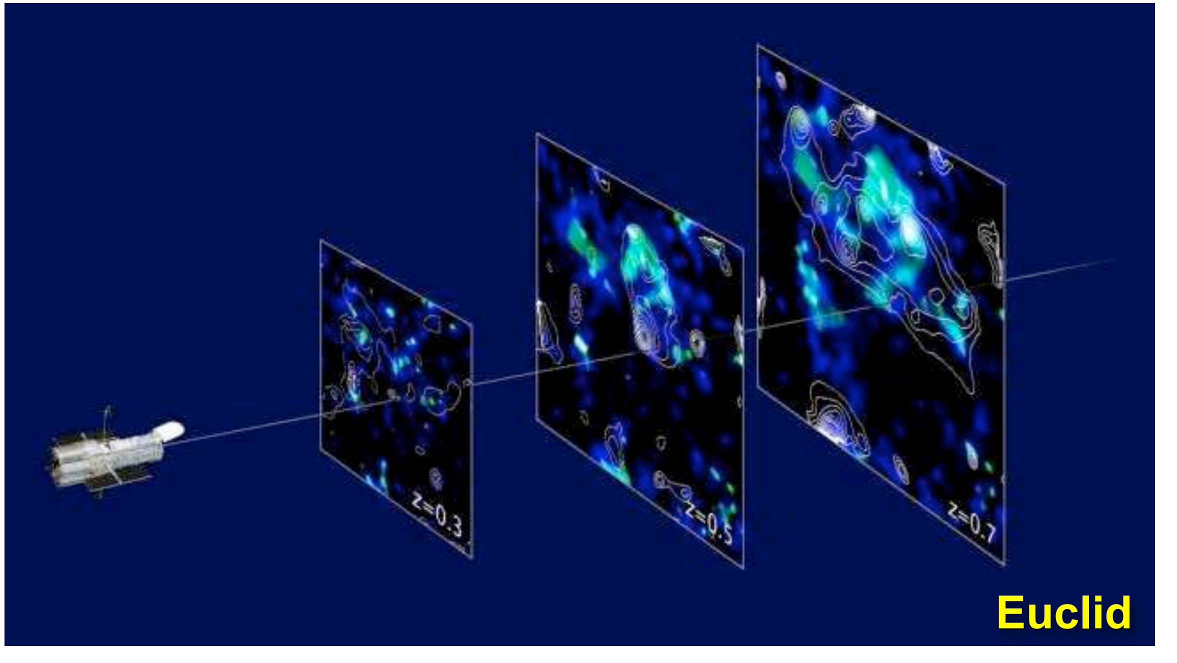
Sparsity in Astrophysics: from Wavelets to Compressed Sensing

J.-L. Starck



What is a good representation for data?

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

Dictionary
(basis, frame)

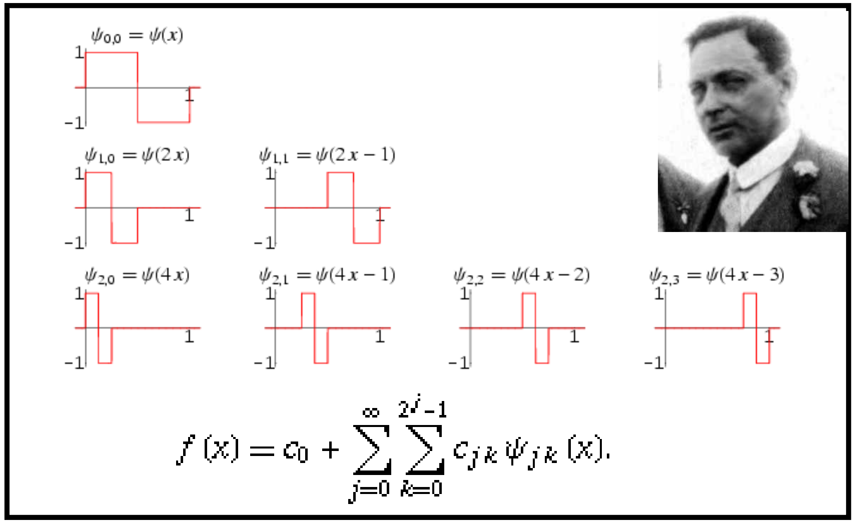
$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

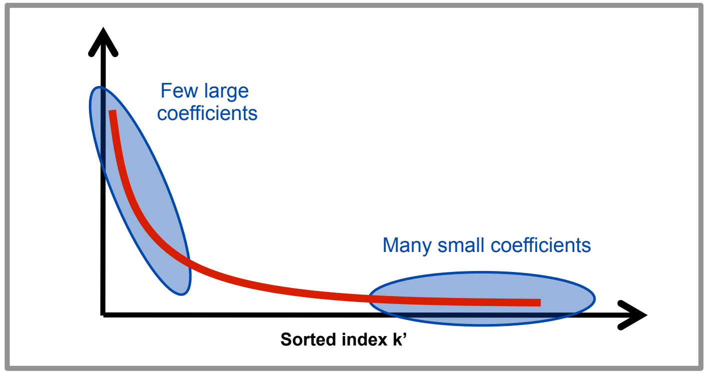
Atoms

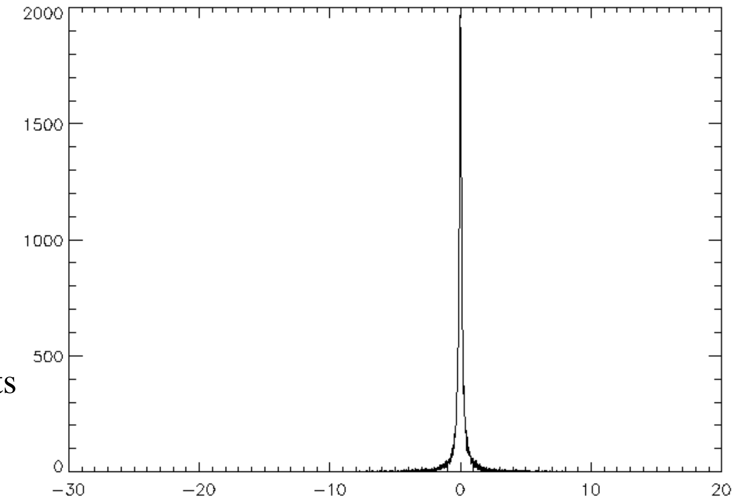
$$s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi\alpha$$

coefficients

Ex: Haar wavelet



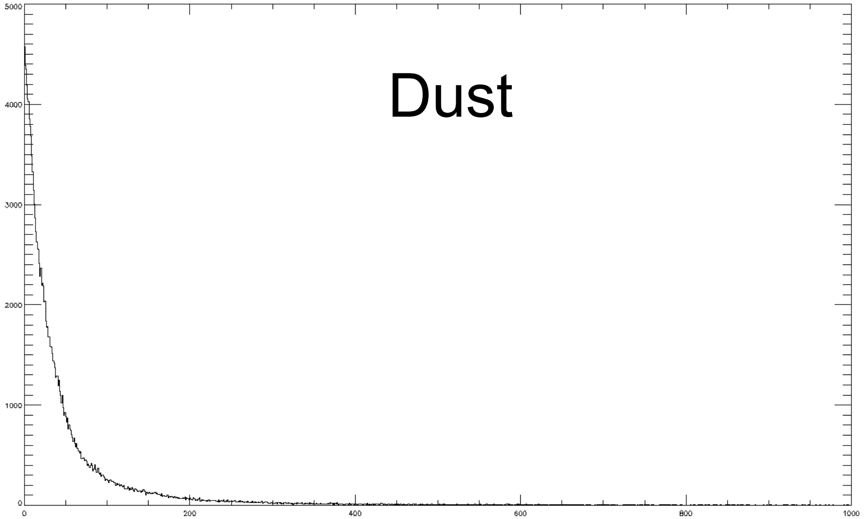


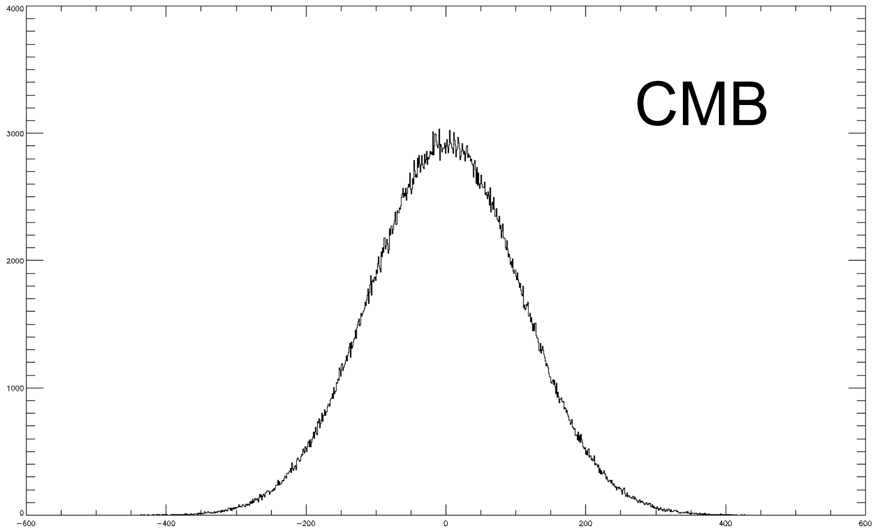


- Fast calculation of the coefficients
- Analyze the signal through the statistical properties of the coefficients
- Approximation theory uses the sparsity of the coefficients

