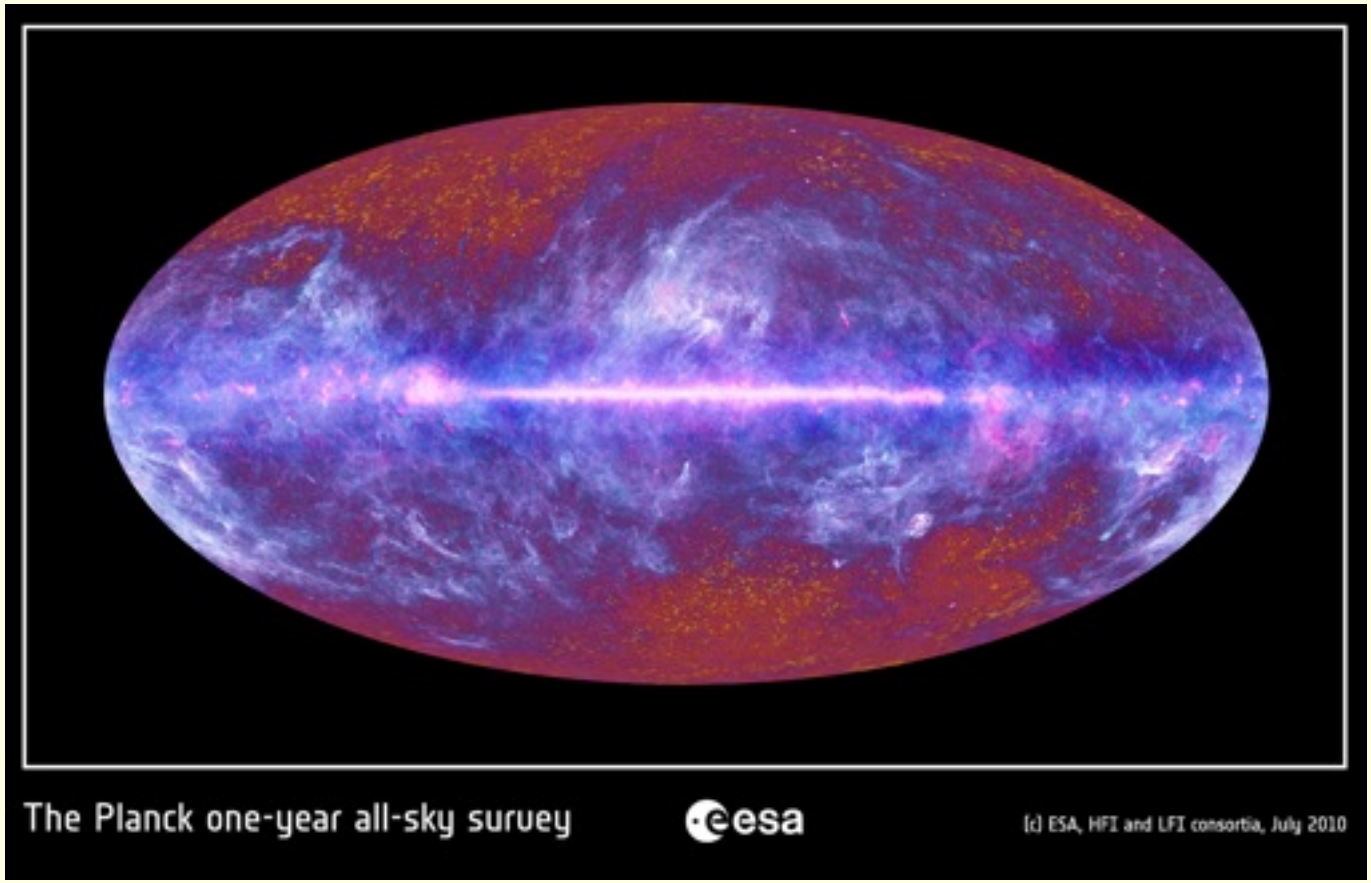
Component separation for the CMB

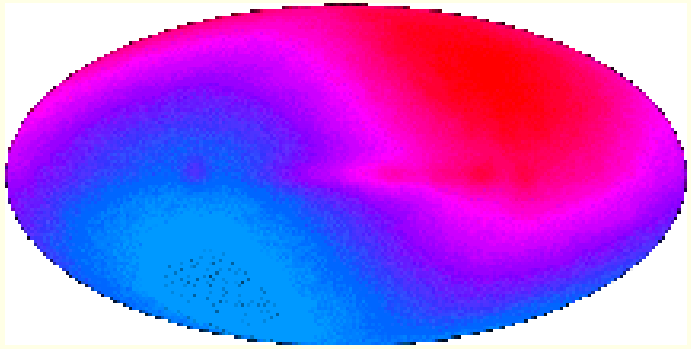
Jean-François Cardoso.
CNRS-LTCI / P.7-APC / IAP



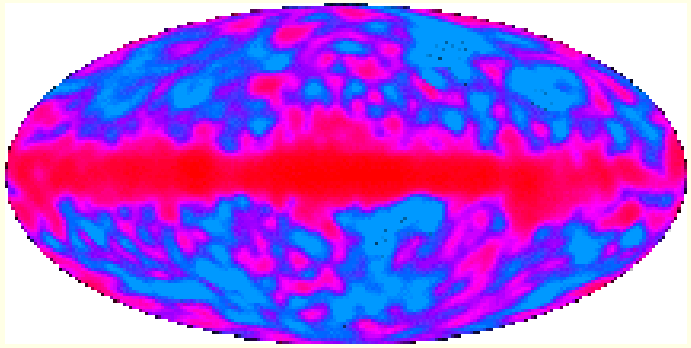
Xth School of Cosmology. Cargese. July 2010

What COBE's DMR saw.

- Temperature map: $T(\theta, \phi) \approx T_o = 2.725K$,
- Very isotropic (good!)
- Shows a $\times 1000$ expansion since recombination.



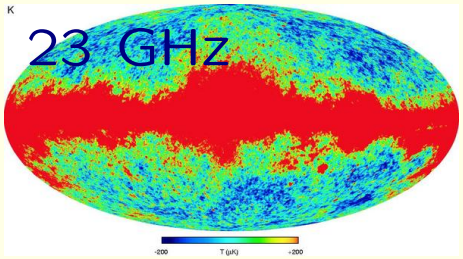
- $T(\theta, \phi) - T_o \approx \langle \vec{d}, \vec{u}(\theta, \phi) \rangle$: dipole anisotropy.
- Quite weak: $\|\vec{d}\|/T_o = O(10^{-3})$.
- Measures COBE's velocity in the CMB sea.

