Sparsity in Astrophysics: from Wavelets to Compressed Sensing

Jean-Luc Starck

CEA, IRFU, Service d'Astrophysique, France

jstarck@cea.fr

http://jstarck.free.fr

Collaborators: S. Beckouche, A. Leonard. D. Donoho, J. Fadili, F. Lanusse, A. Rassat

Sparsity in Astrophysics:
From Wavelets to Compressed Sensing

Sparsity and the Bayesian Controversy Story

What is Sparsity

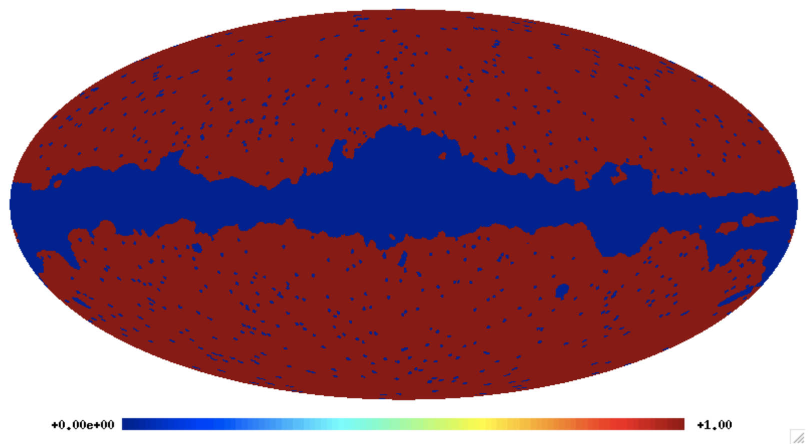
Compressed Sensing

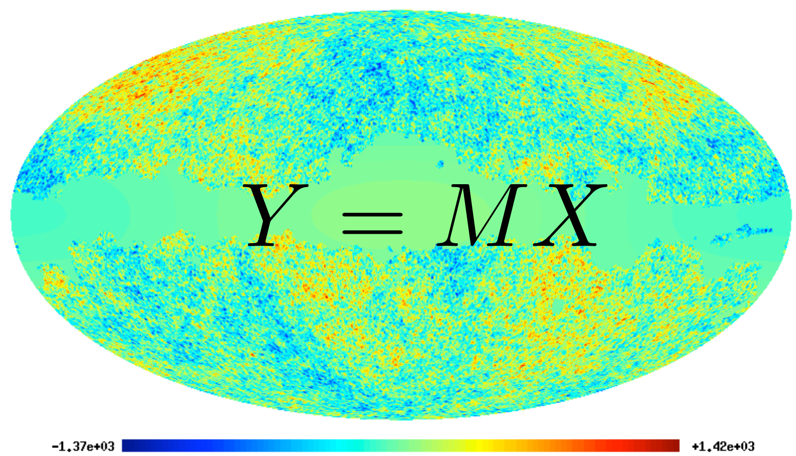
Inverse Problem Tour and Sparse Revovery

Sparsity and 3D Weak Lensing

Interpolation of Missing Data: Sparse Inpainting

Where M is the mask: M(i,j) = 0 ==> missing data
M(i,j) = 1 ==> good data





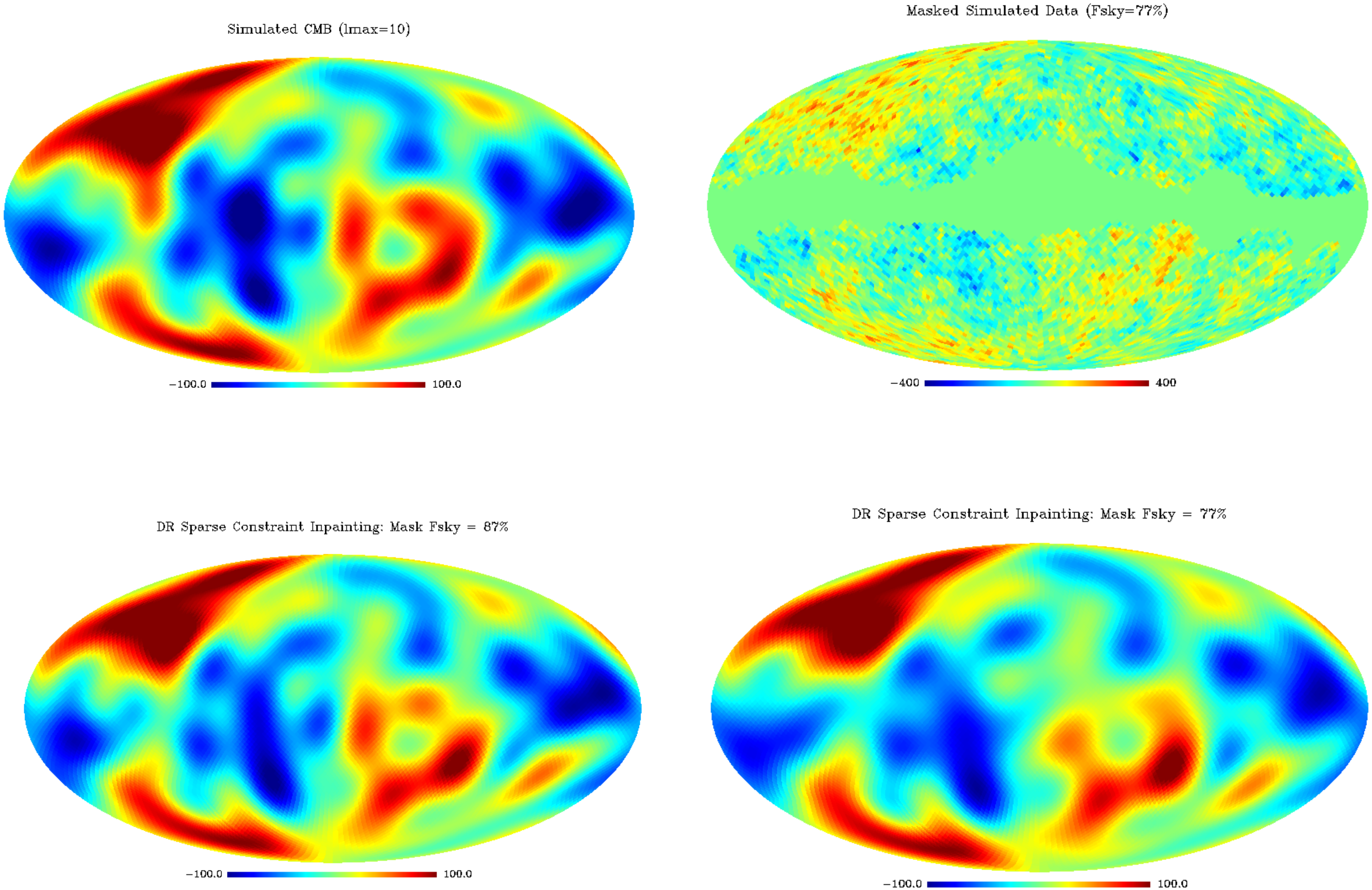
$$\min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \quad Y = M\Phi\alpha$$

$$X = \Phi\alpha \qquad \Phi = \text{Spherical Harmonics}$$

$$\|\alpha\|_1 = \sum_k | \alpha_k |$$

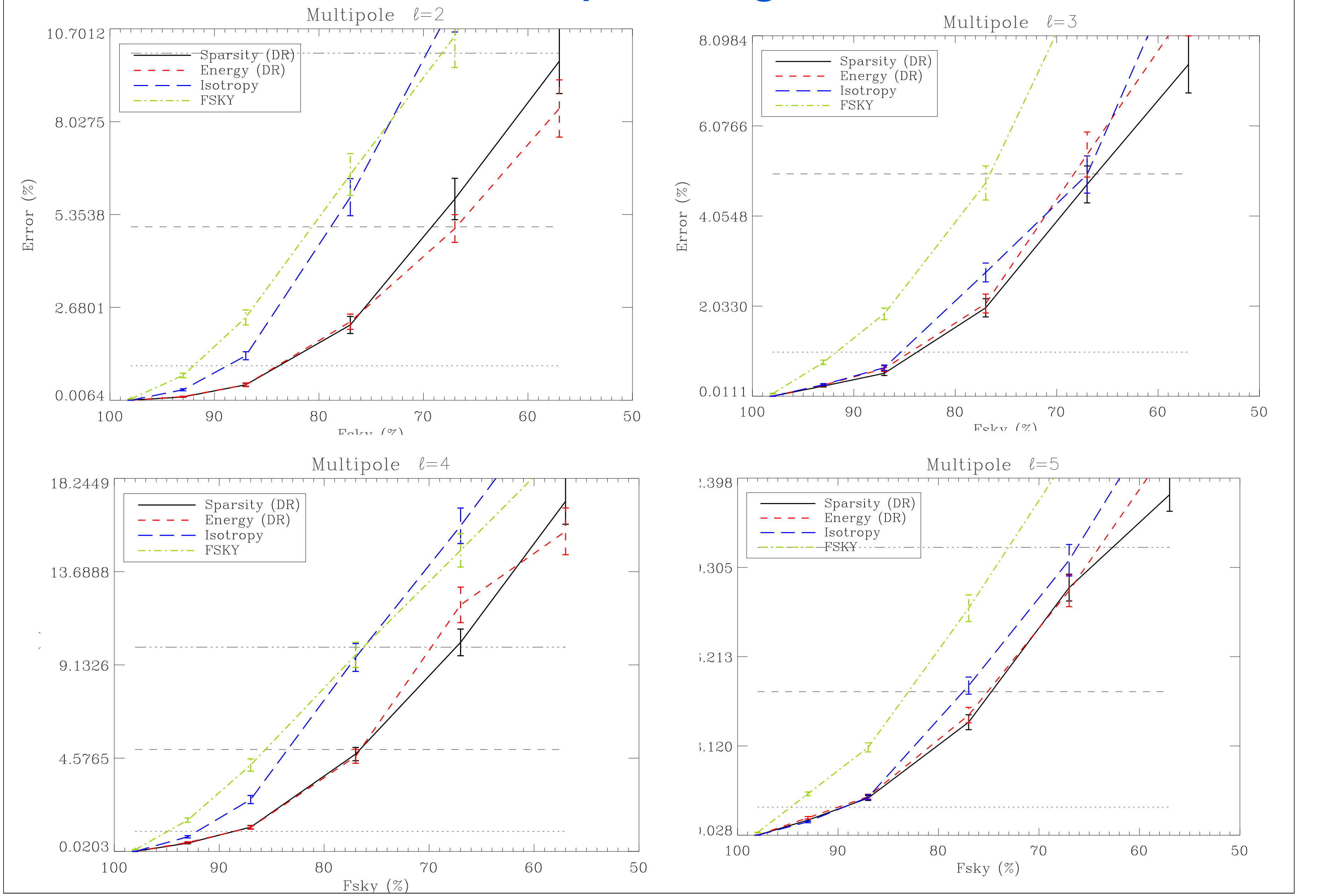
J.-L. Starck, A. Rassat, and M.J. Fadili, "Low-l CMB Analysis and Inpainting", **Astronomy and Astrophysics** , in press.

Large CMB Scale Analysis



J.-L. Starck, A. Rassat, and M.J. Fadili, "Low-l CMB Analysis and Inpainting", **Astronomy and Astrophysics** , in press.

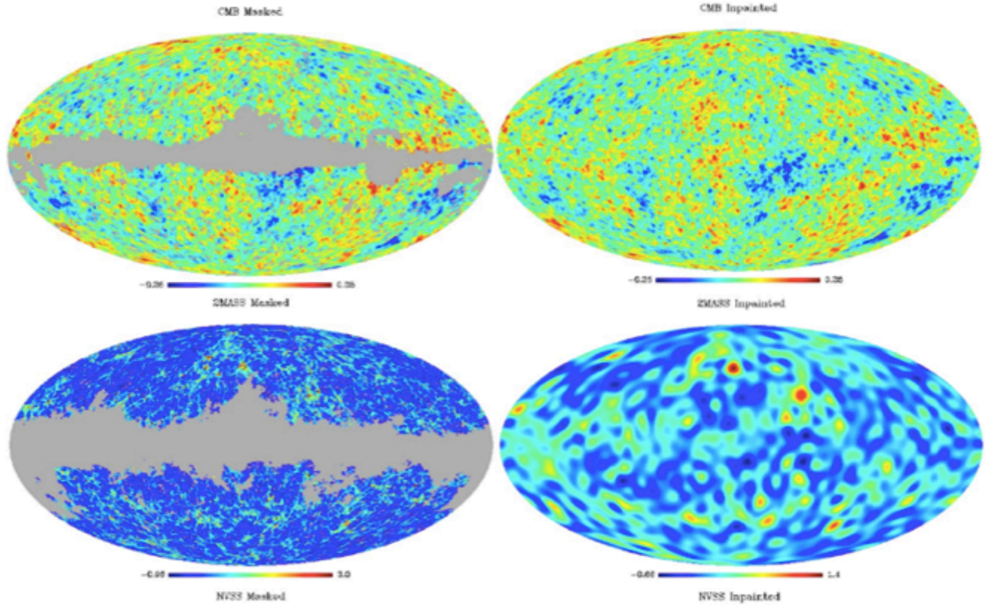
Inpainting



Multipole ℓ=2
Sparsity (DR)
Energy (DR)
Isotropy
FSKY
10.7012
8.0275
5.3538
2.6801
0.0064
Error (%)
100
90
80
70
60
50
Multipole ℓ=3
Sparsity (DR)
Energy (DR)
Isotropy
FSKY
8.0984
6.0766
4.0548
2.0330
0.0111
Error (%)
100
90
80
70
60
50
Multipole ℓ=4
Sparsity (DR)
Energy (DR)
Isotropy
FSKY
18.2449
13.6888
9.1326
4.5765
0.0203
100
90
80
70
60
50
Fsky (%)
Multipole ℓ=5
Sparsity (DR)
Energy (DR)
Isotropy
FSKY
100
90
80
70
60
50
Fsky (%)

Inpainting & CMB ANOMALIES

A. Rassat



=> Low power no longer significant after subtraction of ISW signal

=> Subtracting the ISW effect removes CMB quad/oct anomaly

A. Rassat, J-L. Starck, and F.X. Dupe, "Removal of two large scale Cosmic Microwave Background anomalies after subtraction of the Integrated Sachs Wolfe effect", **Astronomy and Astrophysics** , submitted.

Bayesian Perspective

$$Y = MX = M\Phi\alpha \text{ with } \|\alpha\|_1 \quad \text{minimum}$$

Prior on the solution: $P(\alpha) = e^{-\lambda\|\alpha\|_1}$

Gaussian noise prior: $P(Y/\alpha) = e^{-\|Y-A\Phi\alpha\|_2^2}$

Bayes: $P(\alpha|Y) = P(Y|\alpha)P(\alpha)$

↓ Maximum a Posteriori (MAP)

$$\min_{\alpha} -log\,(P(\alpha|Y)) = \| Y - A\Phi\alpha \|_2^2 + \lambda \| \alpha \|_1,$$

Bayesian Perspective

Prior: $$P(\alpha) = e^{-\lambda \|\alpha\|_1}$$

Severe Critics from Bayesian Cosmologists against CMB Sparse Inpainting

1- Sparsity consists in assuming an anisotropy and a non Gaussian prior, which **does not make sense** for the CMB, which is Gaussian and isotropic.

2- Sparsity **violates** the rotational invariance: The critic here is that a linear combinations of independent exponentials are not independent exponentials.

3- The l_1 norm that is used for sparse inpainting **arose purely out of expediency** because under certain circumstances it reproduces the results of the l_0 norm, (which arises naturally in the context of strict as opposed to weak sparsity) without necessitating combinatorial optimization.

4- There is **no mathematical proof** that sparse regularization preserves/recovers the original statistics.

What is Sparsity?

A signal *s* (*n* samples) can be represented as sum of weighted elements of a given dictionary

Dictionary (basis, frame)

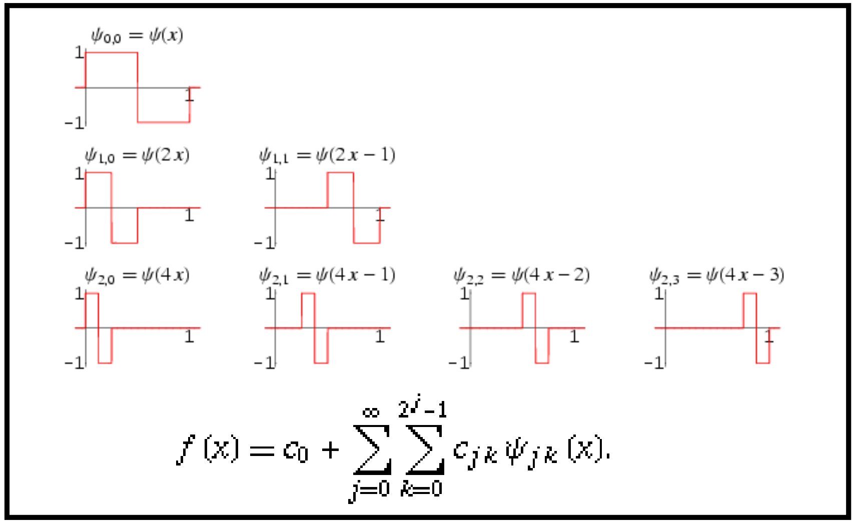
$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

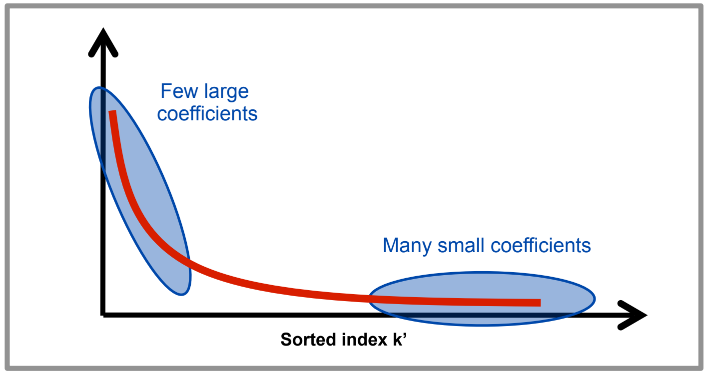
Atoms

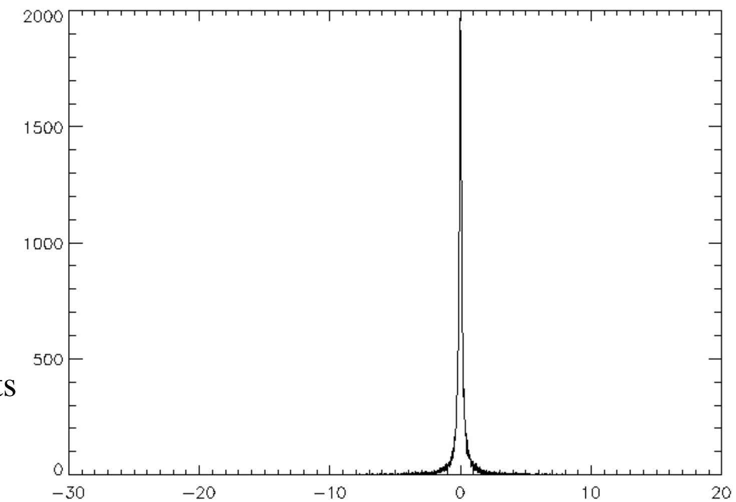
$$s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi\alpha$$

coefficients

Ex: Haar wavelet

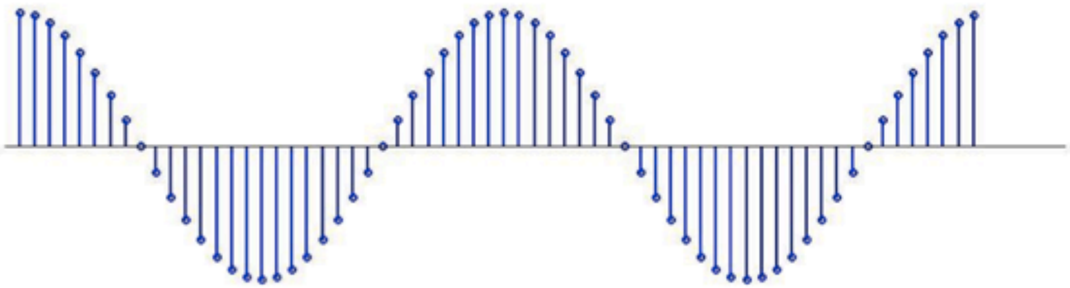






- Fast calculation of the coefficients
- Analyze the signal through the statistical properties of the coefficients
- Approximation theory uses the sparsity of the coefficients

Strict Sparsity: k-sparse signals

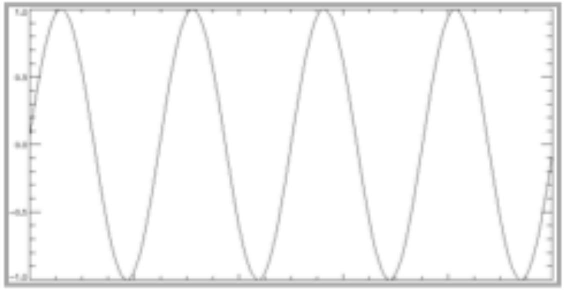


A sine wave in real space…

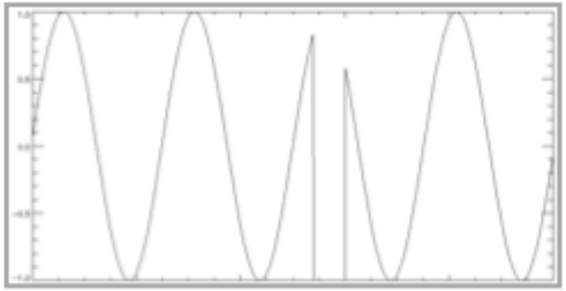
…can be a Dirac in Fourier space.

