empirical spectrum</u>

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

is a sufficient statistic since the likelihood then reads:

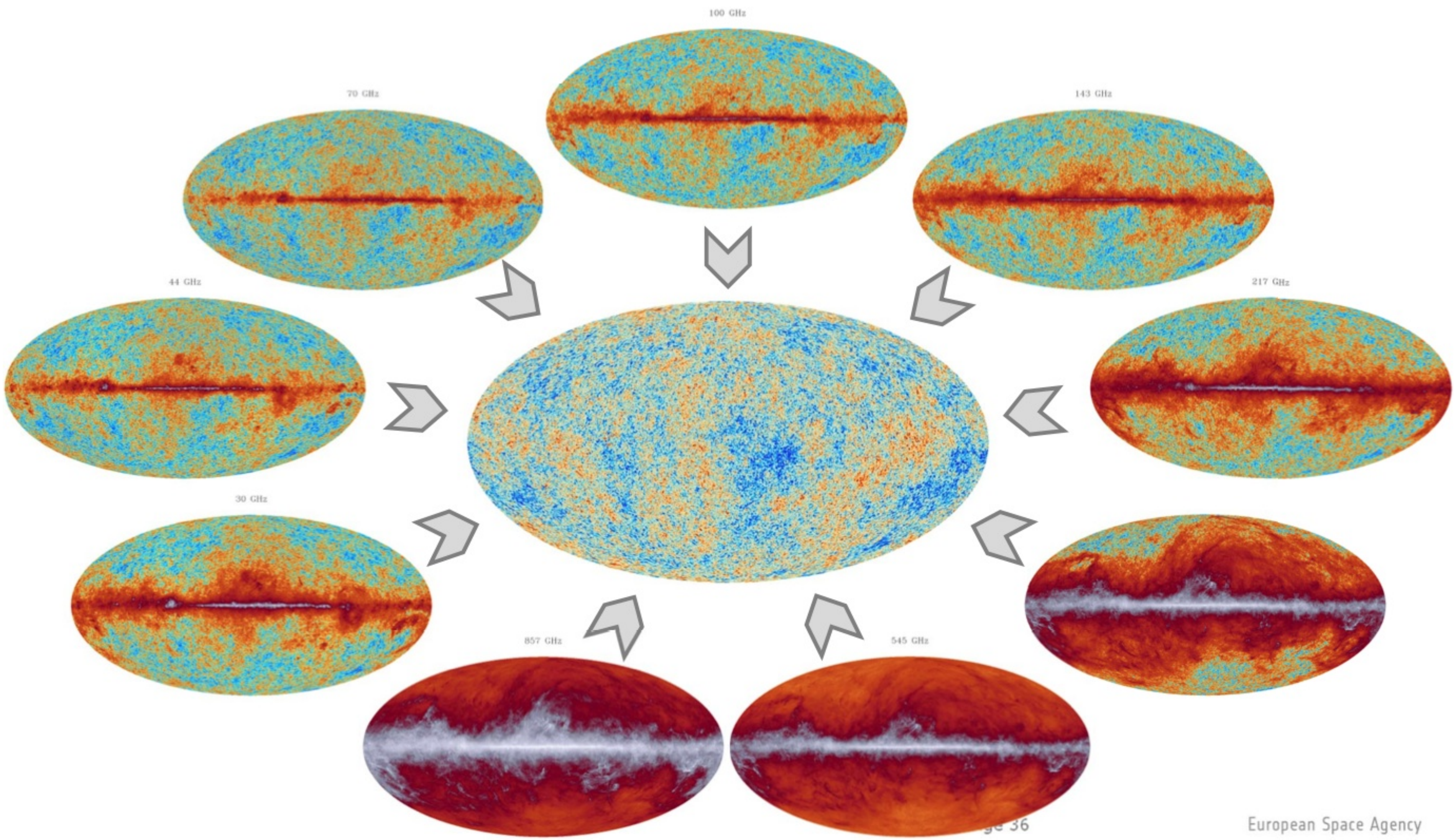
$$-2 \log P(X | \{C_\ell\}) \ = \ \sum_{\ell \geq 0} (2\ell+1) \ \left(\frac{\widehat{C}_\ell}{C_\ell} + \log C_\ell \right) + \mathsf{cst}$$

- Also reads like a spectral mismatch:

$$\left(\frac{\widehat{C}_\ell}{C_\ell} + \log C_\ell \right) = \ k\left(\frac{\widehat{C}_\ell}{C_\ell} \right) + \mathsf{cst'} \qquad k(u) \stackrel{\mathsf{def}}{=} u - \log u - 1 \geq 0$$

Extracting the CMB

Extracting the CMB from Planck frequency channels



How to do it?

Some requirements for producing a CMB map

- The method should be robust, accurate and high SNR (obviously). Special features: data set is expensive and there is ground truth.