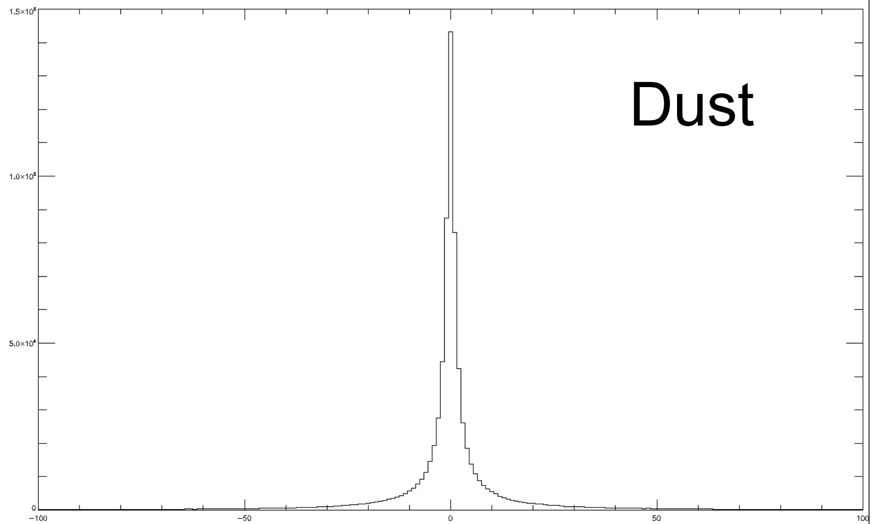
Histogram

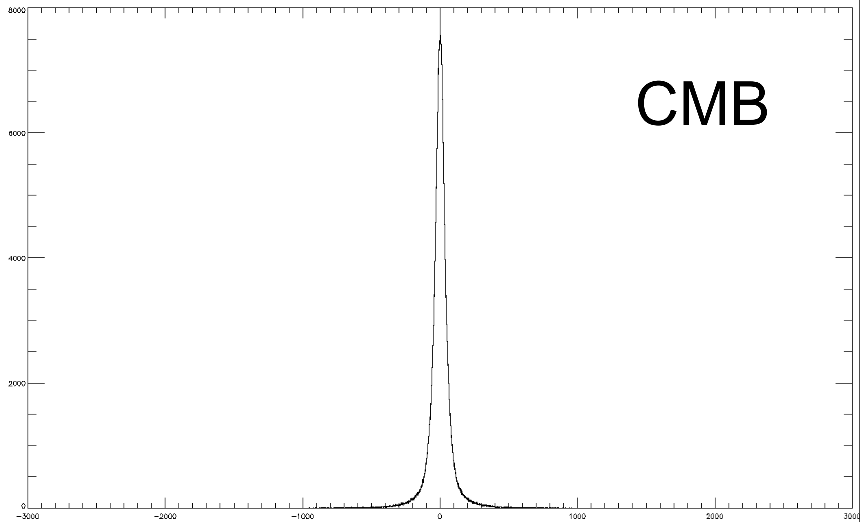
Spatial Domain

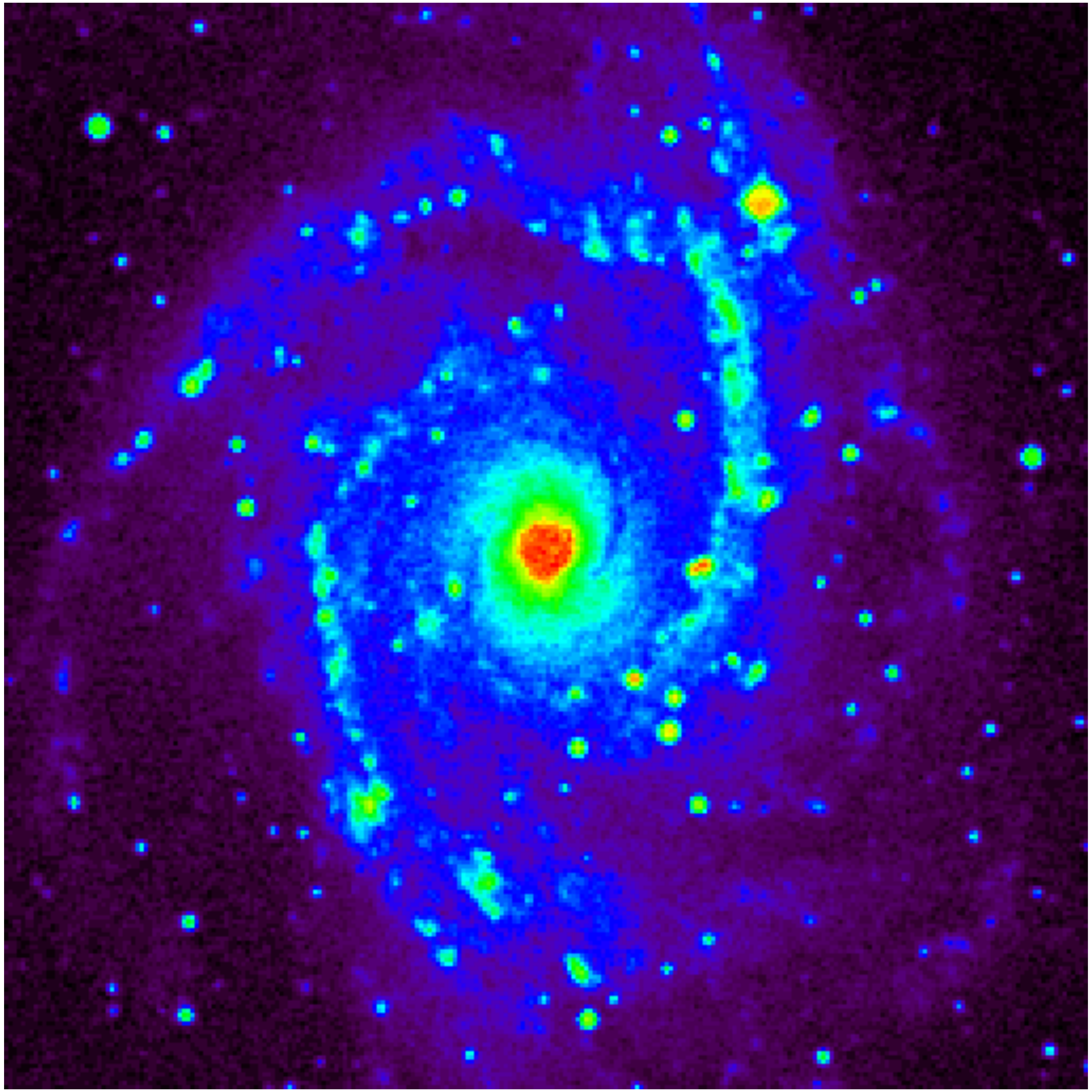




Wavelet Domain







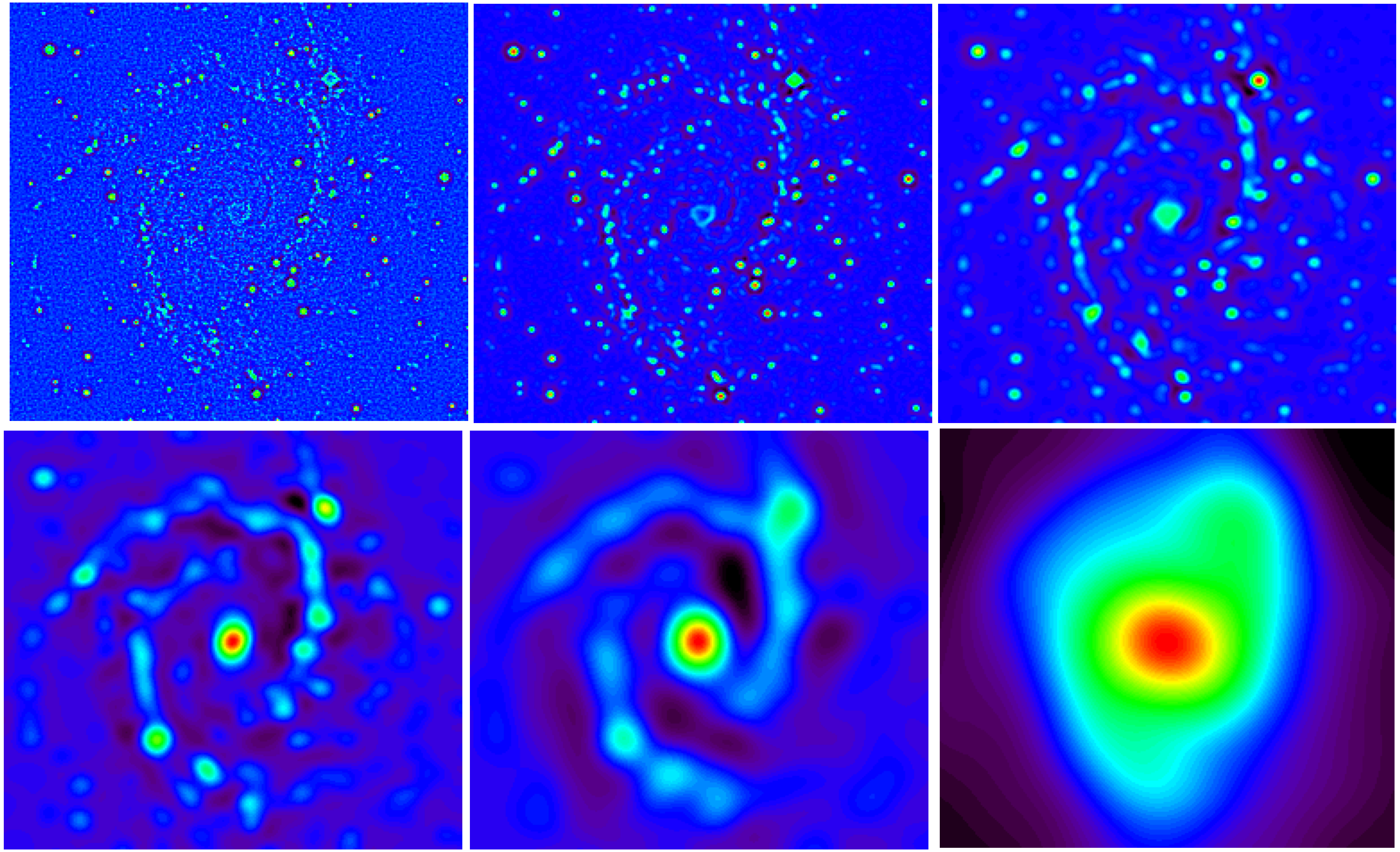
The STARLET Transform

Isotropic Undecimated Wavelet Transform (a trous algorithm)

$\varphi = B_3 - \text{spline}, \quad \frac{1}{2}\psi(\frac{x}{2}) = \frac{1}{2}\varphi(\frac{x}{2}) - \varphi(x)$

$$I(k,l) = c_{J,k,l} + \sum_{j=1}^{J} w_{j,k,l}$$

$h = [1,4,6,4,1]/16, \quad \mathrm{g} = \delta - \mathrm{h}, \quad \tilde{\mathrm{h}} = \tilde{\mathrm{g}} = \delta$



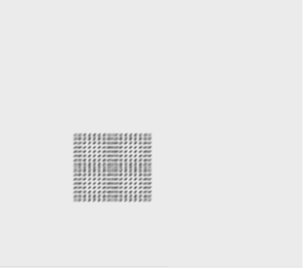
Sparsity Model 1: we consider a dictionary $\Phi = \{\phi_1, \ldots, \phi_K\}$ which has a fast transform/reconstruction operator:

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

<u>Local DCT</u>

Stationary textures

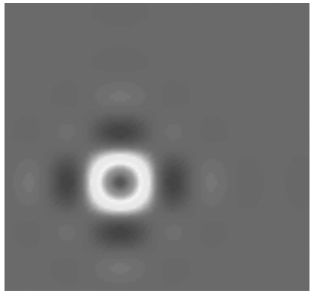
Locally oscillatory



<u>Wavelet transform</u>

Piecewise smooth

Isotropic structures



<u>Curvelet transform</u>

Piecewise smooth,
edge

How to measure sparsity ?

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

Formally, the sparsest coefficients are obtained by solving the optimization problem:

(P0) Minimize $\|\alpha\|_0$ subject to $s = \phi\alpha$

It has been proposed (*to relax and*) to replace the l_0 norm by the l_1 norm (Chen, 1995):

(P1) Minimize $\|\alpha\|_1$ subject to $s = \phi\alpha$

It can be seen as a kind of convexification of (P0).