Sinusoids are sparse in the Fourier domain.

Minimizing the l_0 norm

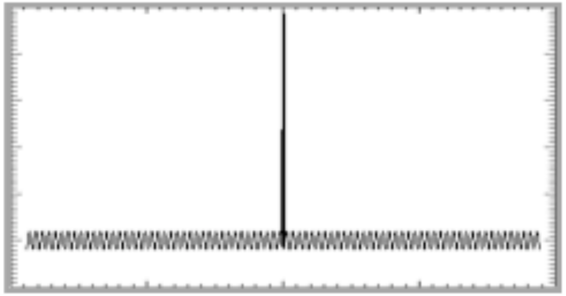


Sine curve



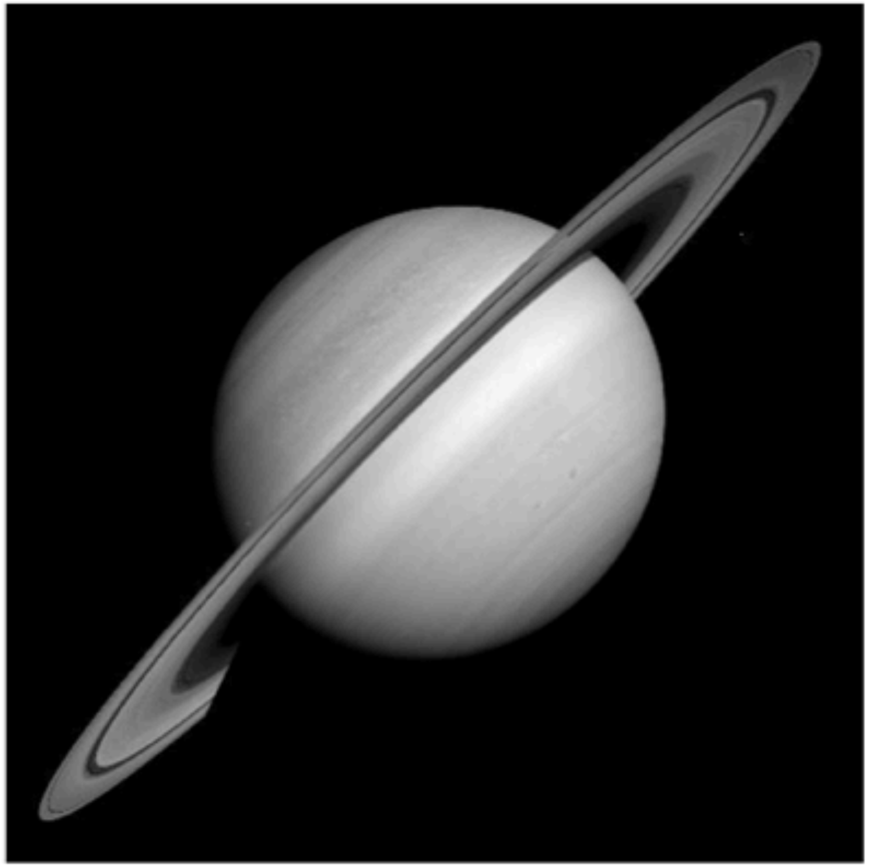
Truncated sine curve

TF of a sine curve



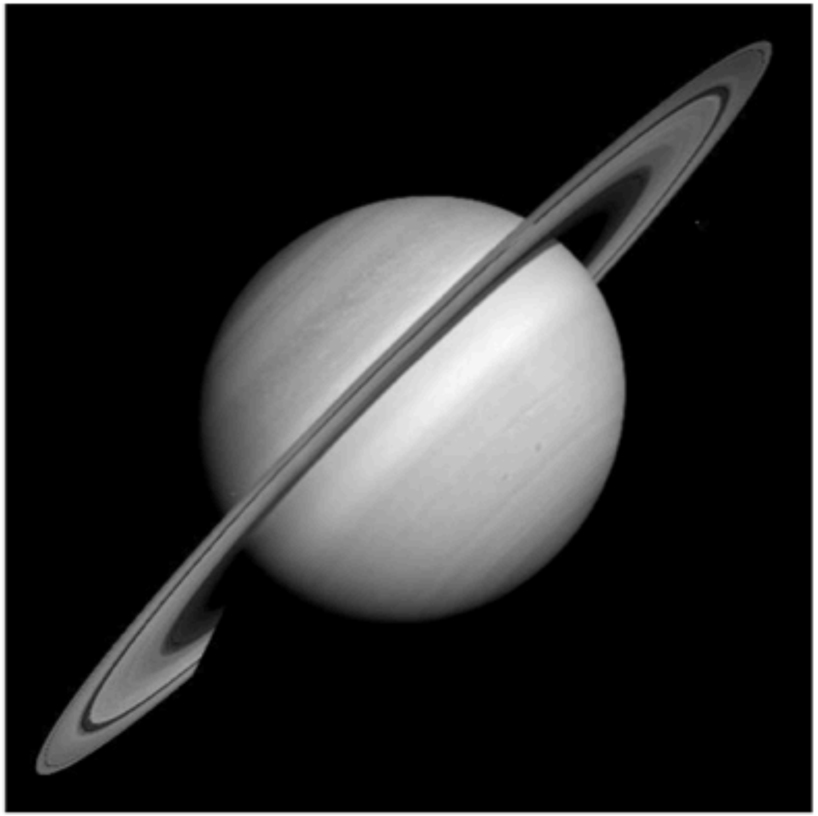
TF of a truncated sine curve

$$\text{with } 0^0 = 0, \qquad \| \alpha \|_0 = \sum_k \alpha_k^0 = \# \{ \alpha_k \neq 0 \}$$



The top 1% of the coefficients concentrate only 8.66% of the energy. Not sparse…

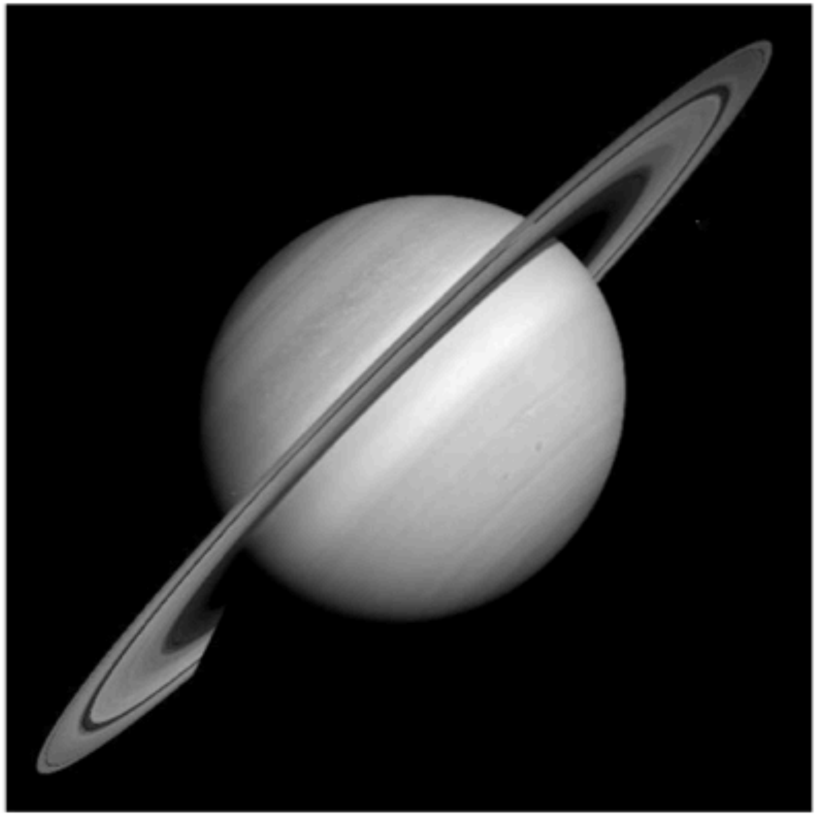
1% largest coefficients in real space
(the others are set to 0)

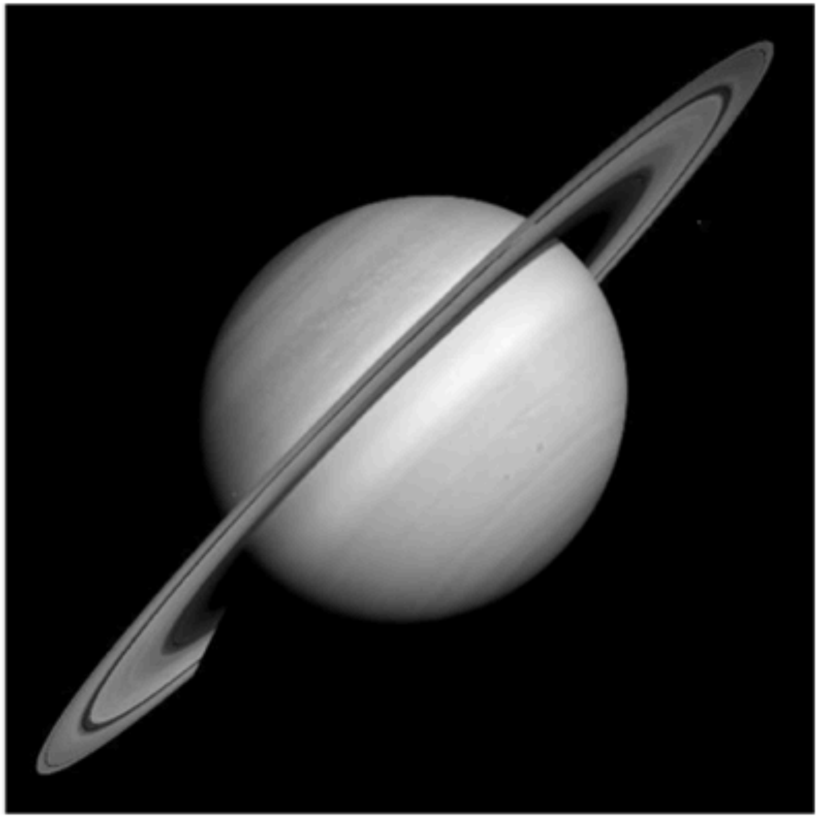


1% largest coefficients in wavelet space
(the others are set to 0)

The wavelet coefficients encode edges and large scale information.

Wavelet transform





1% of the wavelet coefficients concentrate 99.96% of the energy: This can be used as a *prior.*

Reconstruction, after throwing away 99% of the wavelet coefficients

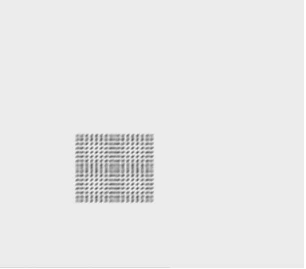
Sparsity Model 1: we consider a dictionary $\Phi = \{\phi_1, \ldots, \phi_K\}$ which has a fast transform/reconstruction operator:

$$s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi\alpha$$

<u>Local DCT</u>

Stationary textures

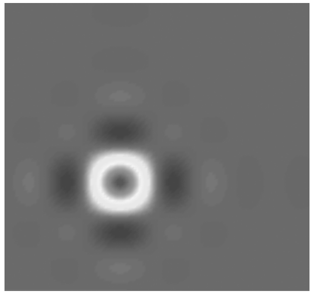
Locally oscillatory



<u>Wavelet transform</u>

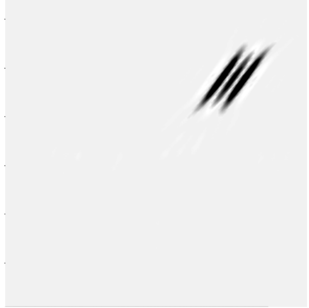
Piecewise smooth

Isotropic structures



<u>Curvelet transform</u>

Piecewise smooth,
edge



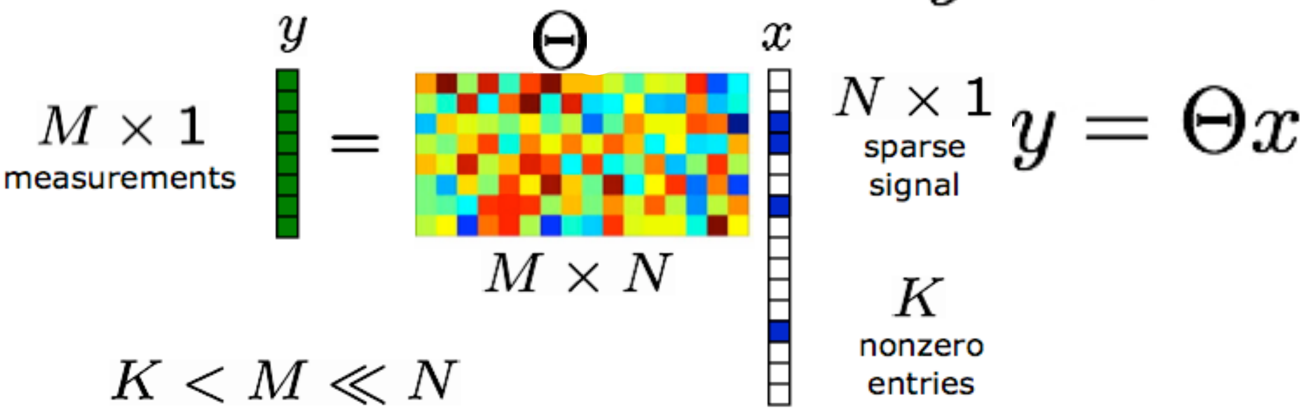
Compressed Sensing

* E. Candès and T. Tao, "Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? ", IEEE Trans. on Information Theory, 52, pp 5406-5425, 2006.
* D. Donoho, "Compressed Sensing", IEEE Trans. on Information Theory, 52(4), pp. 1289-1306, April 2006.
* E. Candès, J. Romberg and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information", IEEE Trans. on Information Theory, 52(2) pp. 489 - 509, Feb. 2006.

A non linear sampling theorem

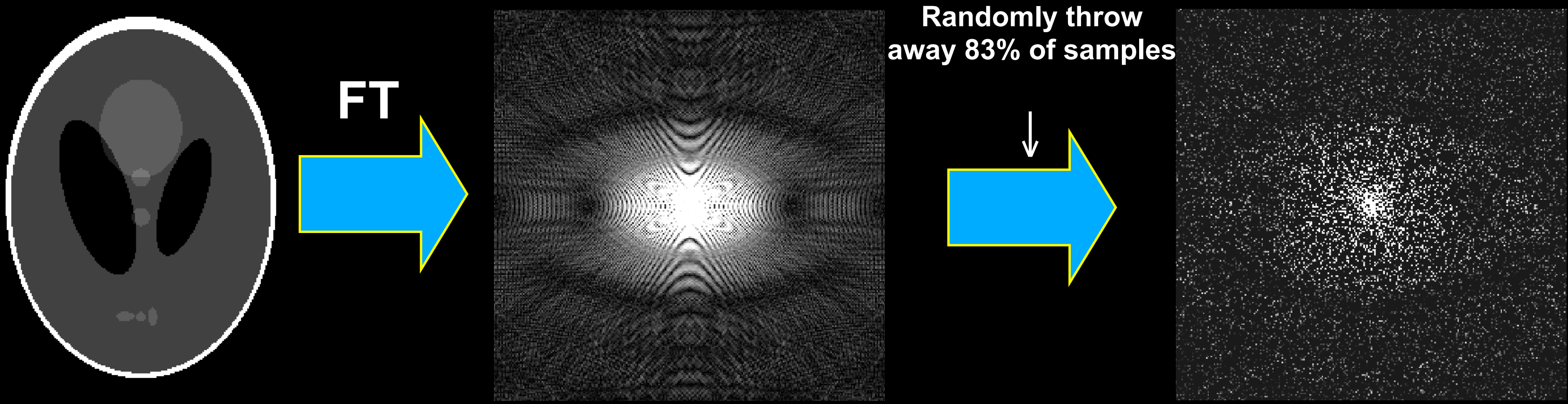
"Signals with exactly K components different from zero can be recovered perfectly from ~ K log N incoherent measurements"

Replace samples with *few linear projections* $y = \Theta x$



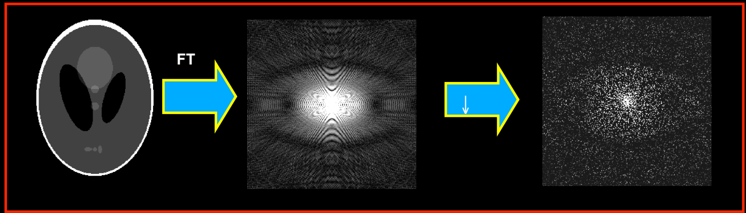
Reconstruction via non linear processing: $$\min_x \|x\|_1 \quad \text{s.t.} \quad y = \Theta x$$

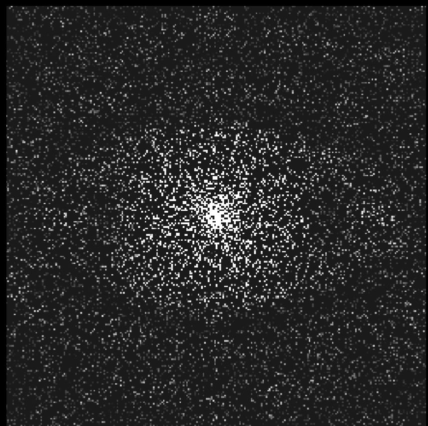
A Surprising Experiment*



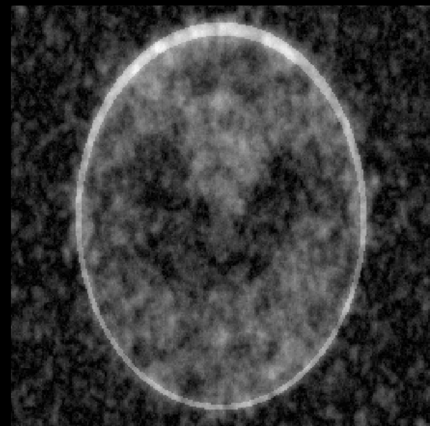
* E.J. Candes, J. Romberg and T. Tao.

A Surprising Result*



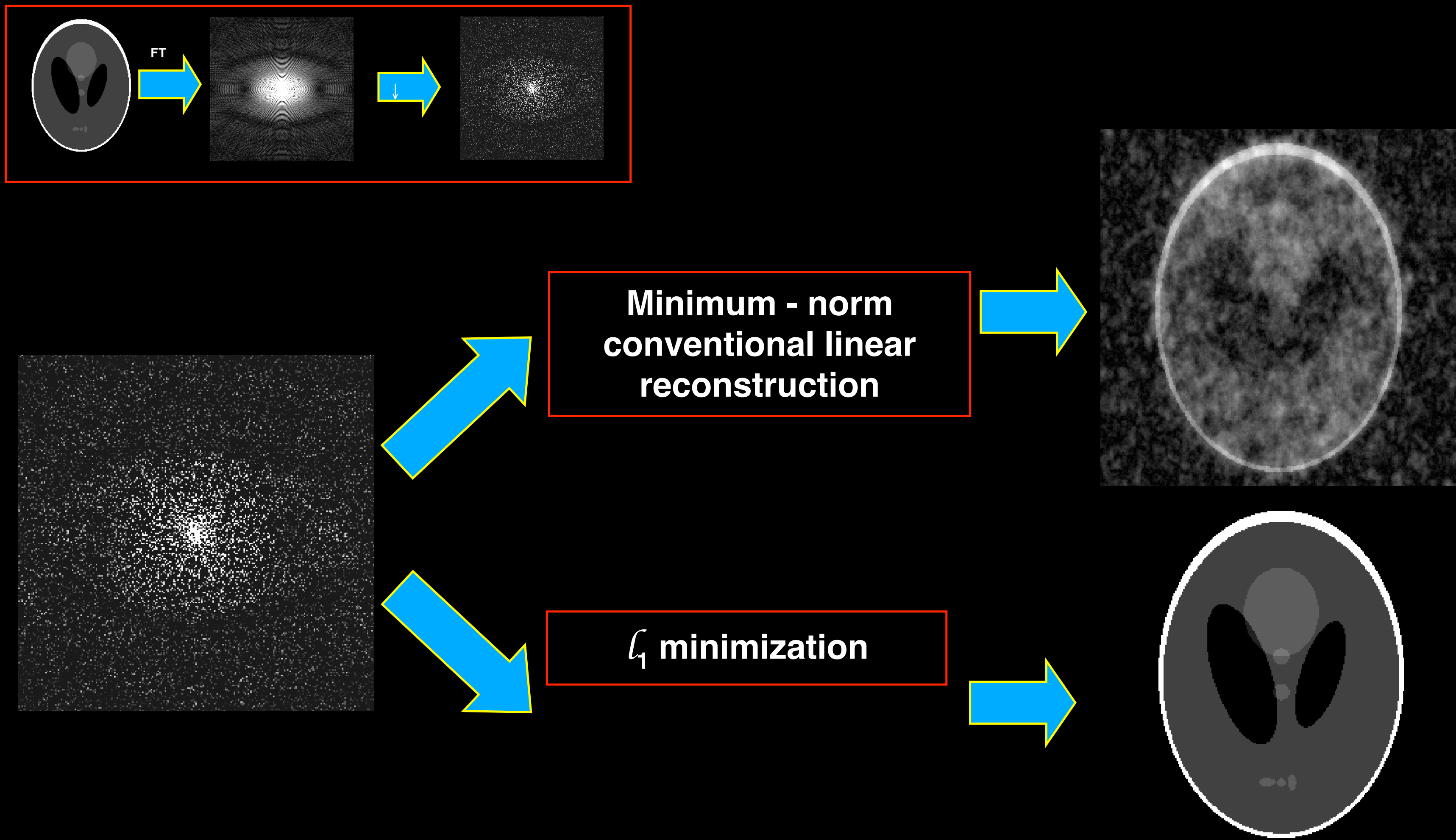


Minimum - norm conventional linear reconstruction



* E.J. Candes, J. Romberg and T. Tao.