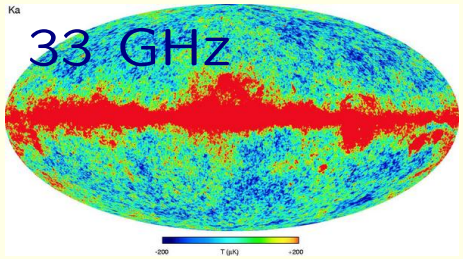


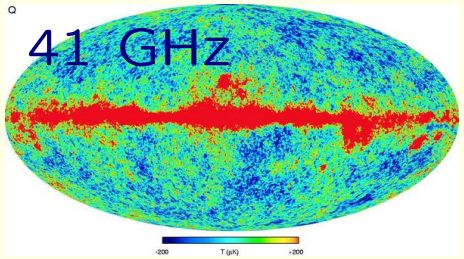
- $\delta T(\theta, \phi) = T(\theta, \phi) - T_o - \langle \vec{d}, \vec{u}(\theta, \phi) \rangle$
- Finally reveals cosmological CMB anisotropies
- A very delicate measure: $\|\delta T\| \approx 10^{-4} K$

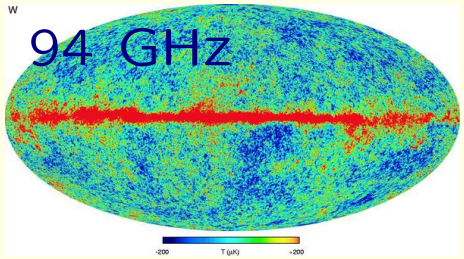
What W-MAP saw.

4 W-MAP channels after subtraction of the monopole and dipole components.



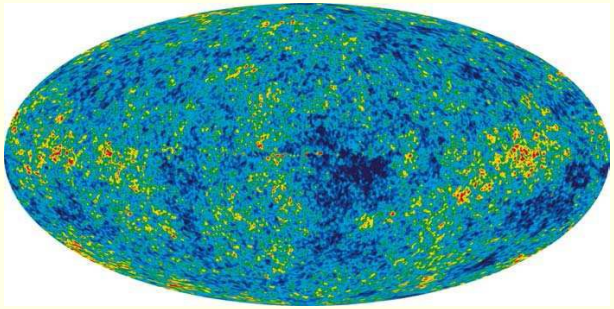




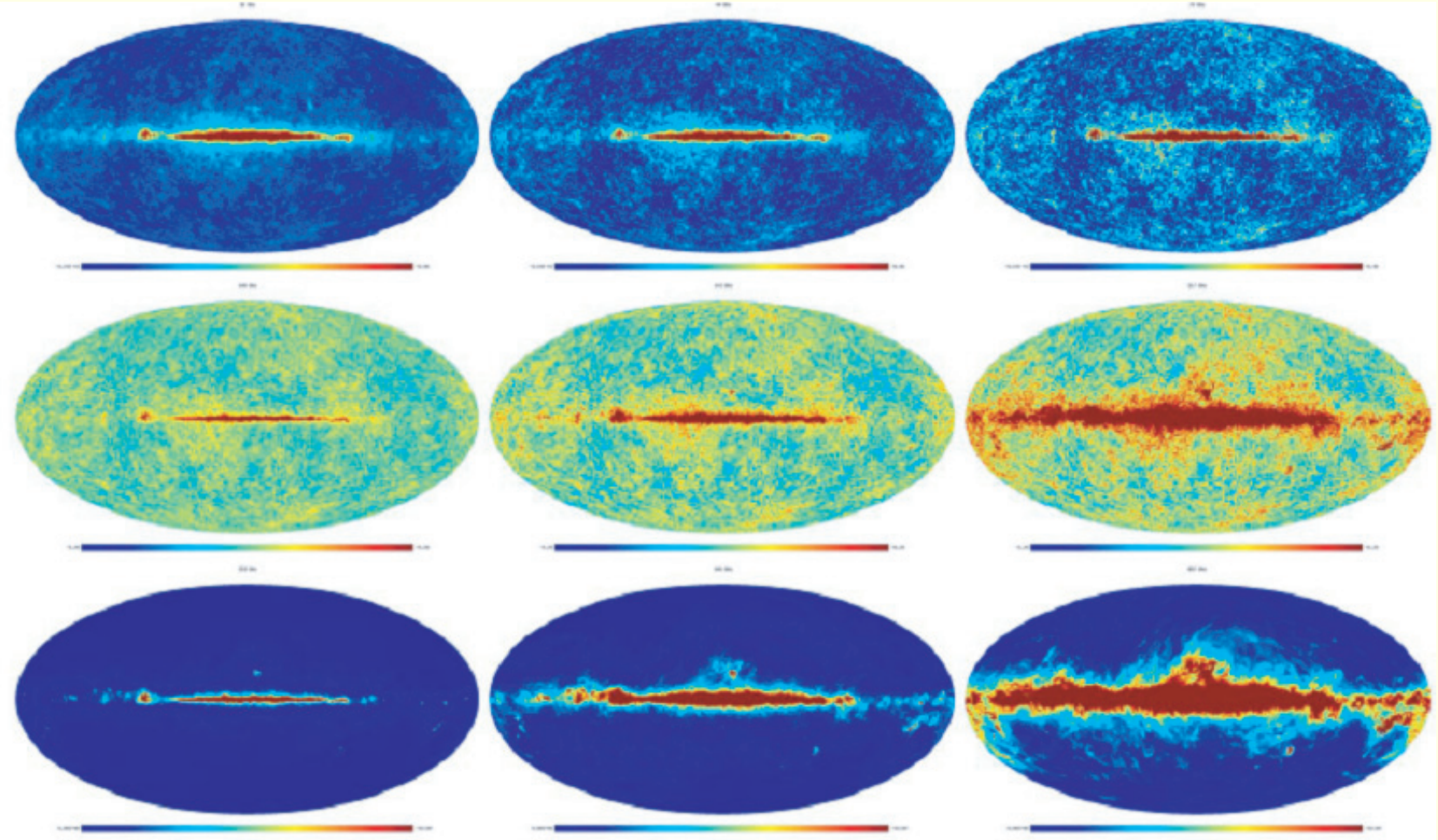


Foreground emissions clearly dominate the low Galactic latitude regions. Homogeneous extra-Galactic foregrounds are also expected everywhere.

The diversity offered by multi-frequency observations allows for CMB cleaning. WMAP ILC map looks pretty good:



What Planck may see (according to the Blue Book)



LFI looks at 30, 44, 70 GHz with radiometers
while HFI looks at 100, 143, 217, 353, 545, 857 GHz with bolometers.