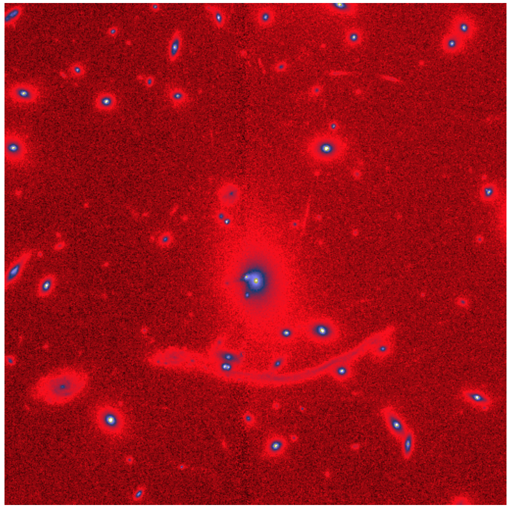
It has been shown (Donoho and Huo, 1999) that for certain dictionary, if there exists a highly sparse solution to (P0), then it is identical to the solution of (P1).

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 :

$$s_k = \Phi_k \alpha_k$$

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

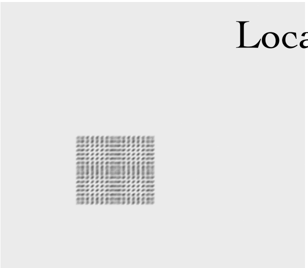
Sparsity Model 1: we consider a dictionary $\Phi = \{\phi_1, \dots, \phi_K\}$ which has a fast transform/reconstruction operator:

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

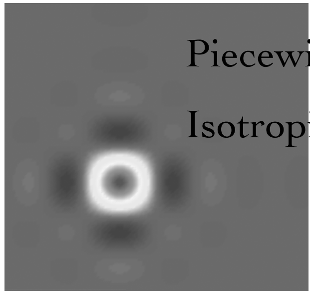
Local DCT

Stationary textures

Locally oscillatory



Wavelet transform



Piecewise smooth

Isotropic structures

Curvelet transform

Piecewise smooth,
edge

Sparsity Model 2: Morphological Diversity:

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

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 (fixed dictionary) : extremely fast.

Advantages of model 2 (union of fixed dictionaries):
- 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 (dictionary learning):
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

INVERSE PROBLEMS

$$Y = HX + N$$

$X = \Phi\alpha$, and α is sparse

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

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

Denoising using a sparsity model

$$Y = X + N$$

Denoising using a sparsity prior on the solution:

X is sparse in Φ, i.e. $X = \Phi\alpha$ where most of α are negligible.

$$\tilde{\alpha} \in \arg\min_{\alpha} \frac{1}{2} \parallel Y - \Phi\alpha \parallel^2 + t \parallel \alpha \parallel_p^p, \quad 0 \leq p \leq 1.$$

p=0

$$\tilde{\alpha} \quad \in \arg\min_{\alpha} \frac{1}{2} \parallel Y - \Phi\alpha \parallel^2 + \frac{t^2}{2} \parallel \alpha \parallel_0$$

==> Solution via Iterative **Hard** Thresholding

$$\tilde{\alpha}^{(t+1)} = \mathrm{HardThresh}_{\mu t}(\tilde{\alpha}^{(t)} + \mu\Phi^T(Y - \Phi\tilde{\alpha}^{(t)})), \mu = 1/\left\|\Phi\right\|^2.$$

$$\tilde{\alpha}_{j,k} = \mathrm{HardThresh}_t(\alpha_{j,k}) = \begin{cases} \alpha_{j,k} & \text{if } \mid \alpha_{j,k} \mid \geq t, \\ 0 & \text{otherwise.} \end{cases}$$

1st iteration solution:

$$\tilde{X} = \Phi \ \mathrm{HardThresh}_t(\Phi^T Y) = \Delta_{\Phi,t}(Y)$$

Exact for Φ orthonormal.

p=1

$$\tilde{\alpha} \quad = \arg\min_{\alpha} \frac{1}{2} \parallel Y - \Phi\alpha \parallel^2 + t \parallel \alpha \parallel_1$$

==> Solution via iterative **Soft** Thresholding

$$\tilde{\alpha}^{(t+1)} = \mathrm{SoftThresh}_{\mu t}(\tilde{\alpha}^{(t)} + \mu\Phi^T(Y - \Phi\tilde{\alpha}^{(t)})), \mu \in (0, 2/\left\|\Phi\right\|^2).$$

$$\tilde{\alpha}_{j,k} = \mathrm{SoftThresh}_t(\alpha_{j,k}) = \mathrm{sign}(\alpha_{j,k})(\mid \alpha_{j,k} \mid - t)_+$$

1st iteration solution:

$$\tilde{X} = \Phi \ \mathrm{SoftThresh}_t(\Phi^T Y) = \Delta_{\Phi,t}(Y)$$

Exact for Φ orthonormal.

Inverse Problems and Iterative Thresholding Minimizing Algorithm

Iterative thresholding with a varying threshold was proposed in (Starck et al, 2004; Elad et al, 2005) for sparse signal decomposition in order to accelerate the convergence. The idea consists in using a different threshold $\lambda^{(n)}$ at each iteration.

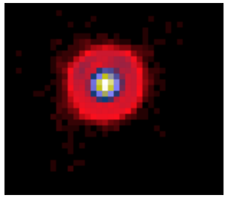
For IST: $$\alpha^{(n+1)} = \mathrm{HT}_{\lambda^{(n)}}\left(\alpha^{(n)} + \Phi^T A^T \left(Y - A\Phi\alpha^{(n)}\right)\right)$$

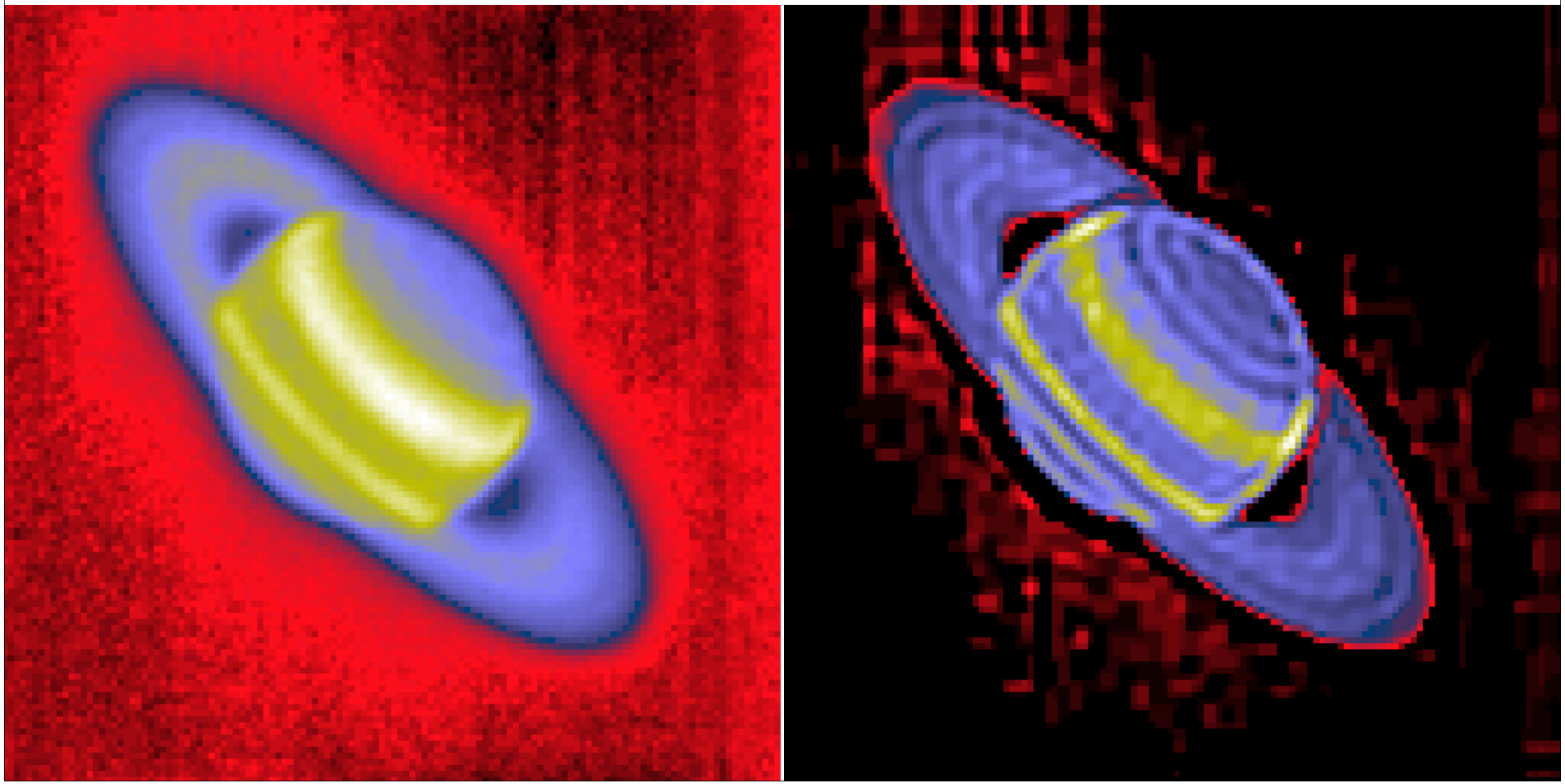
For IHT: $$\alpha^{(n+1)} = \mathrm{ST}_{\lambda^{(n)}}\left(\alpha^{(n)} + \Phi^T A^T \left(Y - A\Phi\alpha^{(n)}\right)\right)$$

More Refs: Vonesch et al, 2007; Elad et al 2008; Wright et al., 2008; Nesterov, 2008 and Beck-Teboulle, 2009; Blumensath, 2008; Maleki et Donoho, 2009 ; Dupe et al, 2010; Dupe et al 2011; Geyre et al, 2011, etc.

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.





