The galactic cocktail
a blind approach to the component separation

Simona Donzelli

the cocktail-party problem

- two people speaking simultaneously: s_1, s_2
- two microphones recording time signals: x_1, x_2

$$\begin{aligned} x_1(t) &= a_{11}s_1 + a_{12}s_2 \\ x_2(t) &= a_{21}s_1 + a_{22}s_2 \end{aligned} \quad \rightarrow \mathbf{x} = \mathbf{A}\mathbf{s}$$

$\mathbf{A}$= mixing matrix

having x_1, x_2 , how to get s_1, s_2 ?

the problem in astrophysics

- superposition of N different emissions

$$\tilde{x}(\boldsymbol{r},\nu) = \sum_{j=1}^{N} \tilde{s}_j(\boldsymbol{r},\nu)$$

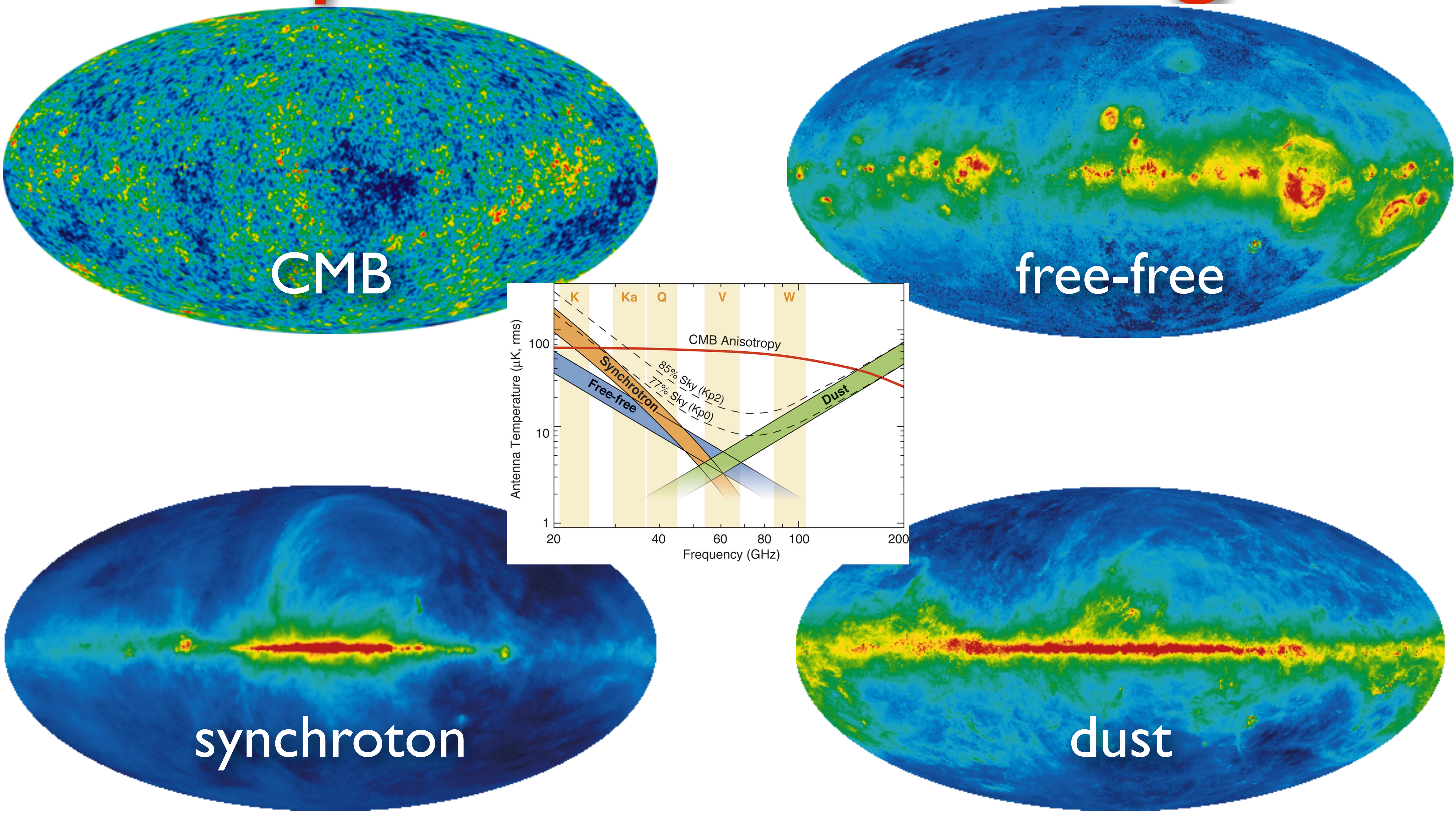
- experiment with M frequency channels

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad \text{(+ noise)}$$

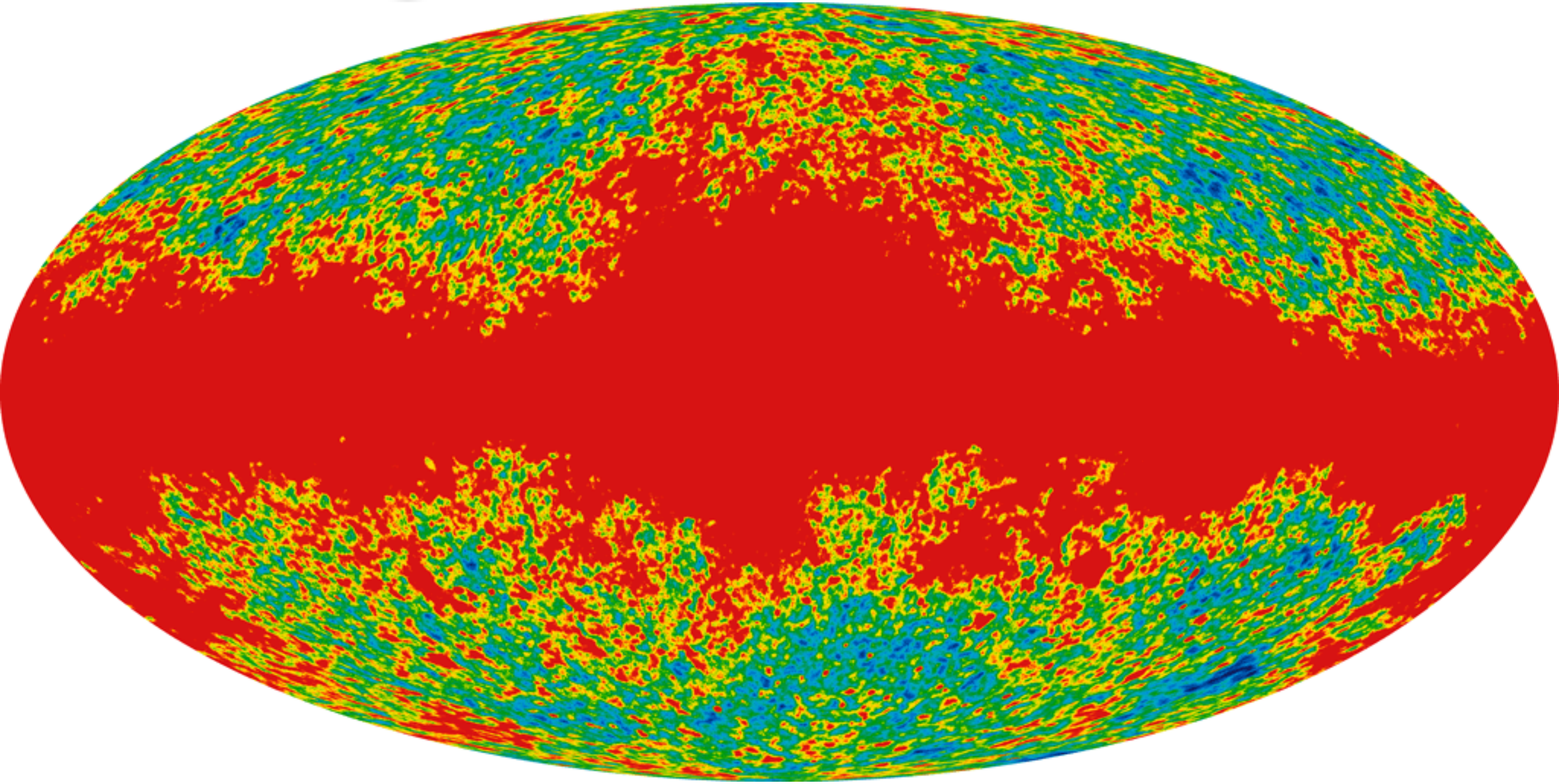
mixing matrix MxN characterized by PSF and frequency response

how to separate the different emissions?