Planck information path and some jargon

1. From the sky to detectors; from a spinning satellite to time-lines. Planck has TOI problems (pun).
2. From time lines to 'phase binned rings'.
3. From rings to spherical maps: map making.
4. From multi-channel spherical maps to a CMB map
5. From a CMB map to its angular spectrum: CMB cleaning or component separation.
6. From the spectrum to the likelihood of cosmological model.
7. From likelihood to (probabilty distribution of) the cosmo. parameters.

The divide and conquer stratgey of steps 4,5,6,7 would be optimal for full sky observations in simple models (nice noise, nice foregrounds).
Otherwise, the 'optimal' processing forbids such a factorization.
Still needs to trade off statistical efficiency for simplicity and CPU cycles.

Foregrounds

The Cosmic Microwave Background is the backgroundest thing there is. Therefore, any other emission must be a foreground. Such as:

- The Cosmic Infrared Background (CIB) is a backgroundish foreground due to distant, unresolved, dusty Galaxies.
- Galaxy clusters seen via the Sunyaev-Zeldovitch (SZ) effect.
- Point sources: radio galaxies, . . .
- Galactic (Milky Way) dust emission.
- Galactic (Milky Way) synchrotron emission.
- Galactic (Milky Way) free-free emission.

$\longrightarrow$ Component separation: sort out all those emissions.
$\longrightarrow$ CMB cleaning: get the cleanest (in some sense) CMB map, do cosmology with it. Pass CMB-free maps to anyone interested (in CIB, Galaxy, SZ clusters,. . .).

Tactics for dealing with foregrounds

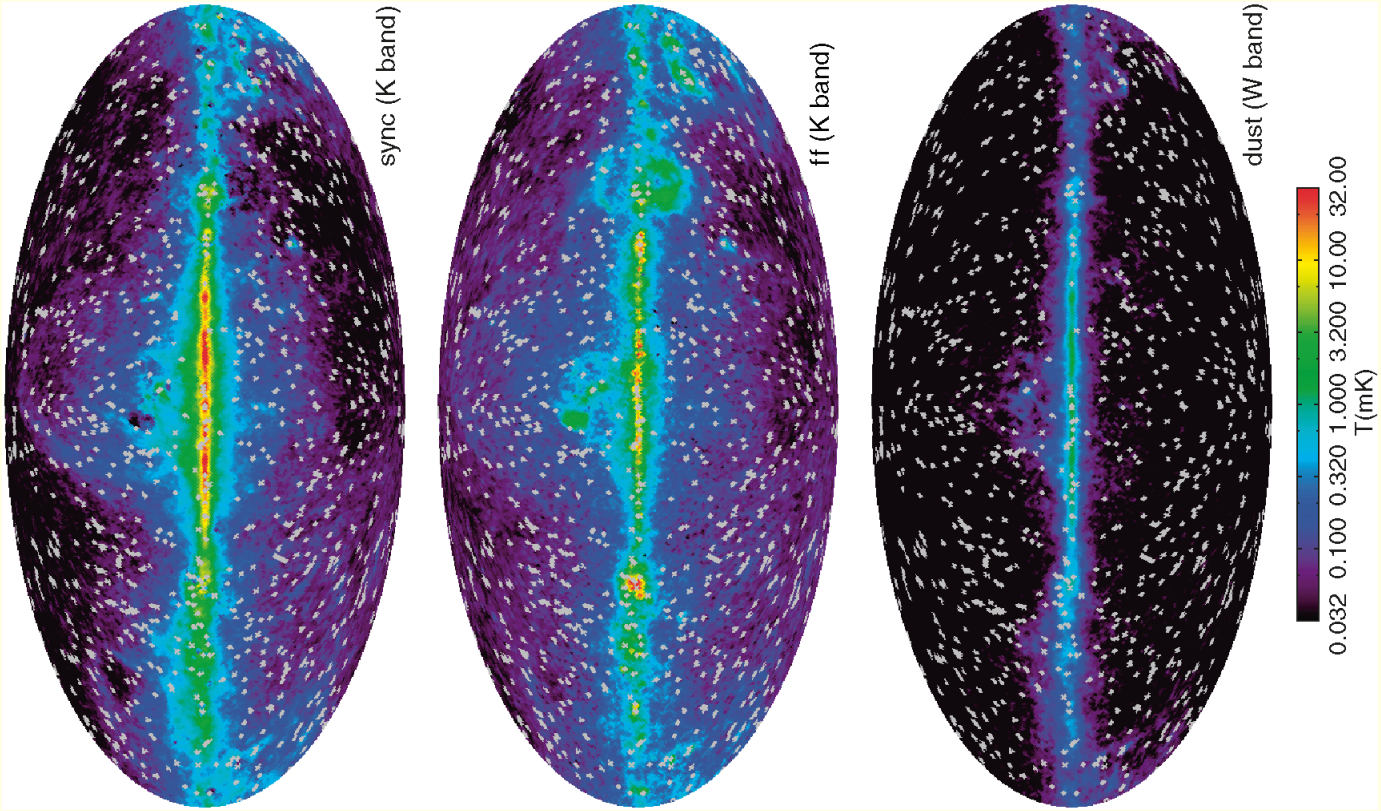
If one wants to do CMB, foregrounds are the ennemies.

Some tactics to minimize foreground annoyances:

1. Observe at frequencies where CMB anisotropies are stronger than the foregrounds
2. Look at regions of the sky where (Galactic) foregrounds are the weakest.
3. Mask out point sources (if at reasonable area loss)
4. Model contribution from diffuse foregrounds
5. Do component separation/CMB extraction:
   predict foregrounds using multi-channel observations
   i.e. exploit foreground coherence across frequency, hoping
   1) to overcome the limitations of the above approaches and
   2) to combine 'optimally' infi from all of the sky and all frequencies.

