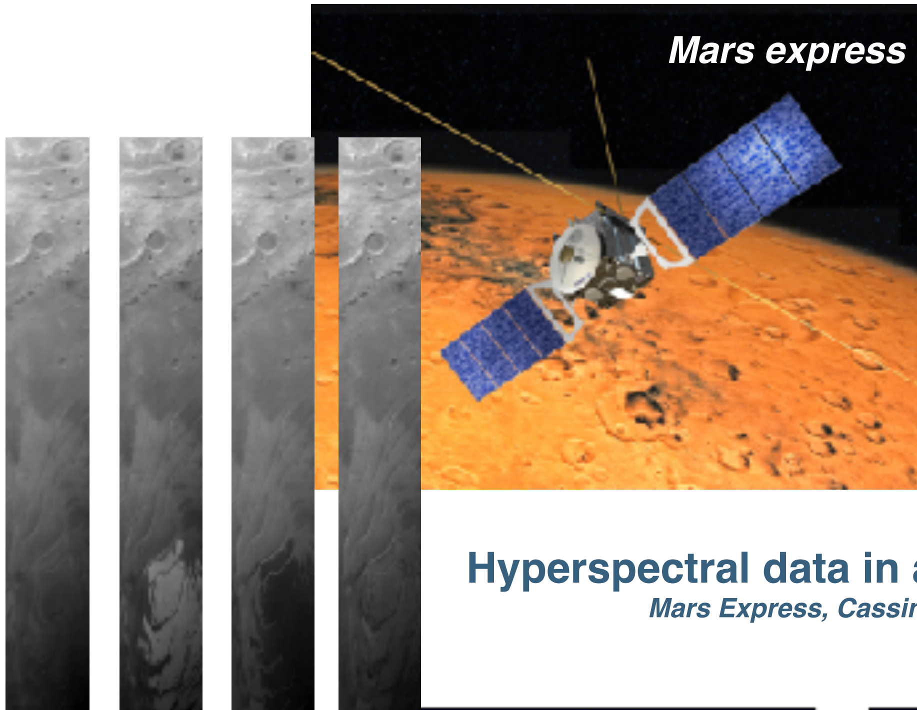
Tackling data analysis challenges in astrophysics with sparse matrix factorization methods

Jérôme Bobin

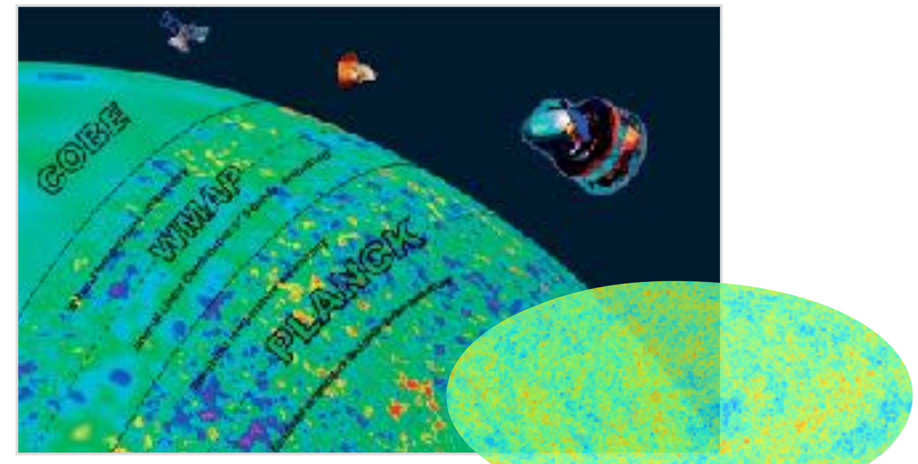
joint work with C.Chenot, J.Rapin, M.Jiang, F.Sureau, J-L Starck

Laboratoire CosmoStat - CEA/Irfu, France

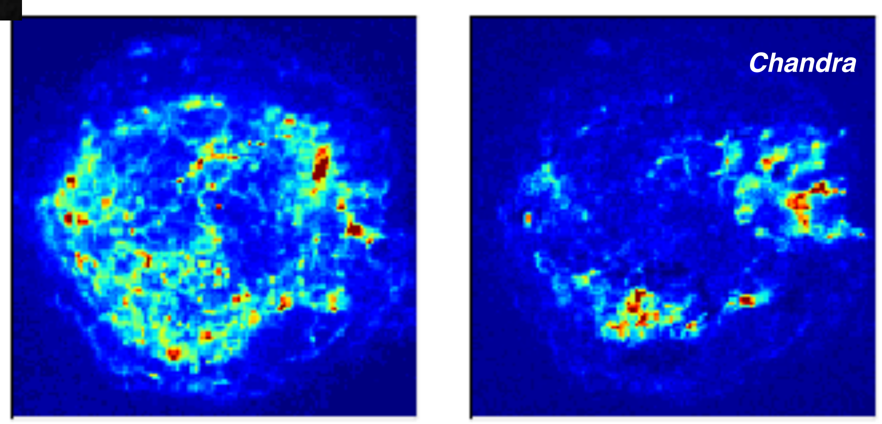
The context: analyzing multivalued data

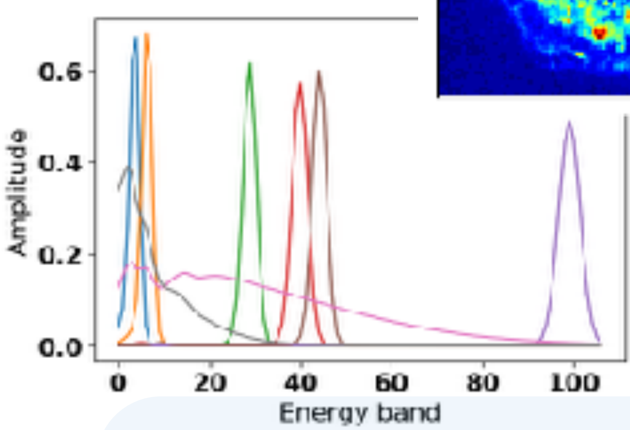


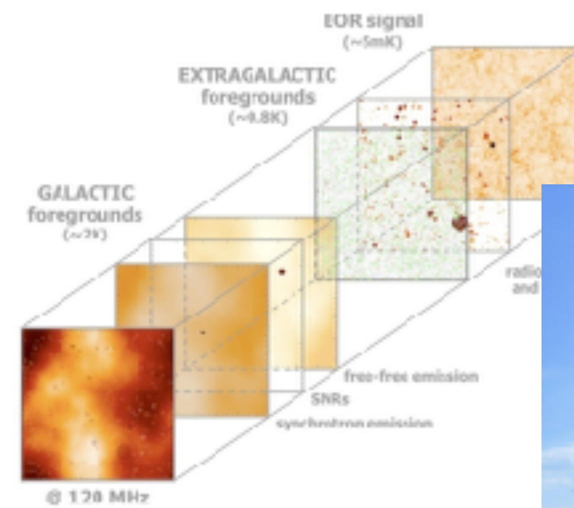
Hyperspectral data in astrophysics
Mars Express, Cassini, etc.



Multispectral data in astrophysics
Planck, Fermi, etc.





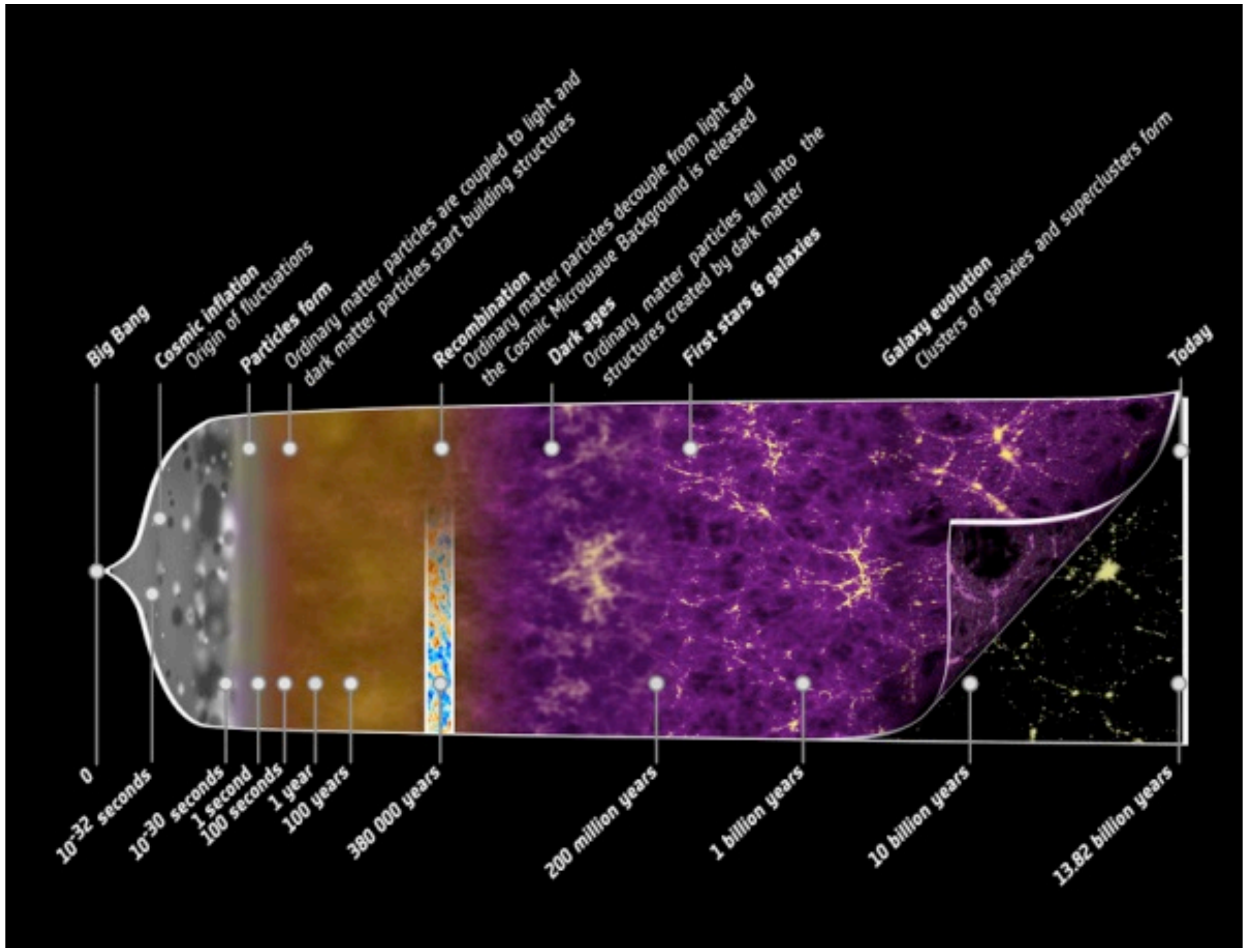


Radio-interferometric data
LoFar, SKA, etc.