- The method should be linear in the data:

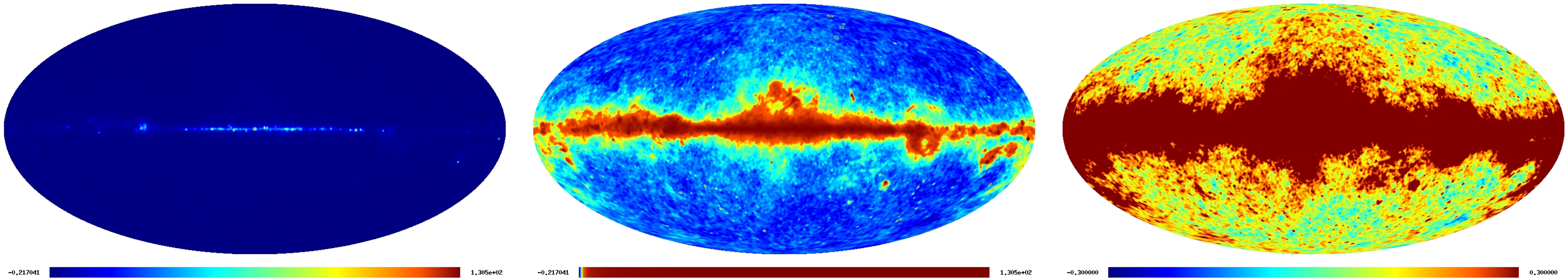
  1. It is critical not to introduce non Gaussianity

  2. Propagation of simulated individual inputs should be straightforward

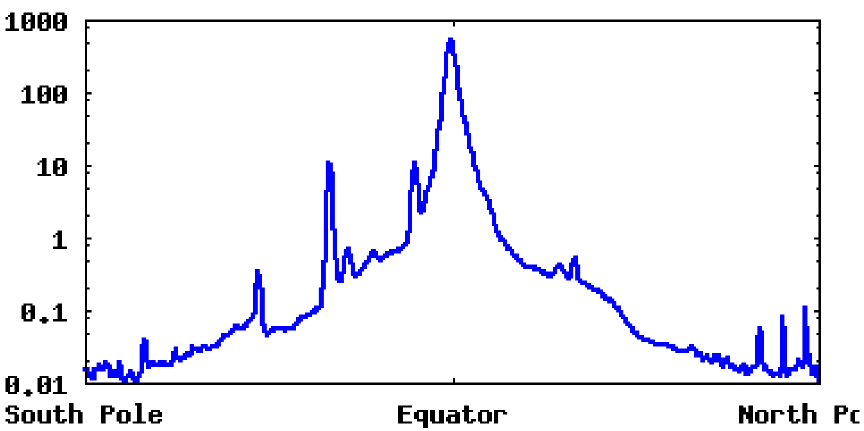
- The method should be able to support wide dynamical ranges, over the sky, over angular frequencies, across channel frequencies.

- (Wo.manpower behind the method should be aplenty).