A Surprising Result



E.J. Candes

Compressed sensing and the Bayesian interpretation failure

The first critic is that the l1 regularization is equivalent to assume that the solution is Laplacian and not Gaussian, which does not make sense in case of CMB analysis.

==> ~~The MAP solution verifies the distribution of the prior.~~

(Nikolova, 2007; Gribonval, 2011, Gribonval, 2012, Unser, 2012)

The beautiful Compressed Sensing counter-example

but x does NOT follow a Laplacian distribution

What Bayesian Perspective Cannot See !!!

For most Bayesian cosmologists, if a prior derives an algorithm, therefore to use this algorithm, we must have the coefficients distributed according to this prior.

But this is simply a false logical chain.

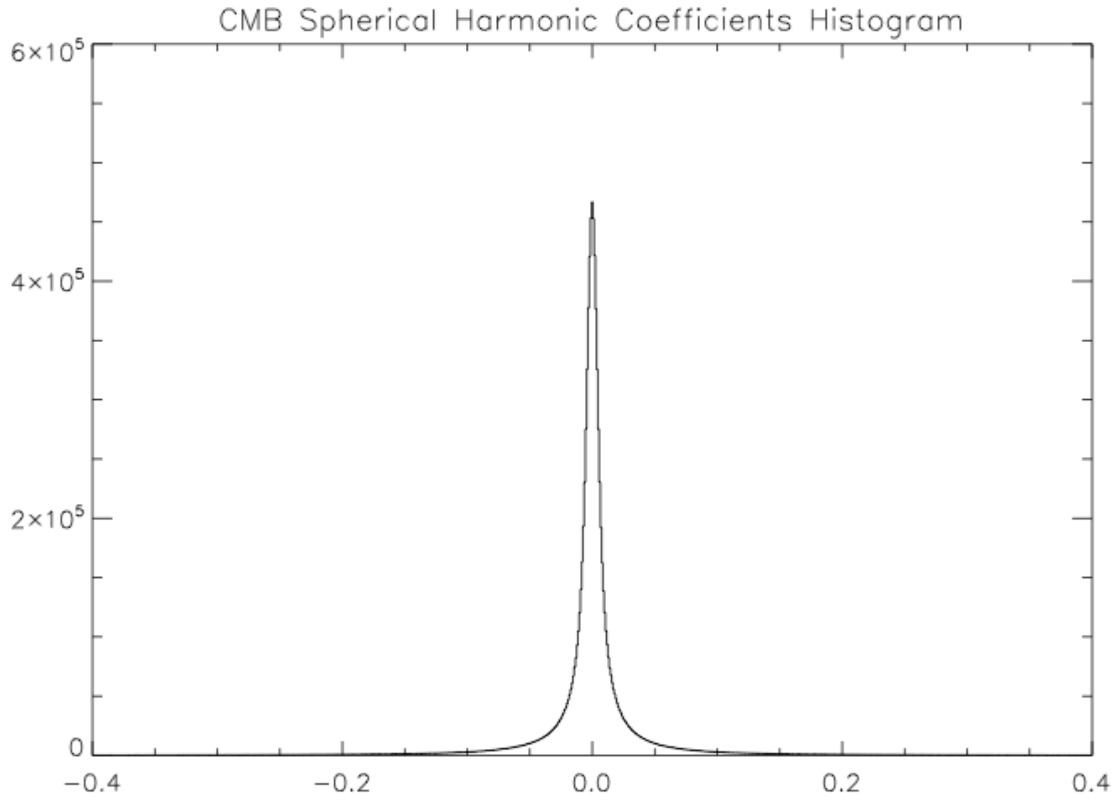
What compressed sensing shows is that:

we can have prior A be completely true, but impossible to use for computation time or any other reason, and can use prior B instead, and get the correct results!

But what is exactly the prior in the sparse analysis ?

Bayesian: each (spherical harmonic) coefficient is a realization of a stochastic process.

Sparsity: we see the data as a function, and the coefficients follows a given distribution. Even if each spherical harmonic coefficient is a realization of Gaussian variable, the distribution of all coefficients is not necessary Gaussian.



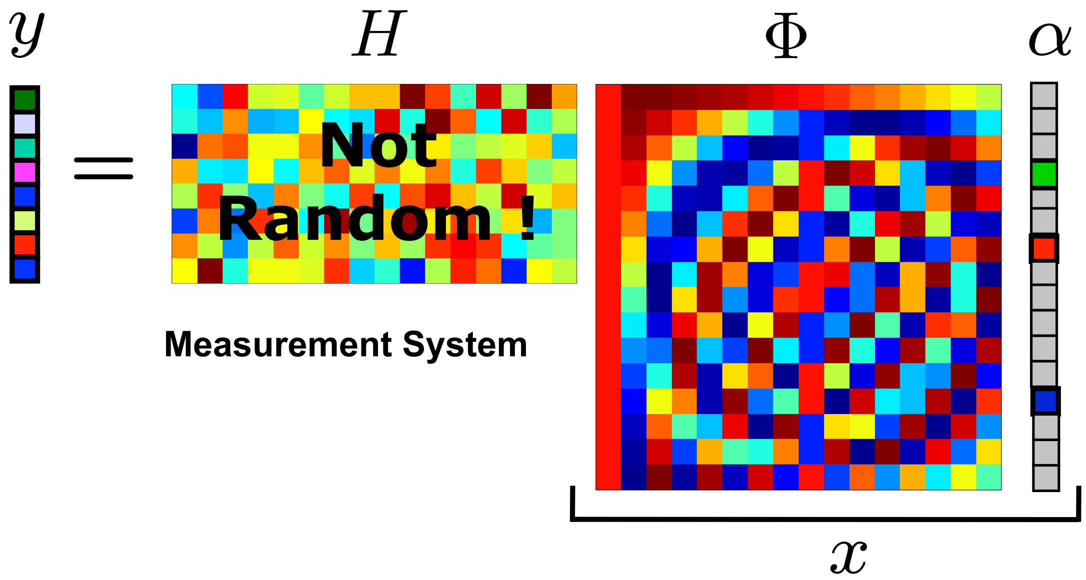
INVERSE PROBLEM TOUR and SPARSE RECOVERY

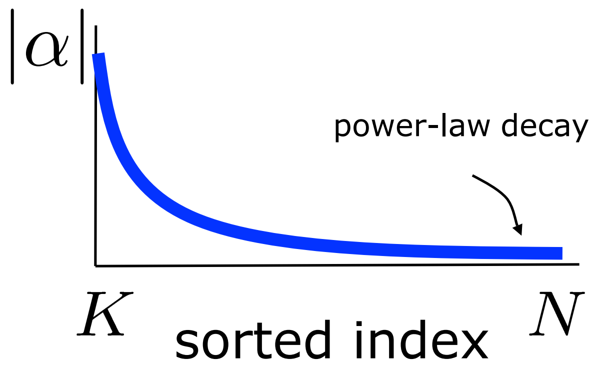
$$Y = HX + N$$

$X = \Phi\alpha$, and α is sparse

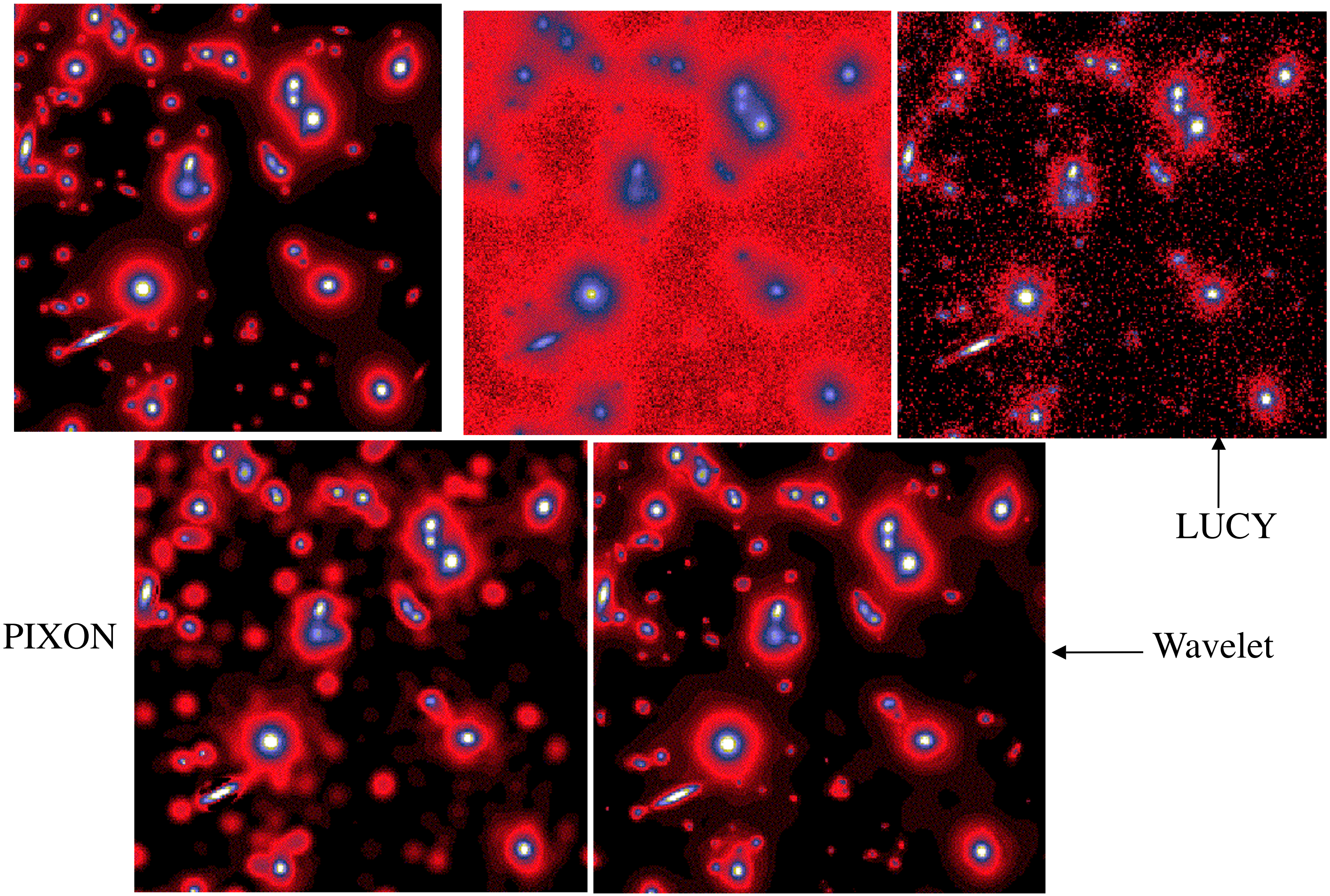
- Denoising
- Deconvolution
- Component Separation
- Inpainting
- Blind Source Separation
- Minimization algorithms
- Compressed Sensing

$$\min_{\alpha} \|\alpha\|_p^p \quad \text{subject to} \quad \|Y - H\Phi\alpha\|^2 \leq \epsilon$$



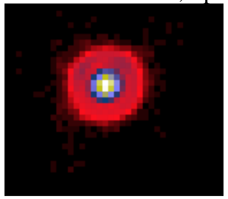


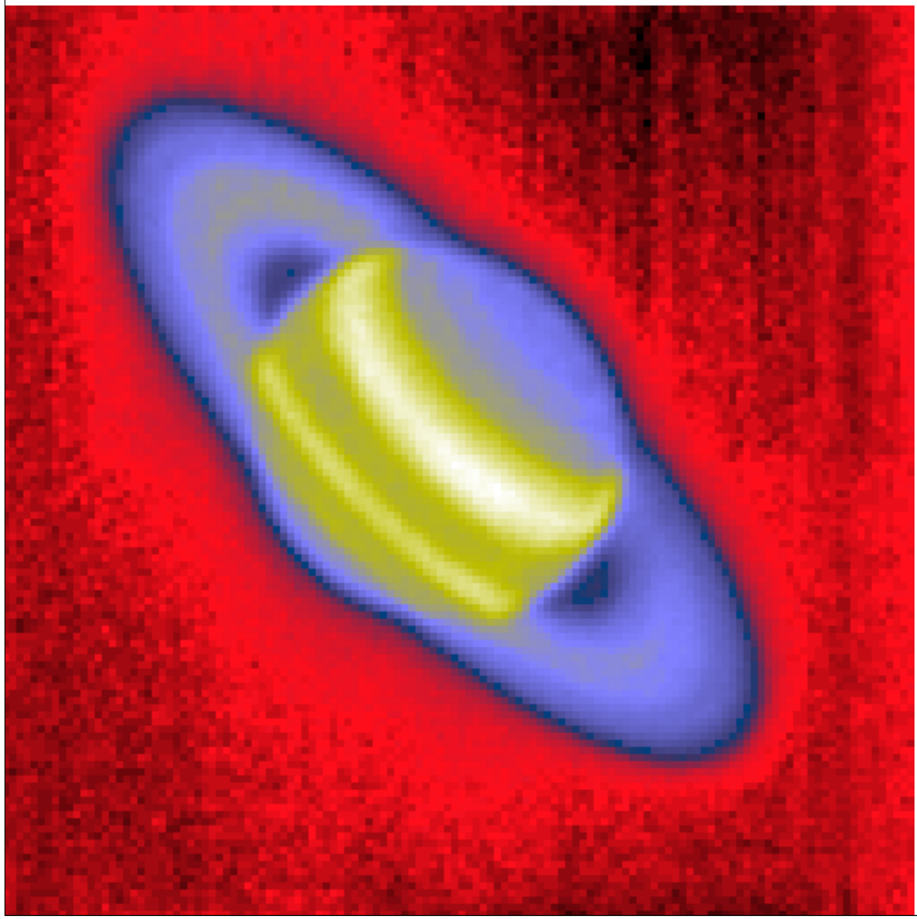
DECONVOLUTION SIMULATION

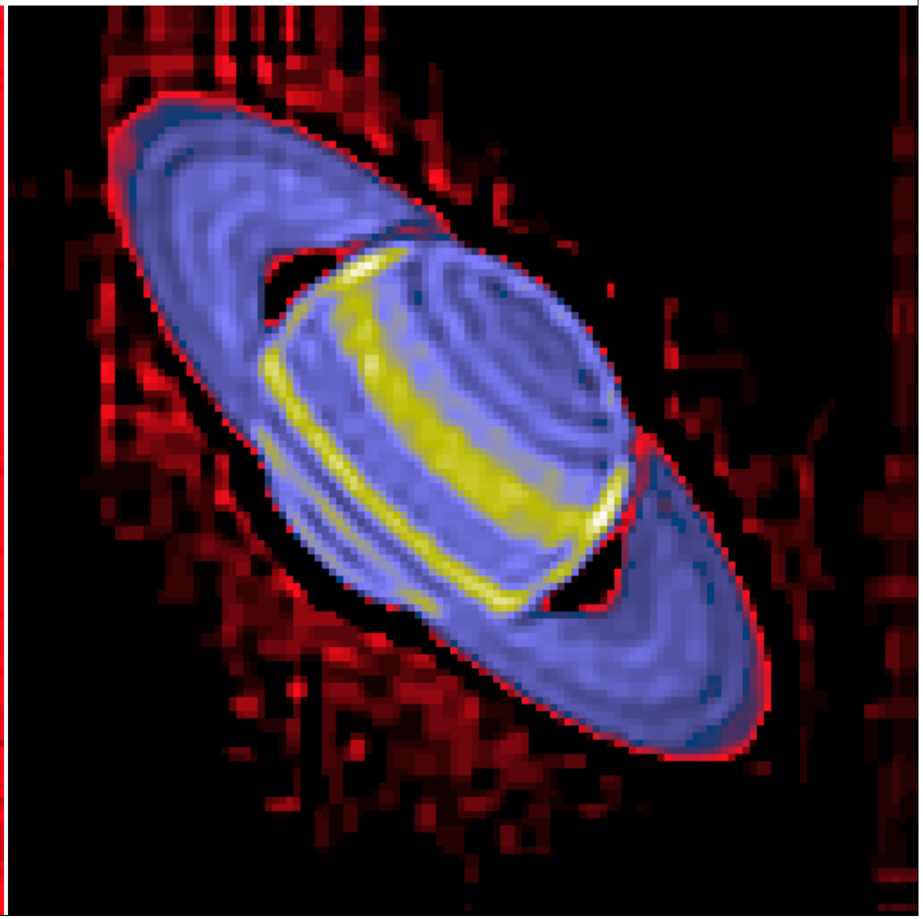


DECONVOLUTION

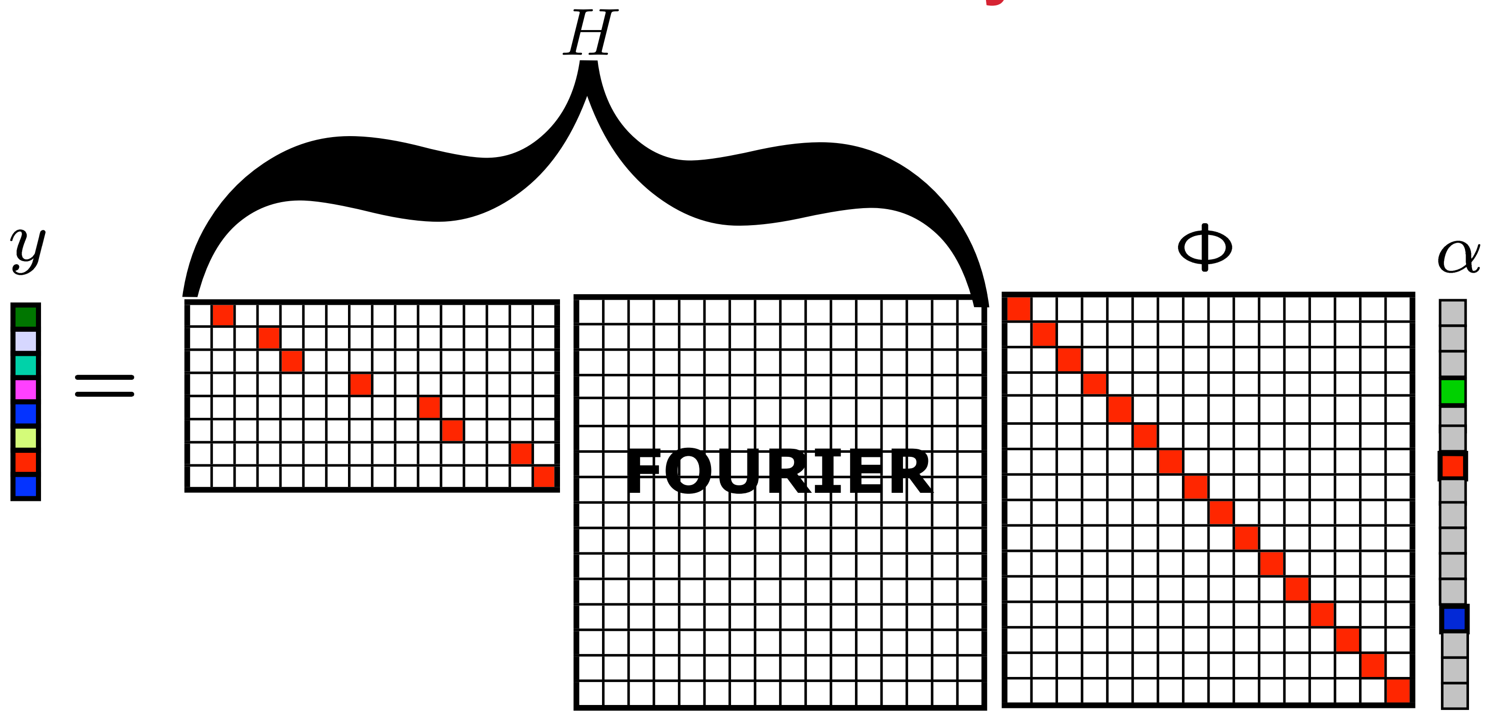
- E. Pantin, J.-L. Starck, and F. Murtagh, "Deconvolution and Blind Deconvolution in Astronomy", in Blind image deconvolution: theory and applications, pp 277--317, 2007.
- J.-L. Starck, F. Murtagh, and M. Bertero, "The Starlet Transform in Astronomical Data Processing: Application to Source Detection and Image Deconvolution", Springer, Handbook of Mathematical Methods in Imaging, in press, 2010.