example: microwaves signals

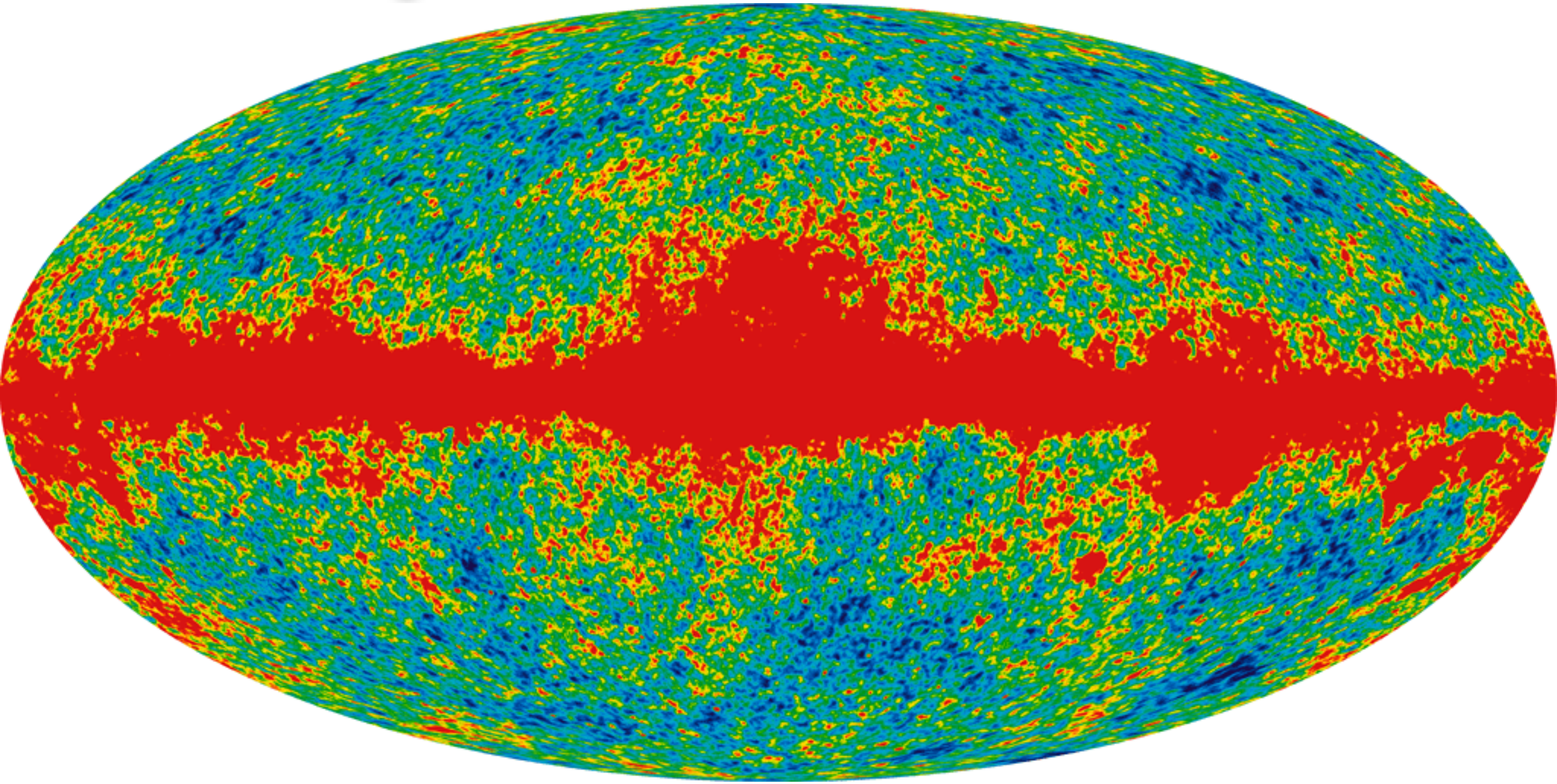


example: microwaves data



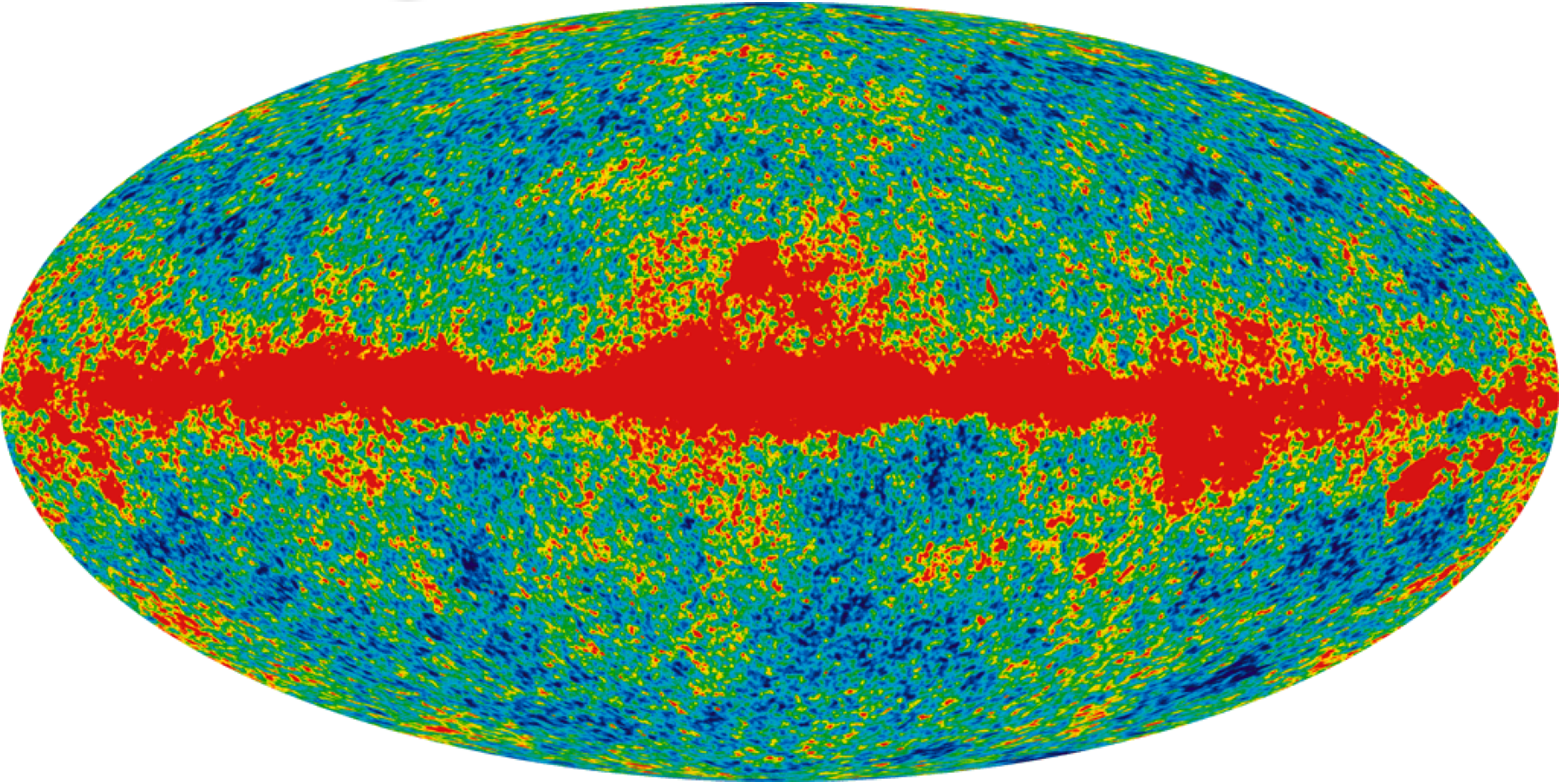
WMAP 7yr - K 23 GHz

example: microwaves data



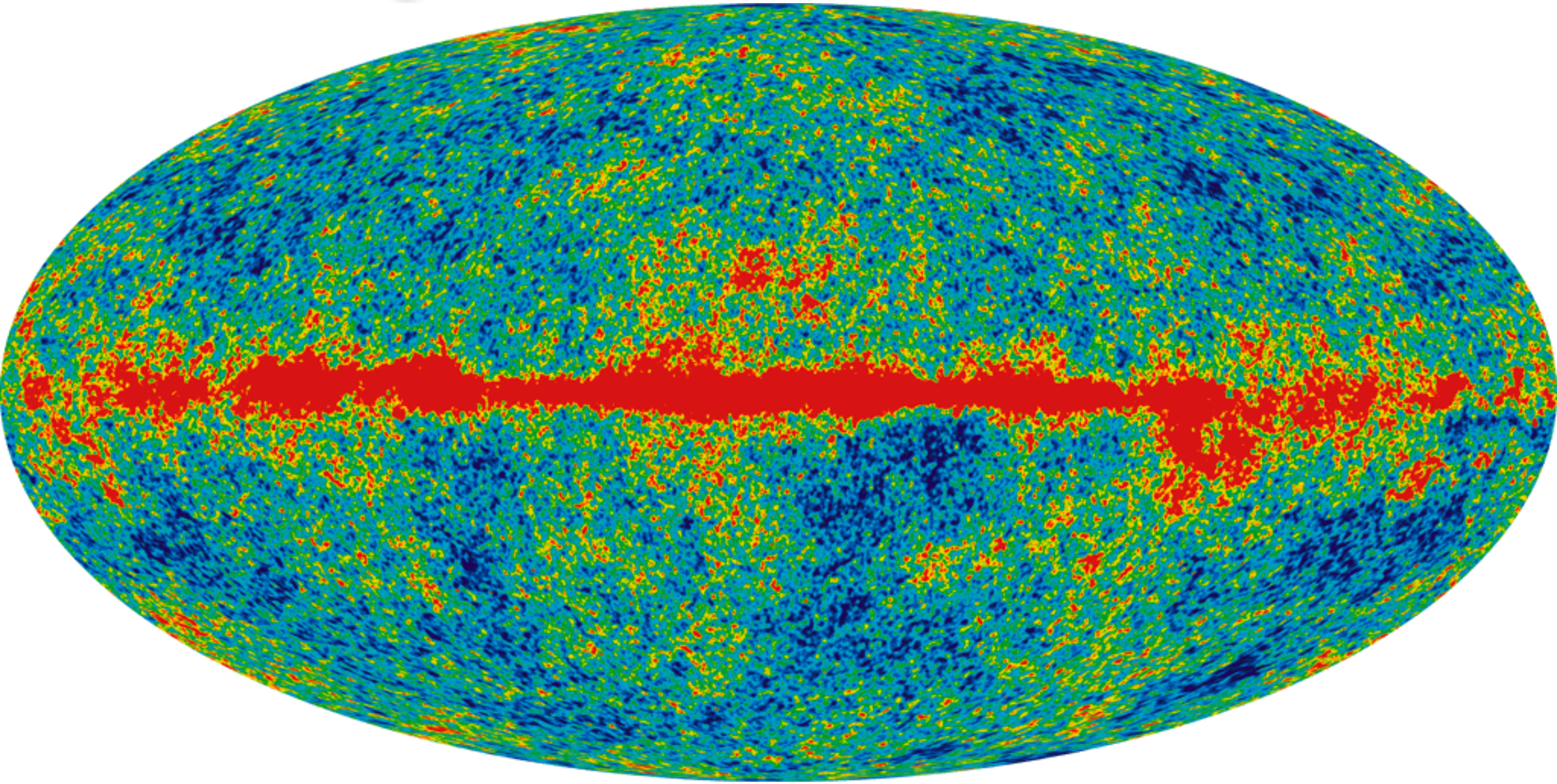