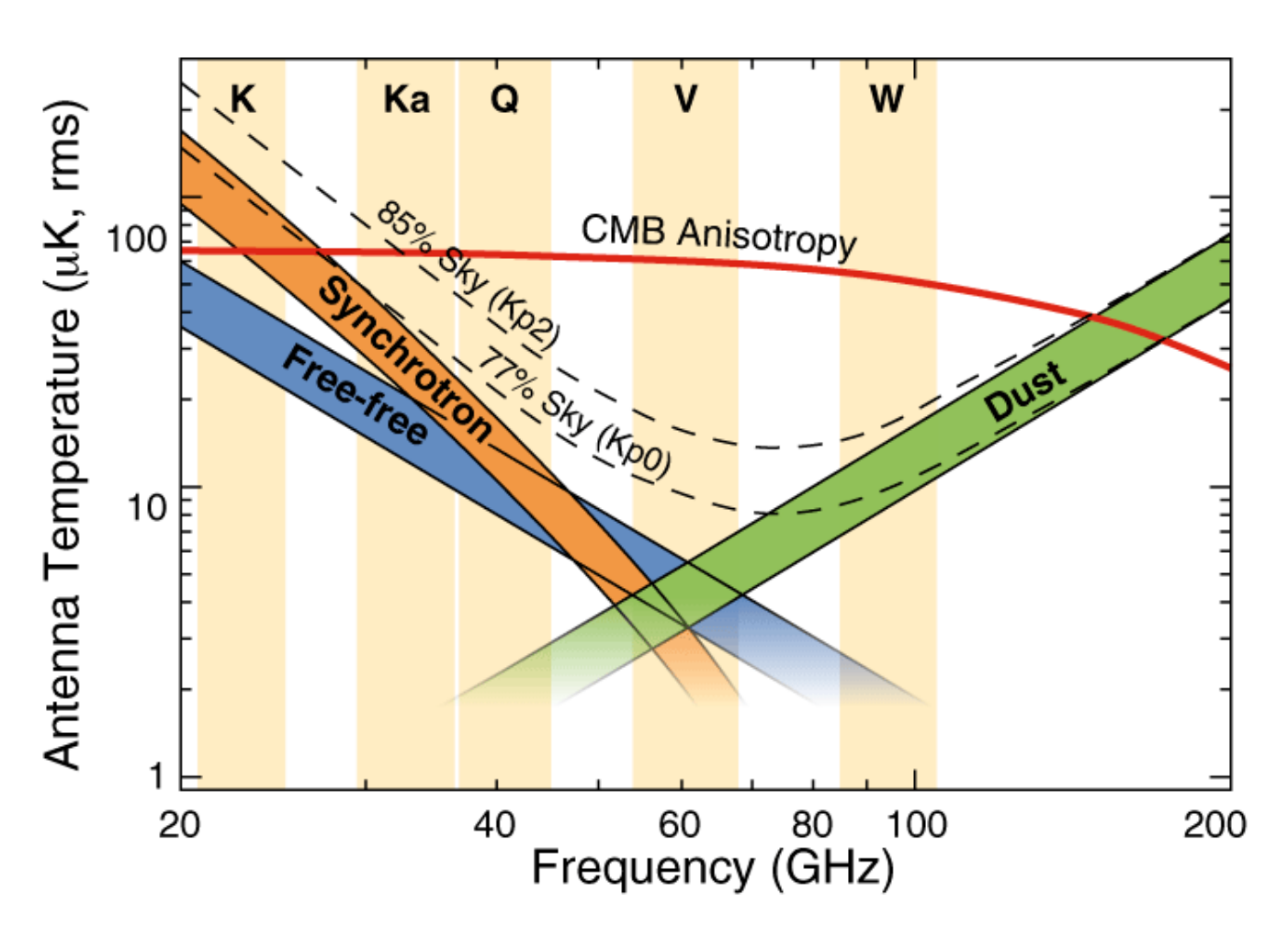
Distribution of Galactic foregrounds. 1) Spatial distribution

Spatial distribution: see page 5 for some ideas about it. Roughly controlled on average by a co-secant law in a simple parallel slab Galactic model.



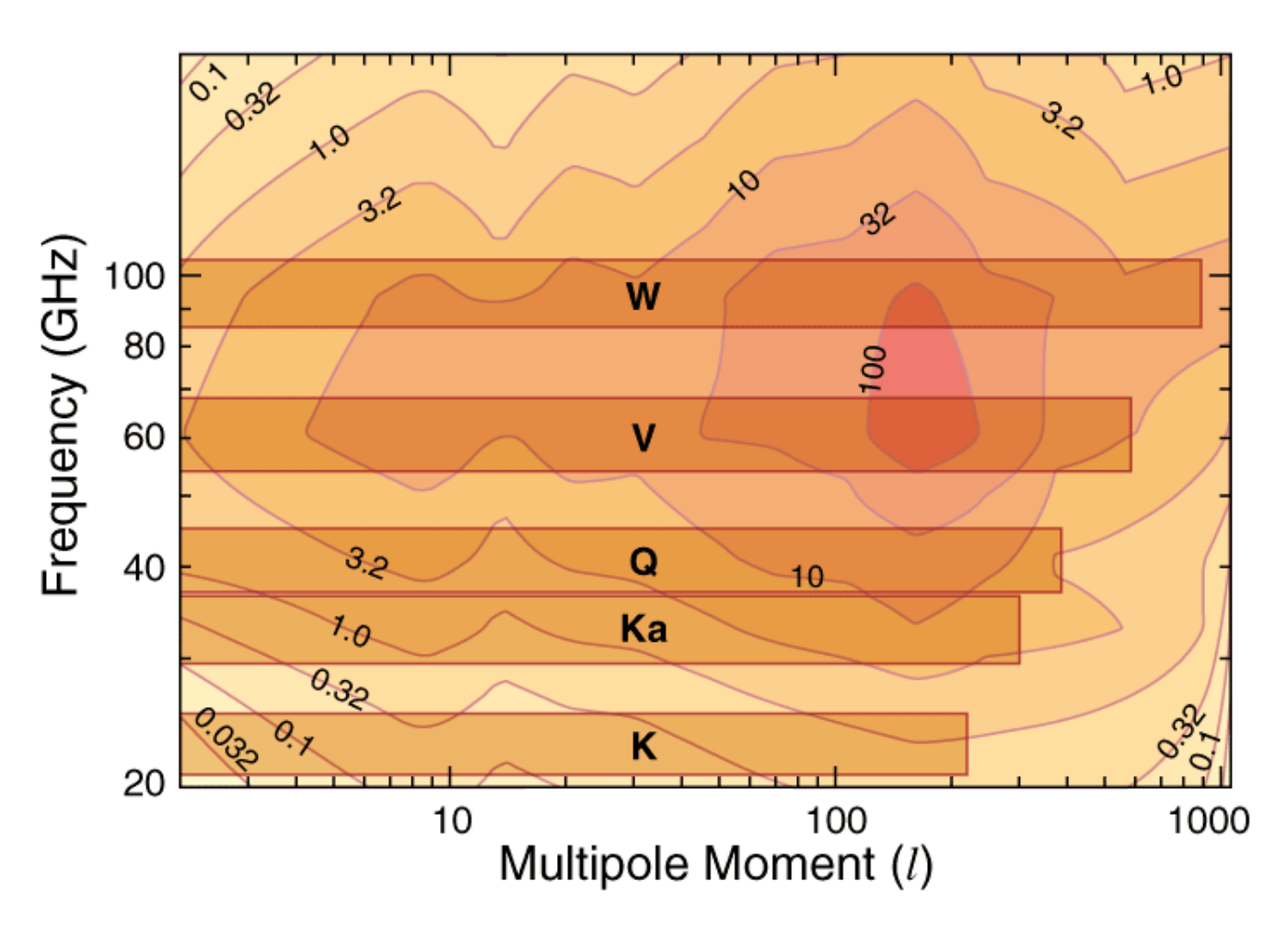
Estimates by WMAP:
for synchrotron and free-free at 22 GHz and for dust at 90 GHz.