Wide dynamics over the sky

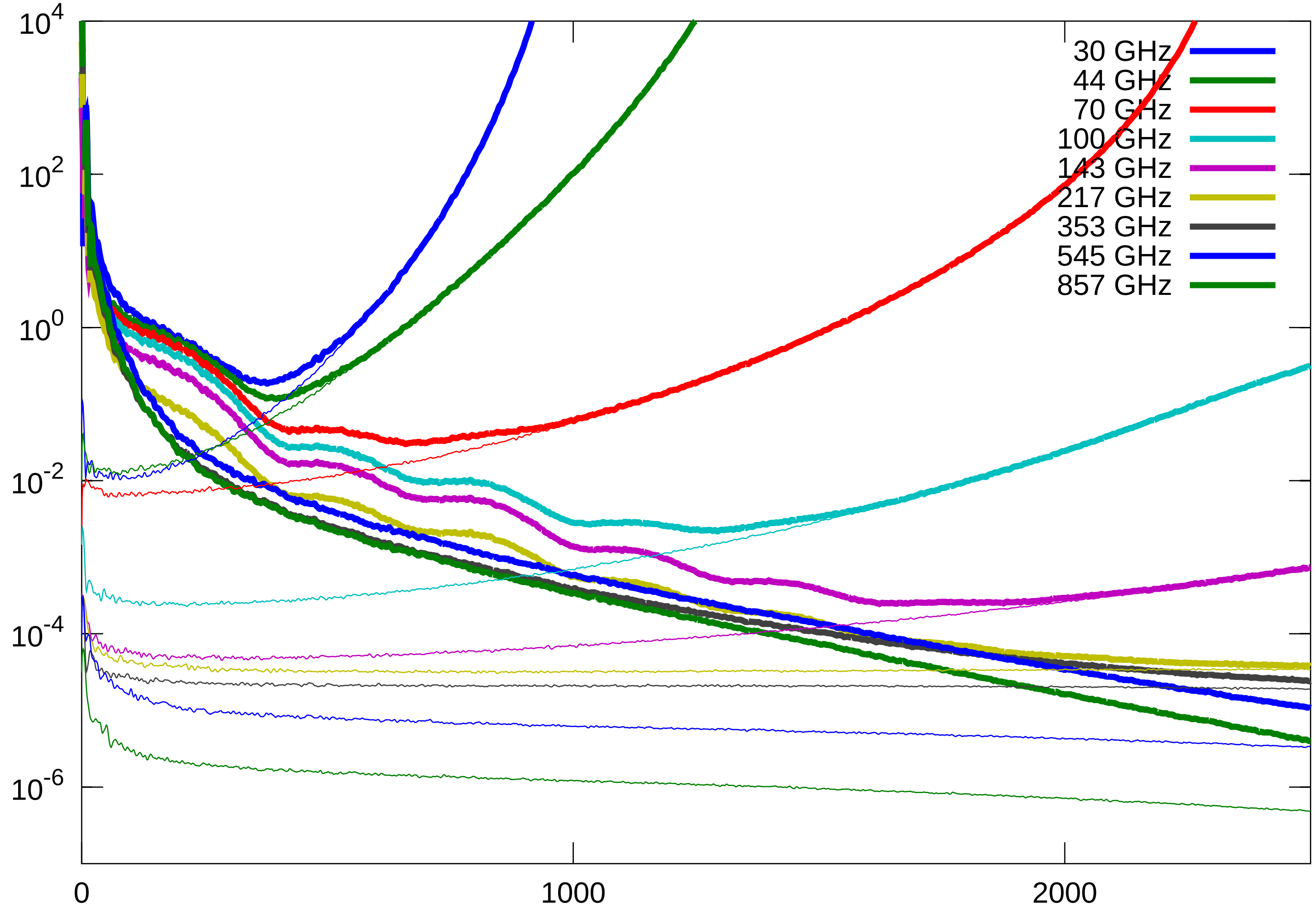


Left: The W-MAP K band. Natural color scale [-200, 130000] μK.
Middle: Same map with an equalized color scale.
Right: Same map with a color scale adapted to CMB: [-300, 300] μK.



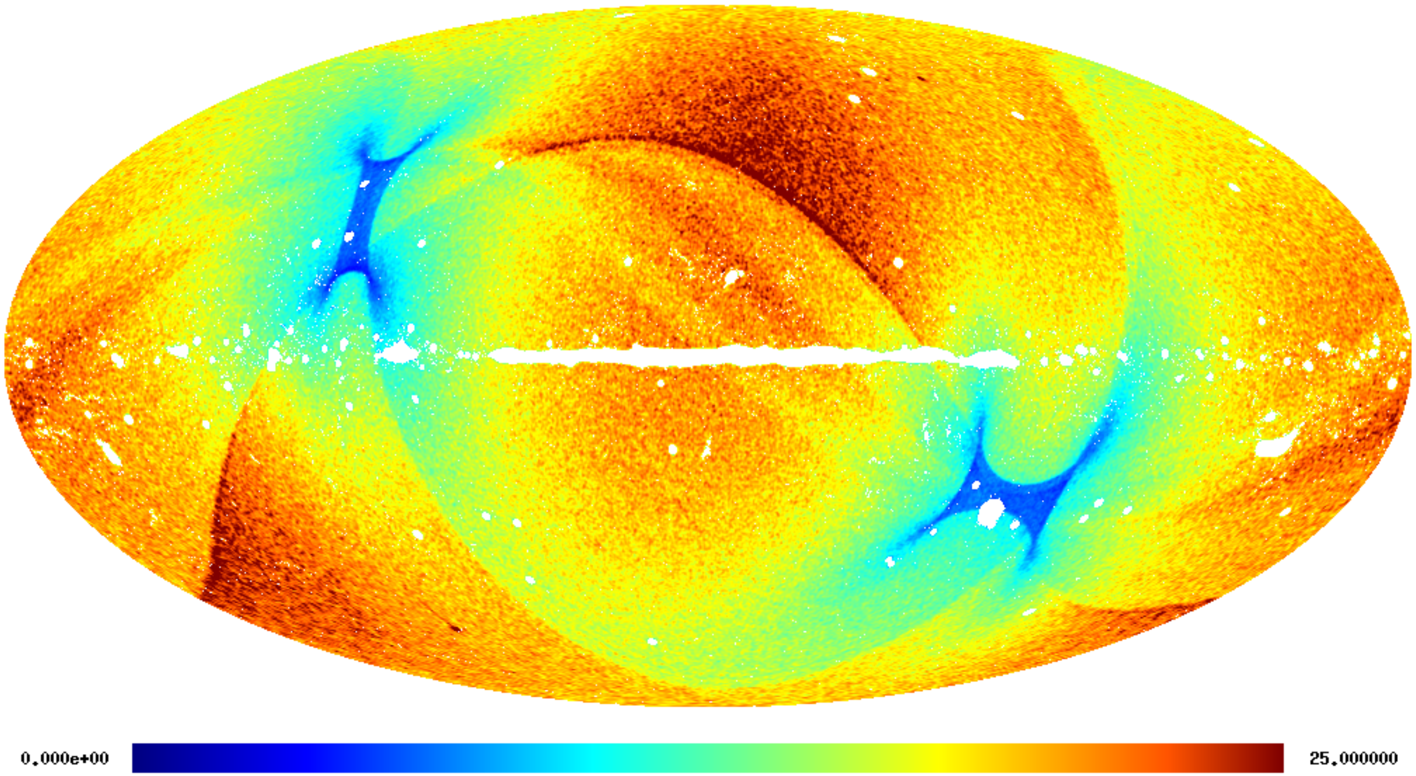
Average power as a function of latitude on a log scale for the same map.