Radio-Interferometry

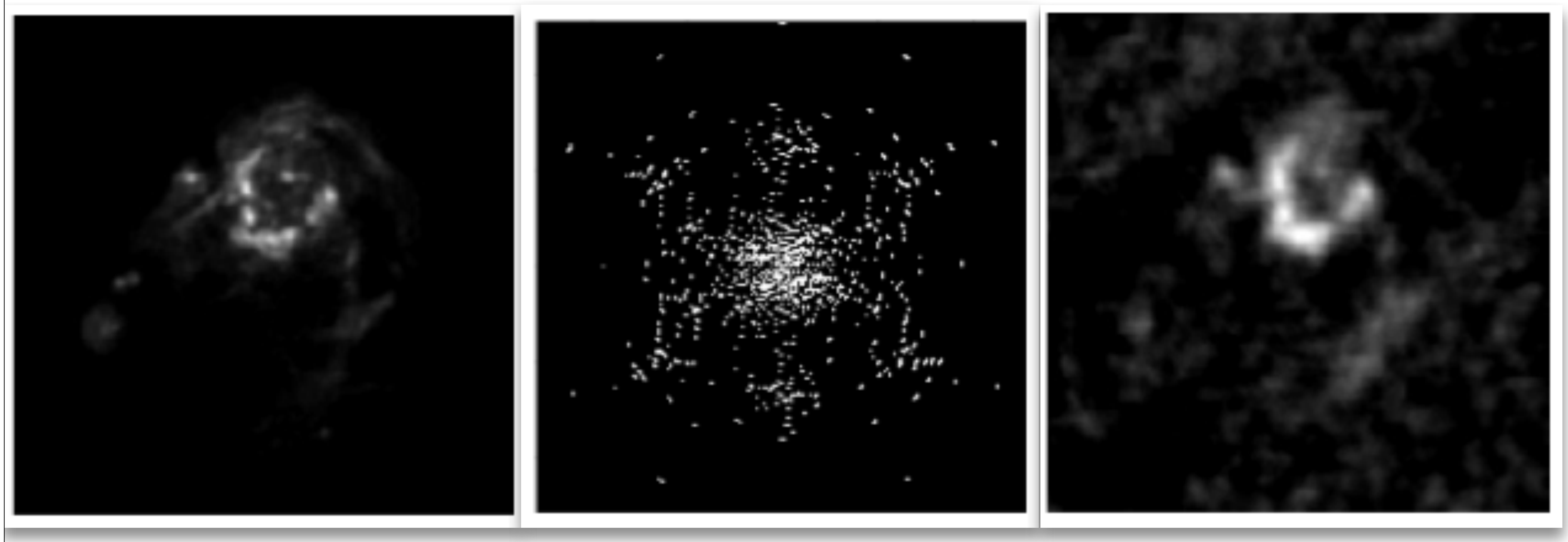


Measurement System

==> See (McEwen et al, 2011; Wenger et al, 2010; Wiaux et al, 2009; Cornwell et al, 2009; Suskimo, 2009; Feng et al, 2011).

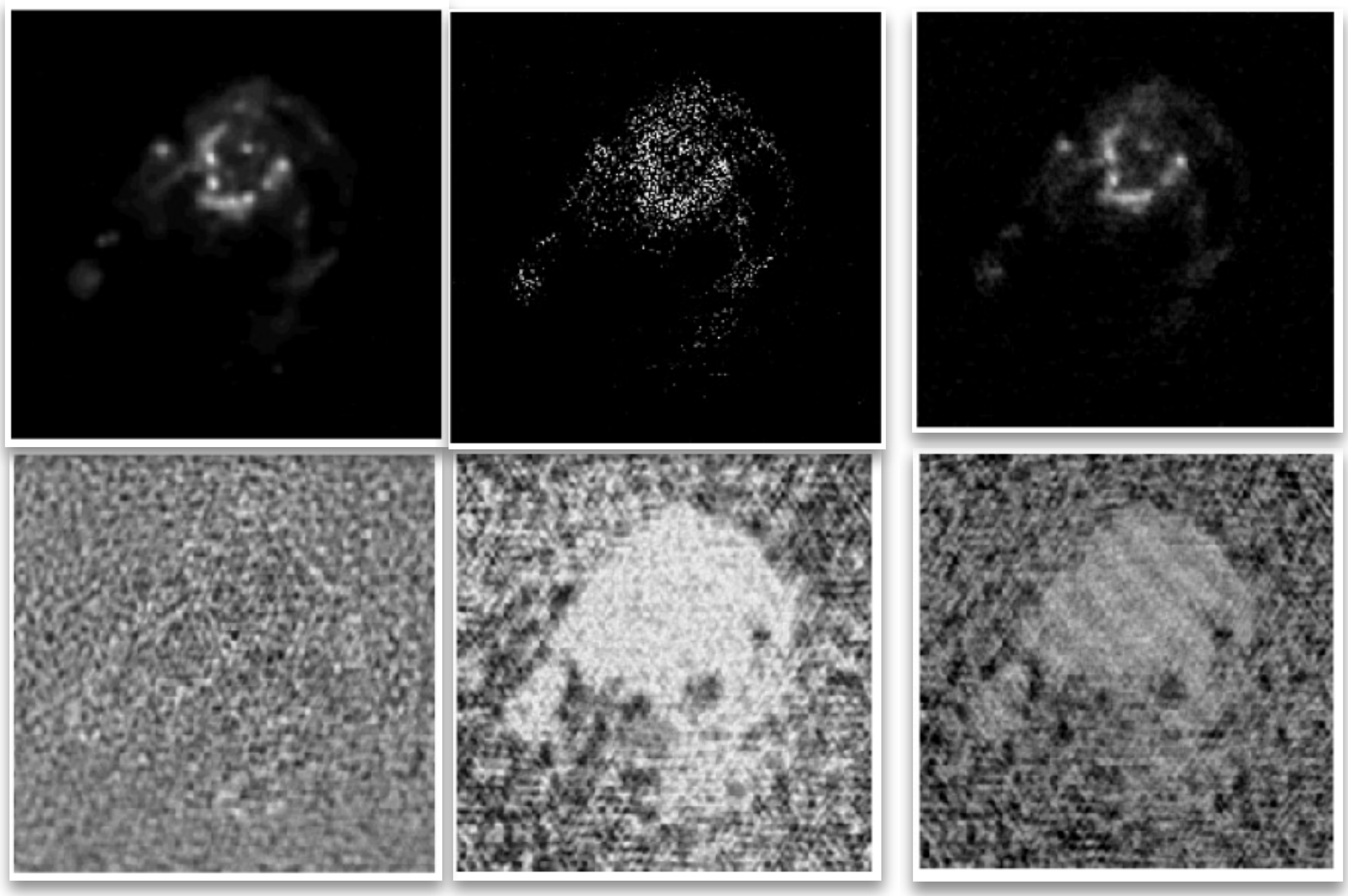
CS-Radio Astronomy

The Applications of Compressive Sensing to Radio Astronomy: I Deconvolution
Feng Li, Tim J. Cornwell and Frank De hoog, ArXiv:1106.1711, Volume 528, A31,2011.



Australian Square Kilometer Array Pathfinder (ASKAP) radio telescope.

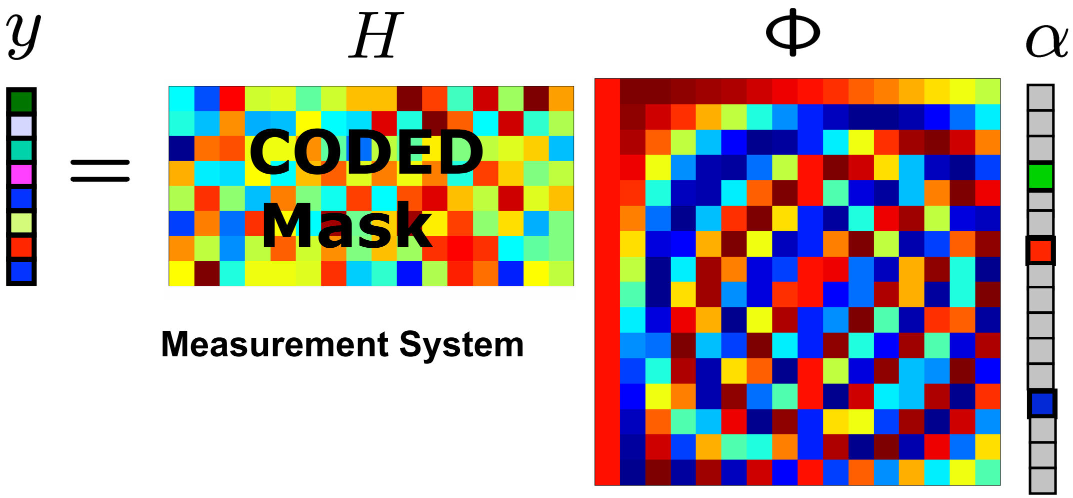
CS-Radio Astronomy

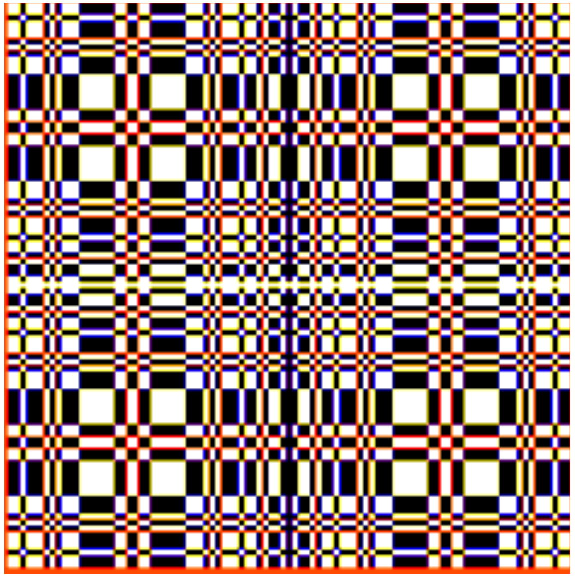


Hogbom CLEAN

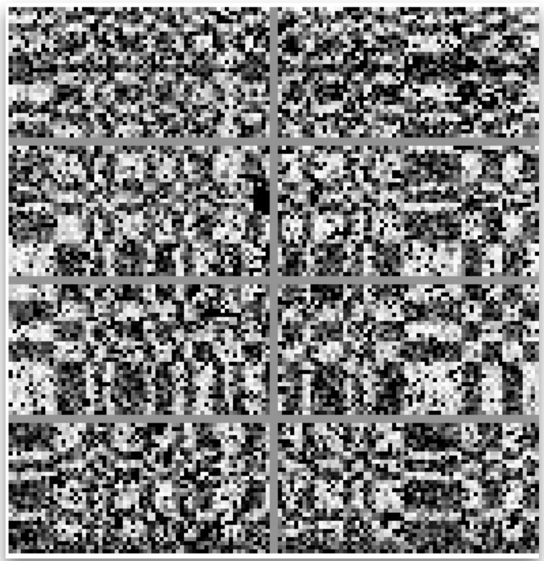
MEM residual

Gamma Ray Instruments (Integral) - Acquisition with coded masks





INTEGRAL/IBIS Coded Mask



Crab Nebula Integral Observation

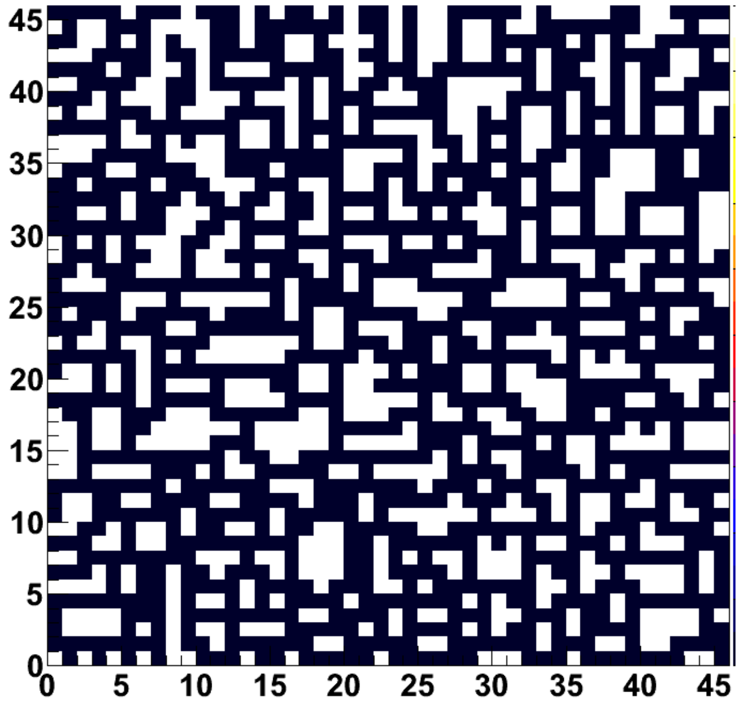
Courtesy I. Caballero, J. Rodriguez (AIM/Saclay)

SVOM (future French-Chinese Gamma-Ray Burst mission)

saclay
irfu

- ***ECLAIRs france-chinese satellite 'SVOM' (launch in 2014-2015)***
  Gamma-ray detection in energy range 4 - 120 keV
  Coded mask imaging (at 460 mm of the detector plane)

Physical mask pattern
(46 x 46 pixels of 11.7 mm)