Distribution of Galactic foregrounds. 2) Frequency distribution



A rough estimate for the 5 WMAP frequencies, with ‘uncertainties’.

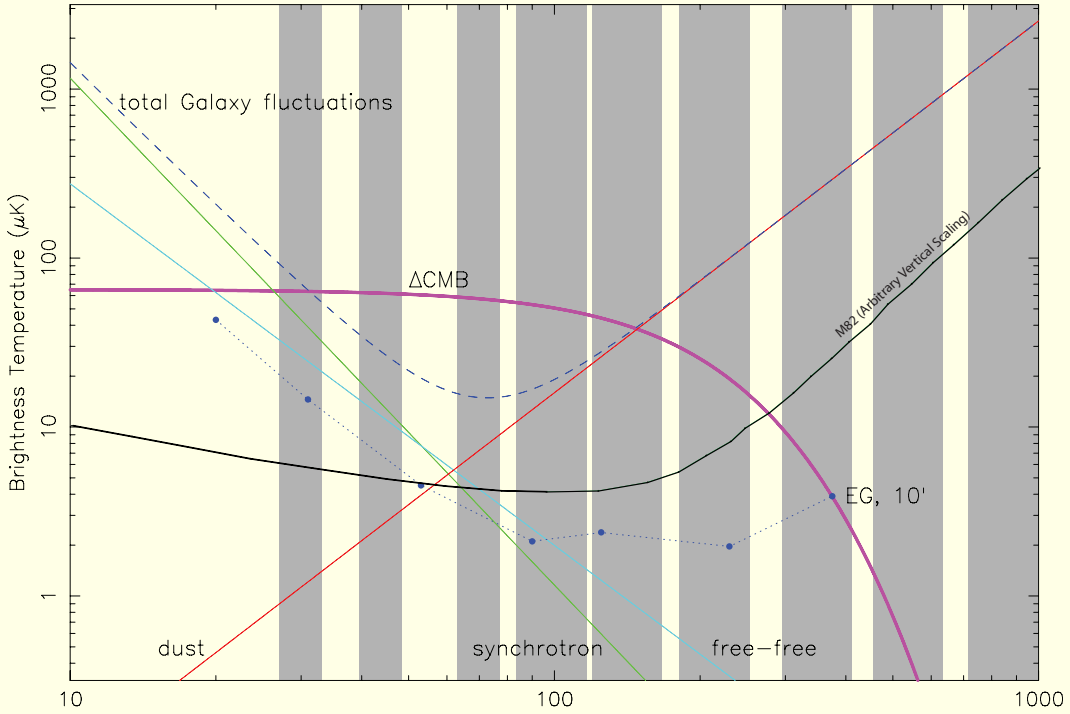
Joint distribution in frequency/multipole space for W-MAP



An observing window through the galactic foregrounds as a function of both frequency and angular scale.

Contours: relative strength of CMB wrt foreground signals.

A more optimistic figure from Planck's Blue Book



Extensive frequency coverage: good for component separation.

The plot is 'optimistic' because it does not show the uncertainties in spectral scaling.

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 be introduced 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

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

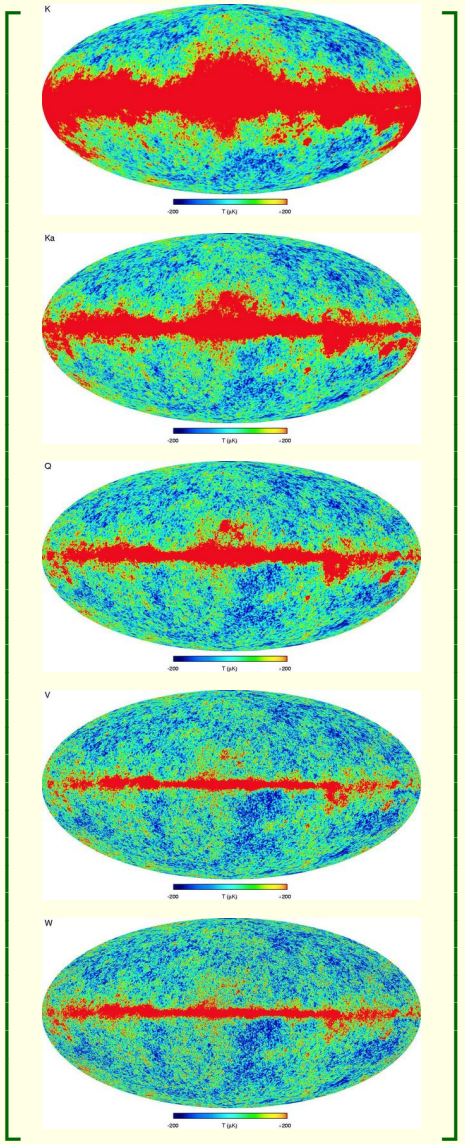
- if there is no noise: $N(\theta,\phi) = 0$ and
- if the mixing matrix A is known perfectly,

then, there is no component separation problem since

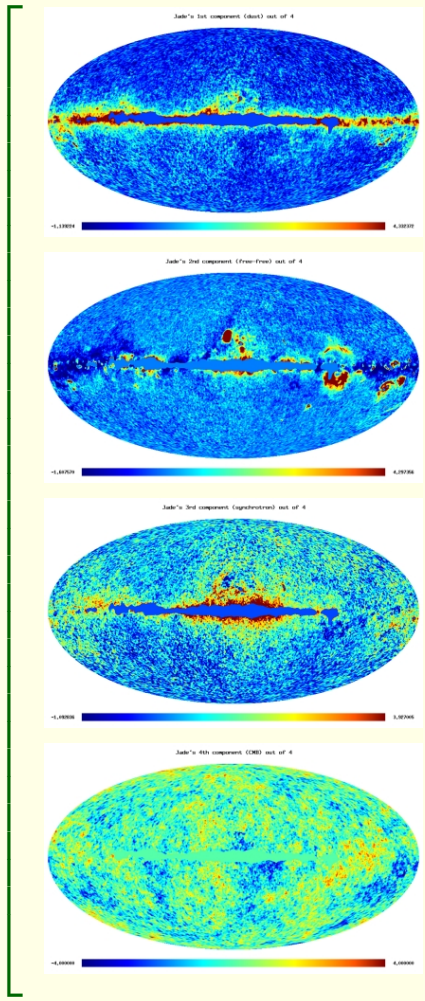
$$S(\theta,\phi) = BX(\theta,\phi) \quad \text{for any } C \times F \text{ matrix } B \text{ such that } BA = I_C$$

The problem is not the separation itself but dealing with the uncertainties and the approximations in the above statement.