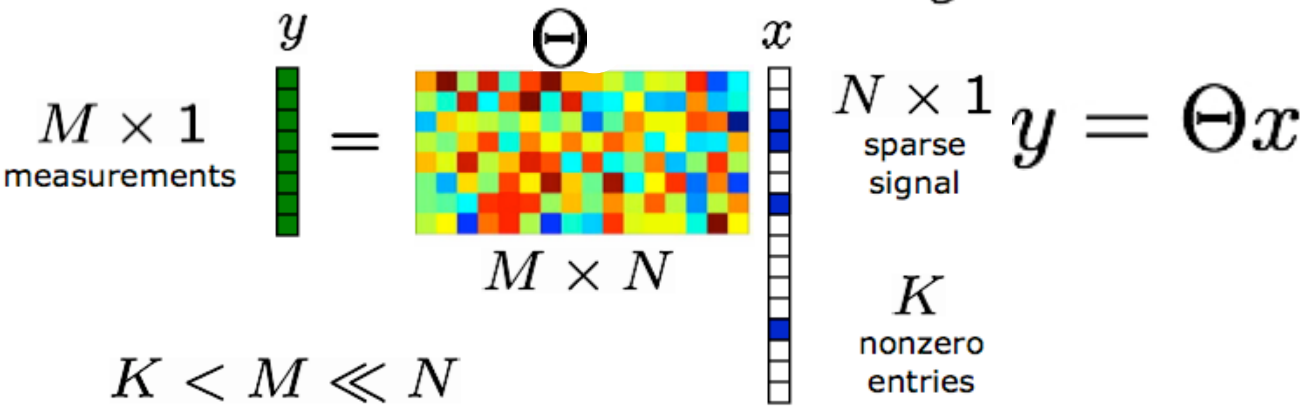
Compressed Sensing

* 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* $y = \Theta x$

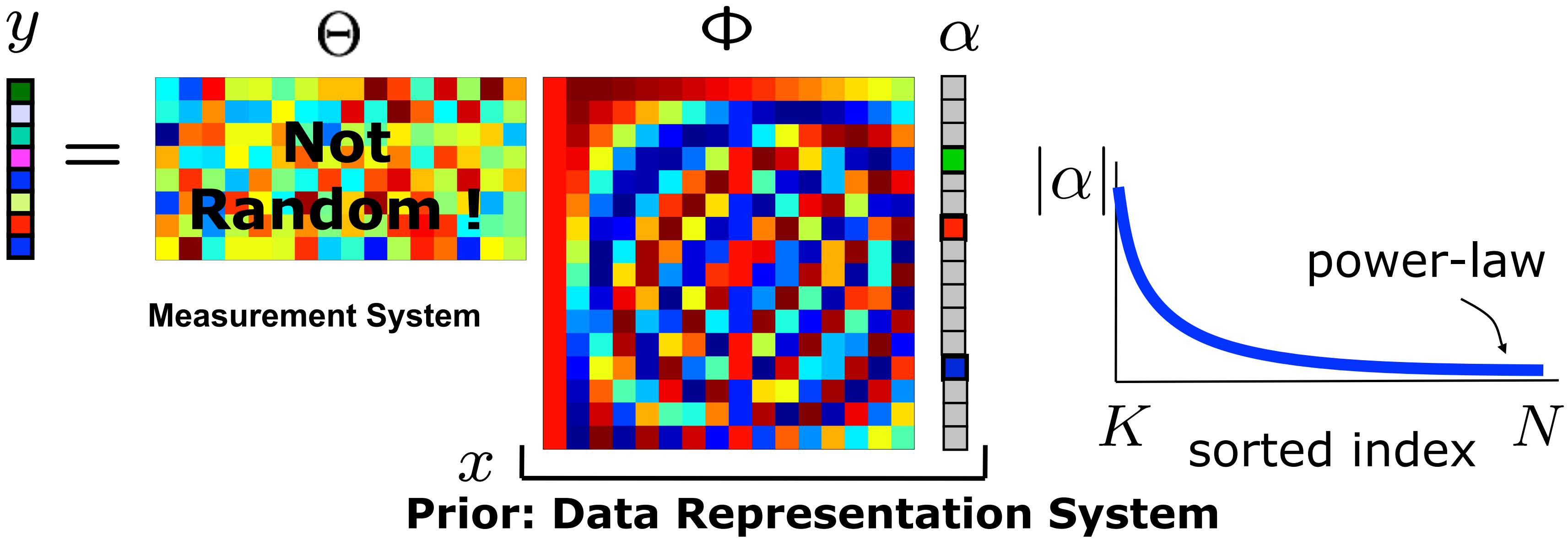


Reconstruction via non linear processing:

$$\min_x \|x\|_1 \quad \text{s.t.} \quad y = \Theta x$$

Soft Compressed Sensing Definition

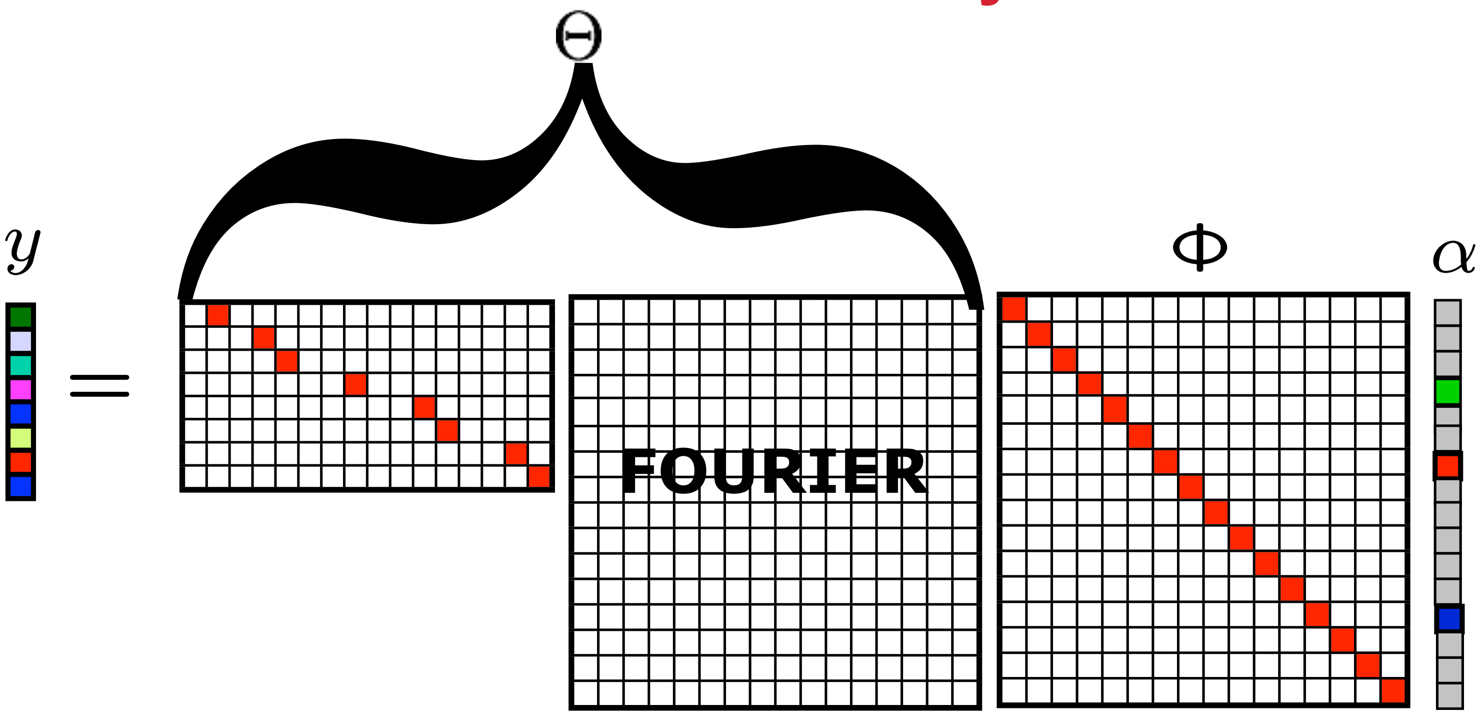
$$Y = \Theta X = \Theta \Phi \alpha$$



Mutual coherence: $\mu_{\Theta,\Phi} = \max_{i,k} \left| \left\langle \Theta_i, \Phi_k, \right\rangle \right|$

Mutual coherence the degree of similarity between the sparsity and measurement systems.

Radio-Interferometry



Measurement System

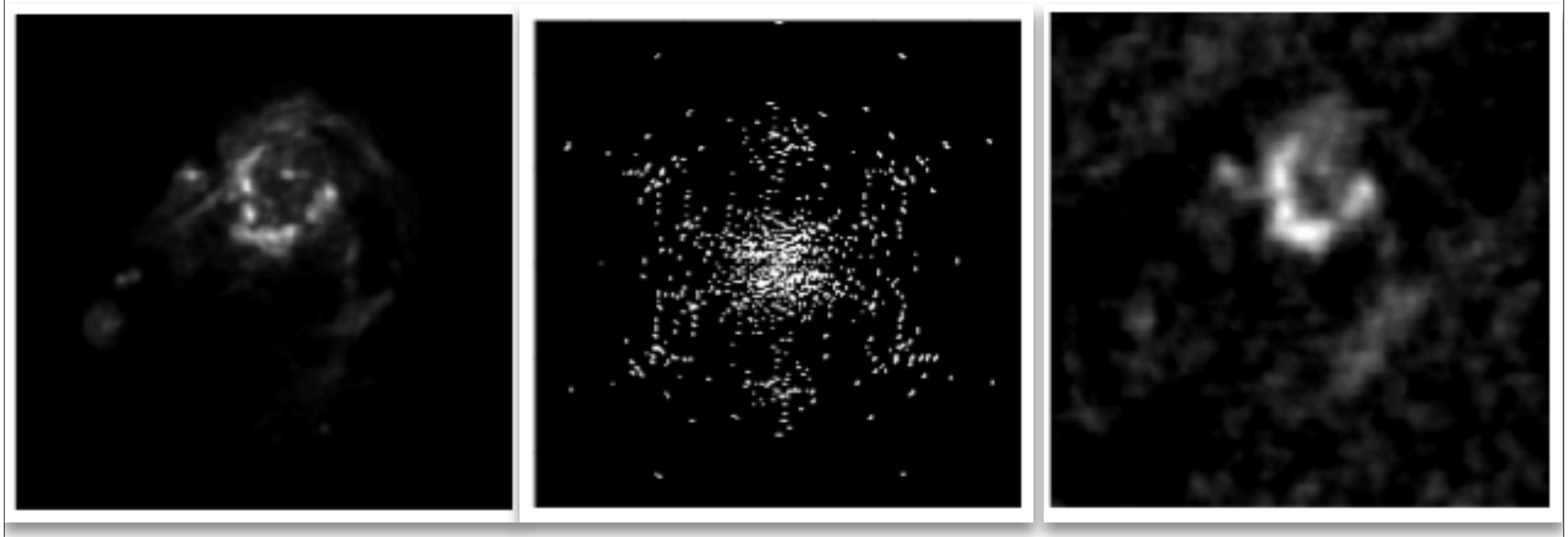
CLEAN Algorithm: $\Phi = \mathrm{Id}$ (Hogbom, 1974)

Wavelet Clean Φ = Wavelet Transform (Wakker, and Schwarz, 1988; Starck et al 1994)