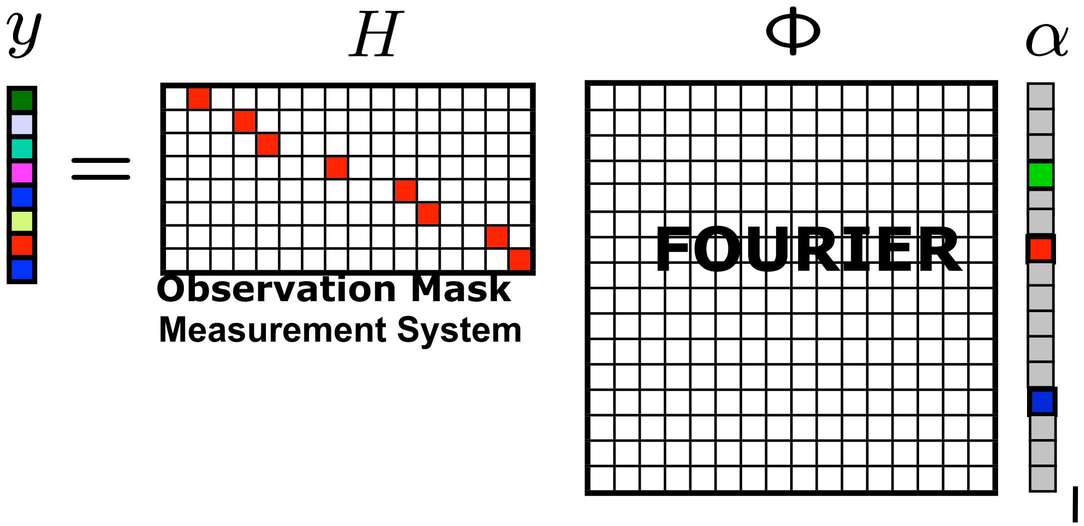


ECLAIR could become the first CS-Designed Astronomical Instrument

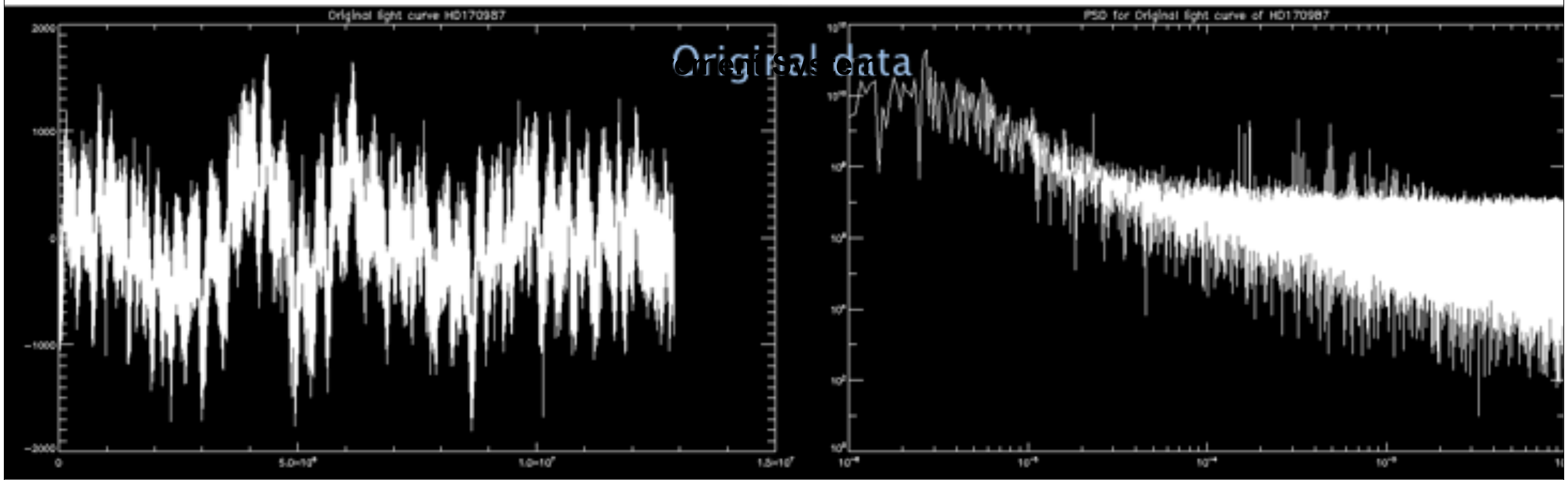
Stéphane Schanne – CEA

Missing Data

Period detection in temporal series

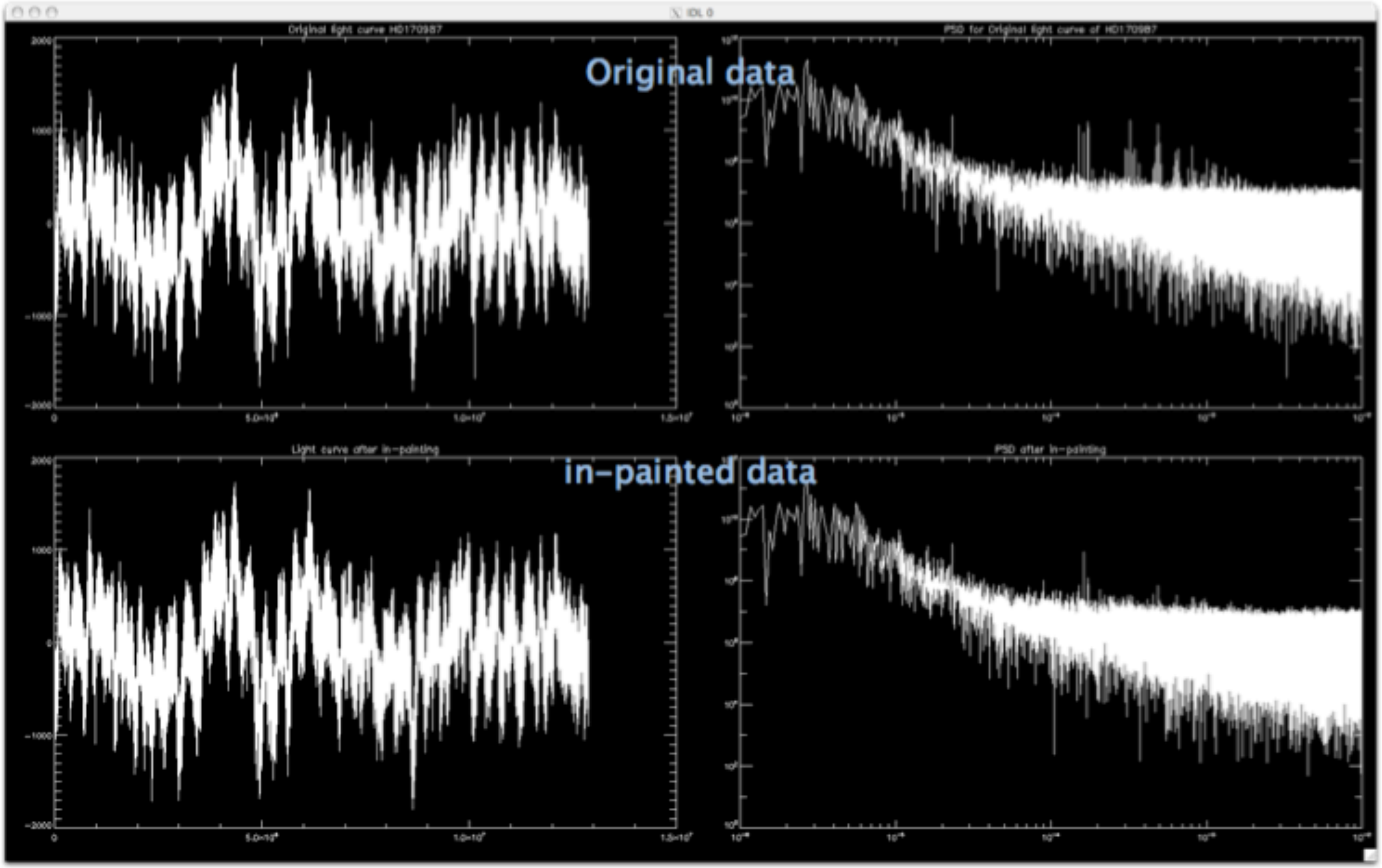


COROT: HD170987



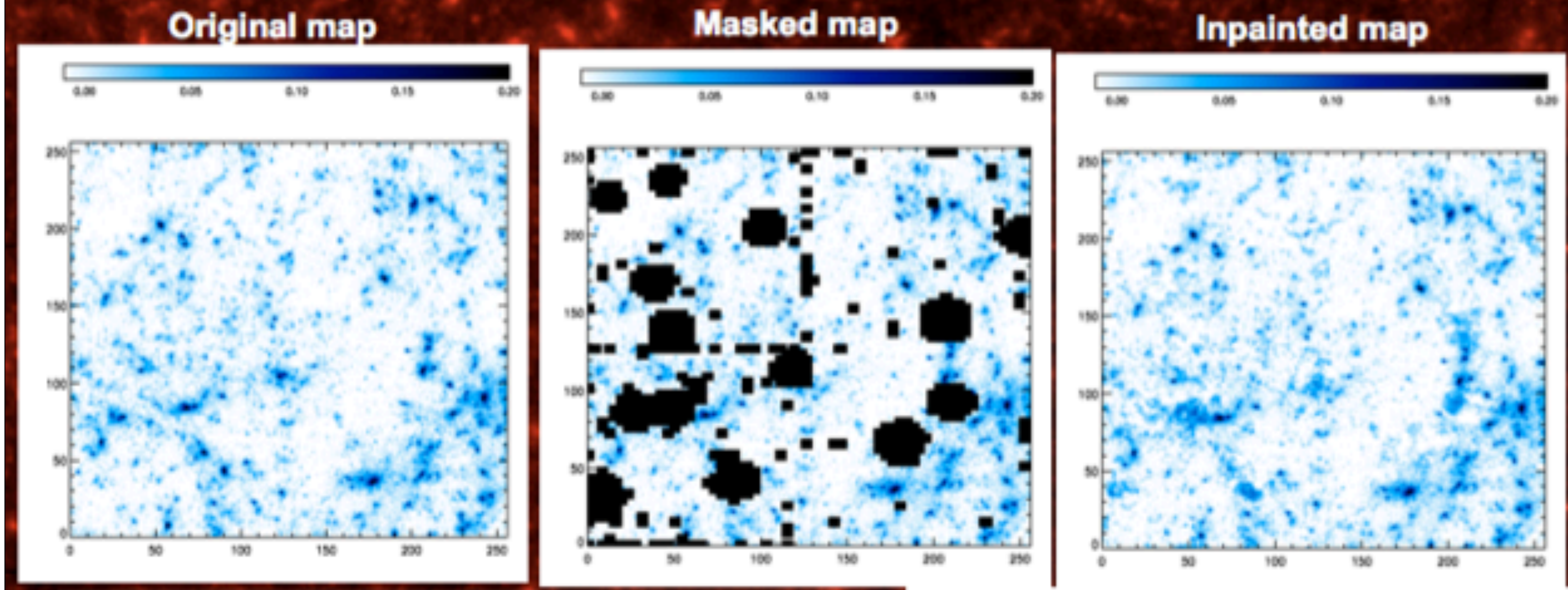
COROT: HD170987 with in-painting

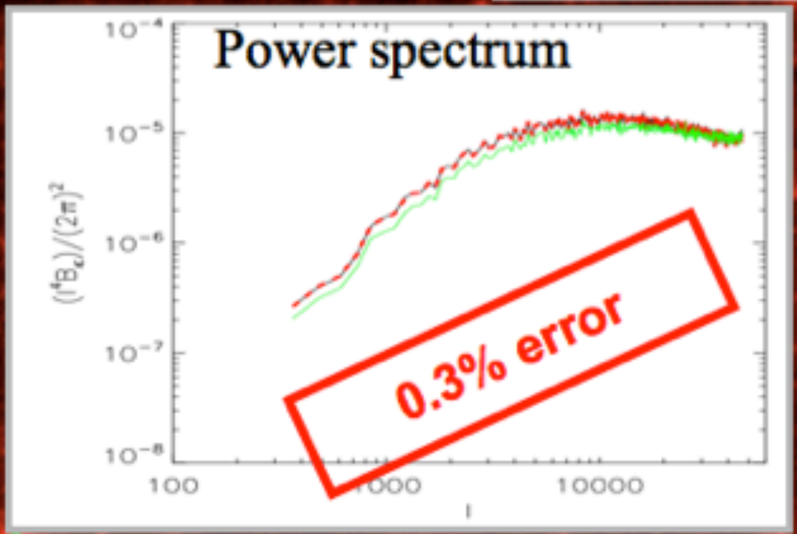
arXiv:1003.5178

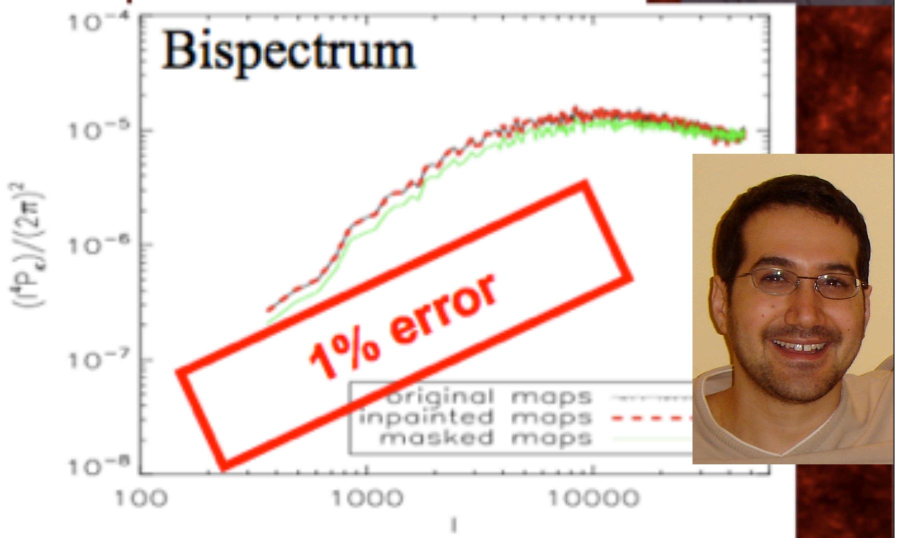


Inpainting :

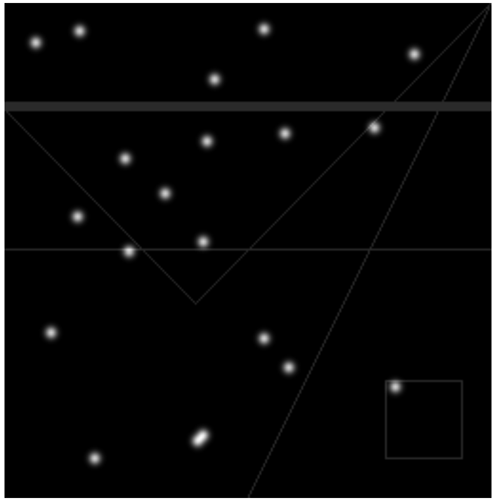
S. Pires, J.-L. Starck, A. Amara, R. Teyssier, A. Refregier and J. Fadili, "FASTLens (FAst STatistics for weak Lensing) : Fast method for Weak Lensing Statistics and map making", MNRAS, 395, 3, pp. 1265-1279, 2009.

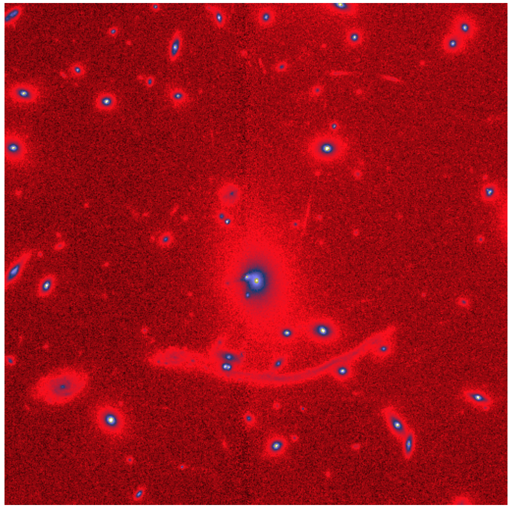






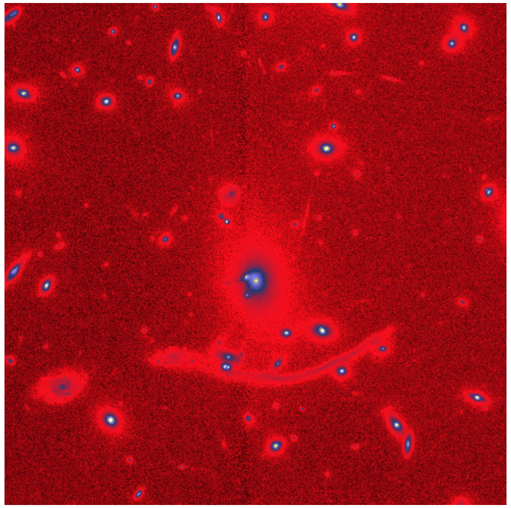
PB: a given transform does not necessary provide a good dictionary for all features contained in the data.





Morphological Diversity

- J.-L. Starck, M. Elad, and D.L. Donoho, ***Redundant Multiscale Transforms and their Application for Morphological Component Analysis***, *Advances in Imaging and Electron Physics, 132, 2004.*
- J.-L. Starck, M. Elad, and D.L. Donoho, ***Image Decomposition Via the Combination of Sparse Representation and a Variational Approach***, *IEEE Trans. on Image Proces., 14, 10, pp 1570--1582, 2005.*
- J.Bobin et al, ***Morphological Component Analysis: an adaptive thresholding strategy***, IEEE Trans. on Image Processing, Vol 16, No 11, pp 2675--2681, 2007.

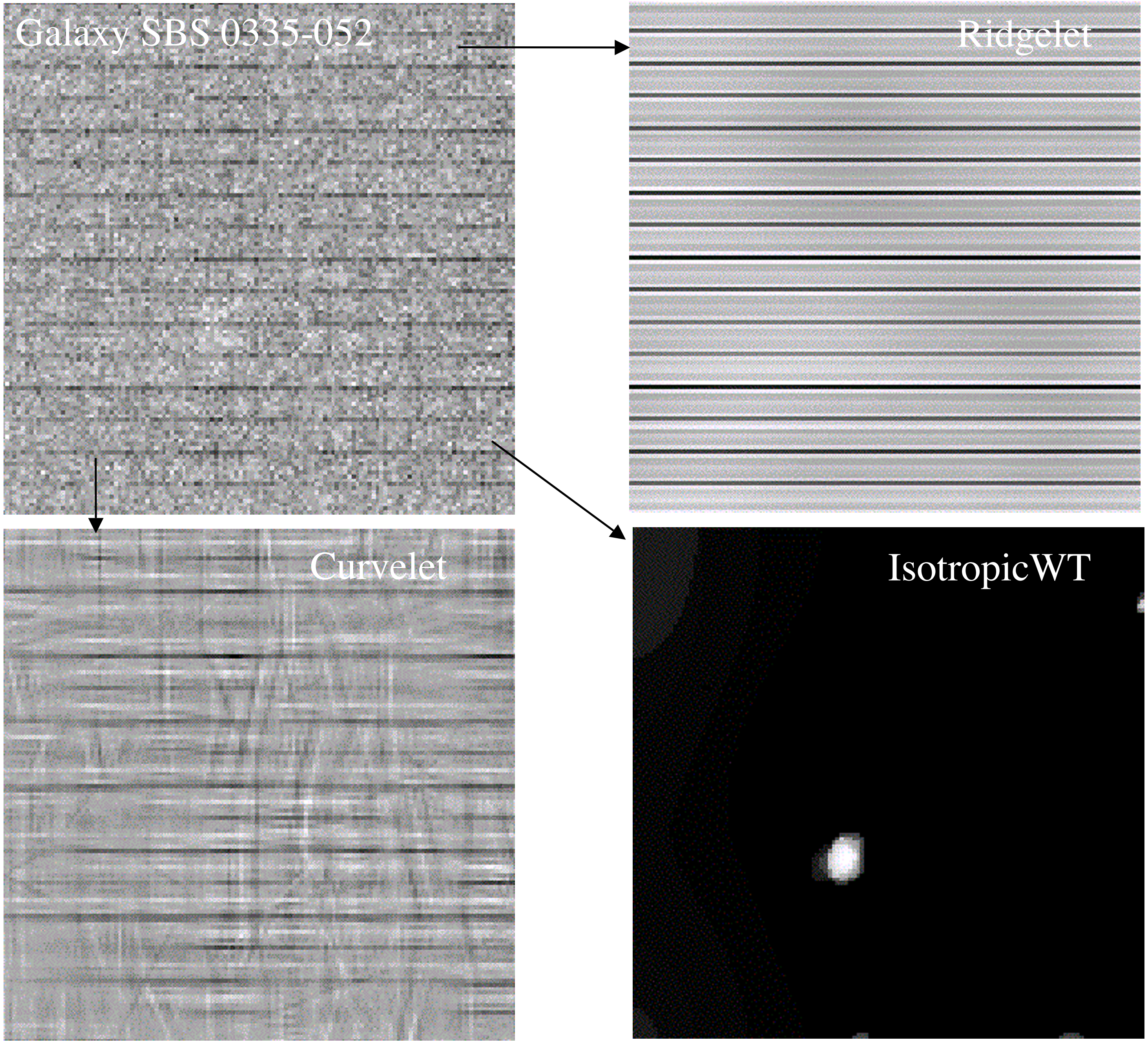


$$\phi = [\phi_1, \ldots, \phi_L], \quad \alpha = \{\alpha_1, \ldots, \alpha_L\}, \quad s = \phi\alpha = \sum_{k=1}^{L} \phi_k \alpha_k$$

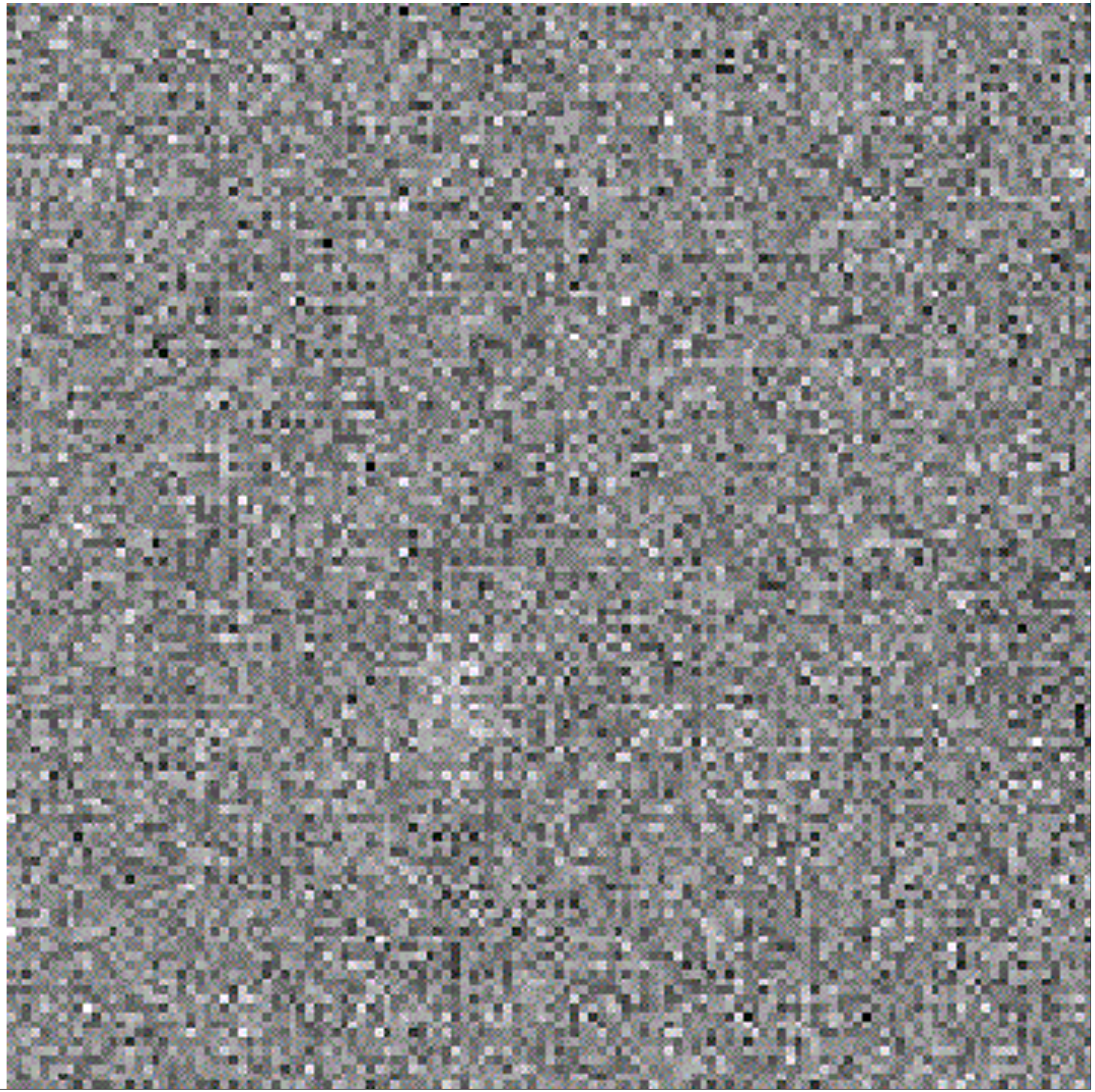
Sparsity Model 2: we consider a signal as a sum of K components s_k, $s = \sum_{k=1}^{K} s_k$ each of them being sparse in a given dictionary :

$$s_k = \Phi_k \alpha_k$$

$$s = \sum_{k=1}^{K} s_k = \sum_{k=1}^{K} \Phi_k \alpha_k = \Phi\alpha$$


Galaxy SBS 0335-052
Ridgelet
Curvelet
IsotropicWT

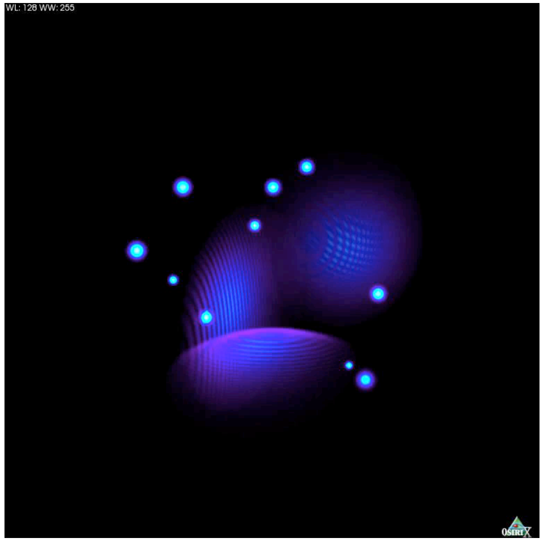
Galaxy SBS 0335-052
10 micron
GEMINI-OSCIR



3D Morphological Component Analysis

A. Woiselle

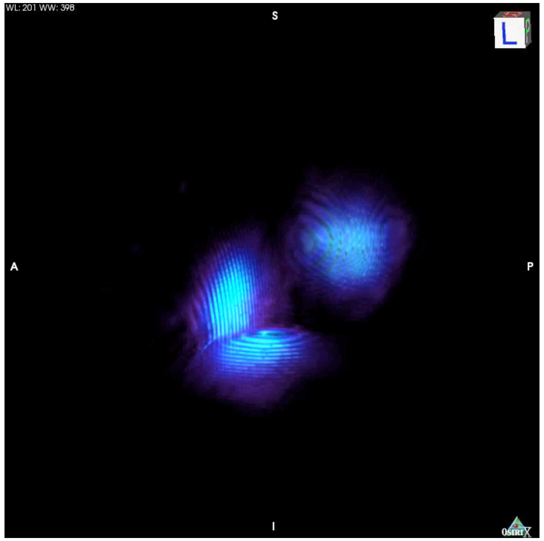
Original (3D shells + Gaussians)



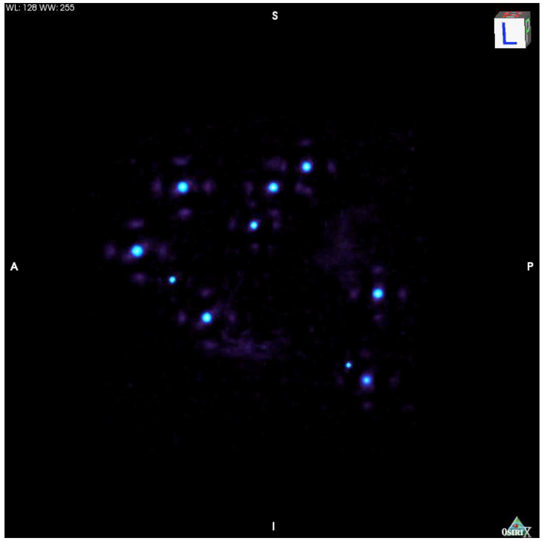
Dictionary

RidCurvelets + 3D UDWT.