WMAP 7yr - Ka 33 GHz

example: microwaves data



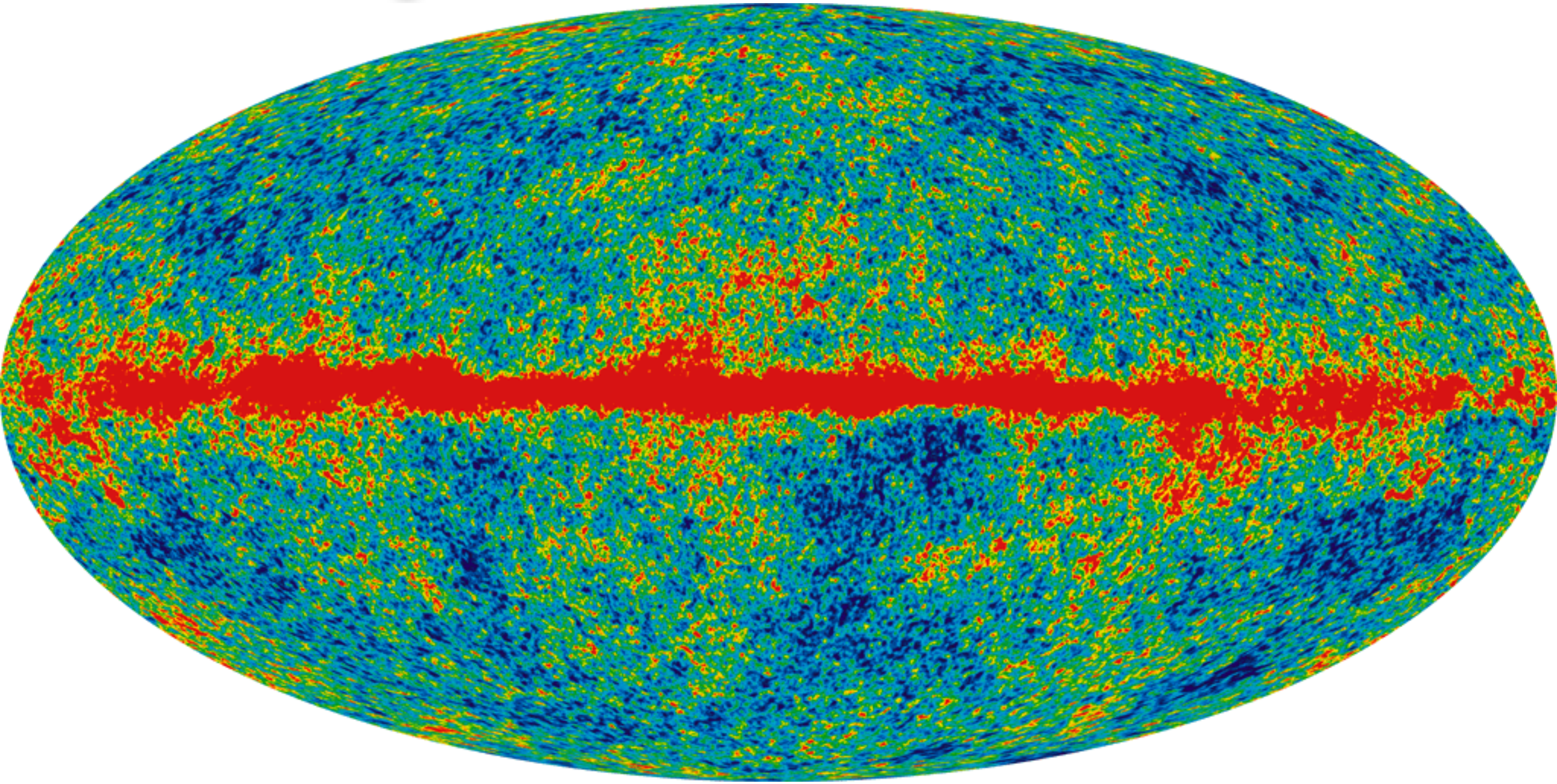
WMAP 7yr - Q 41 GHz

example: microwaves data



WMAP 7yr - V 61 GHz

example: microwaves data



WMAP 7yr - W 94 GHz

ICA: a blind approach

two assumptions:

- the signals s are statistical independent
- all the signals, but at most one, have a non-Gaussian distribution

$$\Downarrow$$

the **I**ndependent **C**omponent **A**nalysis can solve the component separation problem <u>without</u> 'a priori' knowledge on the signals

principle of ICA

solution: find a transformation $\mathbf{W}$ such that the vector $\mathbf{y}=\mathbf{W}\mathbf{x}$ has independent components

we need a measure of independence!