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

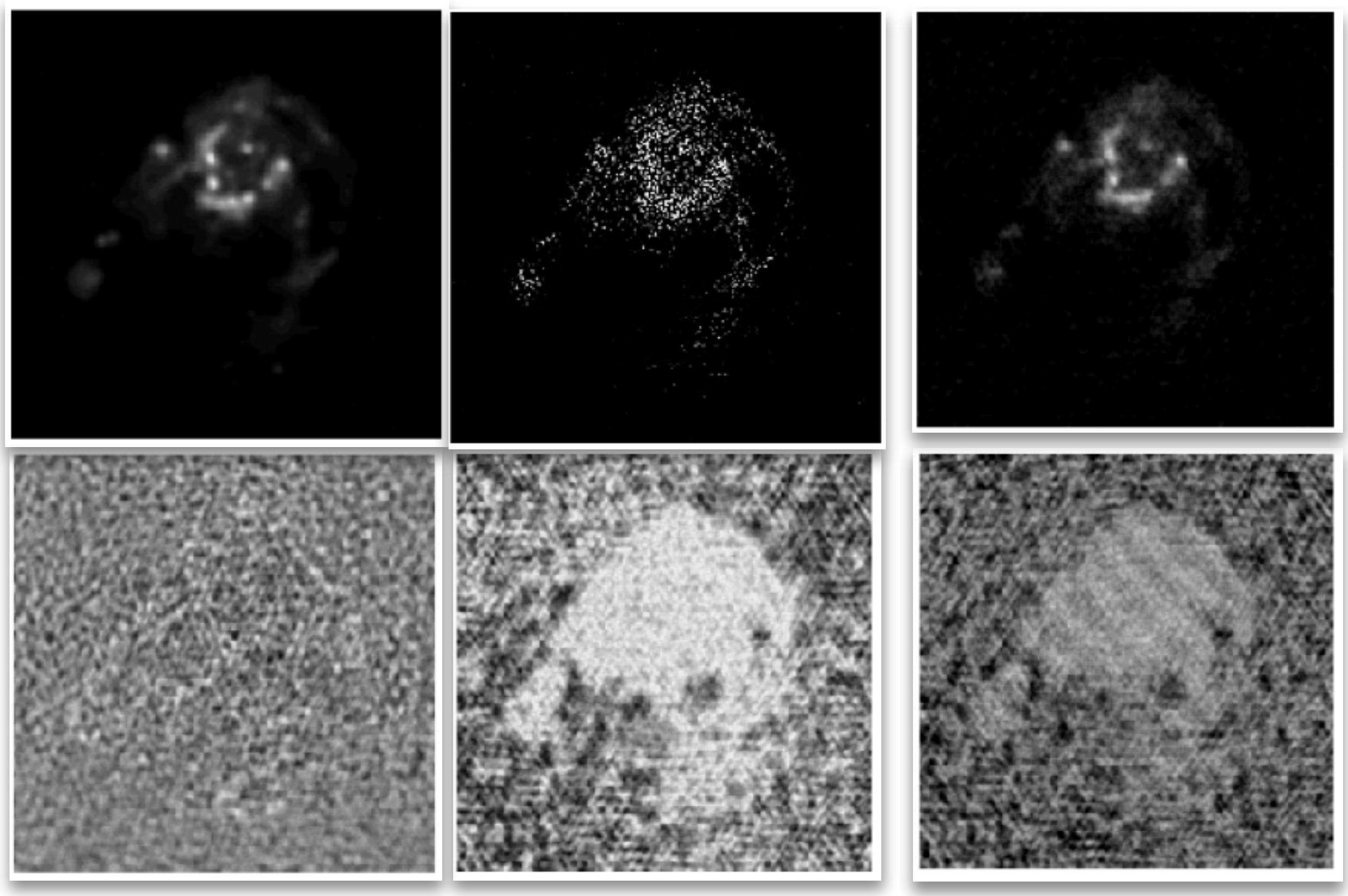
CS-Radio Astronomy

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



Australian Square Kilometer Array Pathfinder (ASKAP) radio telescope.

CS-Radio Astronomy



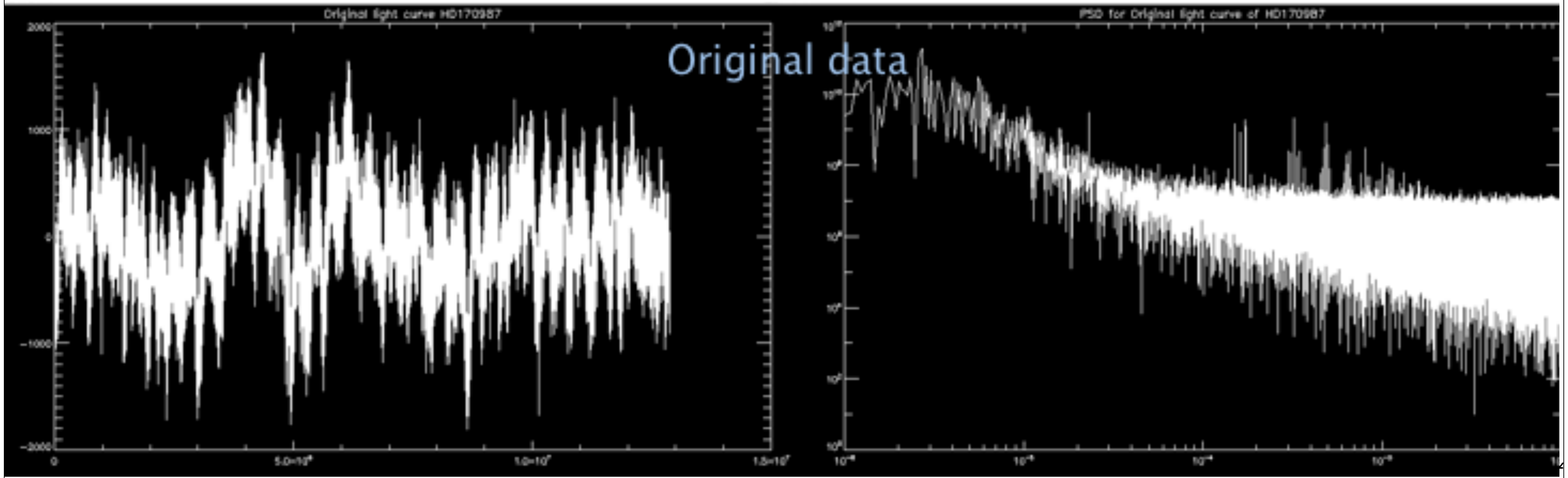
Hogbom CLEAN

MEM residual

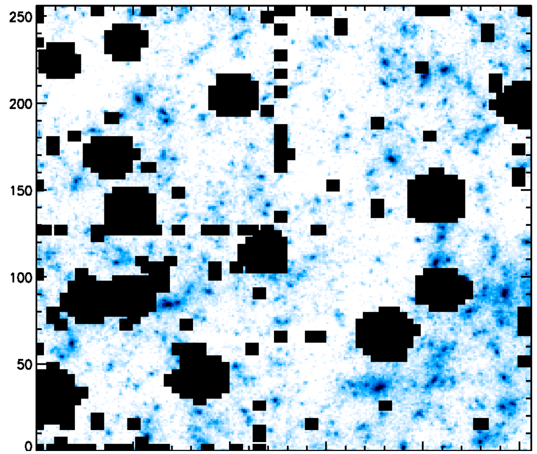
Missing Data

- Period detection in temporal series

COROT: HD170987

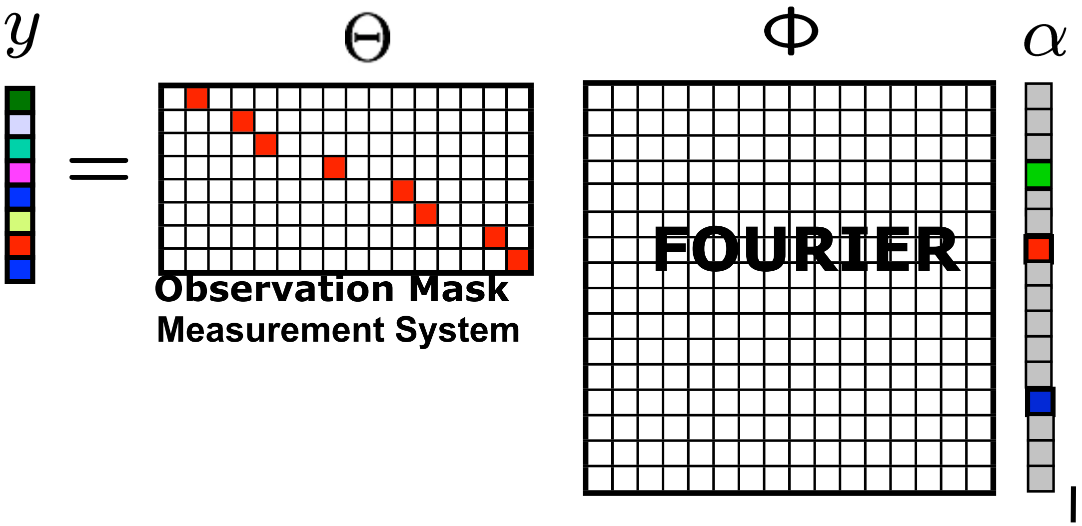


- Bad pixels, cosmic rays, point sources in 2D images, ...