Shells

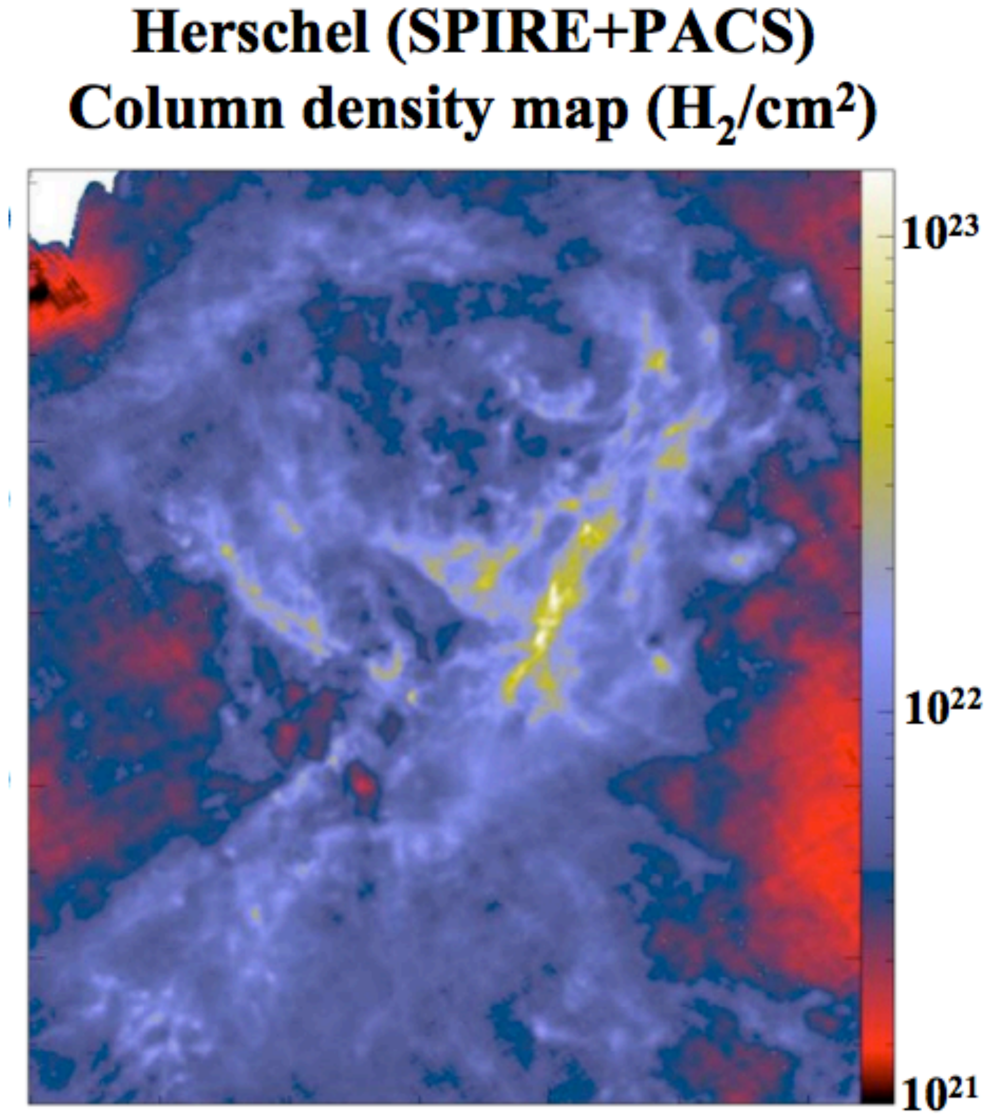


Gaussians



- A . Woiselle, J.L. Starck, M.J. Fadili, "3D Data Denoising and Inpainting with the Fast Curvelet transform", **JMIV**, 39, 2, pp 121-139, 2011.
- A. Woiselle, J.L. Starck, M.J. Fadili, "3D curvelet transforms and astronomical data restoration", **Applied and Computational Harmonic Analysis**, Vol. 28, No. 2, pp. 171-188, 2010.

Revealing the structure of one of the nearest infrared dark clouds (Aquila Main: d ~ 260 pc)



Dense cores form primarily in filaments

Morphological Component Analysis:

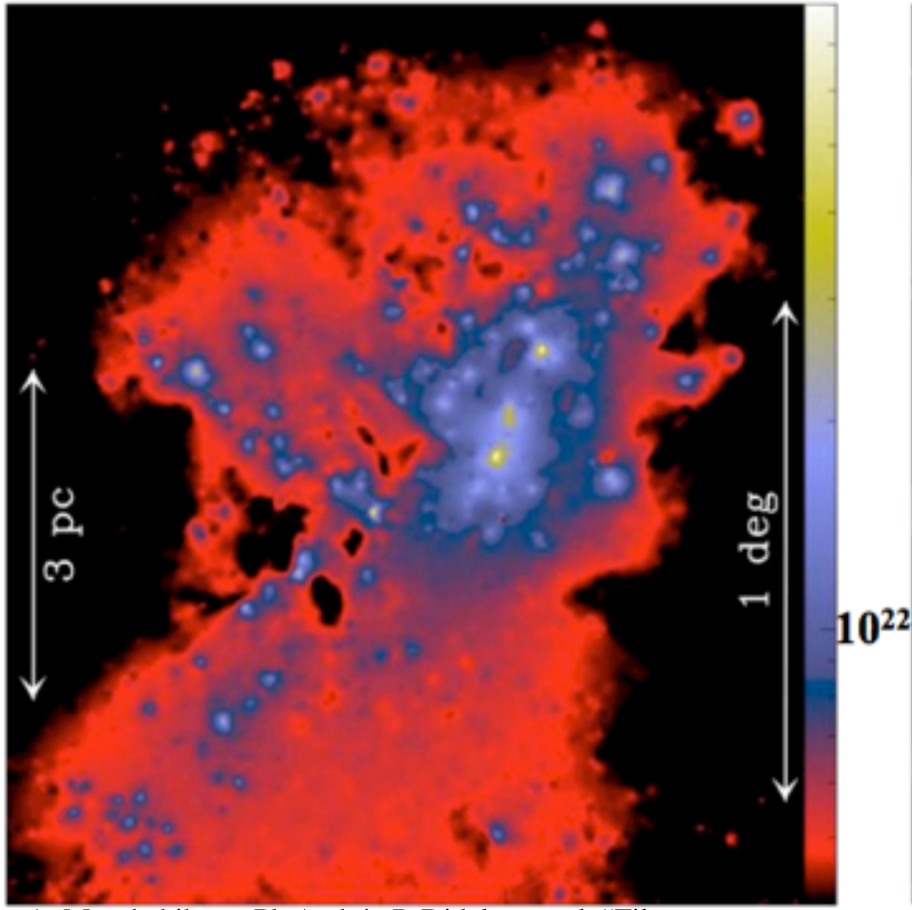
***Herschel* Column density map**

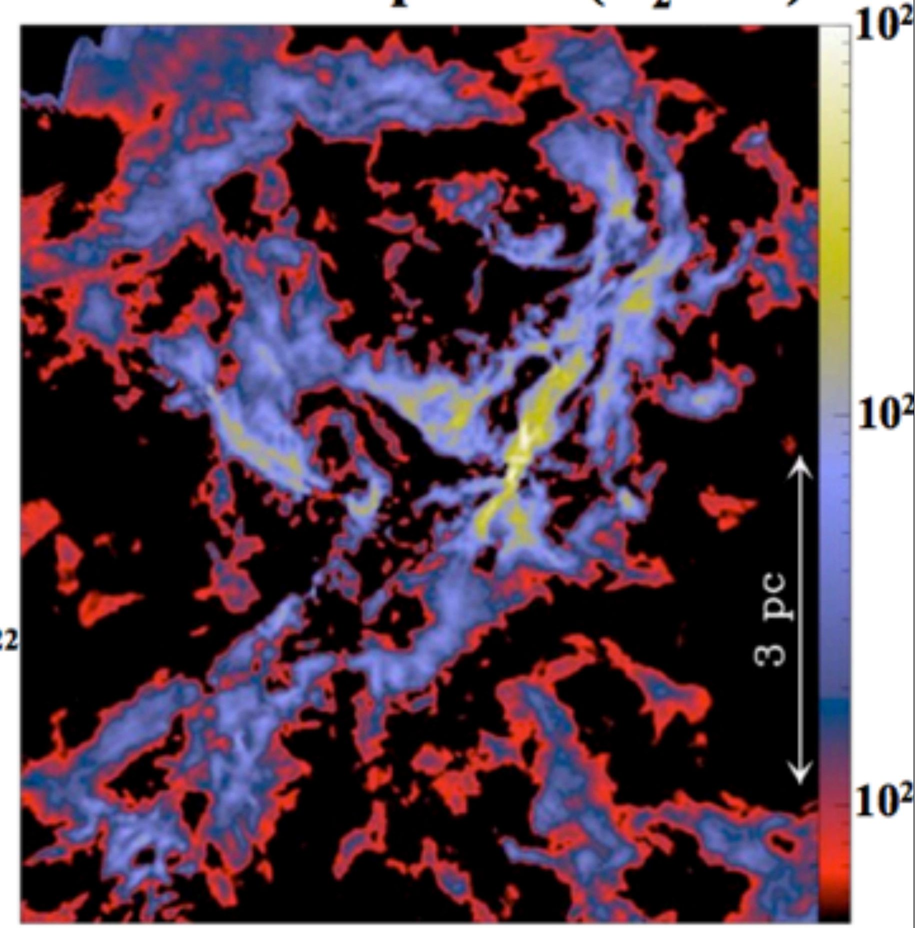
(P. Didelon based on Starck et al. 2003)

=

Cores
Wavelet component (H_2/cm^2) **+**

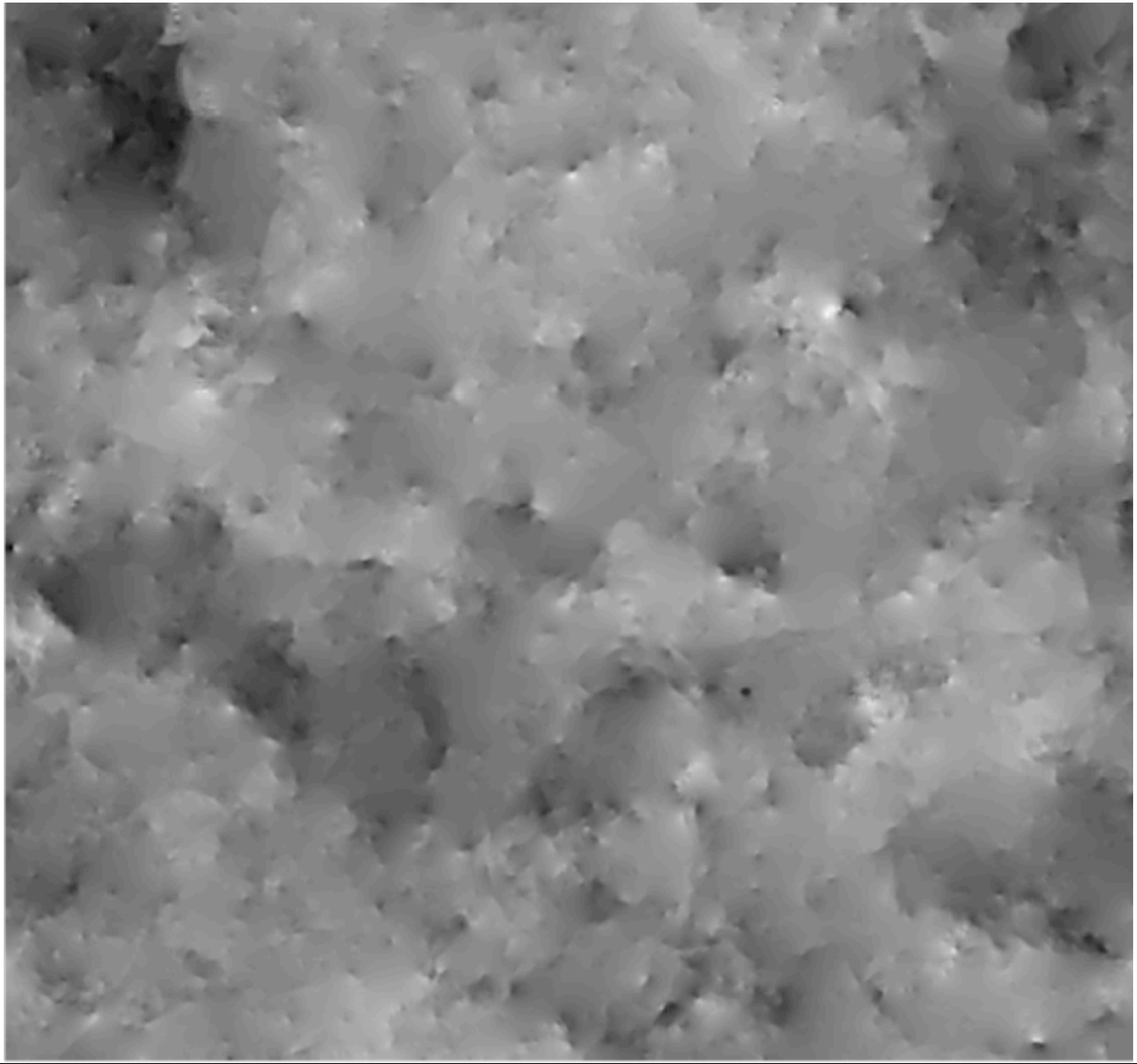
Filaments
Curvelet component (H_2/cm^2)



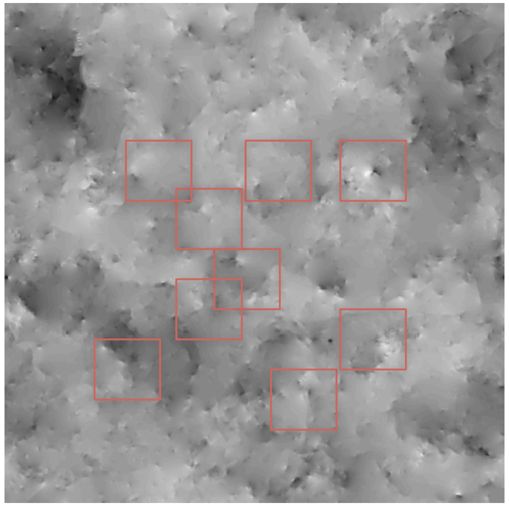


A. Menshchikov, Ph.André. P. Didelon, et al, "Filamentary structures and compact objects in the Aquila and Polaris clouds observed by Herschel", A&A, 518, id.L103, 2010.

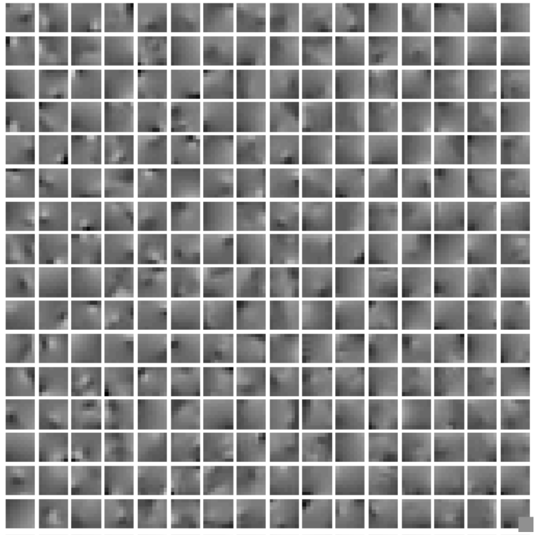
Simulated Cosmic String Map



Dictionary Learning

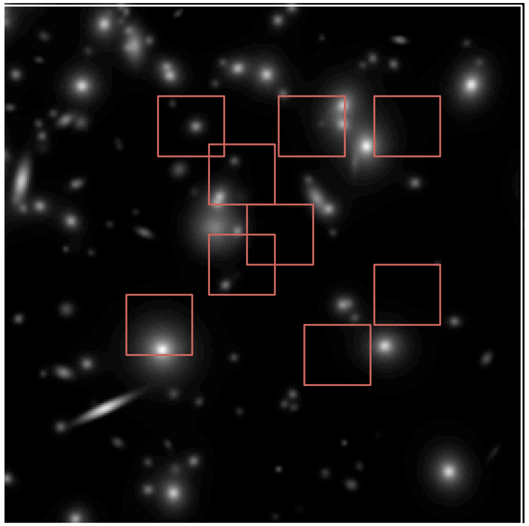


Training basis.



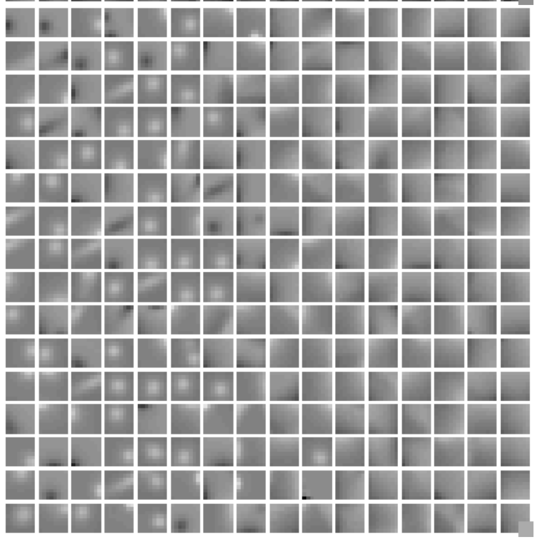
$$\left(\hat{D},\hat{A}\right) = \underset{\substack{D\in C_1 \\ A\in C_2}}{\arg\min}\left(Y = DA\right)$$

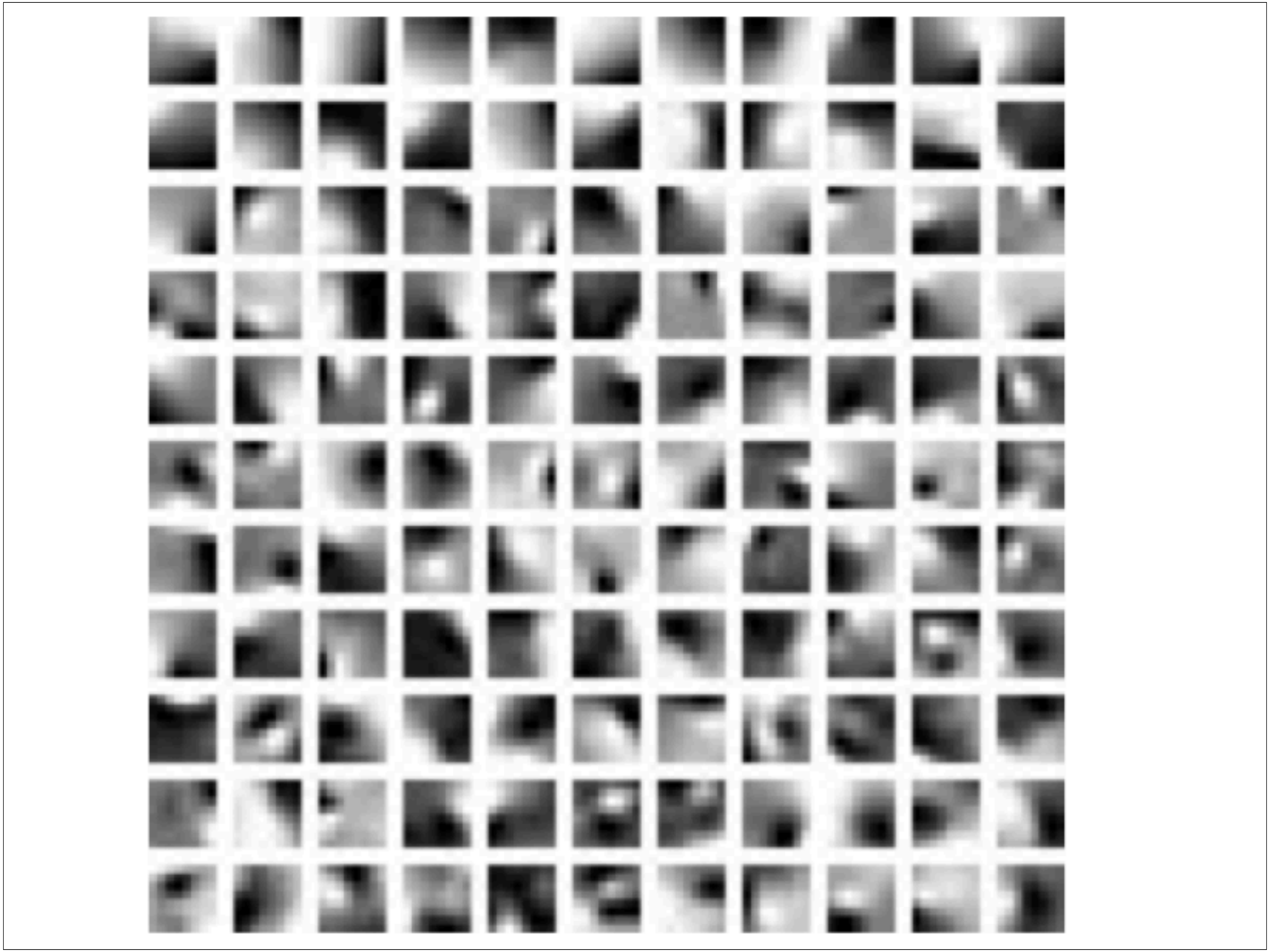
DL: Matrix Factorization problem

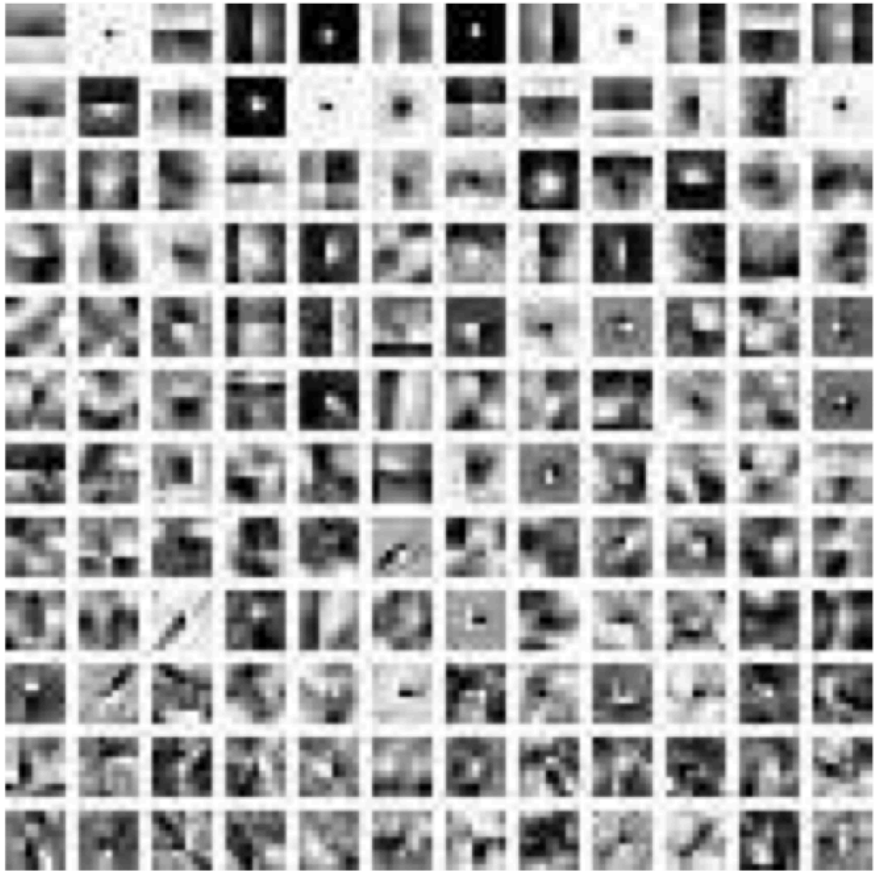


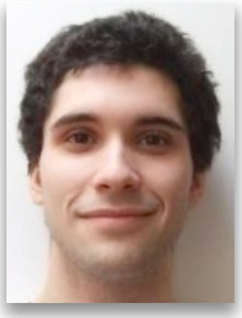
C_1: Constraints on the Sparsifying dictionary D