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



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



$$\begin{matrix} ? \\ ? \\ ? \\ \text{CMB!} \end{matrix} + \begin{bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \\ N_5 \end{bmatrix}$$

JADE finds uncorrelated, maximally non Gaussian components.

Here, results on 5 W-MAP channels degraded to common resolution.

See also AltICA (based on FastICA) by Baccigalupi, Maino *et al.*

Pixel space versus harmonic space

Ideally, we should work jointly on big data matrices X of size $N_{\text{chan}} \times N_{\text{pixels}}$.

In practice, our processors rather work on $N_{\text{chan}} \times 1$ vectors $X(t)$
where t is an index for:
a direction in the sky $t = (\theta, \phi)$
a spherical harmonic coefficient $t = (l, m)$
a wavelet coefficient $t = (j, k)$ (to be discussed later).

In theory, one can move freely from 'pixel space' to harmonic space (and back):

$$a_{\ell m} = \int X(\theta,\phi) \mathcal{Y}^{\star}_{\ell m}(\theta,\phi) \quad \leftrightarrow \quad X(\theta,\phi) = \sum_{\ell} \sum_{m} a_{\ell m} \mathcal{Y}_{\ell m}(\theta,\phi)$$

The basic mixture model

$$X(\theta,\phi) = \mathbf{A} S(\theta,\phi) + N(\theta,\phi)$$

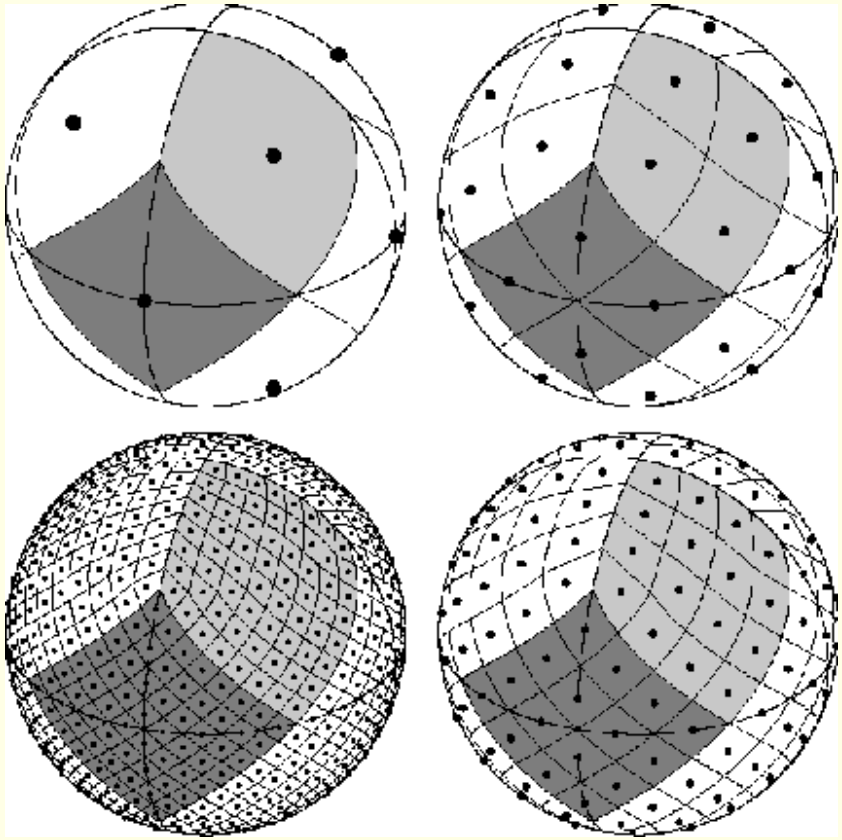
retains the same structure in harmonic space:

$$X_{\ell m} = \mathbf{A} S_{\ell m} + N_{\ell m}$$

but the statistical properties are dramatically different.

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 $N_{\rm pix} = 12N_{\rm side}^2 = 12\cdot 2^{2r}$ pixels.
It offers synthesis and (approximate) analysis up to $\ell_{\rm max} \approx 3\times N_{\rm side}$:

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

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

Wiener filters and friends

In the next slides, we focus in the generic problem of estimating the $C \times 1$ 'component vector' s from F noisy mixtures available on a vector x:

$x = As + n$ The 'mixing matrix' A has size $F \times C$

Ask to yourself (or to me): when is it critical that $F \geq C$?

Often, but not necessarily always, we consider the case of redundant observations: matrix A is 'tall' with linearly independent columns.

Some notations, basic properties

1. The $n \times n$ identity matrix I_n. Also denoted I if clear from context.

2. Transpose: $[A^\dagger]_{ij} = A_{ji}$; trace: $\operatorname{tr} A = \sum_i A_{ii}$;
   Scalar product $\langle A|B\rangle = \sum_i \sum_j A_{ij} B_{ij} = \operatorname{tr} AB^\dagger$;
   Euclidian norm $\|A\|^2 = \sum_i \sum_j A_{ij}^2 = \operatorname{tr} AA^\dagger$ (works for vectors and matrices).

3. Column space of a matrix A denoted $\mathrm{Span}(A)$.

4. Moore-Penrose pseudo inverse $A^\#$.
   For a full-column-rank A, it is $A^\# = (A^\dagger A)^{-1} A^\dagger$.
   It is one of these matrices such that $A^\# A = I$ and
   the unique matrix such that $AA^\#$ is the orthogonal projector onto $\mathrm{Span}(A)$.