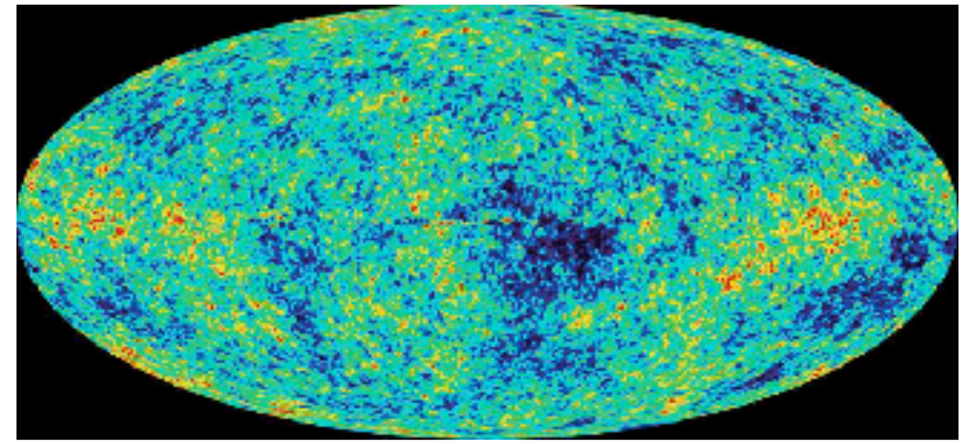
Different scientific fields but …

common problems: mixtures of elementary signals or sources

A key application in cosmology



The Cosmic Microwave Background (CMB) is a relic radiation (with a temperature equals to 2.726 Kelvin) emitted 13 billion years ago when the Universe was about 370 000 years old.

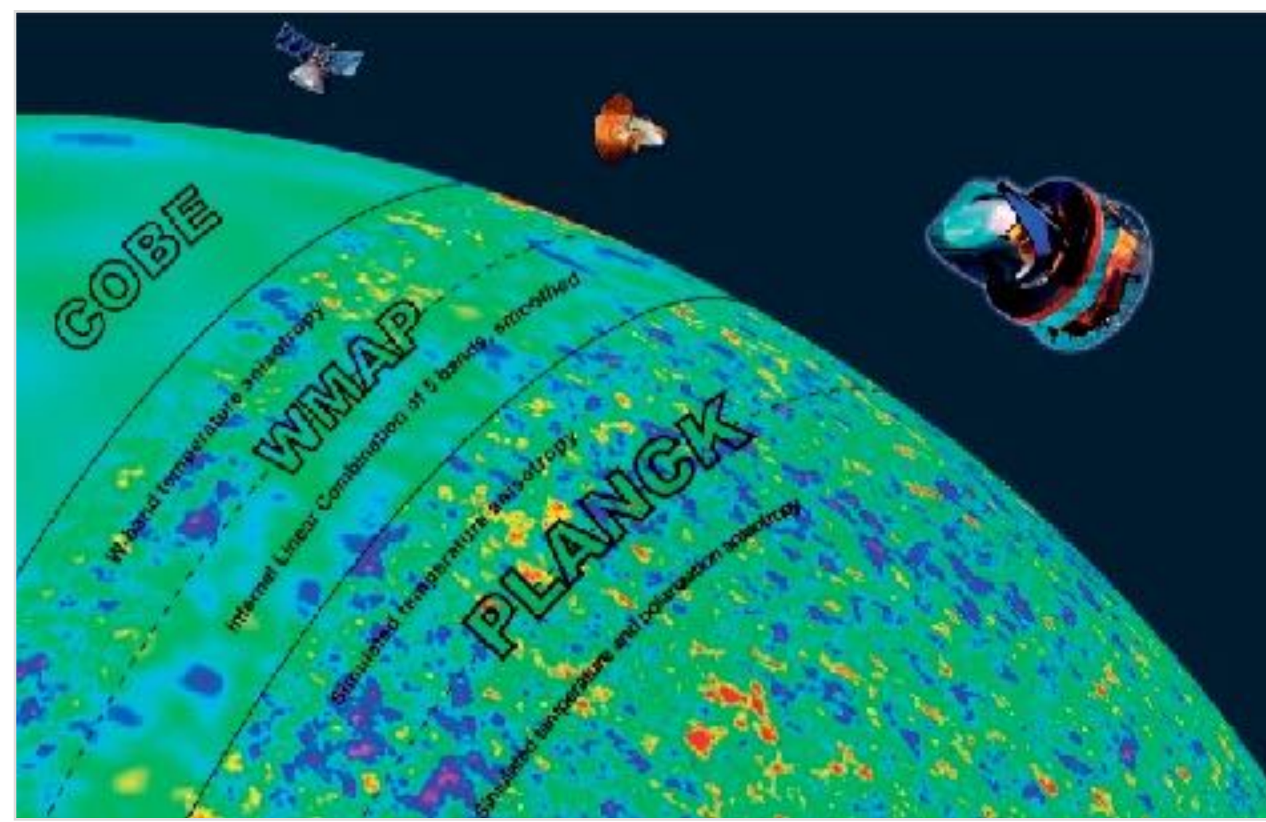


- The CMB is fundamental to study the dawn of our universe !

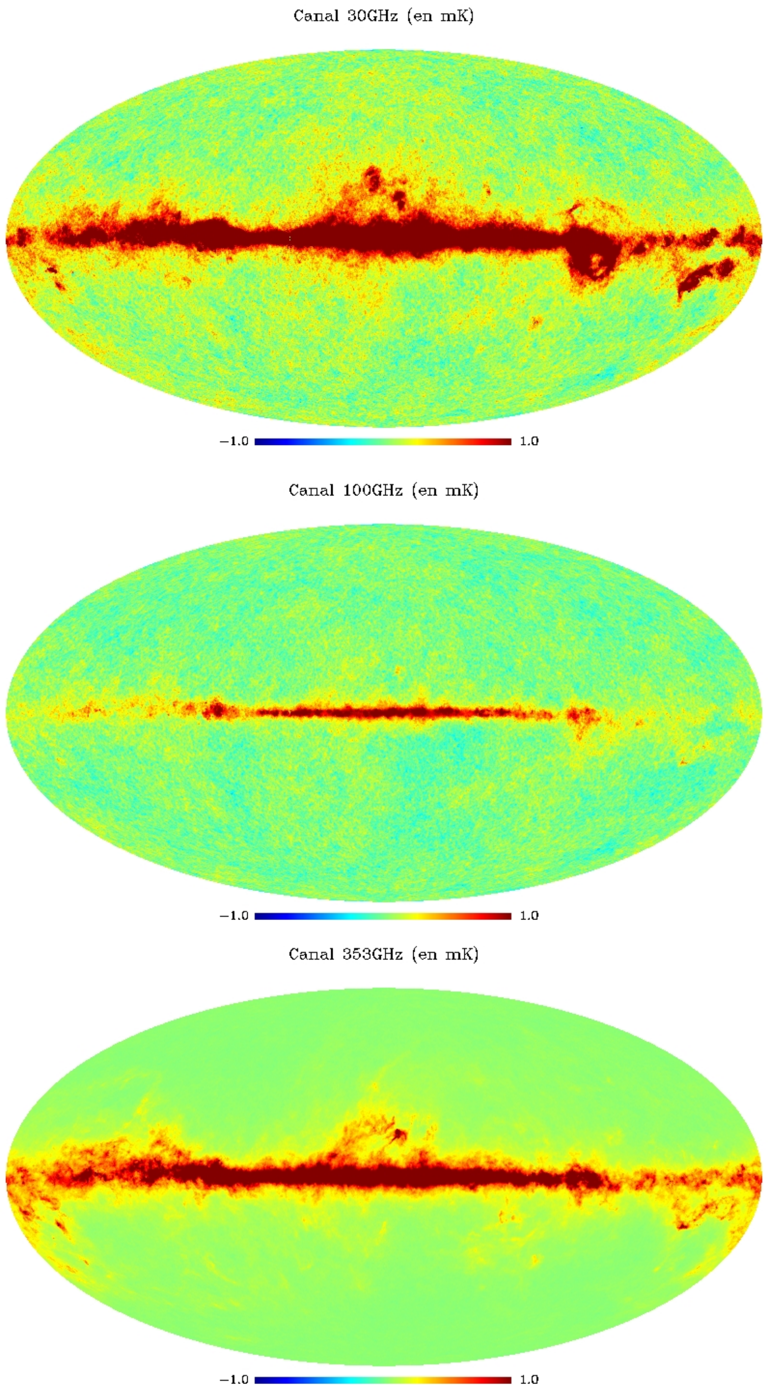
- PLANCK provides full-sky data in 9 channels in the range 30GHz - 857GHz

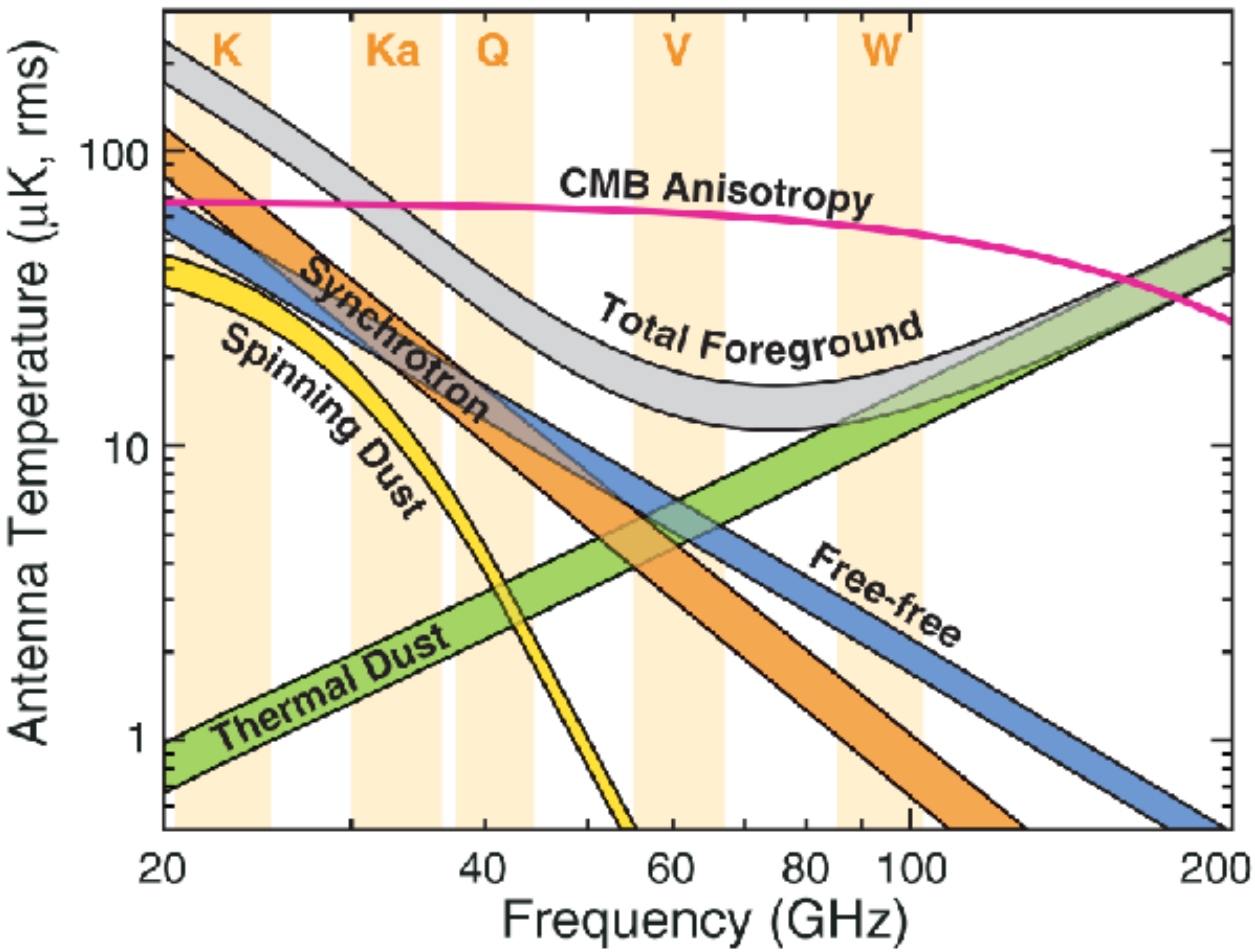
… and 7 are sensitive to polarization (30GHz - 353GHz)