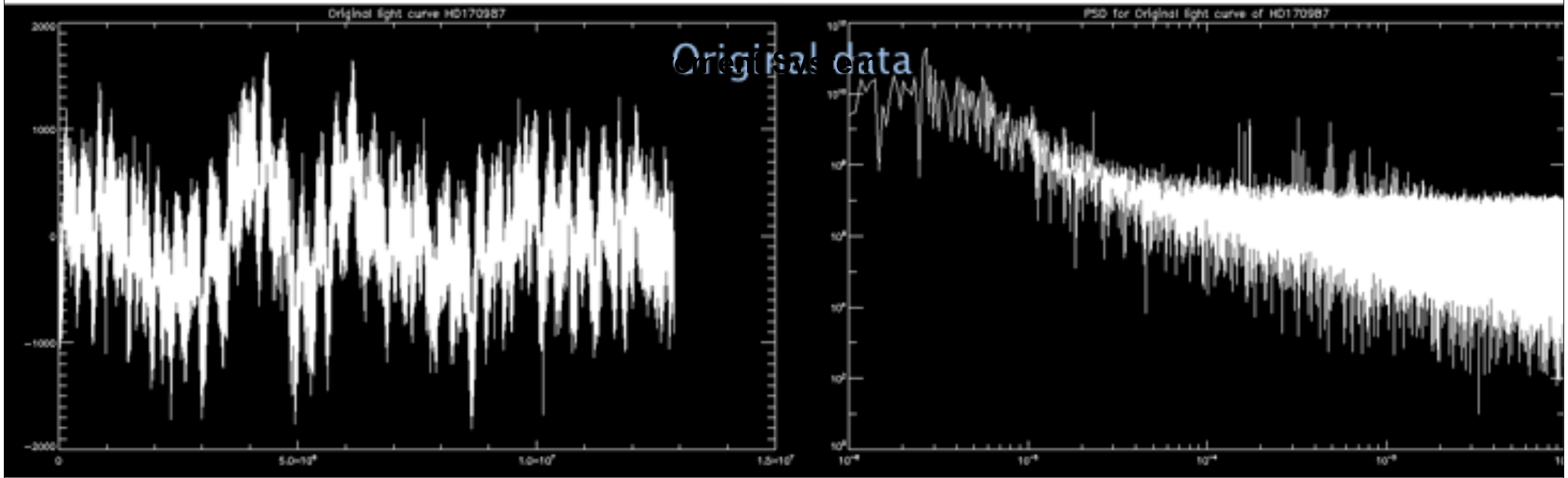


Missing Data

Period detection in temporal series

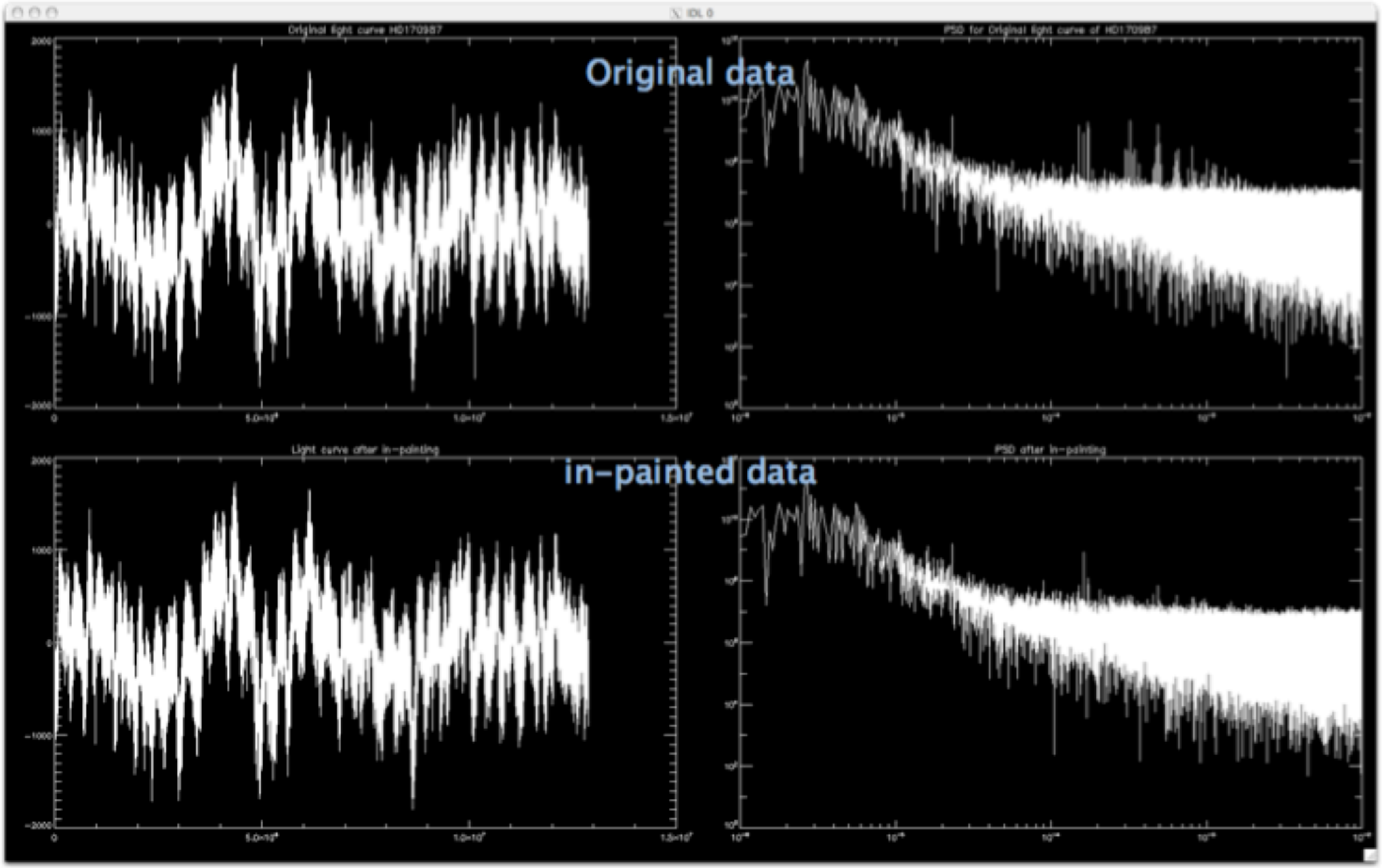


COROT: HD170987



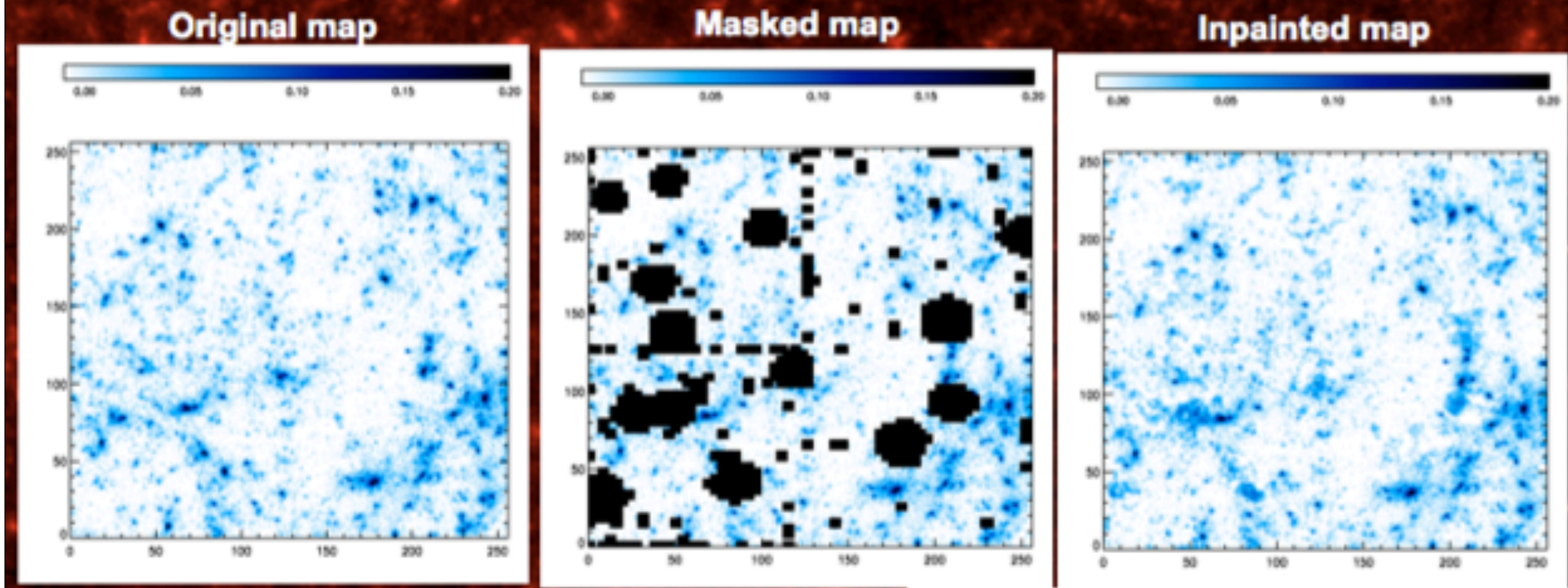
COROT: HD170987 with in-painting

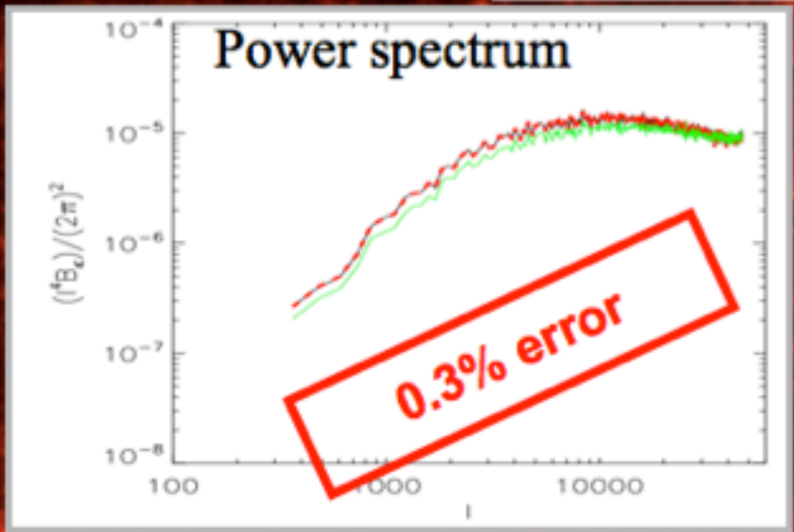
arXiv:1003.5178

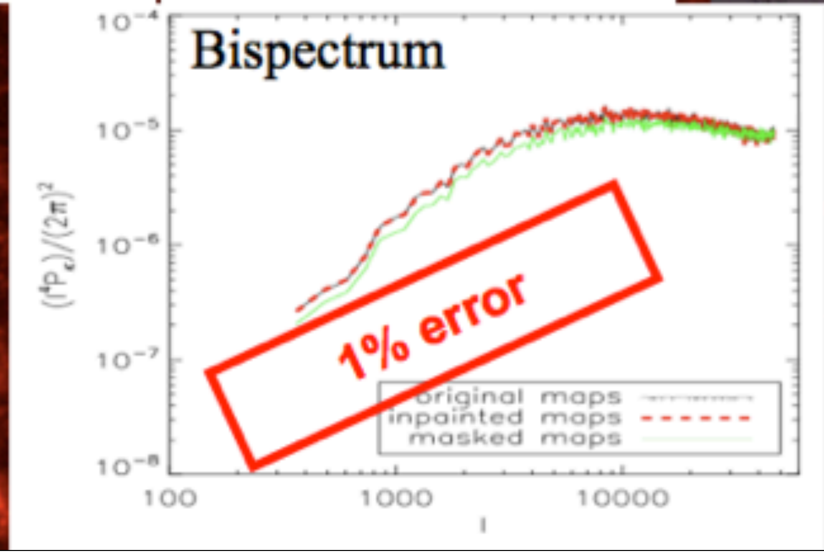


Inpainting :

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.