5. Square root of a non-negative matrix R:
   any matrix W such that $R = WW^\dagger$.

6. Expectation $\mathrm{E}(X)$ of a random variable.
   $\mathrm{E}(X|Y)$: expectation of X conditonned on observing Y.

7. Covariance matrix of a random vector X: $\mathrm{Cov}(X) = \mathrm{E}\, XX^\dagger - \mathrm{E}\, X\ \mathrm{E}\, X^\dagger$.
   Cross-covariance for vectors X and Y: $\mathrm{Cov}(X,Y) = \mathrm{E}\, XY^\dagger - \mathrm{E}\, X\ \mathrm{E}\, Y^\dagger$.

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:

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

The solution is the conditional expectation of X given Y

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

Often called 'the Wiener filter'.

Proof:

The best linear filter

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

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

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

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

This is linear in Y !

Statistical efficiency versus simplicity

For non Gaussian observations, the best processor (in terms of mean squqred 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:

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

The relative reconstruction error

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

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 $\mathsf{E}\, s_p^2 = \sigma^2_{\mathsf{cmb}}$ and $\mathsf{E}\, s_p^2 = \sigma_n^2$ 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

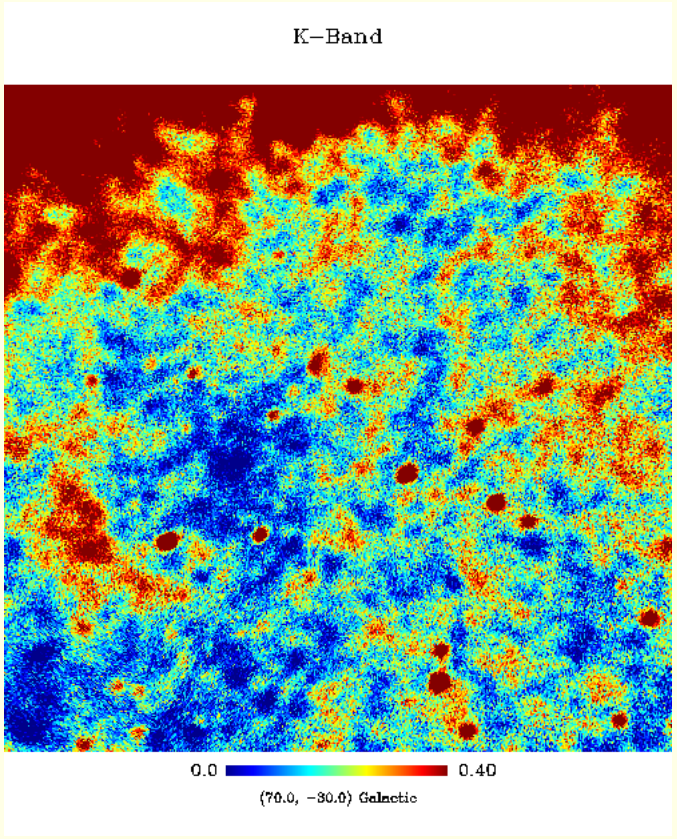
$$\mathsf{E}(s_{\ell m} s_{\ell' m'}) = C(\ell)\ \delta_{\ell\ell'}\,\delta_{mm'} \qquad \mathsf{E}(n_{\ell m} n_{\ell' m'}) \approx \Omega\,\sigma_n^2\ \delta_{\ell\ell'}\,\delta_{mm'} \qquad \Omega = \frac{4\pi}{N_{\mathsf{pix}}}$$

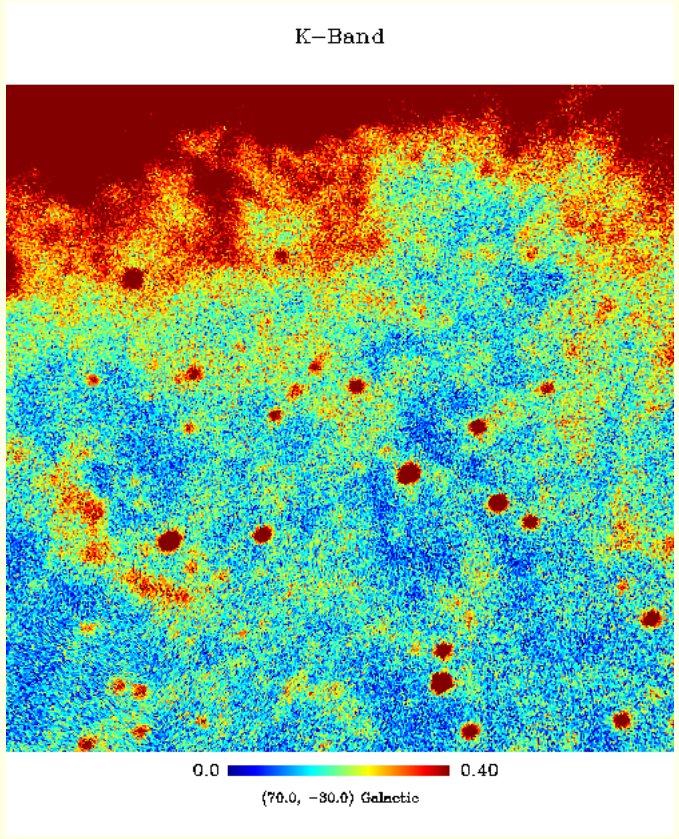
It exposes the SNR contrast and justifies mode-wise processing, namely:

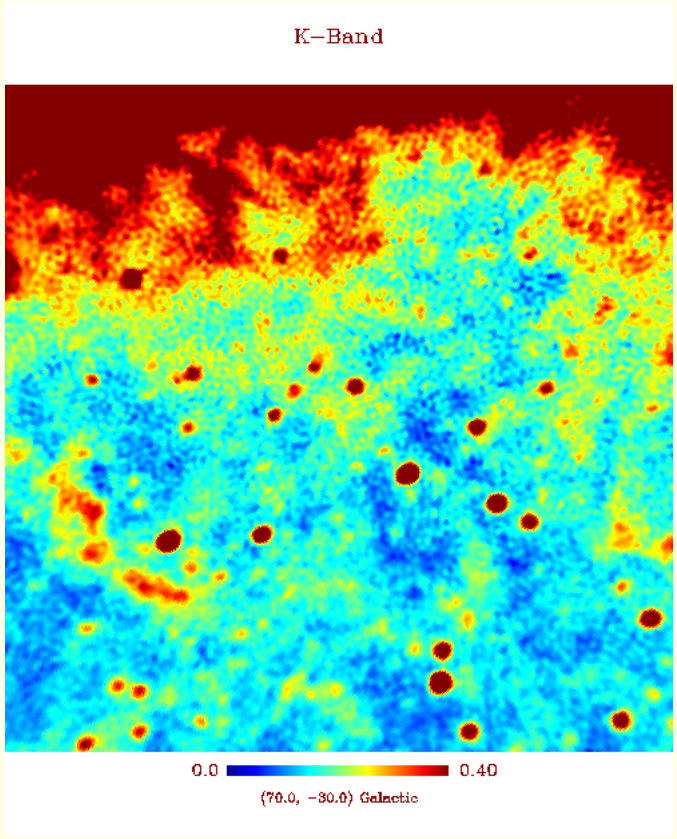
$$\widehat{s}_{\ell m} = x_{\ell m}\ \frac{C(\ell)}{C(\ell) + \Omega\,\sigma_n^2} \qquad \text{The 'Wiener beam'}$$

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.