- High resolution data of (up to 5 arcmin)



CMB estimation as a BSS problem

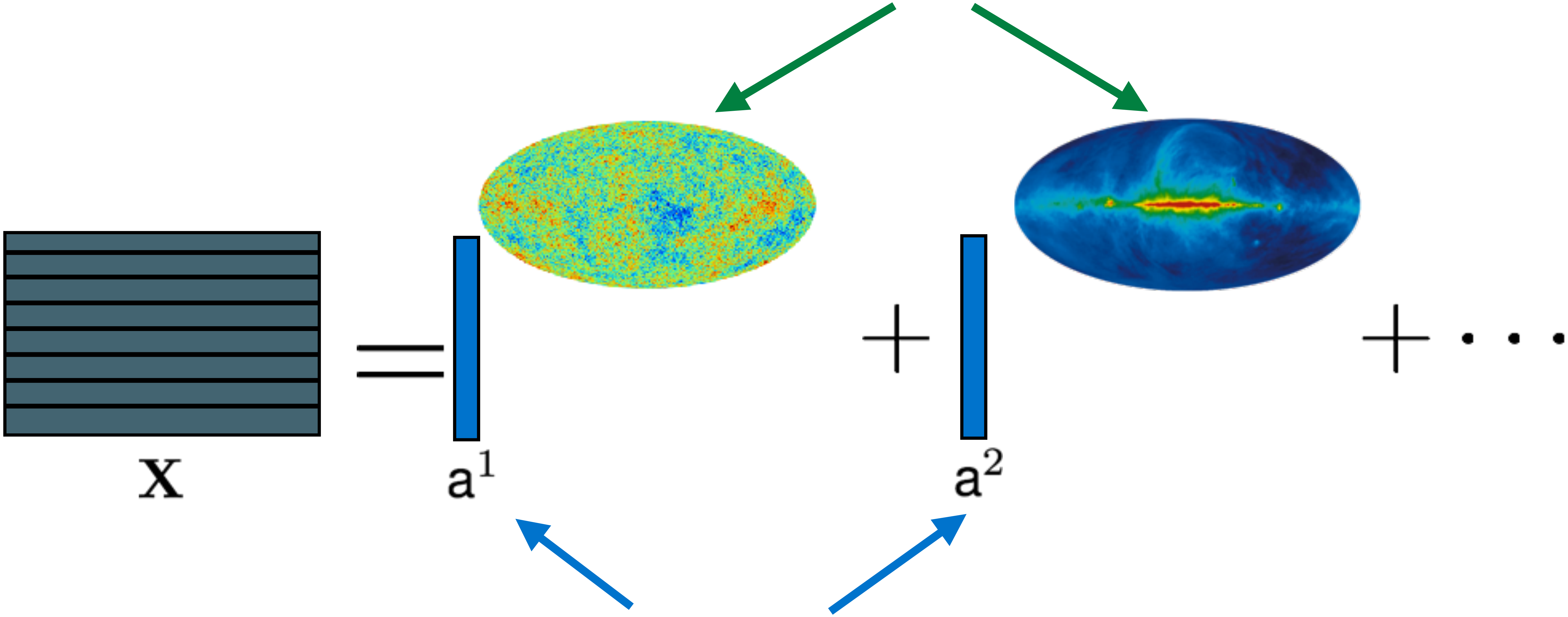




Estimating the CMB is a BSS problem

The linear mixture model

The sources



The mixture weights

e.g. spectral signatures, electromagnetic spectra, etc.

BSS: Blind Source Separation

The source matrix

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}$$

The mixing matrix

Noise

Blind Source Separation:

Estimation both A and S from X only

This is an ill-posed matrix factorization problem

Non-negative Matrix Factorization, Clustering, Classification, Dictionary Learning

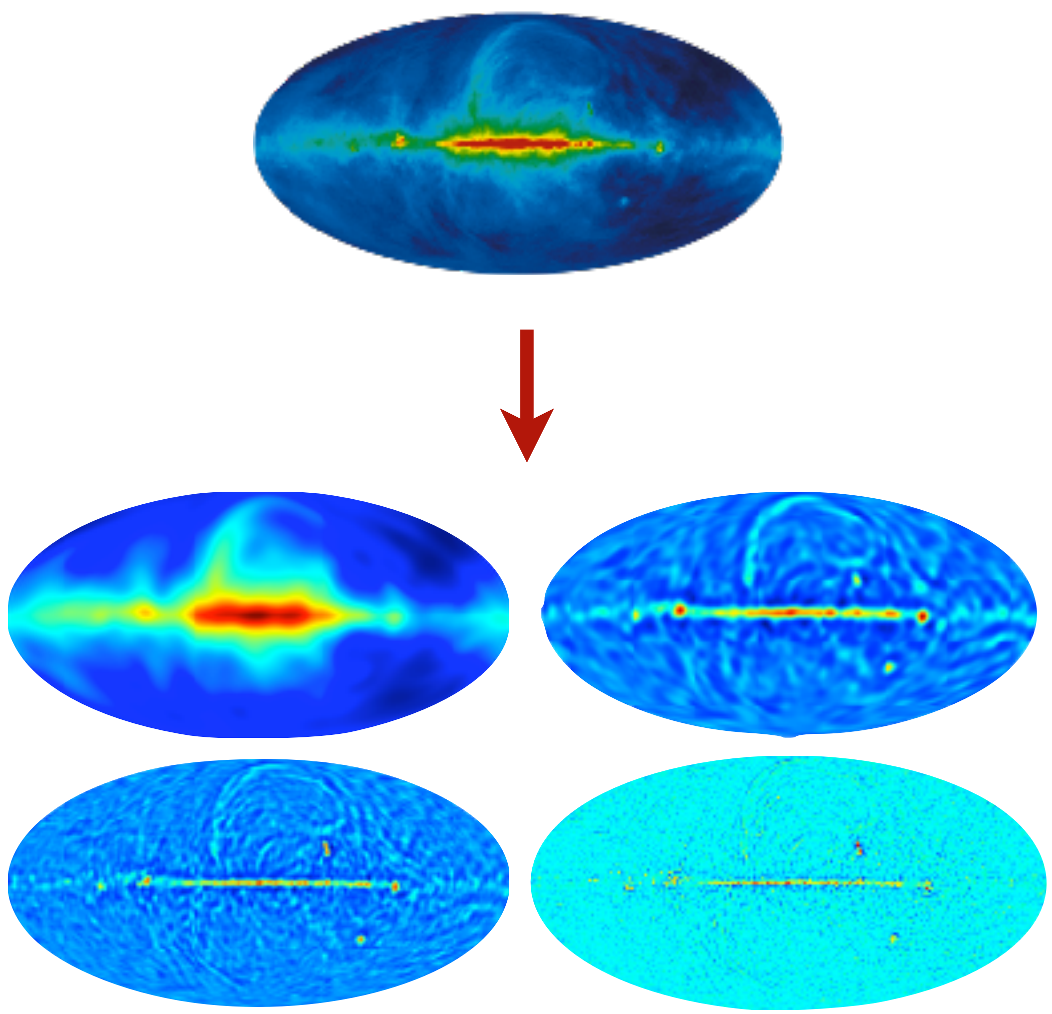
Sparse signal modeling at a glance

Prior information on S and/or A

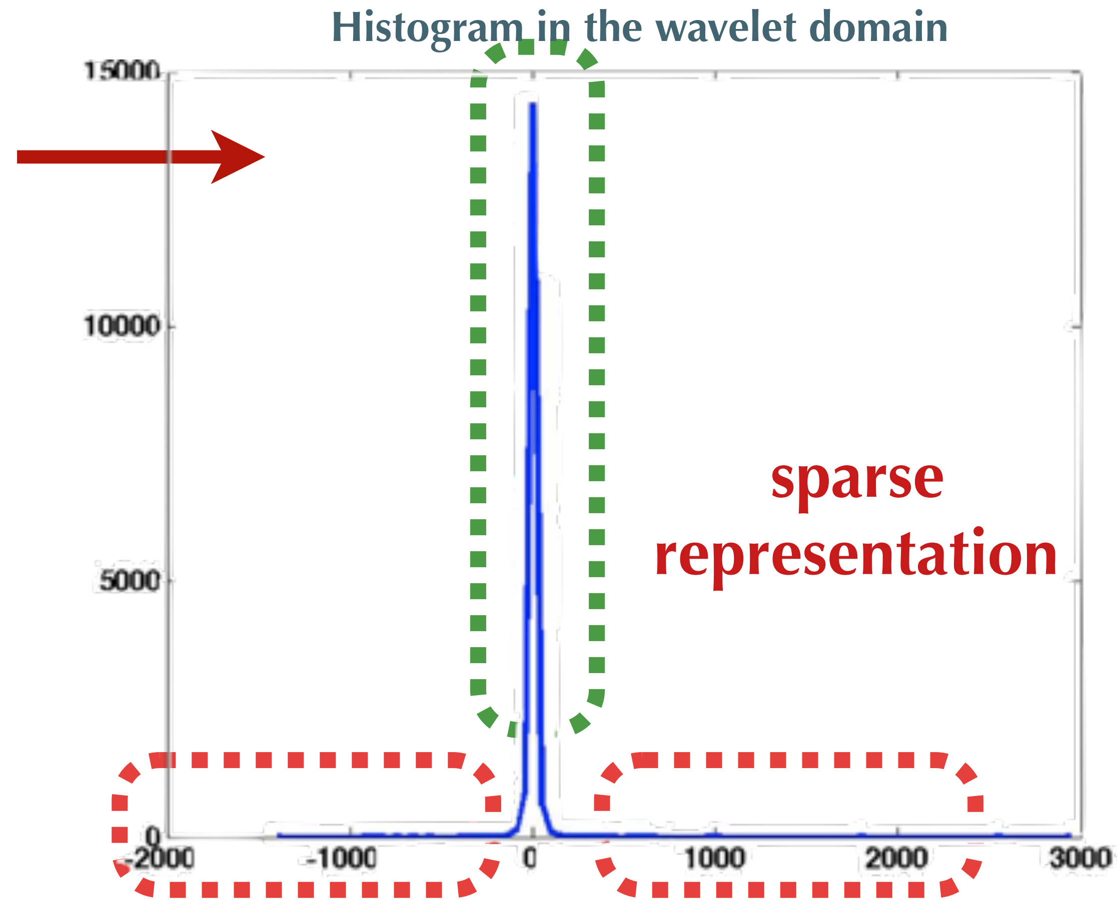
Statistical independence, non-negativity, etc.