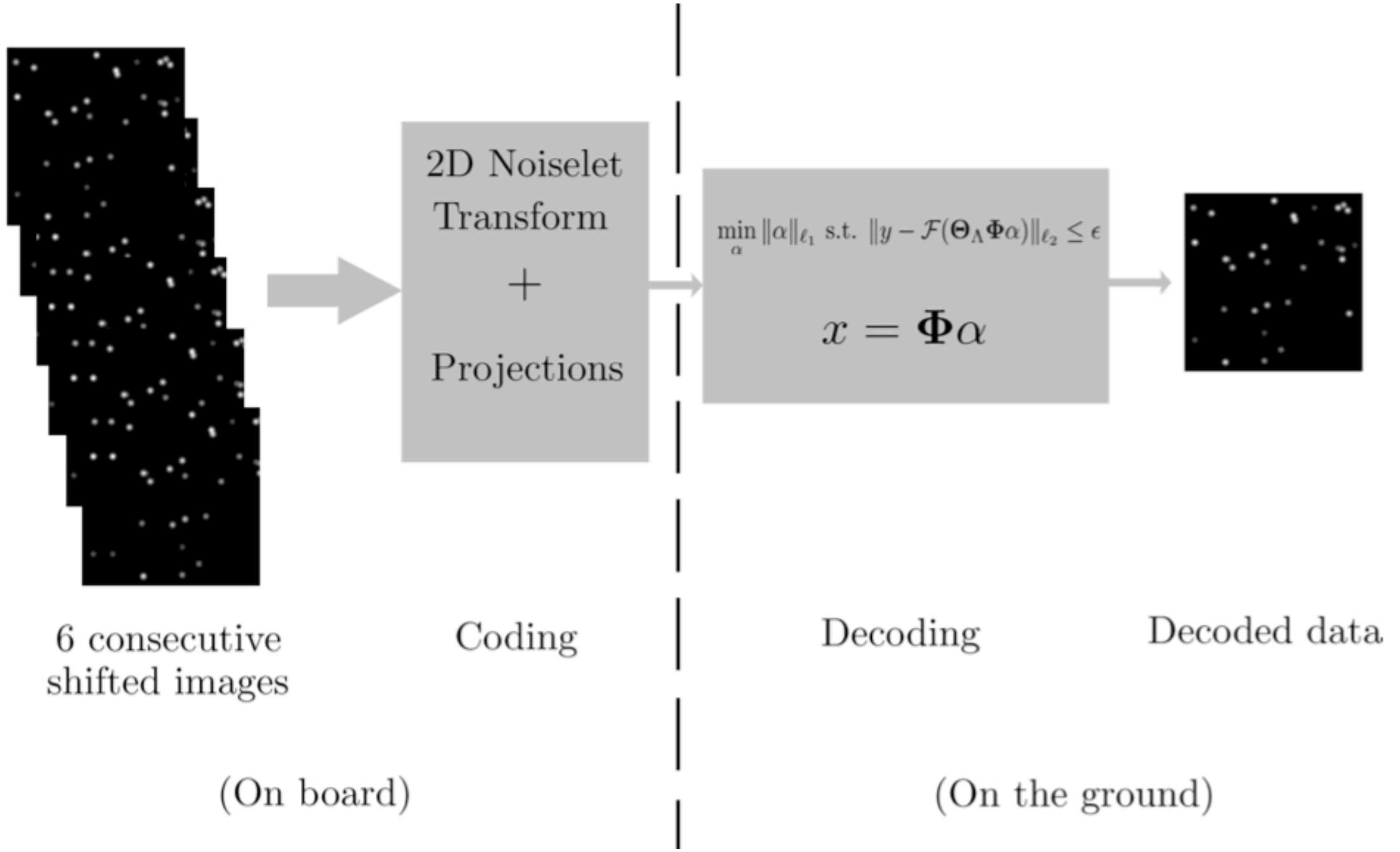






Transfering Spatial Data to the Earth

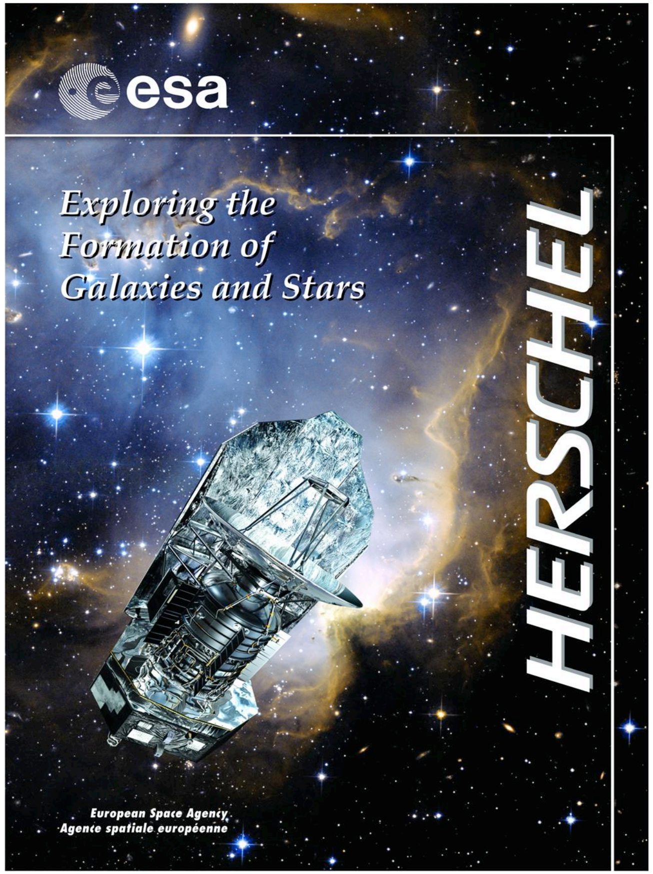
Bobin, J.-L. Starck, and R. Ottensamer, "Compressed Sensing in Astronomy", IEEE Journal of Selected Topics in Signal Processing, Vol 2, no 5, pp 718--726, 2008.

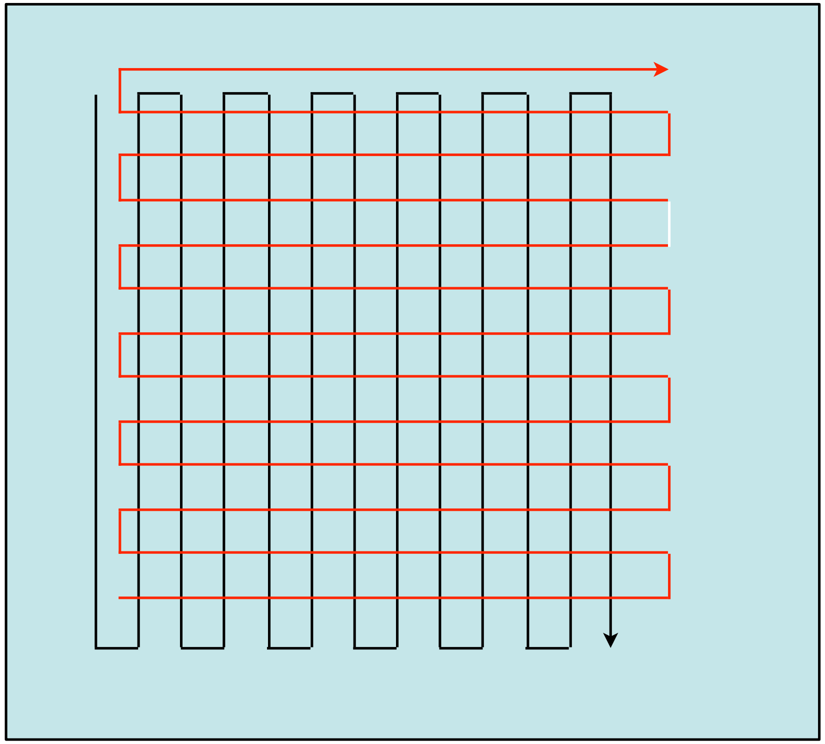


Transfering Data to the Earth

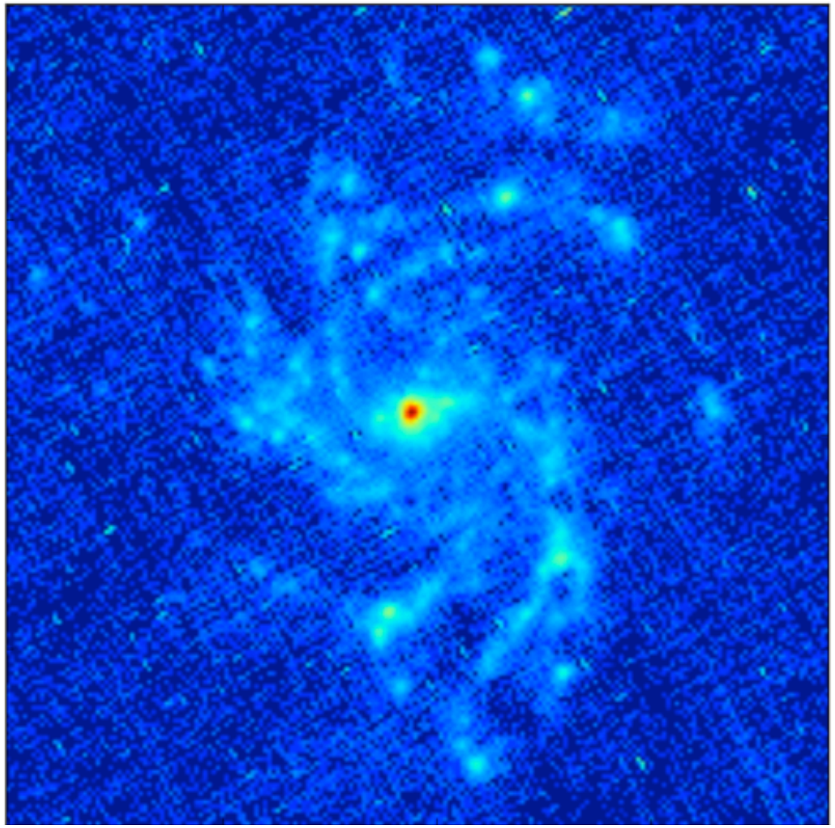
Observed Herschel Data During the Calibration Phase, November 2010, without any compression.

A scan (16 x 16 pixels at 40 Hz during 25 min each, we obtained 60000 images.