Wide spectral dynamics, SNR variations



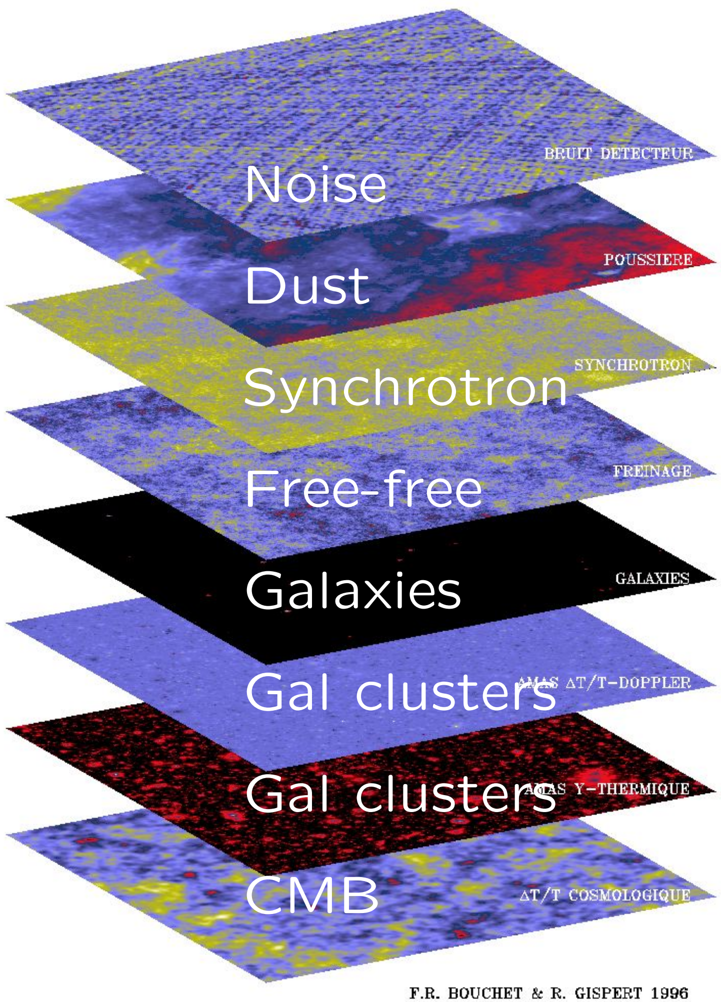
S & N angular spectra in Planck channels (unbeamed) for $f_{\mathrm{sky}} = 0.40$.

And what about the noise ?



Local RMS (μK) of the noise in a reconstructed CMB map.

Foregrounds



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:

$$\mathbf{d} = \begin{bmatrix} d_{30} \\ \vdots \\ d_{857} \end{bmatrix} = \mathbf{As} + \mathbf{n}$$

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.

Three contamination models based on matrices (or lack thereof)

1 Nine Planck channels modeled as noisy linear mixtures of CMB and 6 (say) "foregrounds"

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

2 Interesting limiting case: maximal foreground, no noise, that is,
Planck channels modeled as linear mixtures of CMB and $9 - 1 = 8$ "foregrounds"

$$\begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ \vdots \\ d_9 \end{bmatrix} = \begin{bmatrix} a_1 & F_{11} & \dots & \dots & F_{18} \\ a_2 & F_{21} & \dots & \dots & F_{28} \\ \vdots & \vdots & \dots & \dots & \vdots \\ \vdots & \vdots & \dots & \dots & \vdots \\ a_9 & F_{91} & \dots & \dots & F_{98} \end{bmatrix} \times \begin{bmatrix} s \\ f_1 \\ f_2 \\ \vdots \\ f_8 \end{bmatrix} \quad \text{or} \quad \mathbf{d} = [\,\mathbf{a} \mid \mathbf{F}\,] \begin{bmatrix} s \\ \mathbf{f} \end{bmatrix}$$

3 No foreground/noise model at all, but an unstructured contaminant $\mathbf{g}$:

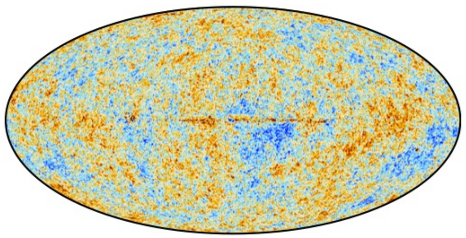
$$\mathbf{d} = \mathbf{a}s + \mathbf{g}.$$

– A linear combination $\widehat{s} = \sum_i w_i d_i = \mathbf{w}^\dagger \mathbf{d}$ of the input channels gives unbiased CMB if $\mathbf{w}^\dagger \mathbf{a} = 1$ and perfect foreground rejection if $\mathbf{w}^\dagger \mathbf{F} = 0$.

Actually, we are after Span($\mathbf{F}$), the foreground subspace.