Sparse signal modeling

Zibulevsky01, Cichocki06, Bobin07



Wavelet transform for spherical data



The building block: GMCA

Gist: looking for the sparsest sources

Regularization params., weight matrix, etc.

$$\min_{\mathbf{A},\mathbf{S}} \underbrace{\|\Lambda \odot \mathbf{S}\mathbf{W}\|_p}_{\text{Sparse regularization}} + \frac{1}{2}\underbrace{\|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2}_{\text{Data fidelity term}}$$

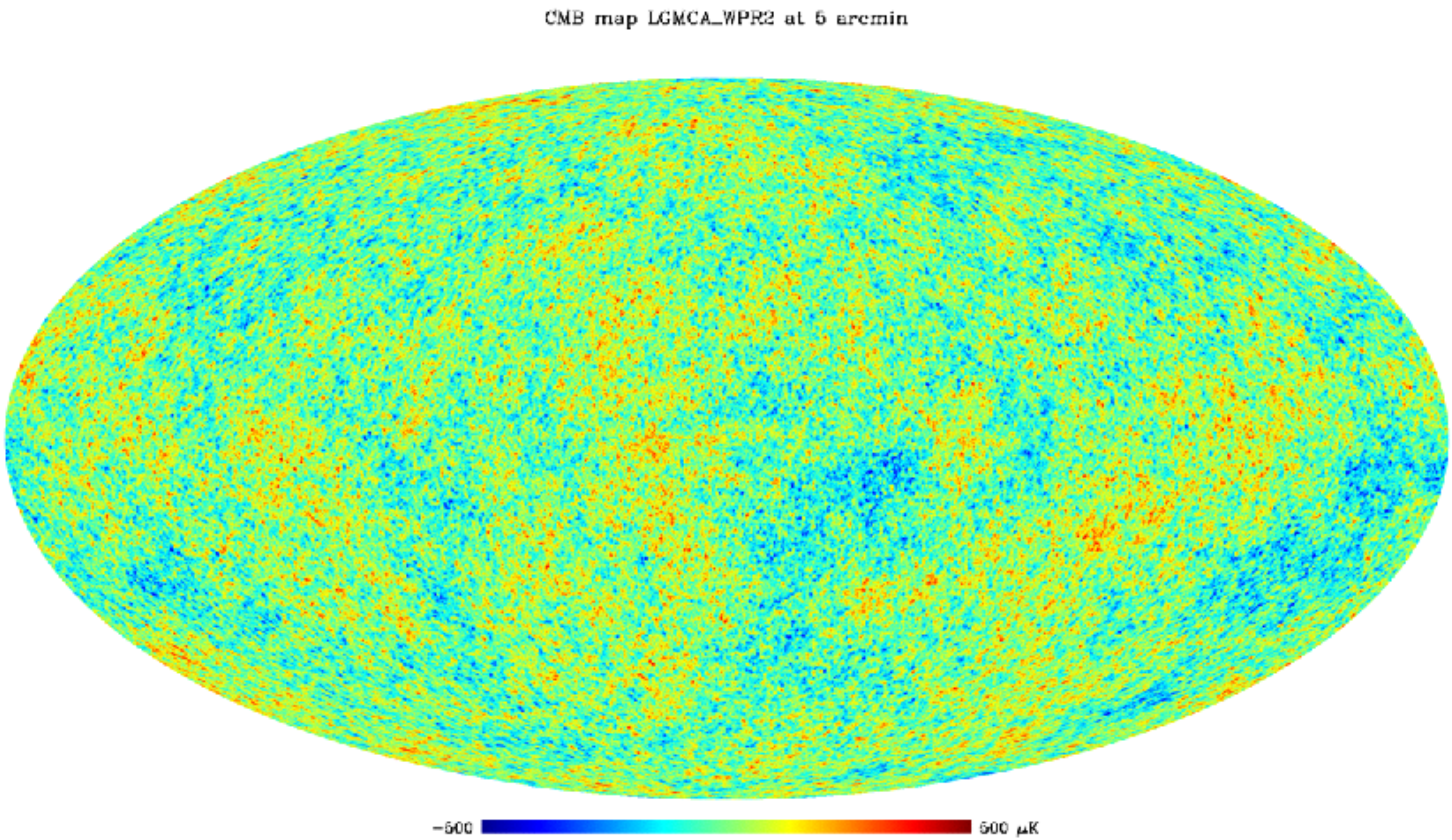
Generalized Morphological Component Analysis (GMCA):

- *S-BSS with redundant sparse representations*
- *Iterative soft/hard thresholding algorithm*
- *Thresholding strategy, robustness to Gaussian noise/local stationary points*
- *No parameters to tune*

Bobin, Starck, Fadili, and Moudden, Sparsity, Morphological Diversity and Blind Source Separation, IEEE Trans. on Image Processing, Vol 16, No 11, pp 2662 - 2674, 2007.

Bobin, Starck, Fadili, and Moudden, Blind Source Separation: The Sparsity Revolution, Advances in Imaging and Electron Physics , Vol 152, pp 221 -- 306, 2008.

Applications to the Planck data

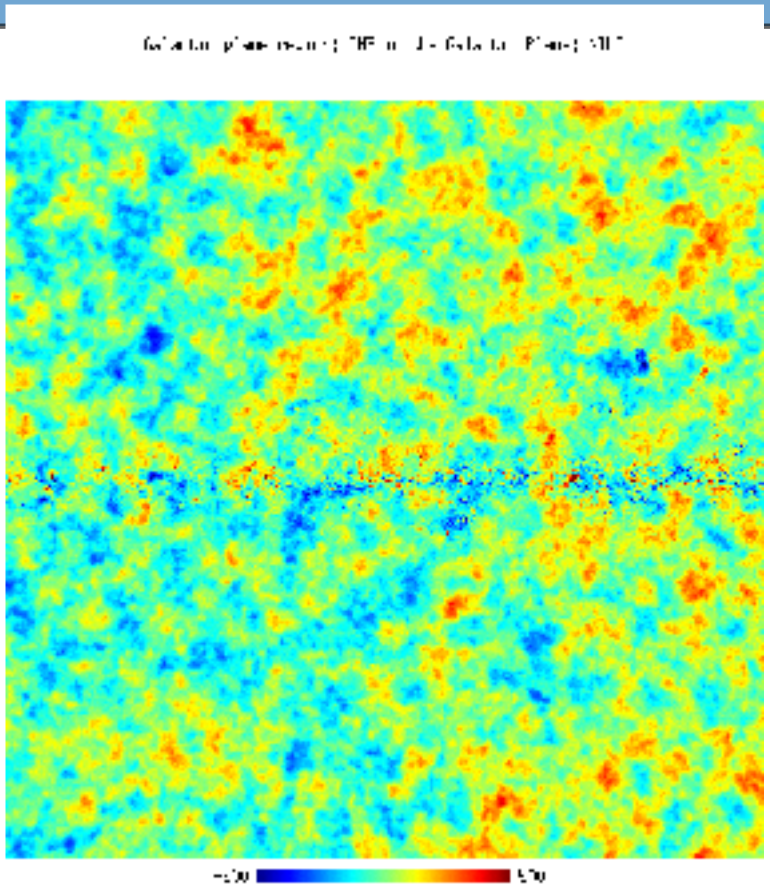


A very clean estimation of the CMB map

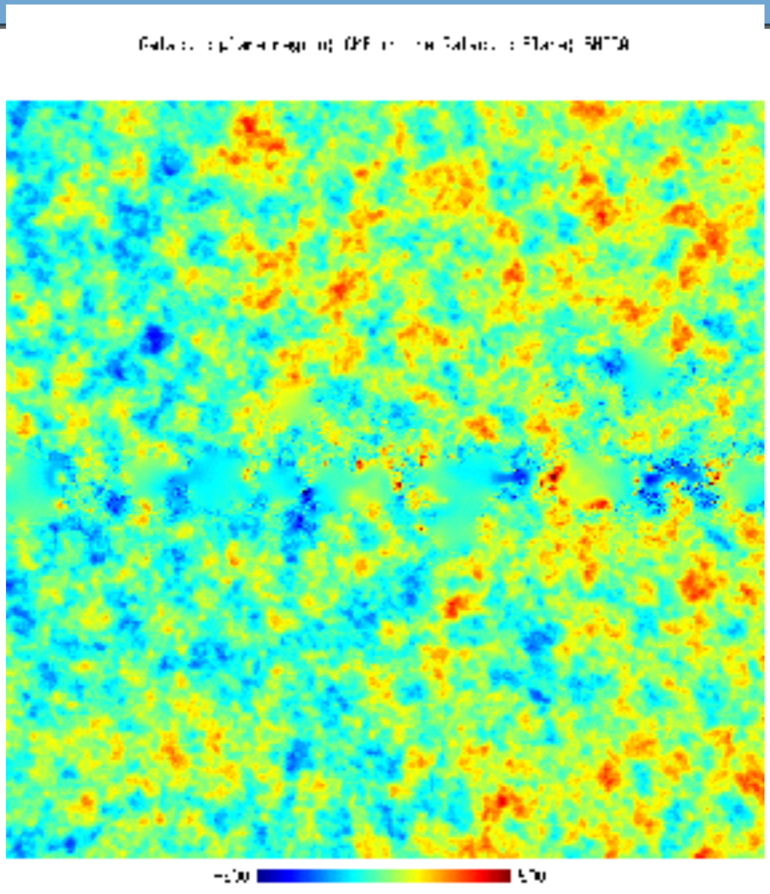
Bobin J., Sureau F., Starck J-L, Rassat A. and Paykari P., Joint Planck and WMAP CMB map reconstruction, A&A, 563, 2014
Bobin J., Sureau F., Starck, CMB reconstruction from the WMAP and Planck PR2 data, A&A, 2016