Central Limit Theorem:

the distribution of a sum of random independent variables tends towards a gaussian distribution

non-Gaussian is independent

$$y = \boldsymbol{w}^T \mathbf{x} = \boldsymbol{w}^T \mathbf{A}\mathbf{s} = \mathbf{z}^T \mathbf{s}$$

- y linear combination of s_i ⇨ more gaussian
- y becomes least gaussian when $y = s_i \quad (\mathbf{z} = z_i)$
- $\boldsymbol{w}^T$ maximizes the non-gaussianity of $\boldsymbol{w}^T \mathbf{x}$

measure of non-gaussianity

entropy of a random variable = degree of information needed to describe the variable, i.e. the coding length (more random, i.e. more unpredictable = larger entropy)

differential entropy H of a random vector $\mathbf{y}$ with density $f(\mathbf{y})$:
$$H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) \mathrm{d}\mathbf{y}$$

a gaussian variable has the largest entropy among all random variables of equal variance [information theory]

<u>negentropy</u>:
$$J(\mathbf{y}) = H(\mathbf{y}_{gauss}) - H(\mathbf{y})$$

negentropy

if $<y>=0$ and $<y^2>=1$ then

$$J(y) \propto [E\{G(y)\} - E\{G(\nu)\}]^2$$

where

- E expected value
- ν Gaussian variable of zero mean and unit variance
- G non quadratic functions

preprocessing for ICA

- centering:

$$\tilde{\mathbf{x}} = \mathbf{x} - E\{\mathbf{x}\} \quad \text{(zero mean)}$$

- (quasi) whitening:

we want $\hat{\mathbf{x}}$ white (uncorrelated component and unit variance)

$$\begin{aligned} \hat{\mathbf{\Sigma}} &= (\mathbf{C} - \mathbf{\Sigma})^{-1/2} \mathbf{\Sigma} (\mathbf{C} - \mathbf{\Sigma})^{-1/2} \\ \hat{\mathbf{x}} &= (\mathbf{C} - \mathbf{\Sigma})^{-1/2} \tilde{\mathbf{x}} \end{aligned}$$

where $\mathbf{C}$ = data covariance, $\mathbf{\Sigma}$ = noise covariance

$\Rightarrow$ mixing matrix $\mathbf{A}$ is orthogonal: $\mathrm{M}(\mathrm{M}\text{-}1)/2$ d.o.f.

the FastICA algorithm

$\mathbf{W}$ estimated *row by row*, i.e. components *one by one*:

1. choose initial vector $\boldsymbol{w}$
2. $\boldsymbol{w}_{new} = E\{\hat{\mathbf{x}} g(\boldsymbol{w}^T \hat{\mathbf{x}})\} - (I + \hat{\boldsymbol{\Sigma}})\boldsymbol{w} E\{\, g'(\boldsymbol{w}^T \hat{\mathbf{x}})\}$
3. $\boldsymbol{w}_{new} = \boldsymbol{w}_{new} / \|\boldsymbol{w}_{new}\|$
4. if not converged back to 2., otherwise new row

- $g = G'$ can be $p(u) = u^3$, $g(u) = u\exp(-u^2)$, $t(u) = \tanh(u)$
- $\mathrm{(k+1)th}$ row must be orthogonal to first k rows: orthogonalization step between 2. and 3.

recovered components

- linear combination

$$y_j = \sum_{i=1}^{M} w_{ji}\, x_{\nu_i} \quad j = 1, \ldots, M$$

- frequency scaling

$$(W^{-1})_{\nu j} / (W^{-1})_{\nu' j}$$

$$\Rightarrow \ \beta = \log[(W^{-1})_{\nu j} / (W^{-1})_{\nu' j}] / \log(\nu/\nu')$$