Four CMB maps in Planck releases

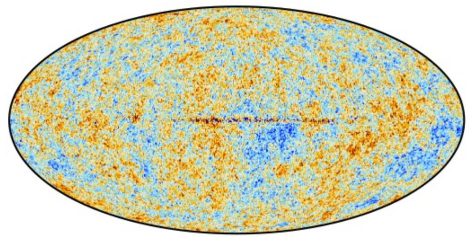
NILC



Wavelet space

$$\mathbf{d} = \mathbf{a}\, s + \mathbf{g}$$

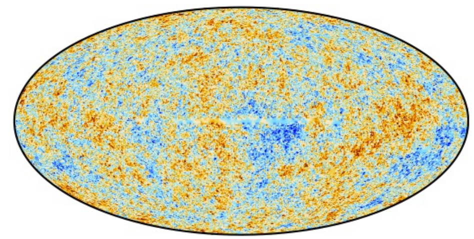
SEVEM



Pixel+Harmonic

$$\mathbf{d} = \mathbf{a}\, s + \mathbf{g}$$

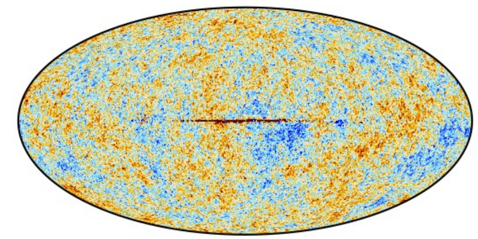
SMICA



Harmonic space

$$\mathbf{d} = [\mathbf{a}|\mathbf{F}] \begin{bmatrix} s \\ \mathbf{f} \end{bmatrix} + \mathbf{n}$$

Commander



Pixel space

$$\mathbf{d} = [\mathbf{a}|\mathbf{F}(\theta)] \begin{bmatrix} s \\ \mathbf{f} \end{bmatrix} + \mathbf{n}$$

- Various filtering schemes (space-dependent, multipole-dependent, or both):

– NILC : Needlet (spherical wavelet) domain ILC.

– SEVEM : Pixel domain, internal template fitting.

– SMICA : Harmonic domain, ML approach, foreground subspace.

– Commander : Pixel domain, Bayesian method with physical foreground models.

The SMICA method

The CMB is statistically independent of the other components.

SMICA = SM + ICA = Spectral Matching Independent Component Analysis

Do **ICA**: Blind estimation of the linear model $\mathbf{d} = [\,\mathbf{a} \mid \mathbf{F}\,] \begin{bmatrix} s \\ \mathbf{f} \end{bmatrix} + \mathbf{n}$,

via **spectral matching**: if all fields are **modeled as Gaussian stationary**, then the likelihood is a mismatch between empirical and theoretical spectra.

Q: **How dare you** do that on sky maps so blatantly non Gaussian/non stationary?

Combining all 9 Planck channels, non parametrically: the ILC

1/ Independent contamination model: Stack the 9 Planck channels into a 9×1 data vector $\mathbf{d} = [d_{30}, d_{44}, \ldots, d_{545}, d_{857}]^\dagger$

$$\mathbf{d}(p) = \mathbf{a}\, s(p) + \mathbf{g}(p) \qquad p = 1, \ldots, N_{\text{pix}}$$

2/ Linear combination: Estimate the CMB signal $s(p)$ in pixel p by weighting the inputs:

$$\widehat{s}(p) = \mathbf{w}^\dagger \mathbf{d}(p)$$

Idea: The variance of independent variables add up. Hence...
Minimize the pixel average $\langle (\mathbf{w}^\dagger \mathbf{d})^2 \rangle_p$ subject to $\mathbf{w}^\dagger \mathbf{a} = 1$, yielding

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

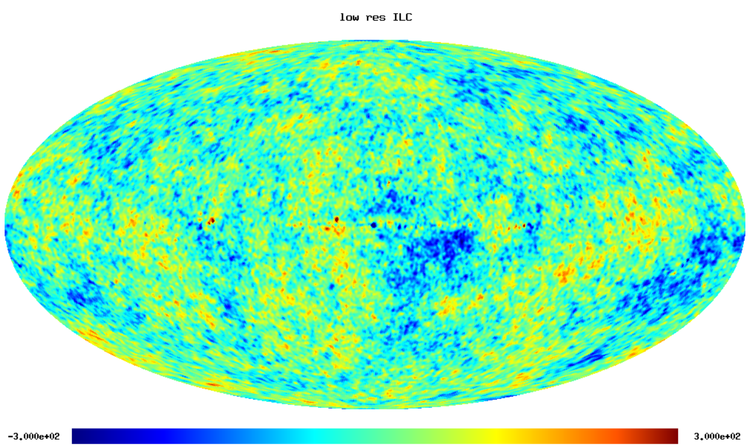
Coined ILC (Internal Linear Combination) by astrophysicists, a.k.a. MVBF (Min. Variance Beam Former) in array processing, BLUE elsewhere.

.

Is the ILC good enough for Planck data ?