Applications to the Planck data

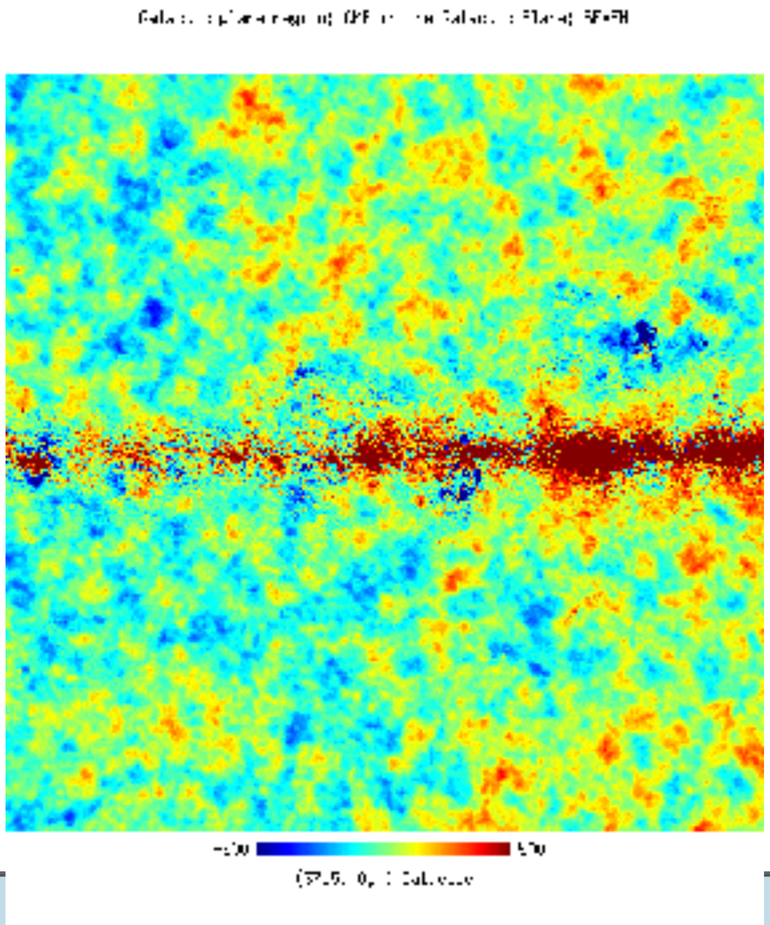
NILC



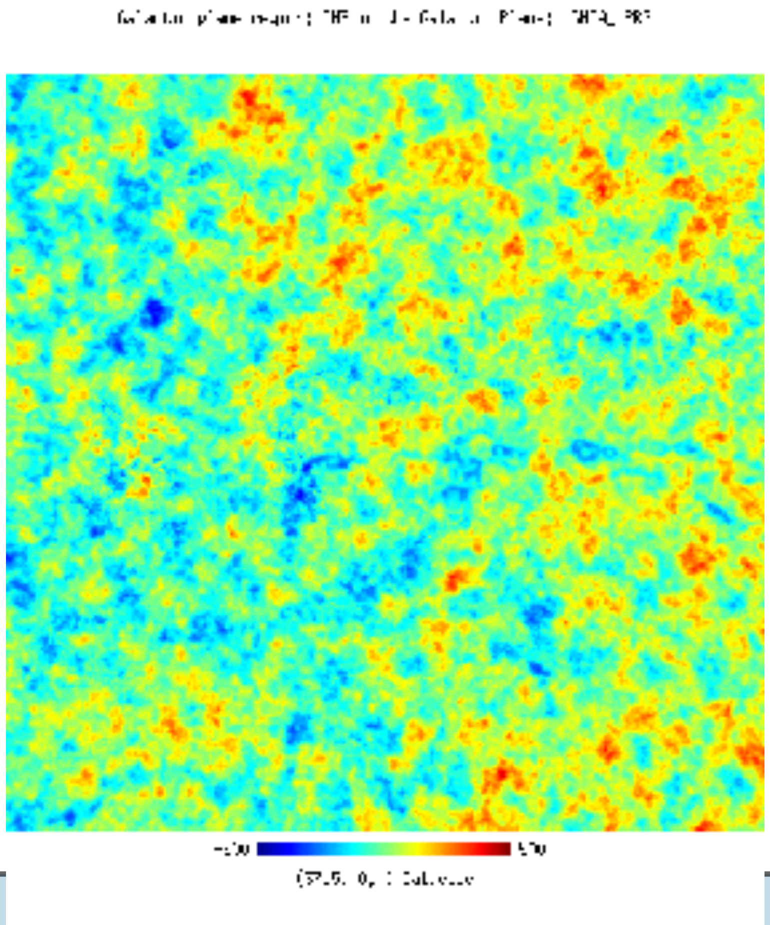
SMICA



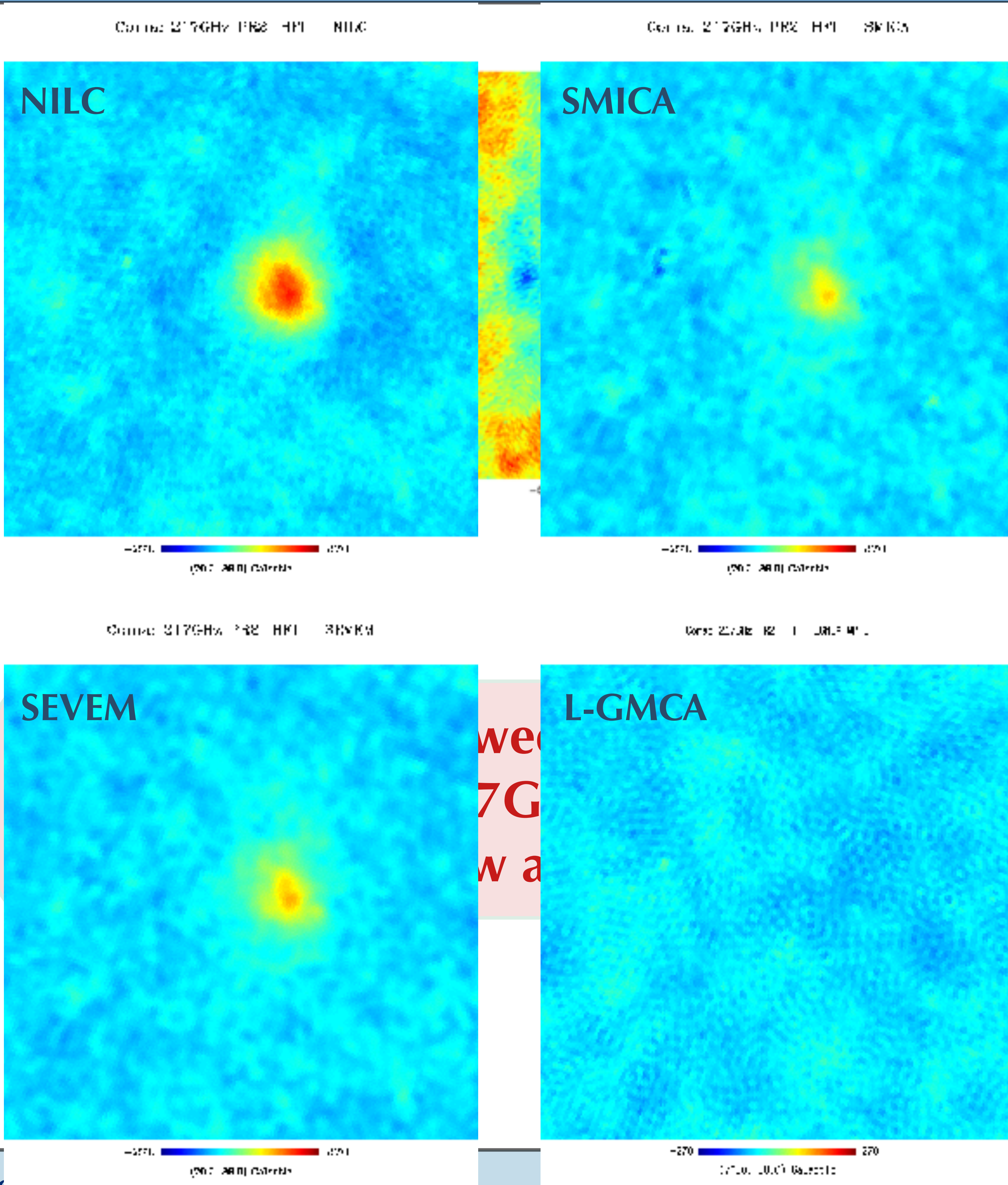
SEVEM



L-GMCA



Applications to the Planck data



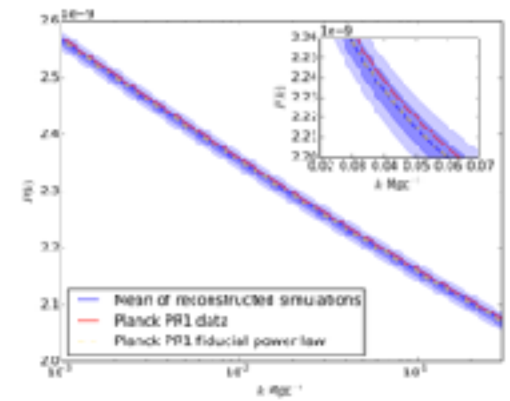
thermal SZ effect vanishes at 217Ghz

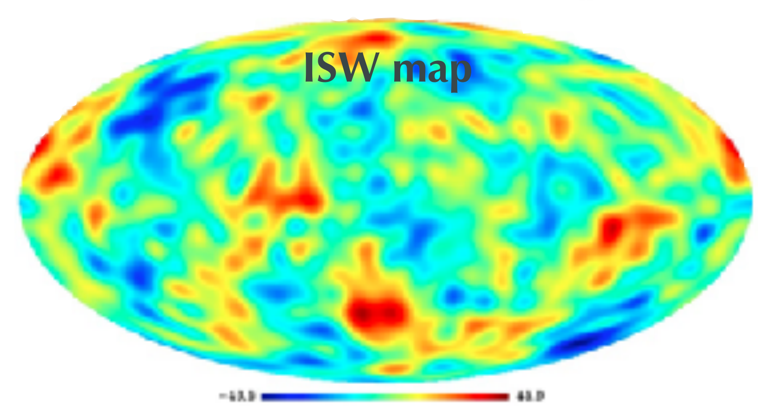
Free of detectable SZ effect

The GMCA CMB map has been used for several cosmological studies

- **Evaluation of primordial power spectrum**

  Lanusse, 2014



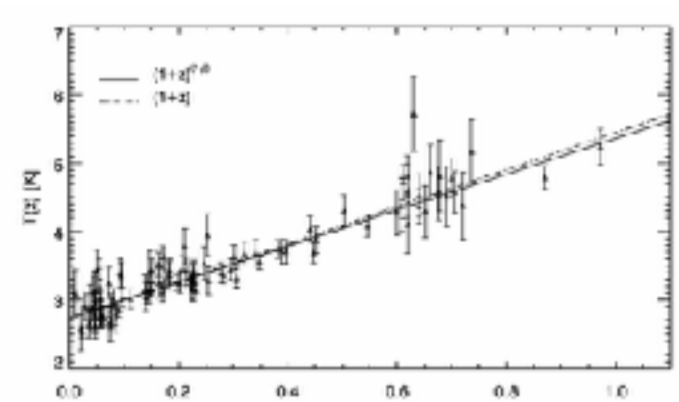


- **Large-scale CMB anomalies**

  Rassat, 2014; Ben-David and Kovetz, 2014
  Aiola, 2014; Notari and Quartin, 2015

- **kSZ studies**

  Luzzi 2014; Hill, 2015



CMB temperature as a function of redshift

A highly flexible framework

- The global linear mixture does not hold true

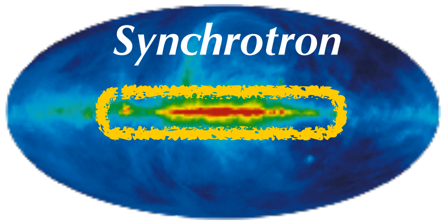
    Local-GMCA: local/multiscale mixture model, handles spectral variabilities