- ambiguities:
  - normalization factor
  - extraction order

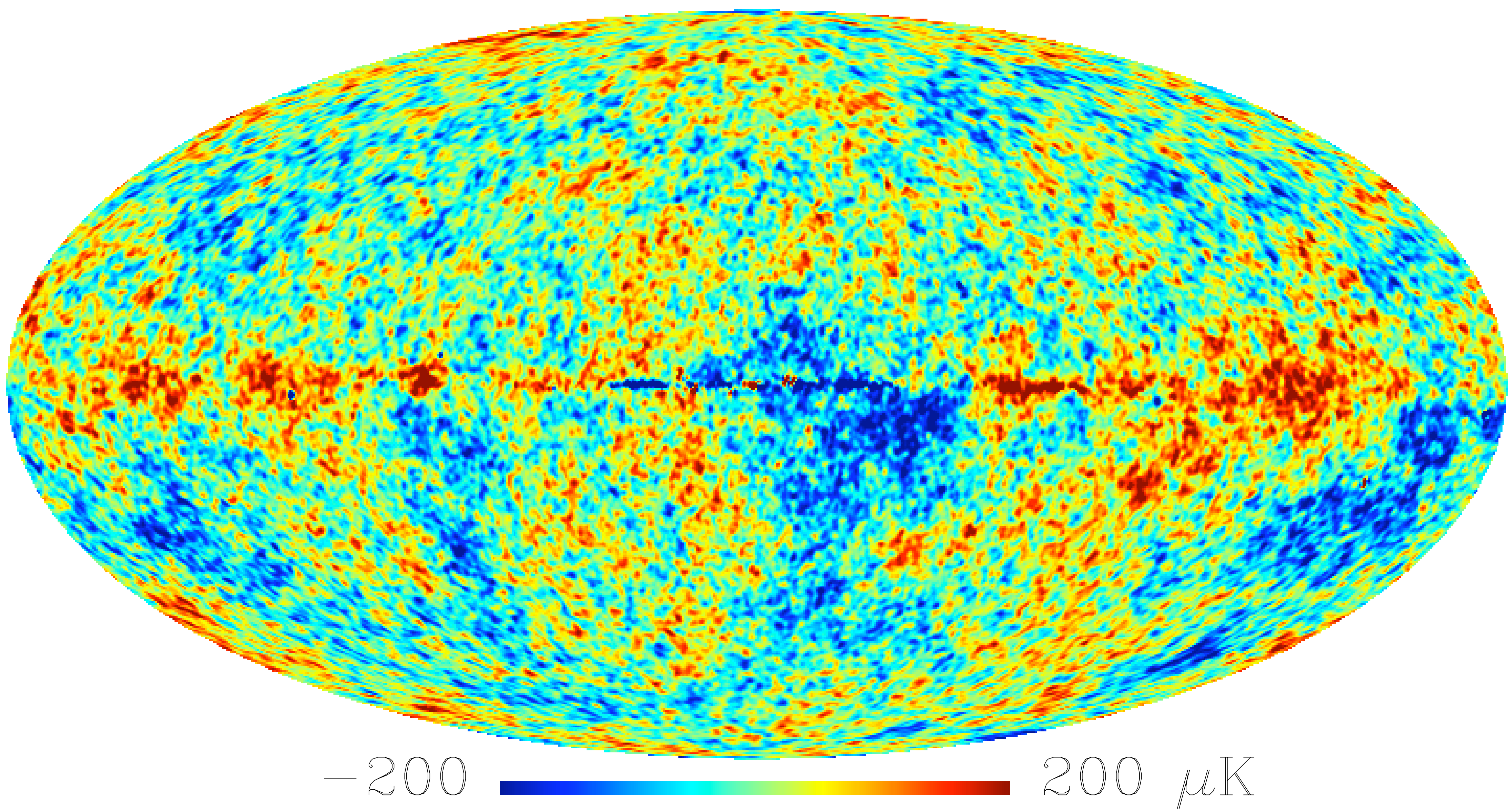
applications of FastICA

- public code: http://research.ics.tkk.fi/ica/fastica/
- demo: http://research.ics.tkk.fi/ica/cocktail/cocktail_en.cgi
- different fields of application (not only astrophysics):
  - Impact of Contrast Functions in Fast-ICA on Twin ECG Separation (Fetal Electrocardiography) - Feb 2011
  - Energy-Efficient FastICA Implementation for Biomedical Signal Separation - Nov 2011
  - Separation of spinal cord motor signals using the FastICA method - Dec 2005
  - Speech Separation in the Vehicle Environment Based on FastICA Algorithm - Feb 2012
  - Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network - Aug 2006
  - etc.....

FastICA in astrophysics

- CMB
- 21-cm reionization signal
- Cosmological Stochastic Background of Gravitational Waves
- spectroscopic analysis of complex mixtures
- infrared emission spectra of interstellar dust
- ....

FastICA on the CMB



application to WMAP 3yr data [Maino et al. MNRAS 374, 1207 (2007)]

references

- FastICA technique:

  "Independent Component Analysis: A Tutorial", A. Hyvarinen (1999) [Neural Networks, 13(4-5):411-430, 2000]

- FastICA in astrophysics:

  Maino et al., MNRAS, 334, 53 (2002)