C_2: Constraints on the Sparse codes



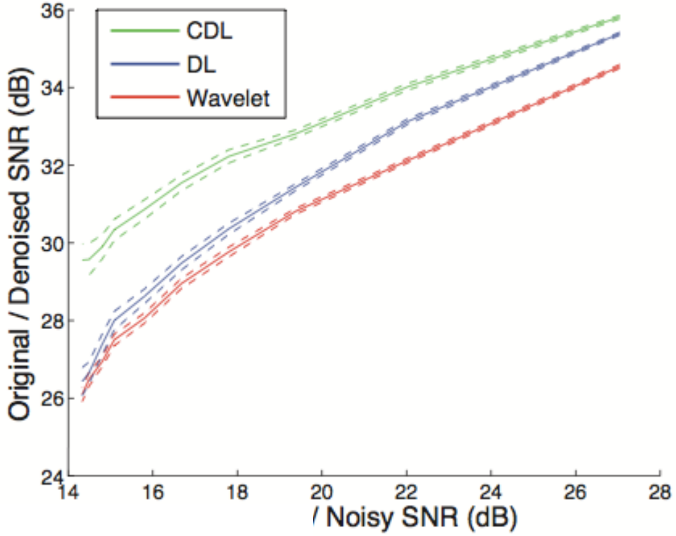


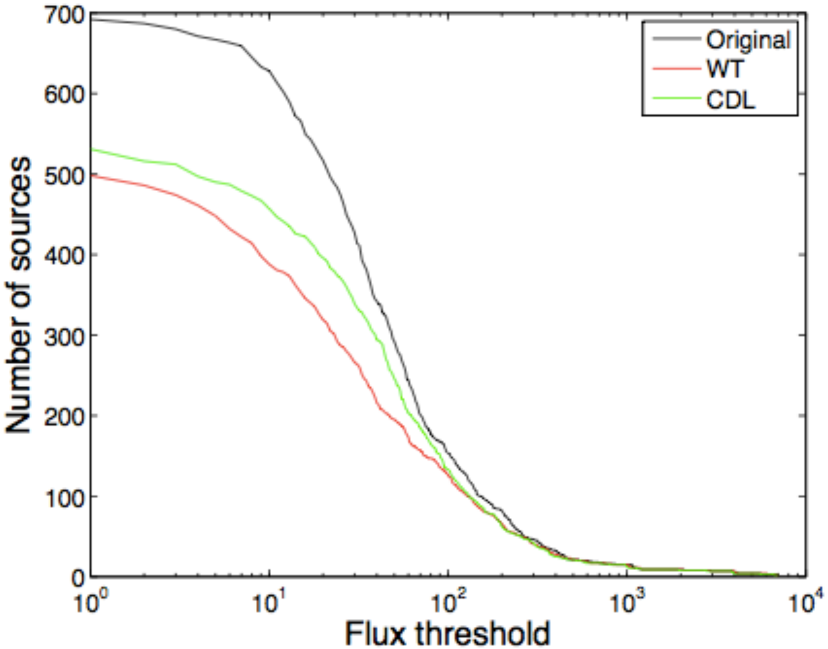



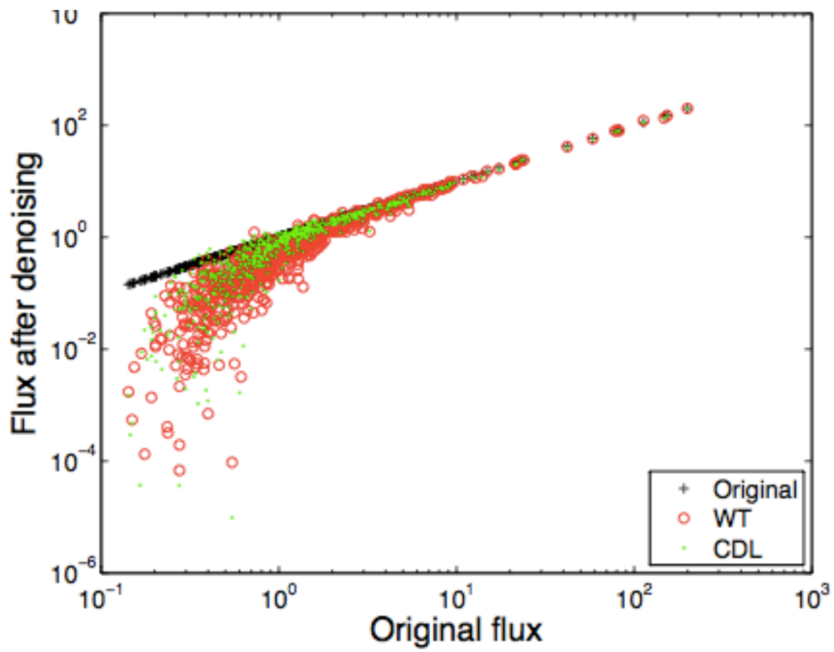


S. Beckouche

Astronomical Image Denoising Using Dictionary Learning, S. Beckouche, J.L. Starck, and J. Fadili, A&A, submitted.







Sparsity Model 1: we consider a dictionary which has a fast transform/reconstruction operator:

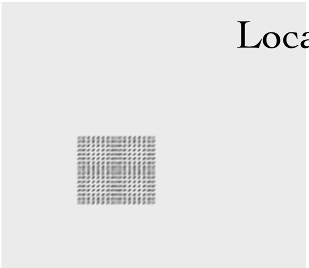
$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

$$s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi\alpha$$

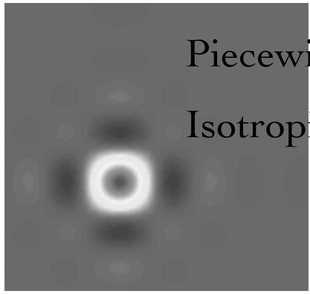
Local DCT

Stationary textures

Locally oscillatory



Wavelet transform



Piecewise smooth

Isotropic structures

Curvelet transform

Piecewise smooth, edge

Sparsity Model 2: Morphological Diversity:

$$\phi = [\phi_1, \ldots, \phi_L], \quad \alpha = \{\alpha_1, \ldots, \alpha_L\}, \quad s = \phi\alpha = \sum_{k=1}^{L} \phi_k \alpha_k$$

Sparsity Model 3: we adapt/learn the dictionary directly from the data

Model 3 can be also combined with model 2:

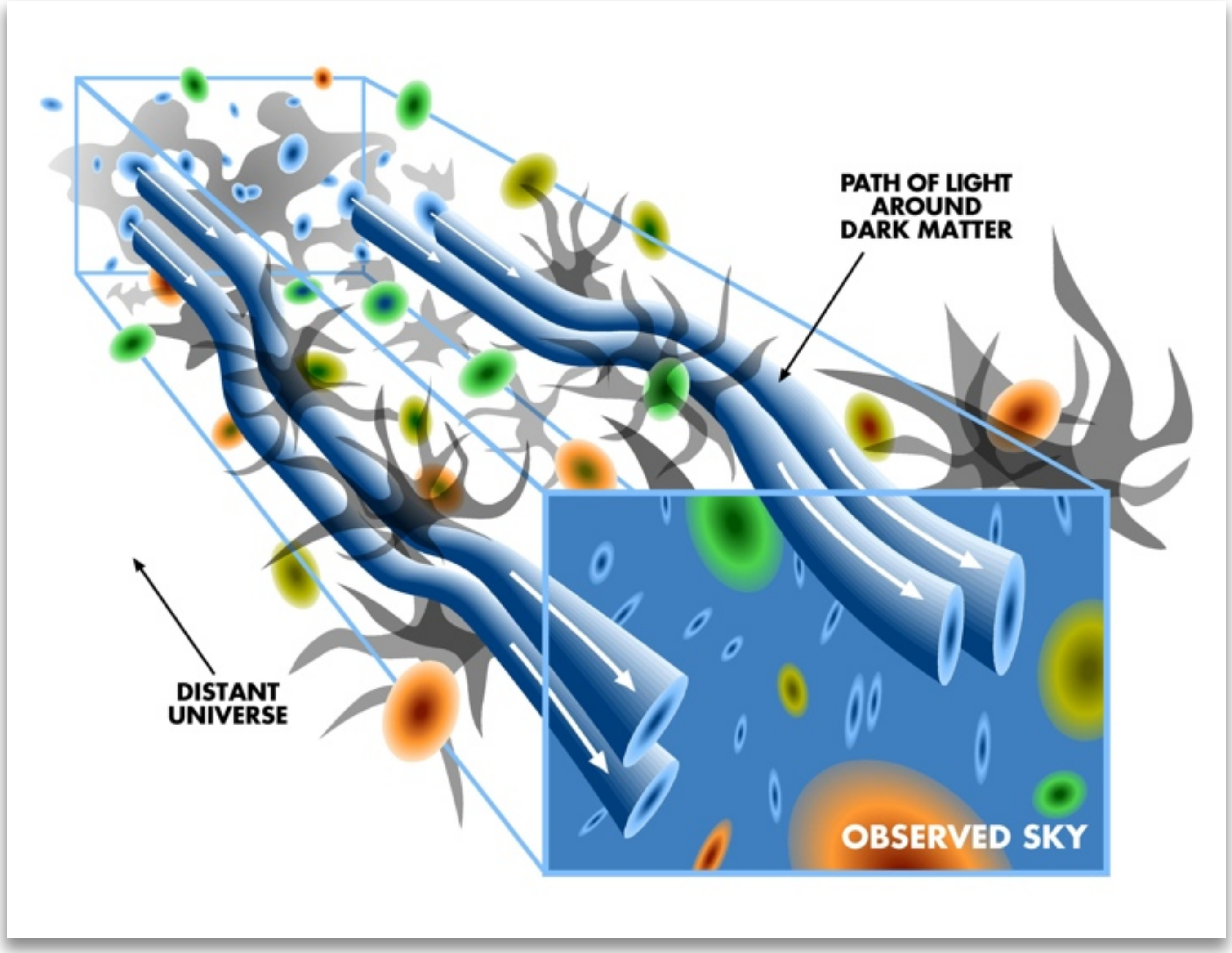
G. Peyre, M.J. Fadili and J.L. Starck, , "Learning the Morphological Diversity", **SIAM Journal of Imaging Science**, 3 (3) , pp.646-669, 2010.

Advantages of model 1: extremely fast.

Advantages of model 2:
- more flexible to model 1.
- The coupling of local DCT+curvelet is well adapted to a relatively large class of images.

Advantages of model 3:
atoms can be obtained which are well adapted to the data, and which could never be obtained with a fixed dictionary.
Drawback of model 3 versus model 1,2:
We pay the price of dictionary learning by being less sensitive to detect very faint features.
Complexity: Computation time, parameters, etc

3D Weak Lensing

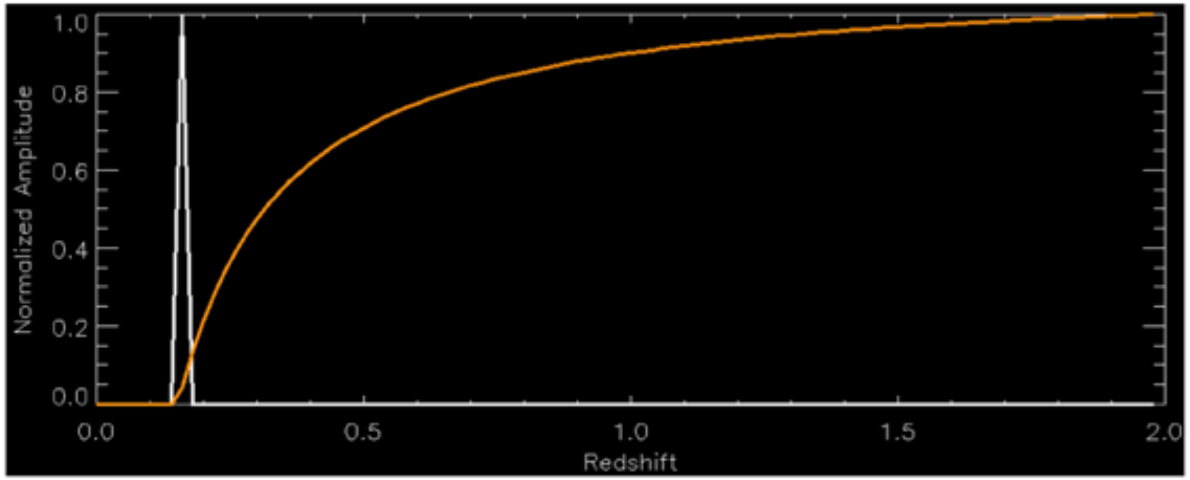


The reconstruction problem

- Along one line of sight, the convergence κ can be linked to the density contrast δ through:

$$\kappa(\theta) = Q\delta(\theta)$$

where Q is the lensing efficiency matrix. Depends on cosmology and binning of the data.



Still Unknown: Convergence κ (in orange)

Even less known: Density contrast δ (in white)

3D Weak Lensing

The convergence κ, as seen in sources of a given redshift bin, is the linear transformation of the matter density contrast, δ, along the line-of-sight (Simon et al 2009):

$$\kappa = Q\delta + N$$

with $\delta(r) \equiv \rho(r)/\overline{\rho} - 1$

$$Q_{i\ell} = \frac{3H_0^2\Omega_M}{2c^2}\int_{w_\ell}^{w_{\ell+1}} dw \frac{\overline{W}^{(i)}(w) f_K(w)}{a(w)}\,, \quad \overline{W}^{(i)}(w) = \int_0^{w^{(i)}} dw' \frac{f_K(w-w')}{f_K(w')}\left(p(z)\frac{dz}{dw}\right)_{z=z(w')}$$

where H_0 is the hubble parameter, Ω_M is the matter density parameter, c is the speed of light, $a(w)$ is the scale parameter evaluated at comoving distance w, and

$$f_K(w) = \begin{cases} K^{-1/2}\sin(K^{1/2}w), & K > 0 \\ w, & K = 0 \\ (-K)^{-1/2}\sinh([-K]^{1/2}w) & K < 0 \end{cases},$$

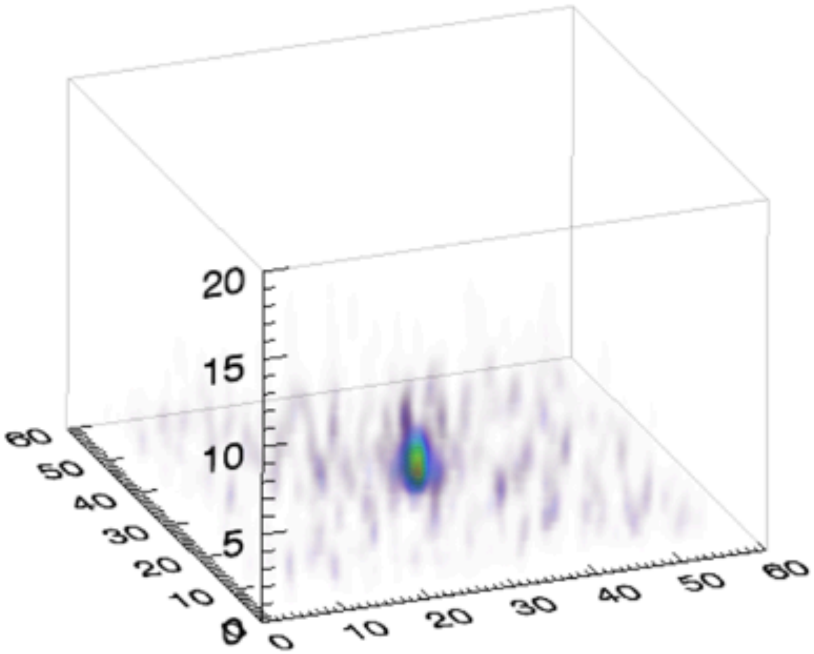
gives the comoving angular diameter distance as a function of the comoving distance and the curvature, K, of the Universe.

The reconstruction problem

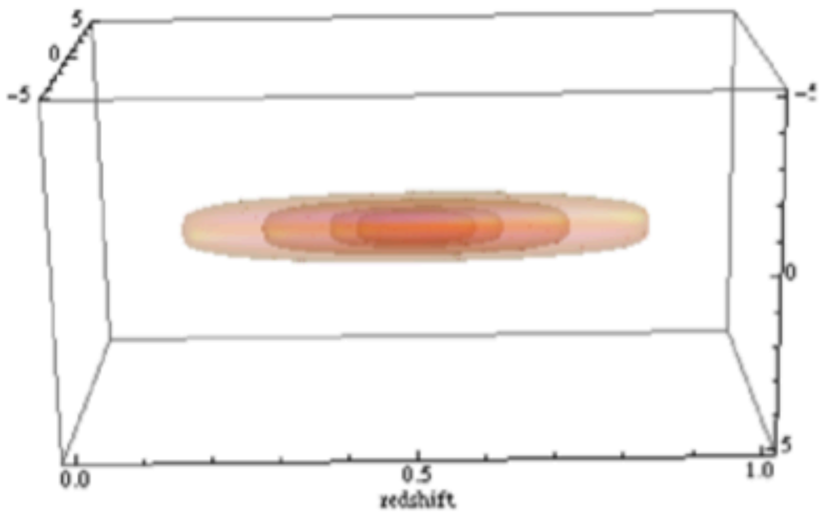
Linear inversion methods

- Structures are smeared along the line of sight
- Bias in the reconstructed redshift
- Amplitude of density contrast heavily damped
- Overall noisy reconstruction

Reconstruction of a single cluster :

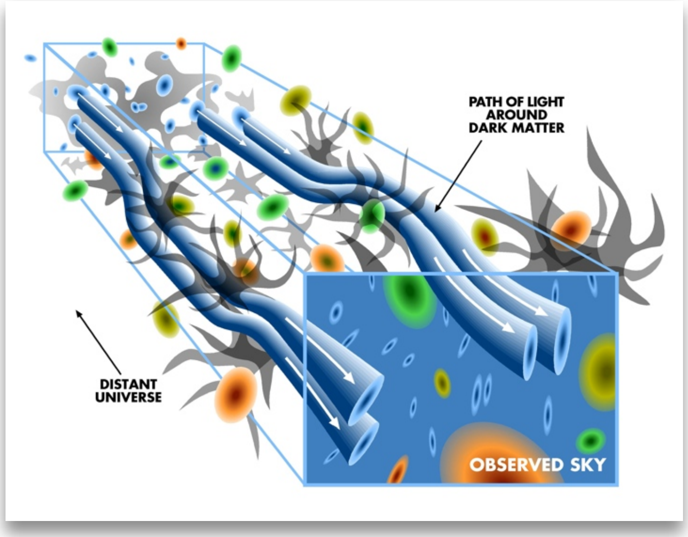


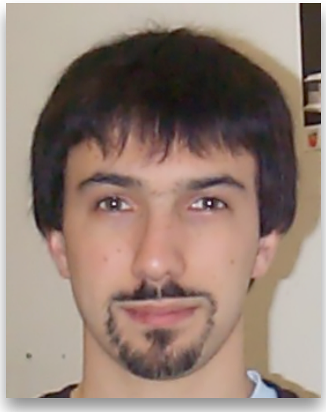
Typical response to a halo (from [Simon et al. (2009)])