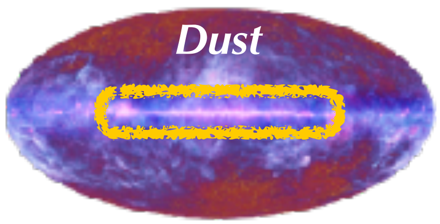
    Bobin J., Sureau F., Starck, CMB reconstruction from the WMAP and Planck PR2 data, A&A, 2016

- Galactic components are partially correlated

    AMCA: robustness w/r to partial correlations

    Bobin J., et al., IEEE Tr. on signal processing, 2015

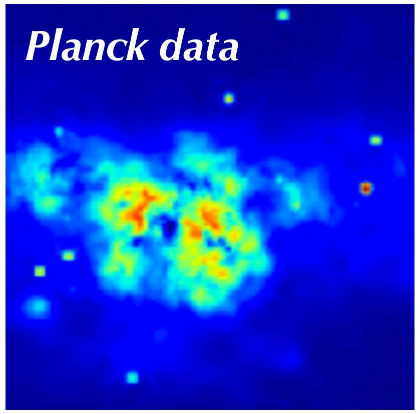




- Many point sources as outliers

    rGMCA: robustness w/r to outliers, based on morphological diversity

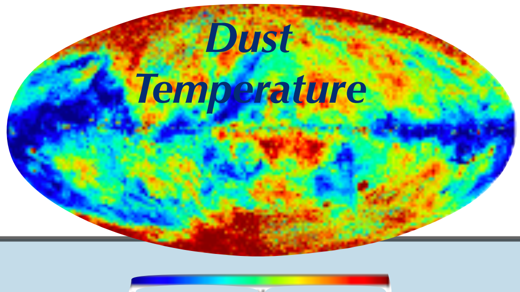
    Chenot, et al., SIAM Imaging Sciences, 2018

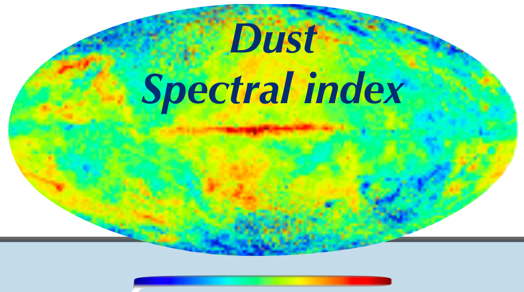


- Accounting for sparse parametric non-linear physical models

    premise: include astrophysical models for a more precise estimation of the galactic sources

    Irfan, et al., MNRAS, 2018





Imaging the dawn of the Universe

LOFAR

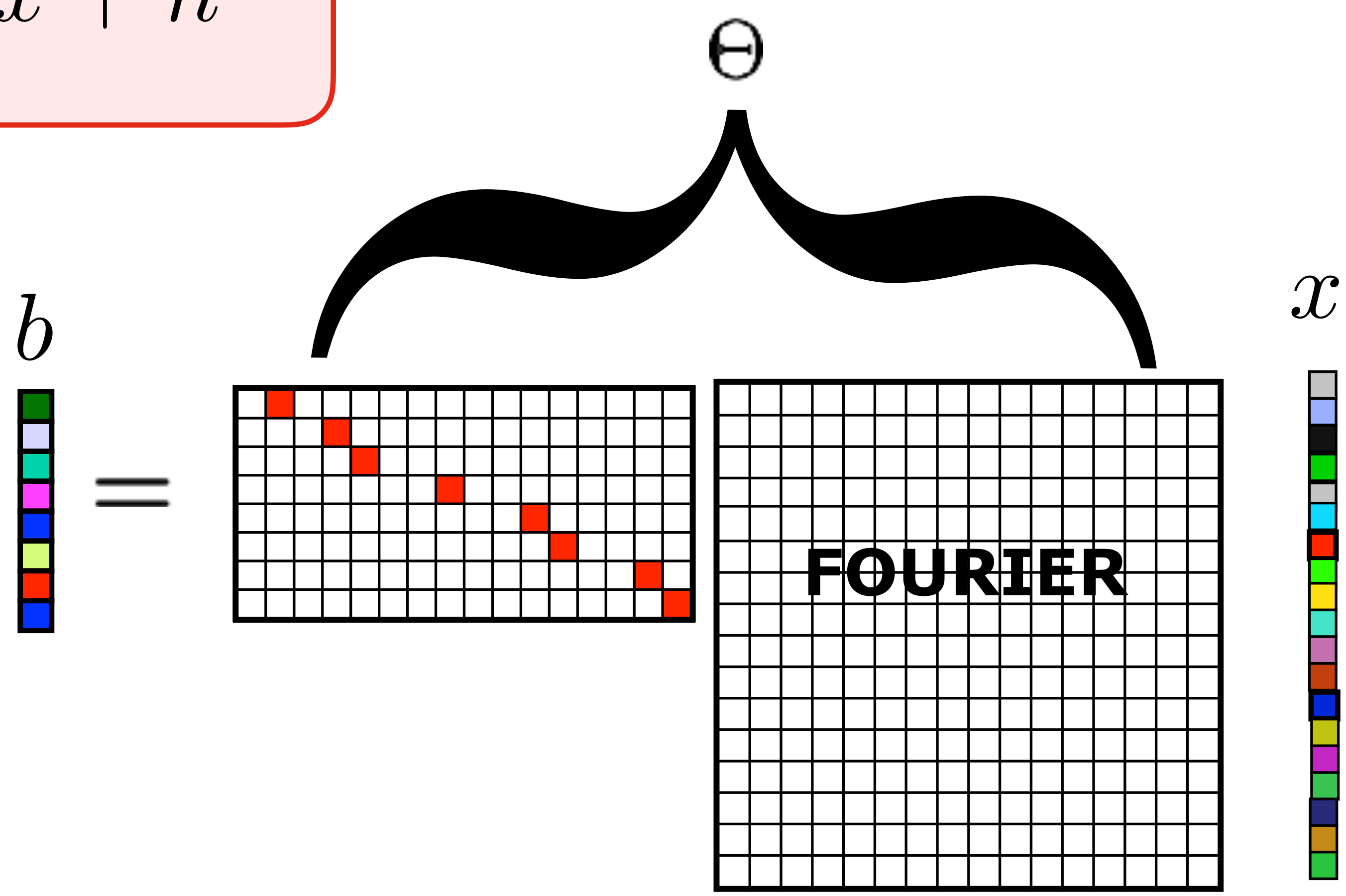
SKA
SQUARE KILOMETRE ARRAY

Radio-interferometric measurements

$$b = \boldsymbol{\Theta} x + n$$

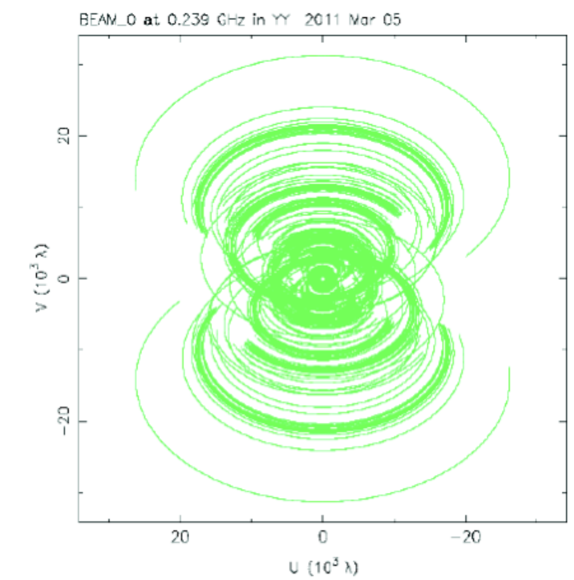
M

observations



N

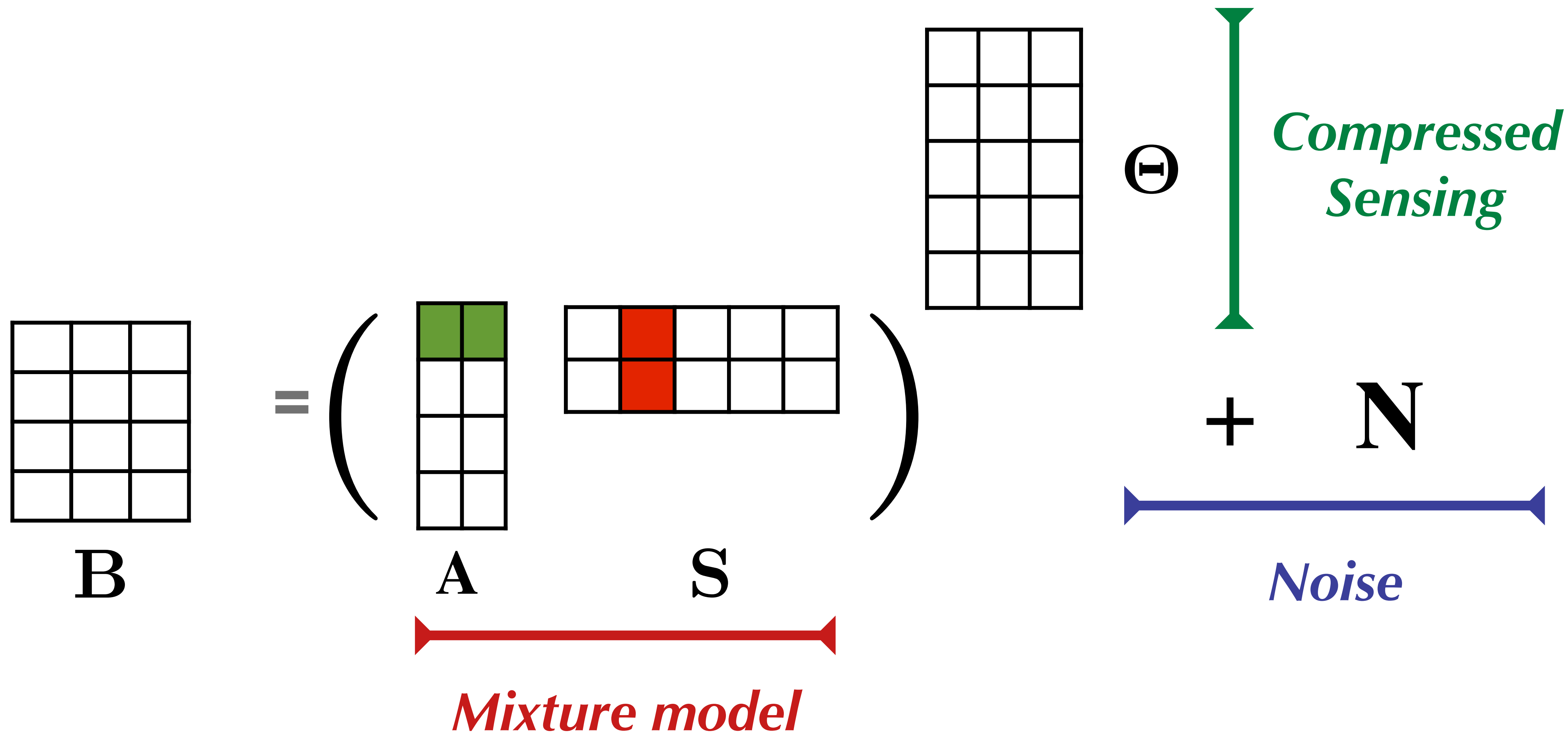
variables



This is a compressed sensing reconstruction problem

Combining CS and BSS

Blind source separation from compressed sensing measurements



$$\forall i; \quad b_i = \left(\sum_j a_{ij} s_j \right) \boldsymbol{\Theta}_i + n_i$$