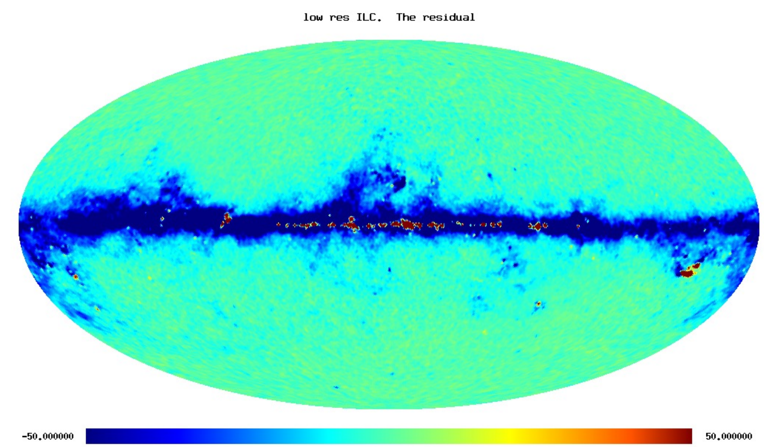
ILC looks good: linear, unbiased, min. MSE, very blind, very few assumptions: knowing $\mathbf{a}$ (calibration) and the CMB uncorrelated from the rest (very true).

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:

- harmonic dependence and
- chance correlations.

Linear filtering in harmonic space

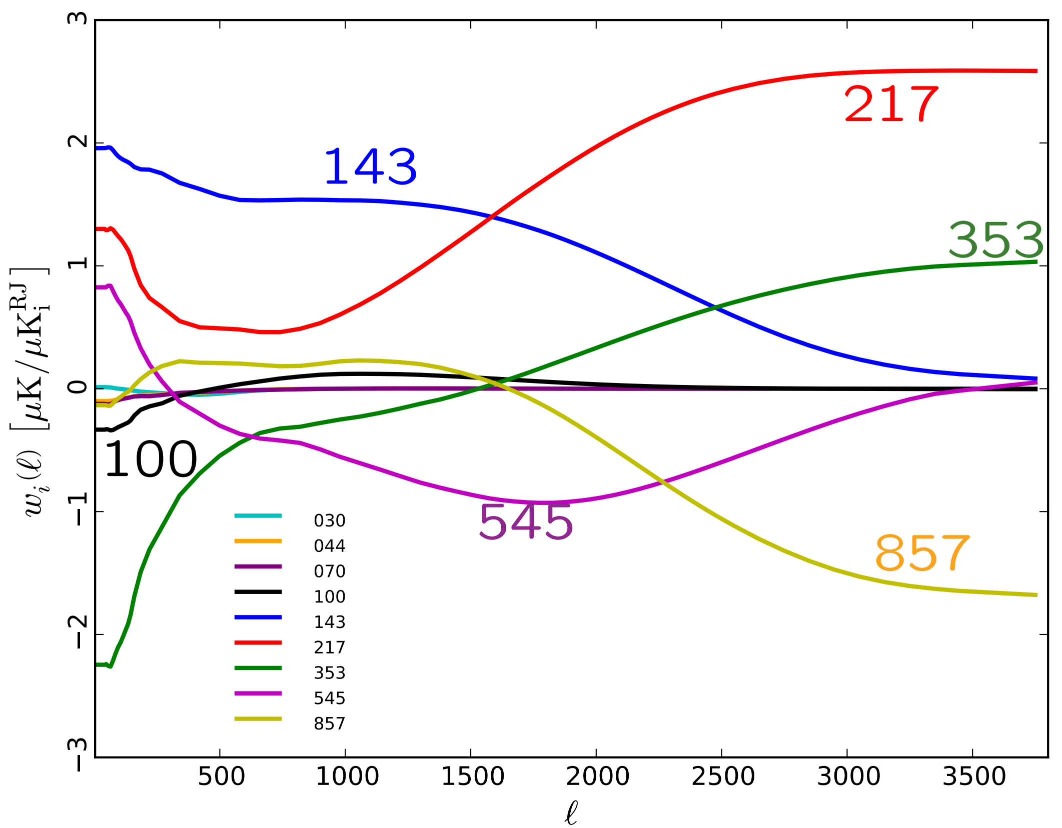
Since resolution, noise and foregrounds vary (wildly) in power over channels and angular frequency, the combining weights should depend on ℓ.

Harmonic ILC (Tegmark):
CMB map synthesized from spherical harmonic coefficients $\widehat{s}_{\ell m}$, obtained as linear combinations:

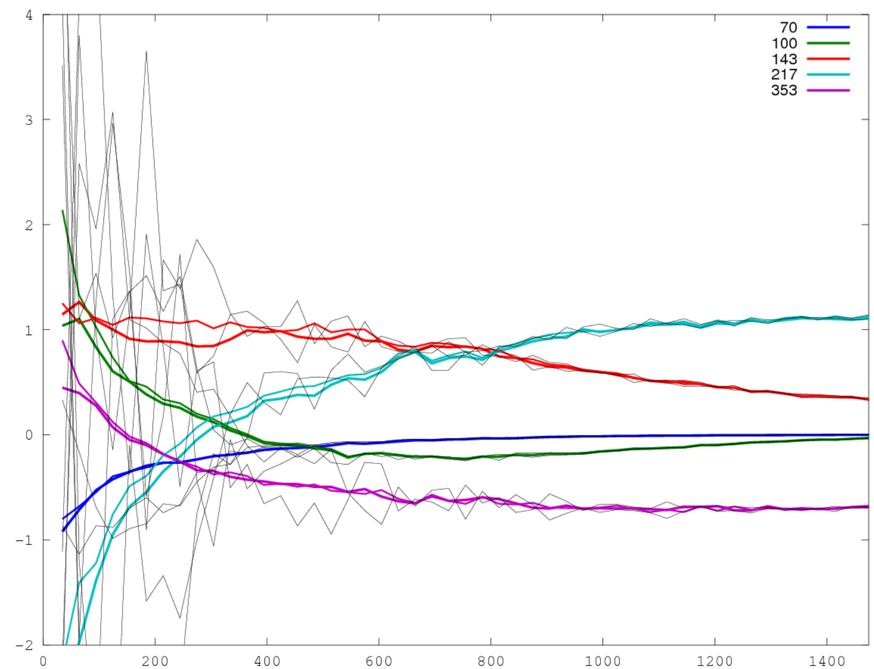
$$\widehat{s}_{\ell m} = \mathbf{w}_\ell^\dagger \, \mathbf{d}_{\ell m} \quad \text{with, again,}$$

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

- At high ℓ, (spectral) matrices $\mathbf{C}_\ell$ well estimated by $\widehat{\mathbf{C}}_\ell = \sum_m \mathbf{d}_{\ell m}\mathbf{d}_{\ell m}^\dagger / \ 2\ell + 1$.
- At lower ℓ, we need to get smarter to fight chance correlation.



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



At low ℓ, the empirical covariance matrices

$$\widehat{\mathbf{C}}_\ell = \frac{1}{2\ell+1} \sum_{-\ell \le m \le \ell} \mathbf{d}_{\ell m} \mathbf{d}_{\ell m}^\dagger$$

do not average over enough modes. Chance correlation hurts.

Chance correlation: the single Universe problem

Simplest illustrative example:

$$\mathbf{d} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} s + \alpha f \\ f' \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} s + \begin{bmatrix} \alpha f \\ f' \end{bmatrix}$$

Form a covariance matrix $\widehat{\mathrm{C}} = \langle \mathbf{d}\mathbf{d}^\dagger \rangle$ and the ILC estimate will yield

$$\hat{s} = \frac{\mathbf{a}^\dagger \widehat{\mathbf{C}}^{-1}\,\mathbf{d}}{\mathbf{a}^\dagger \widehat{\mathbf{C}}^{-1}\,\mathbf{a}} = d_1 - \frac{\langle d_1 d_2 \rangle}{\langle d_2^2 \rangle} \cdot d_2 = s - \underbrace{\frac{\langle s f' \rangle}{\langle f'^2 \rangle} \cdot f'}_{\text{Chance corr.}} + \alpha \underbrace{\left(f - \frac{\langle f f' \rangle}{\langle f'^2 \rangle} \cdot f' \right)}_{\text{Non-rigid scaling}}$$

What is hitting us harder: chance correlation or non-rigid scaling ?

The error due to chance correlation is independent of α.
Same error whether $\alpha = 0$ or $\alpha = 1$ or $\alpha \gg 1$.

A mixing model implies $f = f'$, i.e. rigid scaling.
In such a model, chance correlation (a.k.a. ILC bias) dominates the error.

Similarly...

In the square noise-free case: $\mathbf{d} = \left[\begin{array}{cc} \mathbf{a} & \mathbf{F} \end{array}\right] \left[\begin{array}{c} s \\ \mathbf{f} \end{array}\right]$, the sample covariance is:

$$\langle \mathbf{d}\mathbf{d}^\dagger \rangle = \left[\begin{array}{cc} \mathbf{a} & \mathbf{F} \end{array}\right] \left[\begin{array}{cc} \langle s^2 \rangle & \langle s\mathbf{f}^\dagger \rangle \\ \langle \mathbf{f}s \rangle & \langle \mathbf{f}\mathbf{f}^\dagger \rangle \end{array}\right] \left[\begin{array}{cc} \mathbf{a} & \mathbf{F} \end{array}\right]^\dagger$$

In the absence of chance correlation: $\langle s\mathbf{f} \rangle = 0$, linear algebra says:

$$\mathbf{a}^\dagger \langle \mathbf{d}\mathbf{d}^\dagger \rangle^{-1} \mathbf{F} = 0$$

so that the ILC filter $\mathbf{w} \propto \langle \mathbf{d}\mathbf{d}^\dagger \rangle^{-1} \mathbf{a}$ would ensure $\mathbf{w}^\dagger \mathbf{F} = 0$,
that is, perfect orthogonality to the foreground subspace.

If the sample average $\langle s\mathbf{f} \rangle$ is not identical to the ensemble average $\mathbb{E} s\mathbf{f} = 0$ (chance correlation), then contamination is unavoidable.

High accuracy CMB requires good determination of the foreground subspace.
How to do it with a single sky ?