CS-Weak Lensing

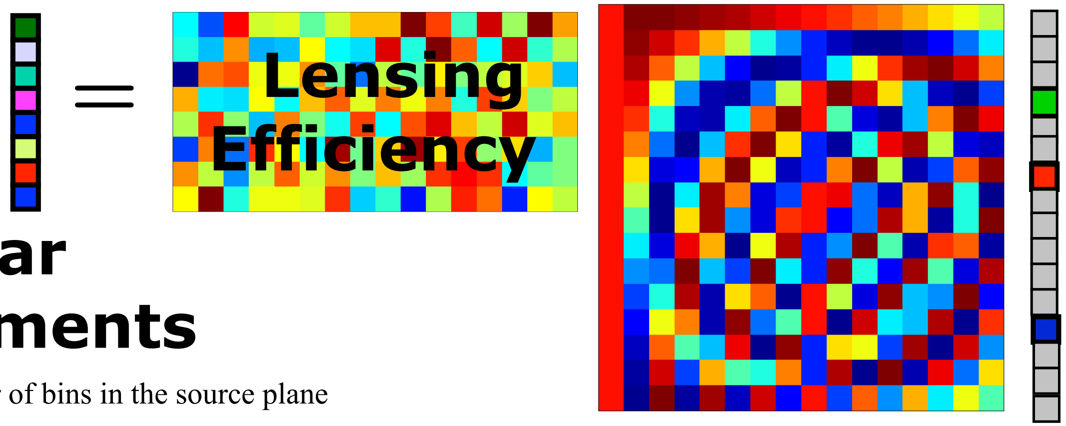
A. Leonard





F.X. Dupe

y Q Φ α



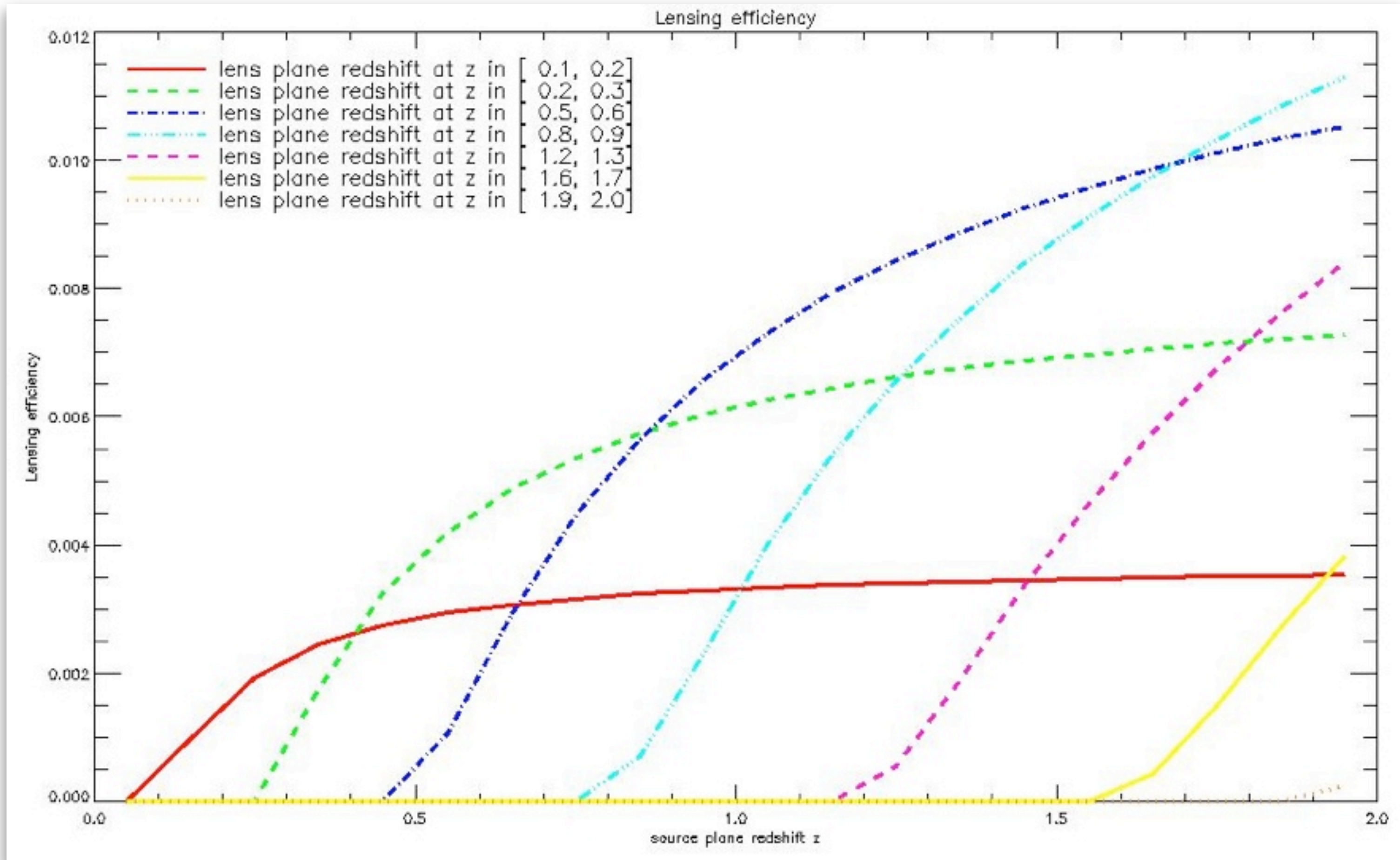
Shear Measurements

M measurements:: number of bins in the source plane

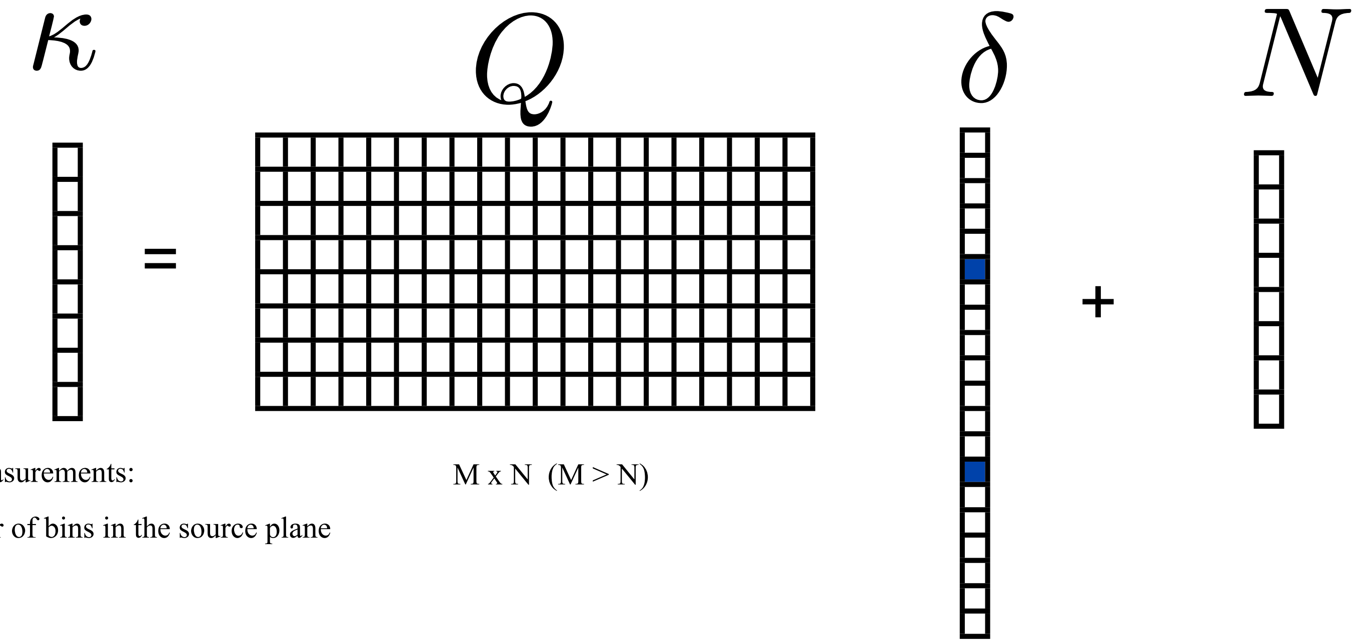
N redshift bin for the density contrast

3D Dark Matter Map Reconstruction

3D Weak Lensing



3D Weak Lensing



M measurements:

number of bins in the source plane

M x N (M > N)

N redshift bin for the density contrast

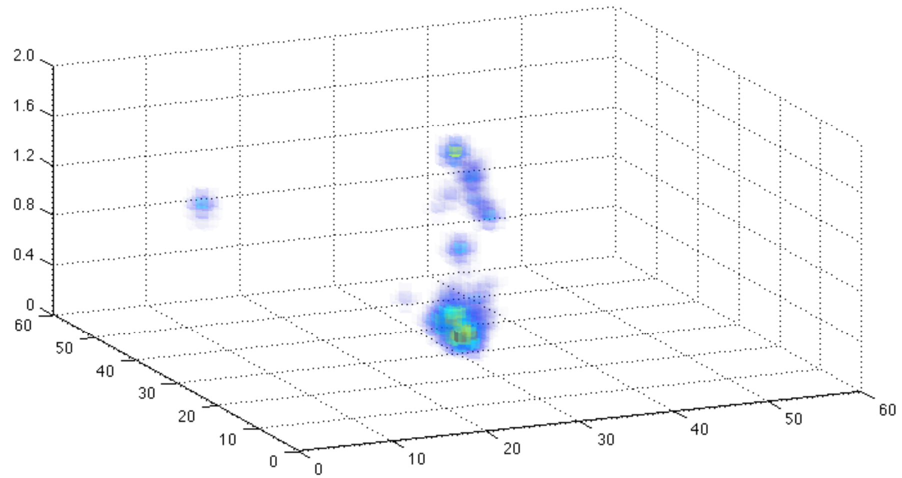
δ is sparse.

Q spreads out the information in δ along κ bins.

More unkown than measurements

3D Weak Lensing

$$\min_{\delta} \parallel \delta \parallel_1 \quad s.t. \ \frac{1}{2} \parallel \gamma - Q\delta \parallel^2_{\mathbf{\Sigma}^{-1}} \leq \epsilon$$



Reconstructions of two clusters along the line of sight, located at a redshift 0.2 and 1.0 (data binned into Nsp = 20 redshift bins, but aim to reconstruct onto Nlp = 25 redshift bins).

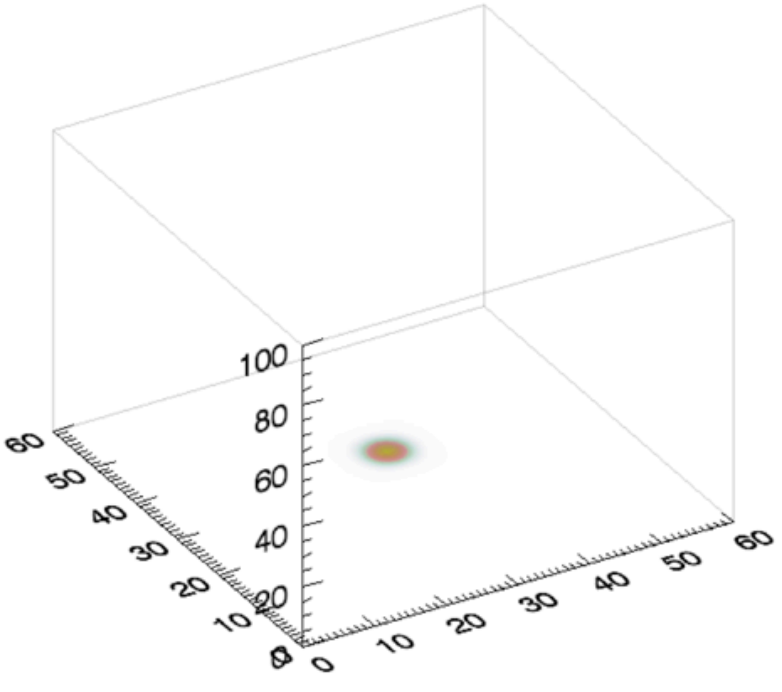
A. Leonard, F.-X. Dupe, J.-L. Starck, "A compressed sensing approach to 3D weak lensing", Astronomy and Astrophysics, arXiv:1111.6478, A&A, 539, A85, 2012.

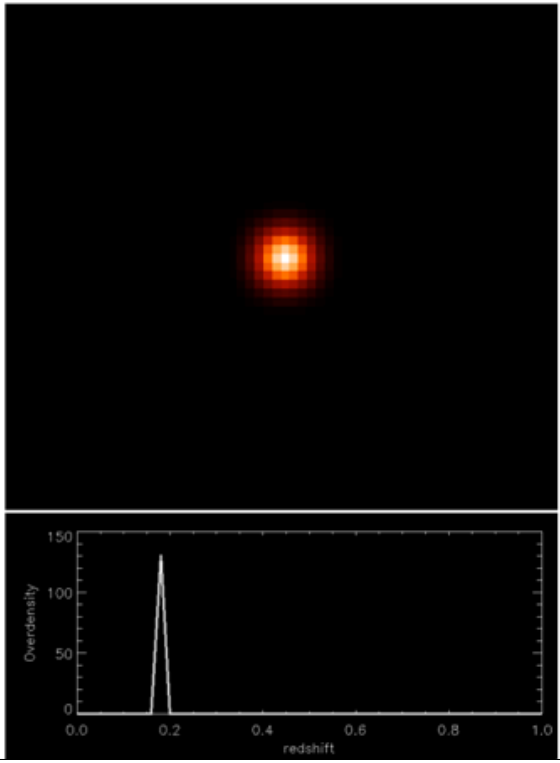
Full 3D Weak Lensing

$$\min_{\alpha} \parallel \alpha \parallel_1 \quad s.t. \ \frac{1}{2} \parallel \gamma - Q\Phi\alpha \parallel^2_{\mathbf{\Sigma}^{-1}} \leq \epsilon$$

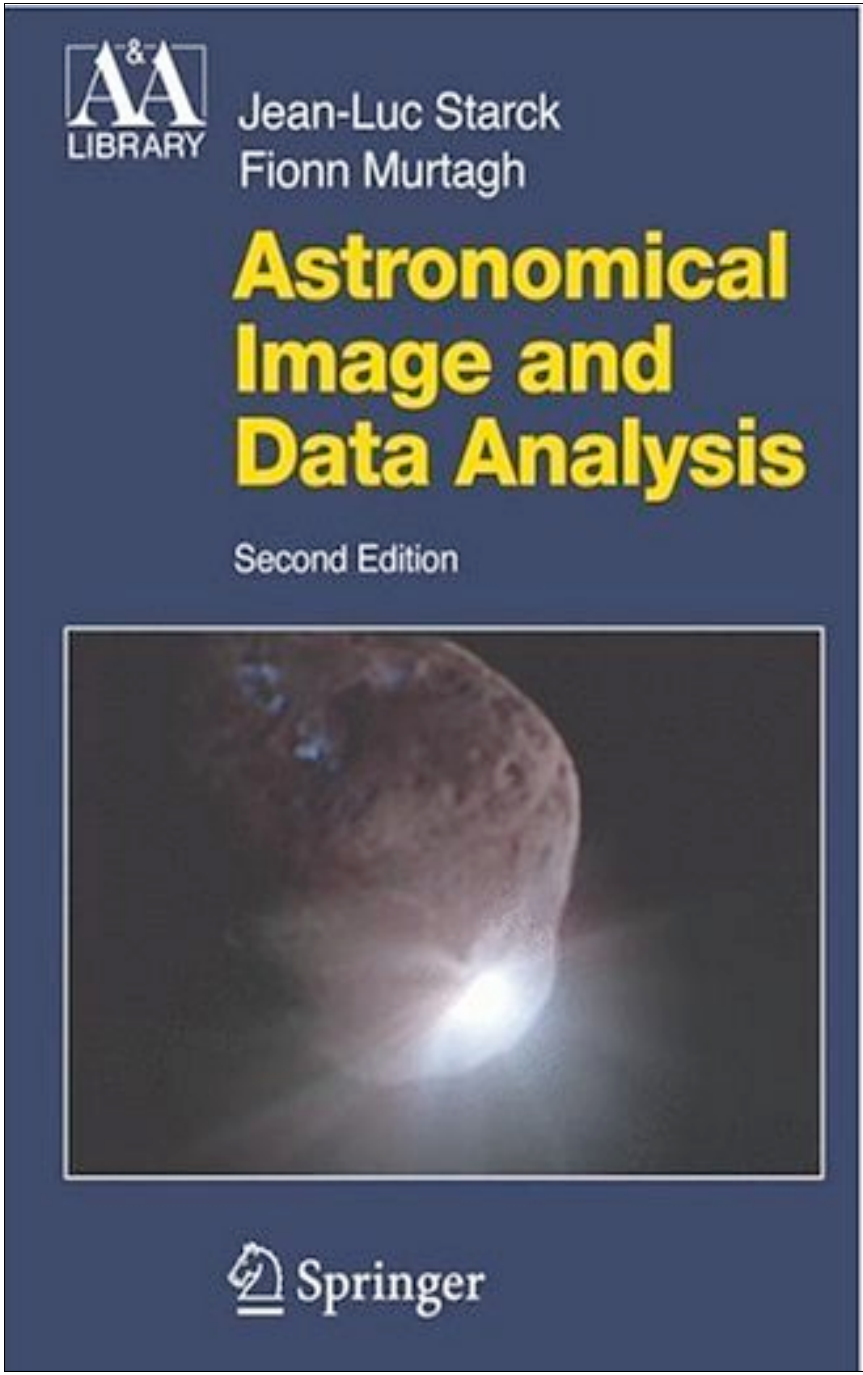
$\delta = \Phi\alpha$ Φ = 2D Wavelet Transform on each redshift bin

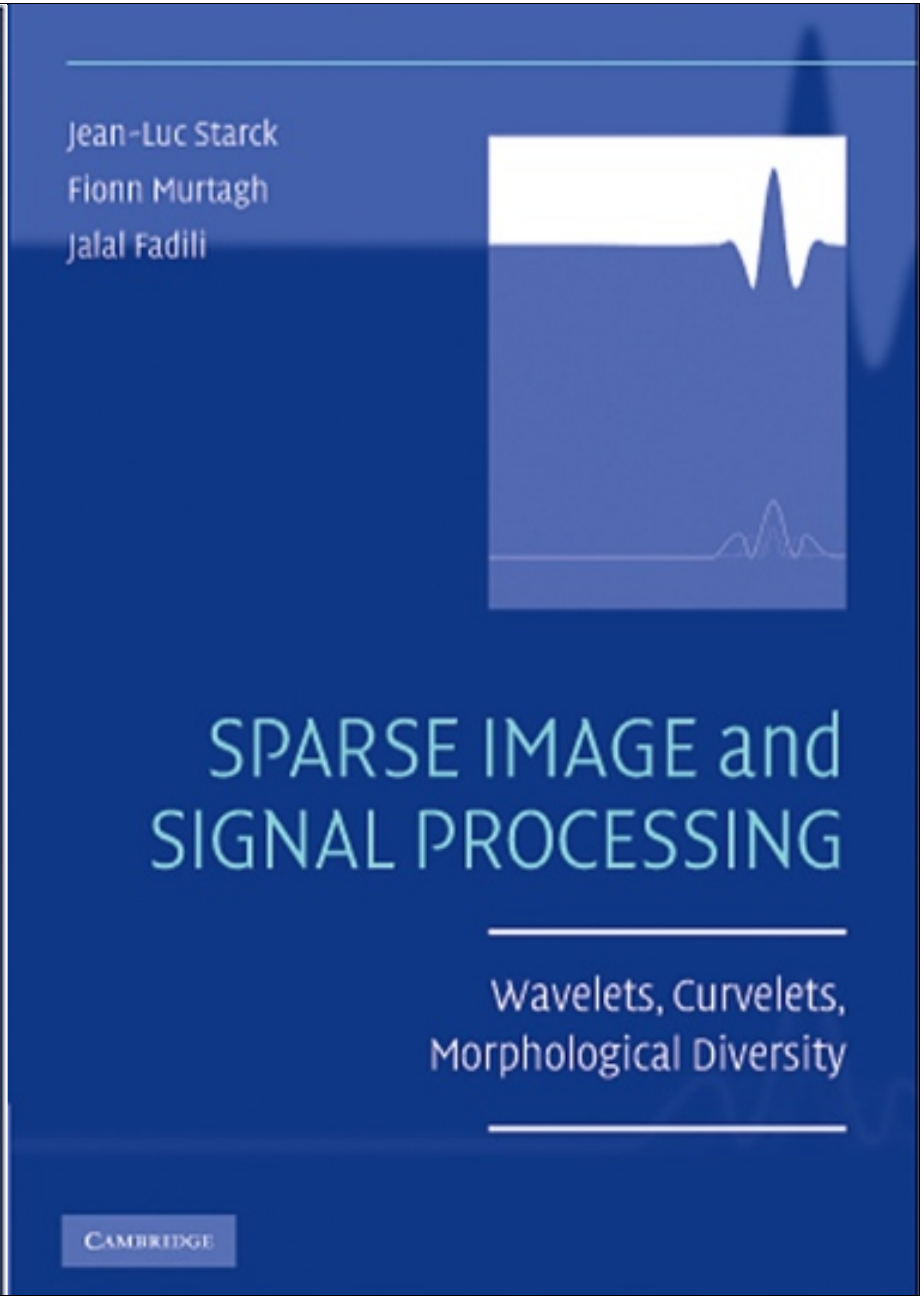
Example of the reconstruction of a cluster at redshift z=0.2 using 5 times more reconstruction bin:





- Sparsity is very efficient for
  - Inverse problems (denoising, deconvolution, etc).
  - Inpainting
  - Component Separation (LOFAR, WMAP, PLANCK).
- Be very careful with Bayesian interpretation.
- Perspectives
  - CMB
  - Weak lensing
    - Test the 3D reconstruction algorithm on a simulated weak lensing survey from nbody simulations.
    - Apply the algorithm to real data (COSMOS, CFHTLS) with all the added fun (non-Gaussian noise, photometric redshift errors, missing data...)


A&A LIBRARY
Jean-Luc Starck
Fionn Murtagh
Astronomical Image and Data Analysis
Second Edition
Springer


Jean-Luc Starck
Fionn Murtagh
Jalal Fadili
SPARSE IMAGE and SIGNAL PROCESSING
Wavelets, Curvelets, Morphological Diversity
CAMBRIDGE