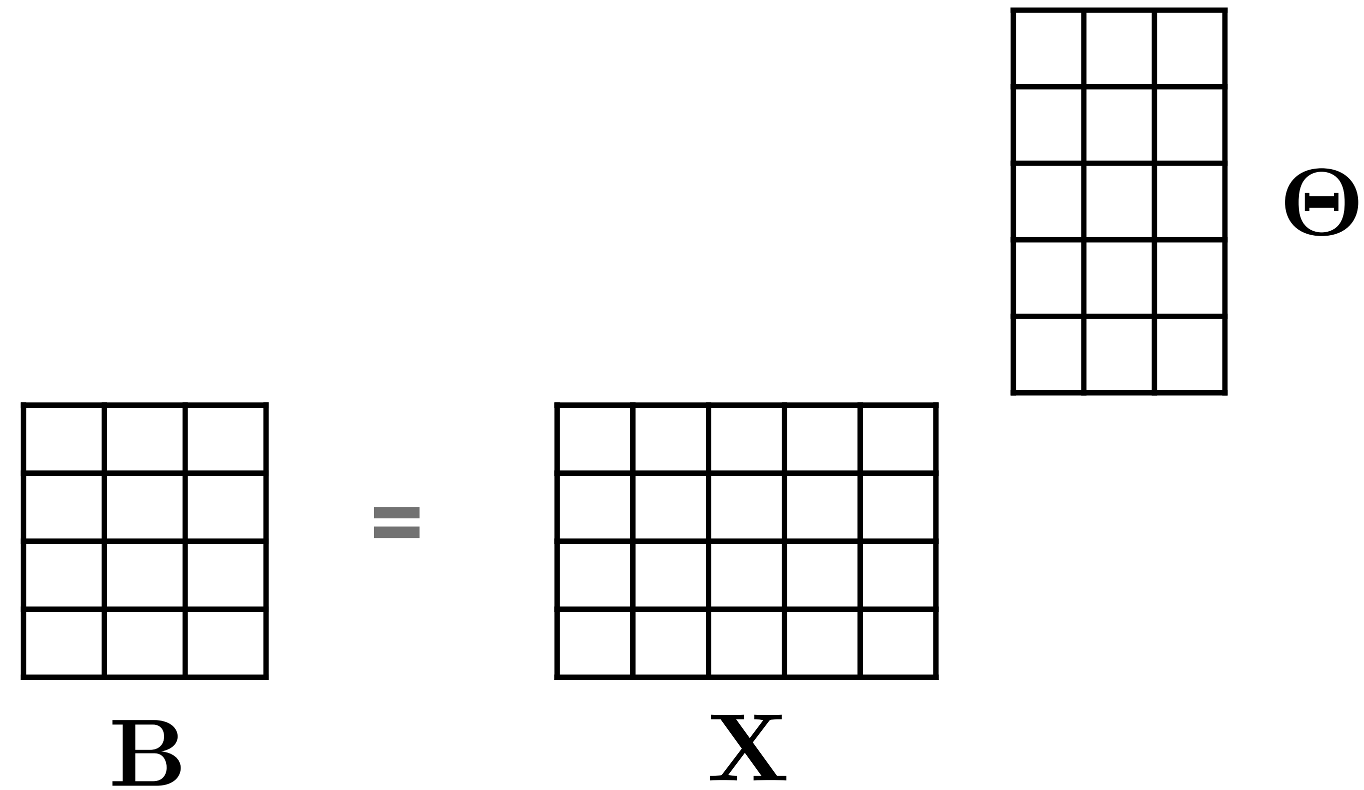
Combining CS and BSS

A naive approach would consist in solving independently each problem:

Multichannel CS

$$\min_{\mathbf{X}} J(\mathbf{X}) + \sum_i \|b_i - \mathbf{X}_i \boldsymbol{\Theta}_i\|_2^2$$

Standard L1 minimization
Matrix completion …

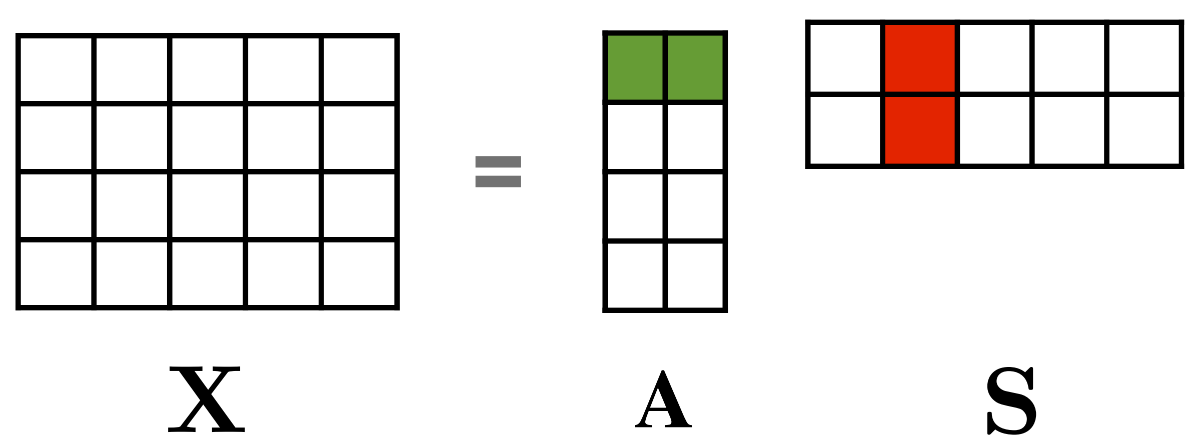


Blind source separation

$$\min_{\mathbf{A},\mathbf{S}} K(\mathbf{S}) + \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2$$

$$K(\mathbf{S}) = \|\boldsymbol{\Lambda} \odot (\mathbf{S}\boldsymbol{\Phi}^T)\|_p$$

Positivity …



The DecGMCA algorithm

The DecGMCA aims at solving the multi-convex problem:

$$\min_{\mathbf{A},\mathbf{S}} \quad \|\mathbf{\Lambda} \odot (\mathbf{S}\mathbf{\Phi}^T)\|_p + \sum_i \left\| b_i - \left(\sum_j a_{ij} s_j \right) \mathbf{\Theta}_i \right\|_2^2$$

Iteratively alternates between:

$$\min_{\mathbf{S}} \quad \|\mathbf{\Lambda} \odot (\mathbf{S}\mathbf{\Phi}^T)\|_p + \sum_i \left\| b_i - \left(\sum_j a_{ij} s_j \right) \mathbf{\Theta}_i \right\|_2^2$$

$$\min_{\mathbf{A}} \sum_i \left\| b_i - \left(\sum_j a_{ij} s_j \right) \mathbf{\Theta}_i \right\|_2^2$$

Ming et al, Joint Multichannel Deconvolution and Blind Source Separation, SIAM Imaging Science, 2017.

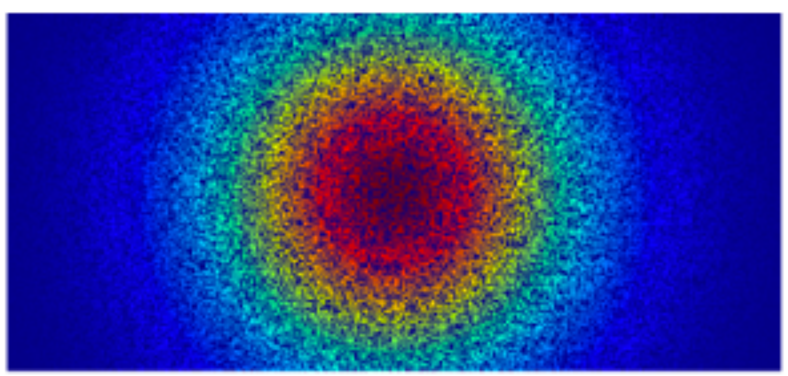
Application to radio-interferometric data

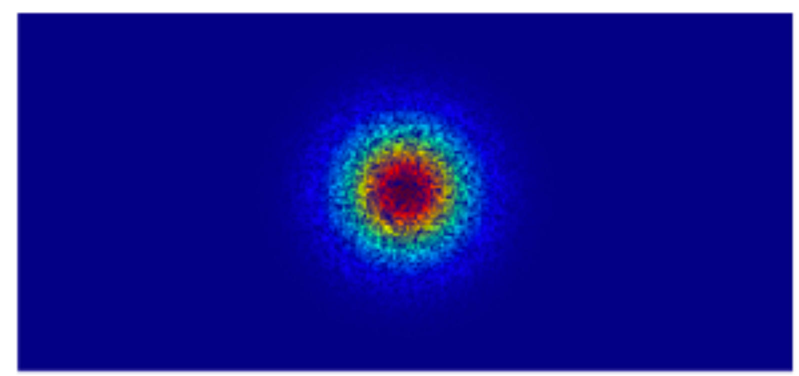
Combines CS and deconvolution:

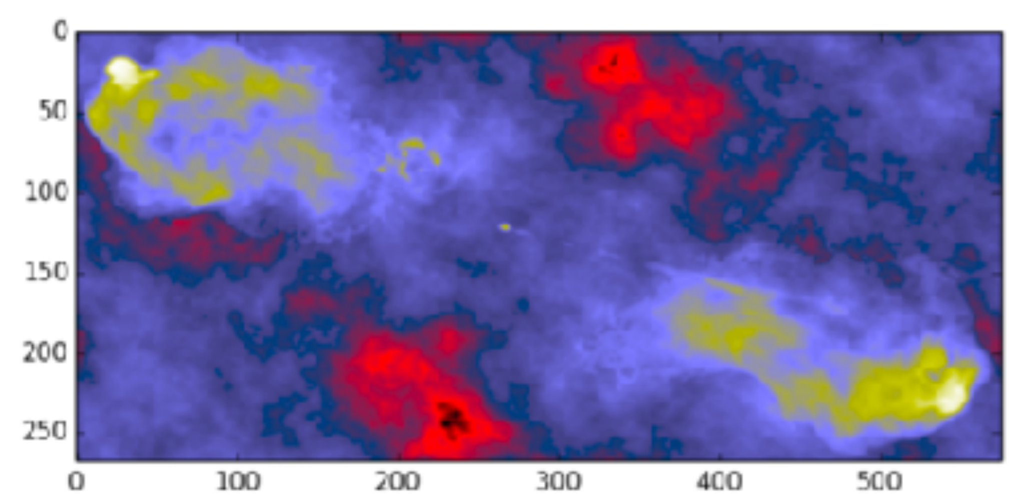
- **incomplete** measurement in the Fourier domain