Note: We are not seeing stationary processes here.
But recall that we are applying the best linear filter.
Doing better is vastly more complicated and hard to characterize.

Wiener filter for Gaussian vectors

Things get 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 $\mathrm{Cov}(s) = \mathbf{S}$ and $\mathrm{Cov}(n) = \mathbf{N}$. 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 $R_{xx} = \mathrm{Cov}(As + n) = A\mathbf{S}A^\dagger + \mathbf{N}$ so

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

So the best reconstruction of the observations is

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

Compare to the scalar case.

The alternate form of Wiener and the high SNR limit

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

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

The second form makes it clear that, in the high SNR limit, thats is when $A^\dagger\mathbf{N}^{-1}A \ll \mathbf{S}^{-1}$, 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 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 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 such that $\mathsf{E}\,\|Wx - s\|^2$ is minimum under the 'unbiasedness' constraint, that is, $WA = I$.

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

Solution: form the Lagrangian:

$$\mathcal{L}(W, \Lambda) = \mathsf{E}\,\|Wx - s\|^2 + \mathrm{tr}\,\Lambda^\dagger(WA - I)$$

and solve to find:

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

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

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

and two limits: the BLUE W_{u} (enforcing unbiasedness) and the high SNR Wiener W_∞:

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

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.

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

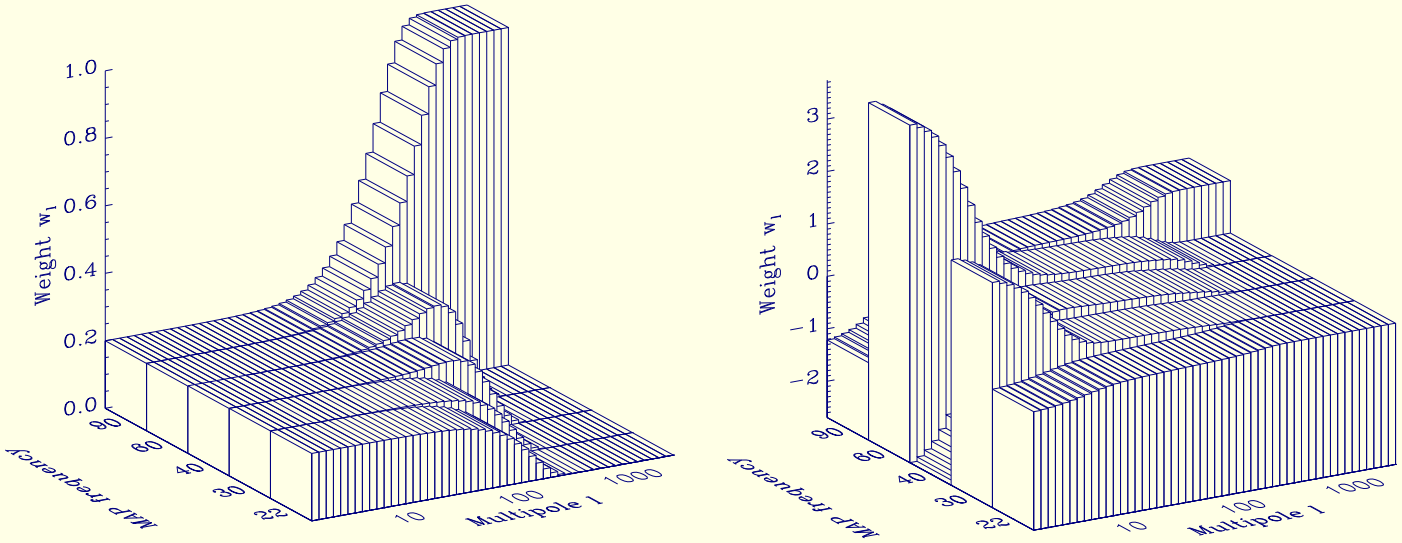
Then, let's rephrase in terms of BLUE output. Define

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

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 $1/(1 + \mathrm{SNR}^{-1})$:

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

Internal linear combination (ILC)



From Tegmark '99

Linear 'unbiased' combination of the frequency channels in harmonic space:

$$\widehat{s}^{\mathsf{cmb}}_{\ell m} = \sum_{i=1}^{5} w^i_\ell x^i_{\ell m}$$

Figure assumes CMB units

Left: fighting only (homogeneous) white noise (see the beam effect).
Right: fighting everything, noise and foregrounds.

The ILC and its Wiener version

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

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

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

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

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

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

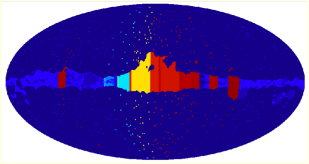
4. Estimation of missing quantities.
   - BLUE: $\mathbf{X}$ estimated by a sample average $\widehat{\mathbf{X}}$ in the appropriate domain.
   - Wiener: What about $C_\ell, \mathbf{N}_\ell$? Estimation using Planck jackknives (woohoo!)

ILC and localization

- Pixel-based ILC (bad!)

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

with the data covariance matrix estimated globally over the sky.

- Localize by working on sky domains  and stitching.

- Localize in harmonic space (remember Tegmark's figure)

$$\widehat{s}_{\ell m} = \frac{\mathbf{a}^\dagger \widehat{\mathbf{X}}_\ell^{-1} x_{\ell m}}{\mathbf{a}^\dagger \widehat{\mathbf{X}}_\ell^{-1}\mathbf{a}} \qquad \widehat{\mathbf{X}}_\ell = \mathrm{Smooth}\left[\frac{1}{2\ell+1}\sum_m x_{\ell m} x_{\ell m}^\dagger\right]$$

- Localize in both space and multipole using wavelets. Needlet ILC.

'Optimally' expose the local SNR condition.

And now for something different

Blind component separation

a.k.a. ICA: Independent component analysis.