SMICA and the foreground subspace model

- Planck channels modeled as noisy linear mixtures of CMB and 6 "foregrounds":

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

i.e. foregrounds coherent enough to be confined to a low dimensional subspace, $\mathrm{Span}(\mathbf{F})$, but otherwise unconstrained: any spectrum, color, correlation...

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

– A very blind model (all non-zero parameters are free!) but still identifiable.

- Only $\mathrm{Span}(\mathbf{F})$, the foreground subspace, is needed to suppress the foregrounds. It is jointly determined by all multipoles involved in SMICA likelihood:

Gaussian + stationary $\rightarrow \sum_\ell (2\ell+1)\Big[\mathrm{trace}\ \widehat{\mathrm{C}}_\ell \mathbf{C}_\ell(\theta)^{-1} + \log\det \mathbf{C}_\ell(\theta)\Big]$.

When the stationary Gaussian likelihood is OK

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

$$[\,\mathbf{a} \mid \mathbf{G}\,]^{-1}\,[\,\mathbf{a} \mid \mathbf{F}\,] = \begin{bmatrix} 1 & \alpha^\dagger \mathbf{X} \\ 0 & \mathbf{X} \end{bmatrix} \qquad \text{for some matrix } \mathbf{X} \text{ and vector } \alpha.$$

Applying $\mathbf{T}$ to data yields a (presumably) dirty CMB and $n-1$ 'templates':

$$\begin{bmatrix} y \\ \mathbf{g} \end{bmatrix} \stackrel{\text{def}}{=} \mathbf{T}\mathbf{d} = \begin{bmatrix} s + \alpha^\dagger \mathbf{X}\mathbf{f} \\ \mathbf{X}\,\mathbf{f} \end{bmatrix}$$

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

.

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

Two ingredients for a likelihood of $\mathbf{d} = \mathbf{A} \begin{bmatrix} s \\ \mathbf{f} \end{bmatrix}$:

$$\underbrace{P\left(\begin{bmatrix} s \\ \mathbf{f} \end{bmatrix}\right) = P_s(s) \cdot P_{\mathbf{F}}(\mathbf{f})}_{\text{1) Statistical independence}} \quad \text{and} \quad \underbrace{\begin{bmatrix} y \\ \mathbf{g} \end{bmatrix} \overset{\text{def}}{=} \mathbf{Td} \overset{\text{model}}{=} \begin{bmatrix} s + \alpha^\dagger \mathbf{X}\mathbf{f} \\ \mathbf{X}\,\mathbf{f} \end{bmatrix}}_{\text{2) (Pre-processed) mixing model}}.$$

After just a little bit of work:

$$p(\mathbf{d}|\mathbf{A}) = \underbrace{p_S(y - \alpha^\dagger \mathbf{g})}_{\text{CMB}} \cdot \underbrace{p_{\mathbf{F}}(\mathbf{X}^{-1}\mathbf{g})|\det(\mathbf{X})|^{-N_{\text{pix}}}}_{\text{foregrounds}} \cdot \underbrace{|\mathbf{T}|^{N_{\text{pix}}}}_{\text{Constant}}.$$

Thus a maximum likelihood solution for the signal of interest is

$$\hat{s}^{\text{ML}} = y - \hat{\alpha}^\dagger \mathbf{g} \qquad \hat{\alpha} = \arg\max_{\alpha}\ p_S(y - \alpha^\dagger \mathbf{g})$$

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

- Hence, in the high SNR limit, a statistical model is needed only for the component of interest, the CMB, whenever $\mathbf{a}$ is known (calibration).

More explicitly

Do we have a good probability model to use in

$$\widehat{\alpha} = \arg\max_{\alpha} \; p_S(y - \alpha^\dagger \mathbf{g})$$

for the CMB ? Of course! It is trivial in harmonic space:

$$-2 \log p_S(y - \alpha^\dagger \mathbf{g}) = \sum_\ell \sum_m \frac{(y_{\ell m} - \alpha^\dagger g_{\ell m})^2}{C_\ell} + \text{cst}$$

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

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

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

Wisdom of the likelihood

Two covariance matrices:

$$\widehat{\mathbf{C}}_H = \sum_\ell \sum_m \mathbf{d}_{\ell m} \mathbf{d}^\dagger_{\ell m} / C_\ell \qquad \widehat{\mathbf{C}}_P = \langle \mathbf{d}\mathbf{d}^\dagger \rangle_p, = \sum_\ell \sum_m \mathbf{d}_{\ell m} \mathbf{d}^\dagger_{\ell m}$$

- The $1/C_\ell$ weight equalizes the variance of the harmonic modes of the CMB.
- The $1/C_\ell$ weight can be replaced with anything similar.
- Pixel-based covariance $\widehat{\mathbf{C}}_P$ dominated by a small number of effective modes.

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.

It's like having 5.60 extra Universes to average over.

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

Conclusions

Trying to make the best out of a single Universe realization

ICA: Key idea about the foregrounds : one subspace to rule them all (out).

Robustness by doing without a complete foreground model (subspace only).

Can be made not too naive statistically for CMB extraction: spectral matching.

Targetting the CMB makes modeling foreground distribution irrelevant
in the high SNR (large scales, low ℓ) limit.

Future: data-driven foreground models when the SNR is not so great.