- Each observation has a **different resolution**

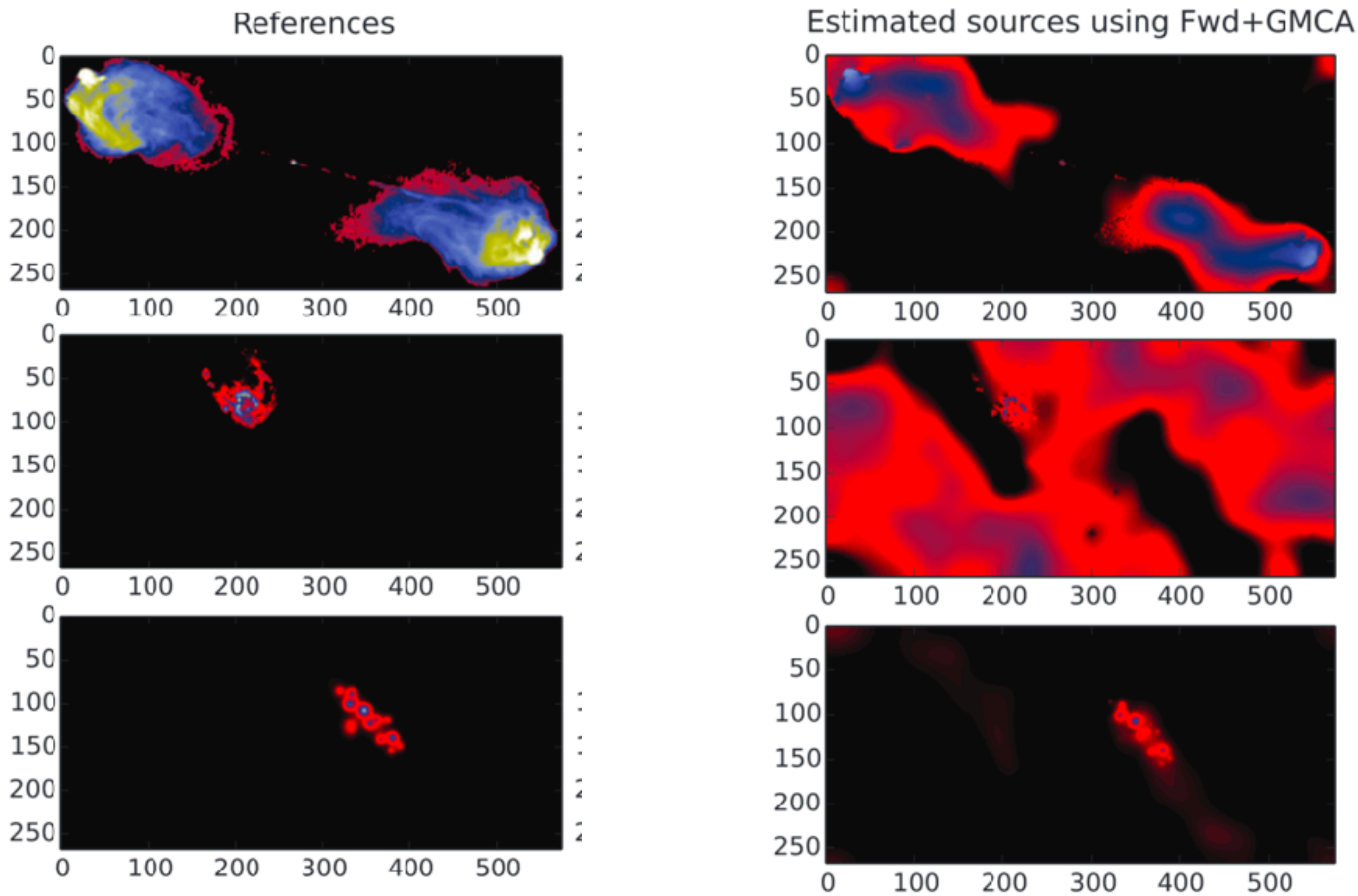
$$\forall i; \quad b_i = \left(\sum_j a_{ij} s_j \right) \mathbf{H}_i \mathbf{\Theta}_i + n_i$$



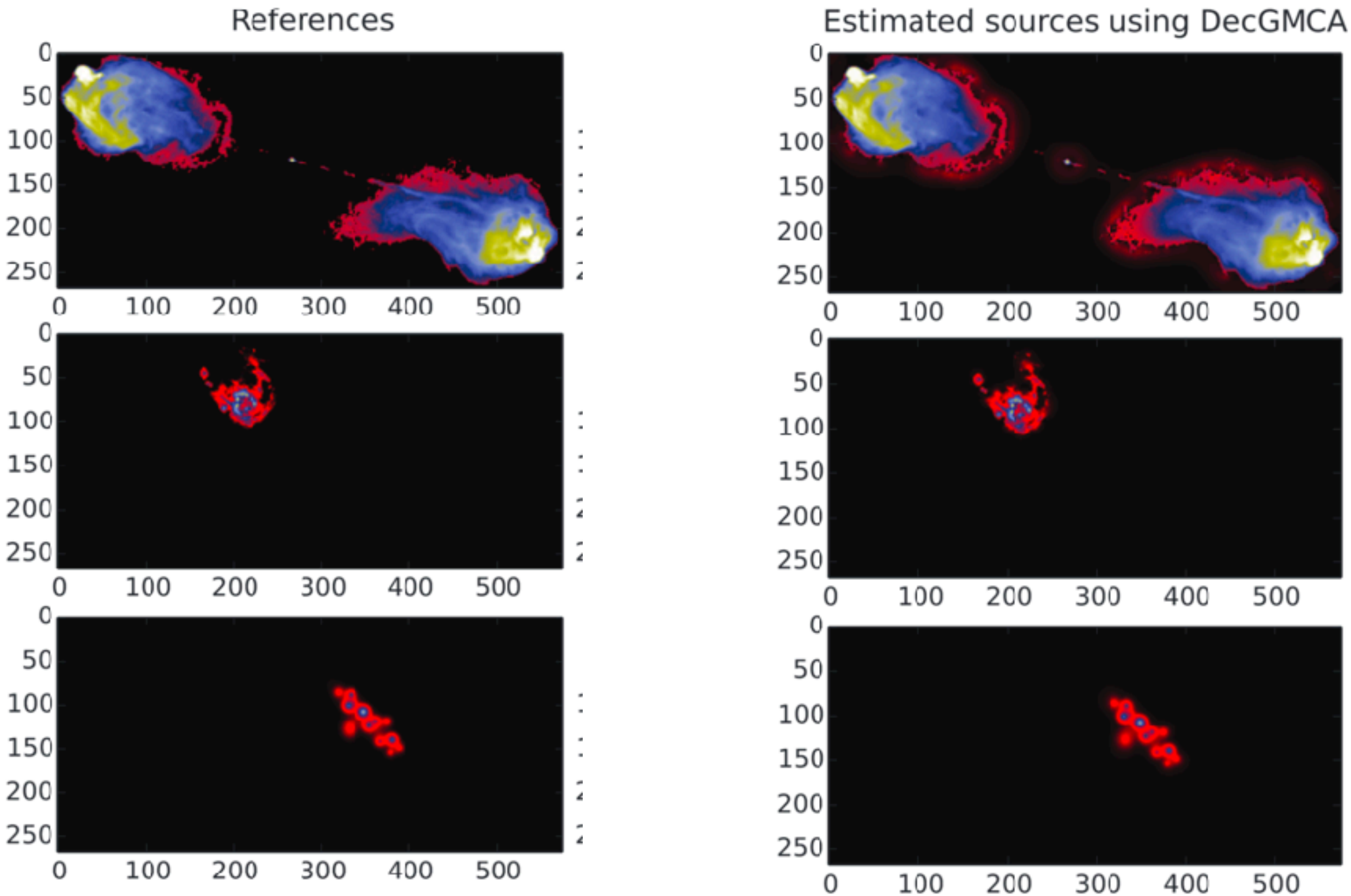




Application to radio-interferometric data



Application to radio-interferometric data



Potential links with GW data analysis

Regularization params., weight matrix, etc.

$$\min_{\mathbf{A},\mathbf{S}} \underbrace{\|\Lambda \odot \mathbf{SW}\|_p}_{\text{Sparse regularization}} + \frac{1}{2} \underbrace{\|\mathbf{X} - \mathbf{AS}\|_F^2}_{\text{Data fidelity term}}$$

- Strong connections with dictionary learning
- Learn elementary waveforms that yield a sparse decomposition
- Preliminary application to GW denoising *Torres-Forné, et al., 2016*

<u>Extensions:</u> robustness w/r glitches, account for missing data, etc.

Potential links with GW data analysis

- Signature unmixing will be challenging for the LISA data processing

$$x = \sum_{p}^{P} \sum_{k}^{K_p} \alpha_{pk} \phi_{pk} + n$$

$$= \underbrace{\sum_{p}^{P} \alpha_p \mathbf{\Phi}_p + n}$$

Sparse combination
waveforms of different categories
(EMRI, MBHB, etc.)

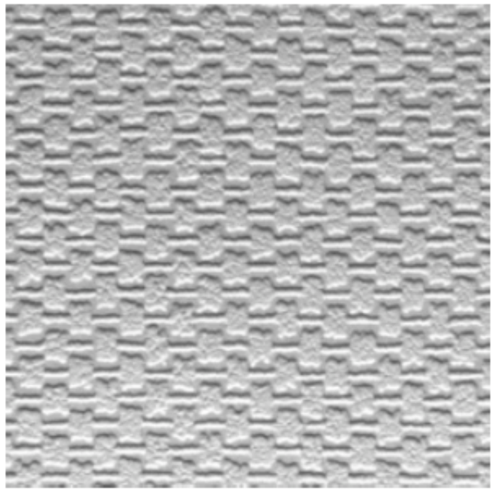
Waveforms from different categories are sparse in different domains

Potential links with GW data analysis

Analogy in image processing: Morphological Component Analysis

$=$

$+$



Starck, et al., 04

Bobin, et al., 07

Curvelets

DCT/L-DCT

$$x = \sum_{i=1}^{K} \mathbf{\Phi}_i \alpha_i \qquad \varphi_1 = \mathbf{\Phi}_1 \alpha_1 \qquad \varphi_2 = \mathbf{\Phi}_2 \alpha_2$$

$$\min_{\alpha_1, \cdots, \alpha_K} \sum_{i=1}^{K} \|\alpha_i\|_{\ell_p} \quad \text{s.t.} \quad x = \sum_{i=1}^{K} \mathbf{\Phi}_i \alpha_i$$

where $\forall i = 1, \cdots, K; \; \varphi_i = \mathbf{\Phi}_i \alpha_i$

Take-away messages

- A highly flexible framework to tackle Sparse MF problems

- Highly reliable algorithms in real-world applications in astrophysics

- Potential connections to tackle GW unmixing problems

→ Exploit sparsity and morphological diversity

- **pyGMCALab:** *python implementation of GMCA and its extensions*

- **As part of ISAP package:** *GMCA (C++) and L-GMCA (IDL)*

Codes are made publicly available at www.cosmostat.org

Thanks !