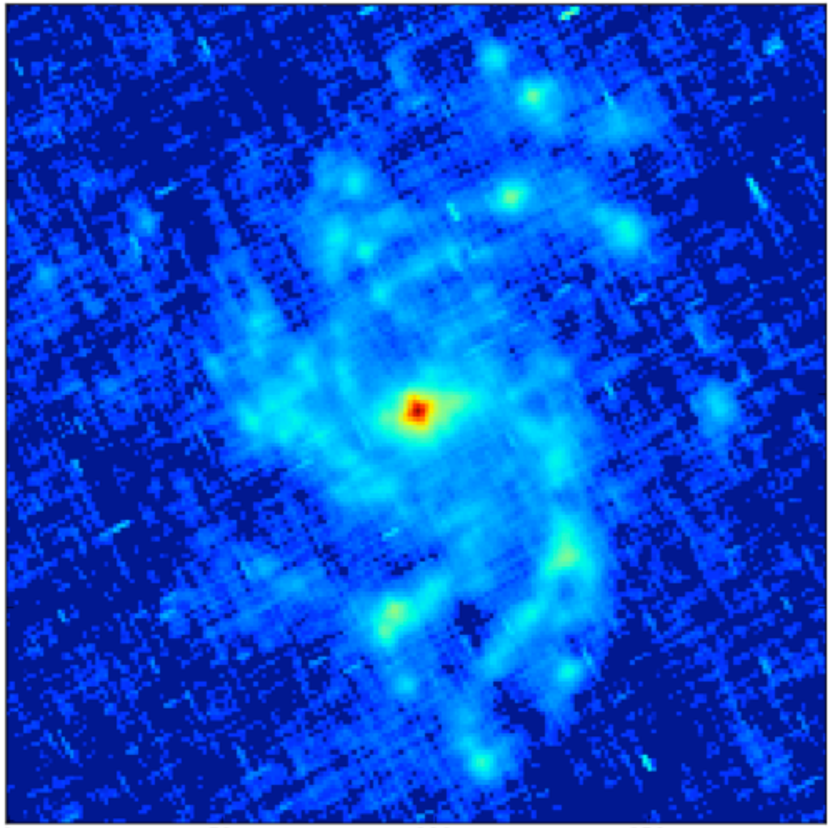




Map from Uncompressed Data



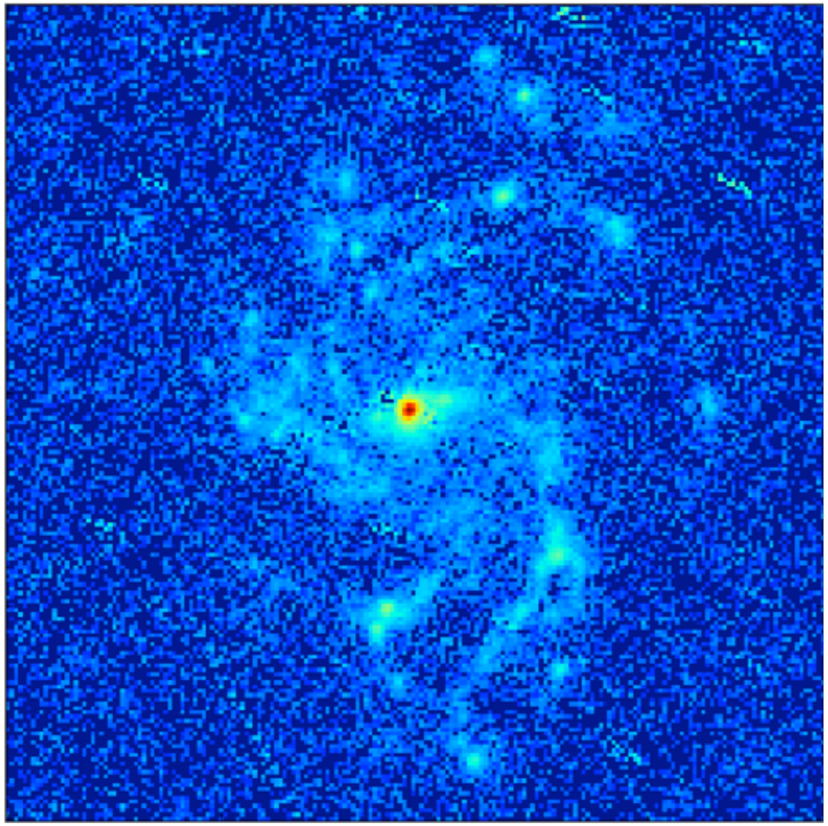
Official Pipeline Reconstruction: Averaging

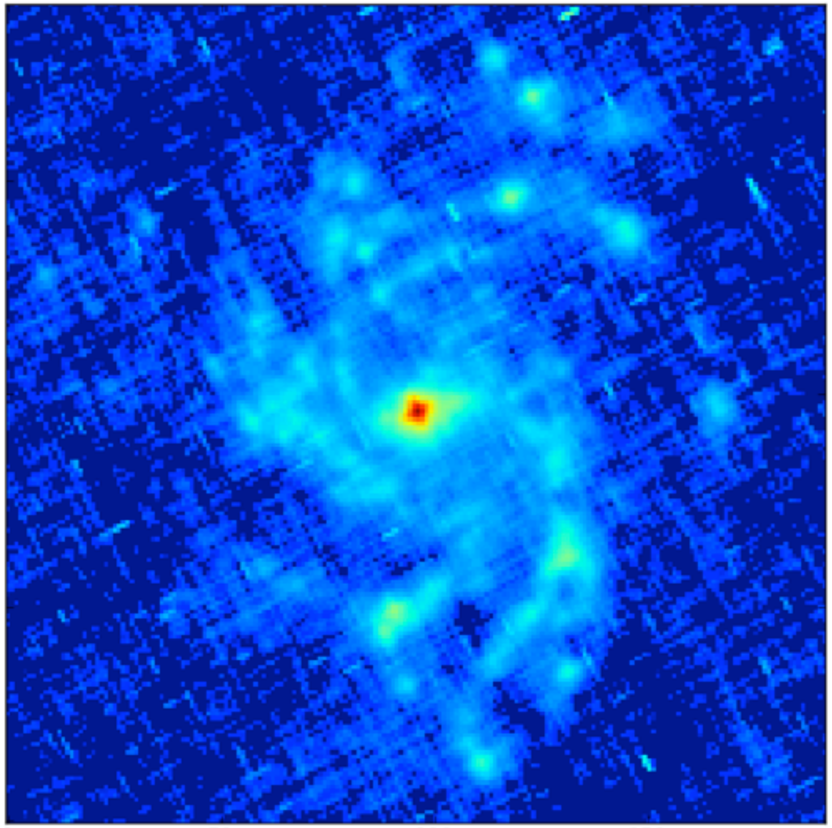


N. Barbey, M.Sauvage, J.-L. Starck, and R. Ottensamer, "Feasibility and performances of compressed-sensing and sparse map-making with Herschel/PACS data", **Astronomy and Astrophysics**, 527, 102 , 2011.

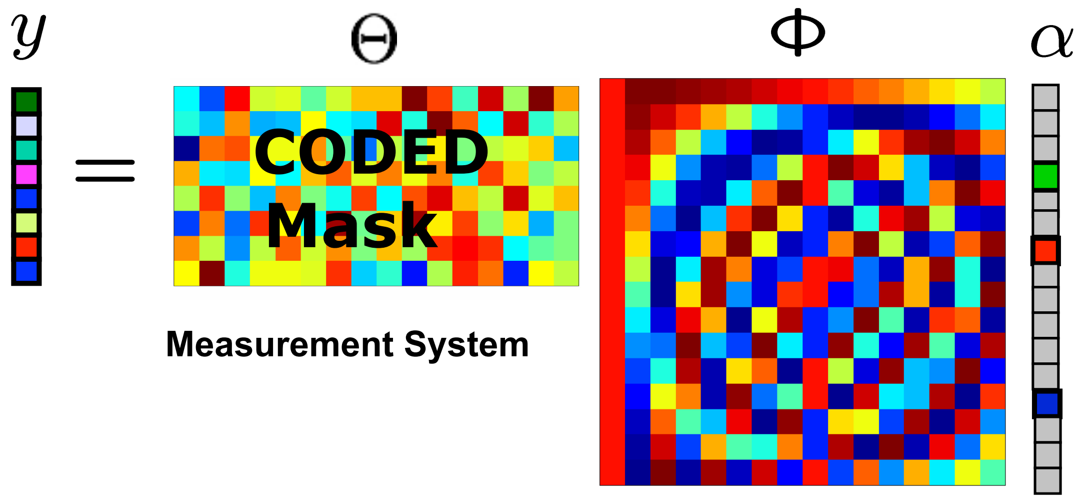
Compressed Sensing Reconstruction

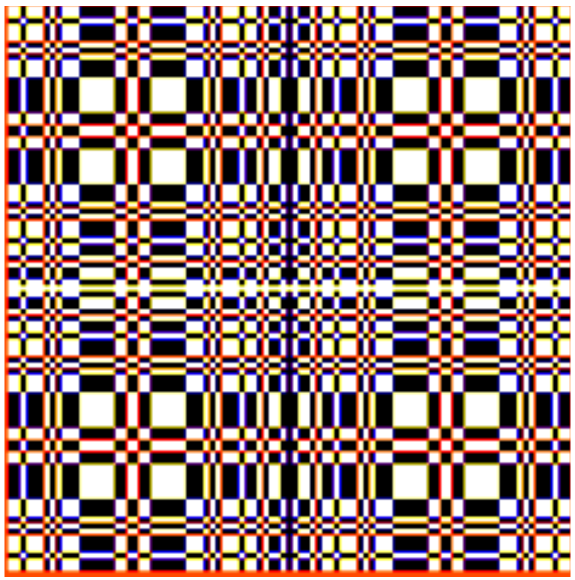
Official Pipeline Reconstruction: Averaging



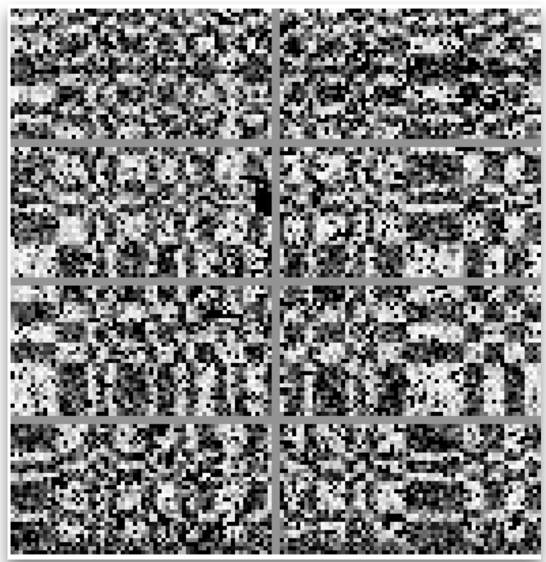


Gamma Ray Instruments (Integral) - Acquisition with coded masks





INTEGRAL/IBIS Coded Mask



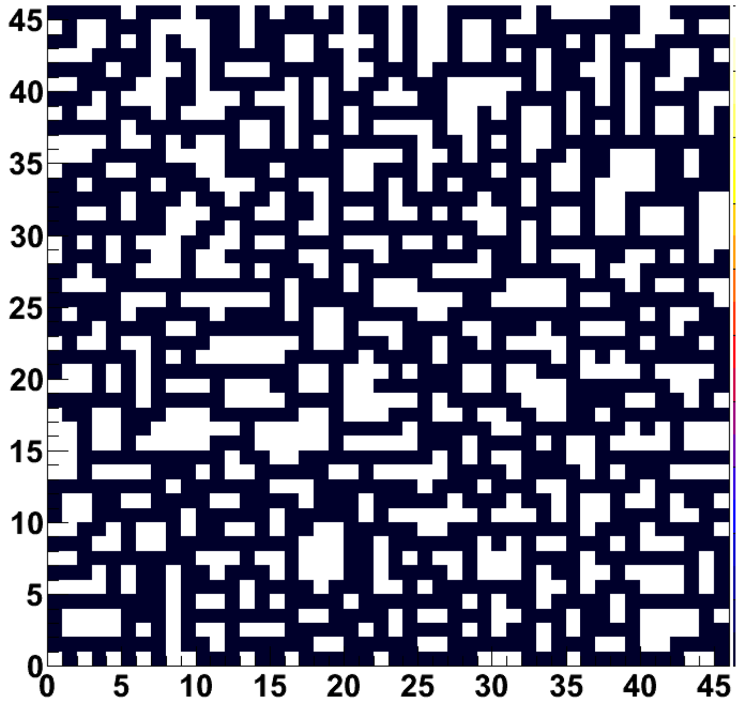
Crab Nebula Integral Observation

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

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

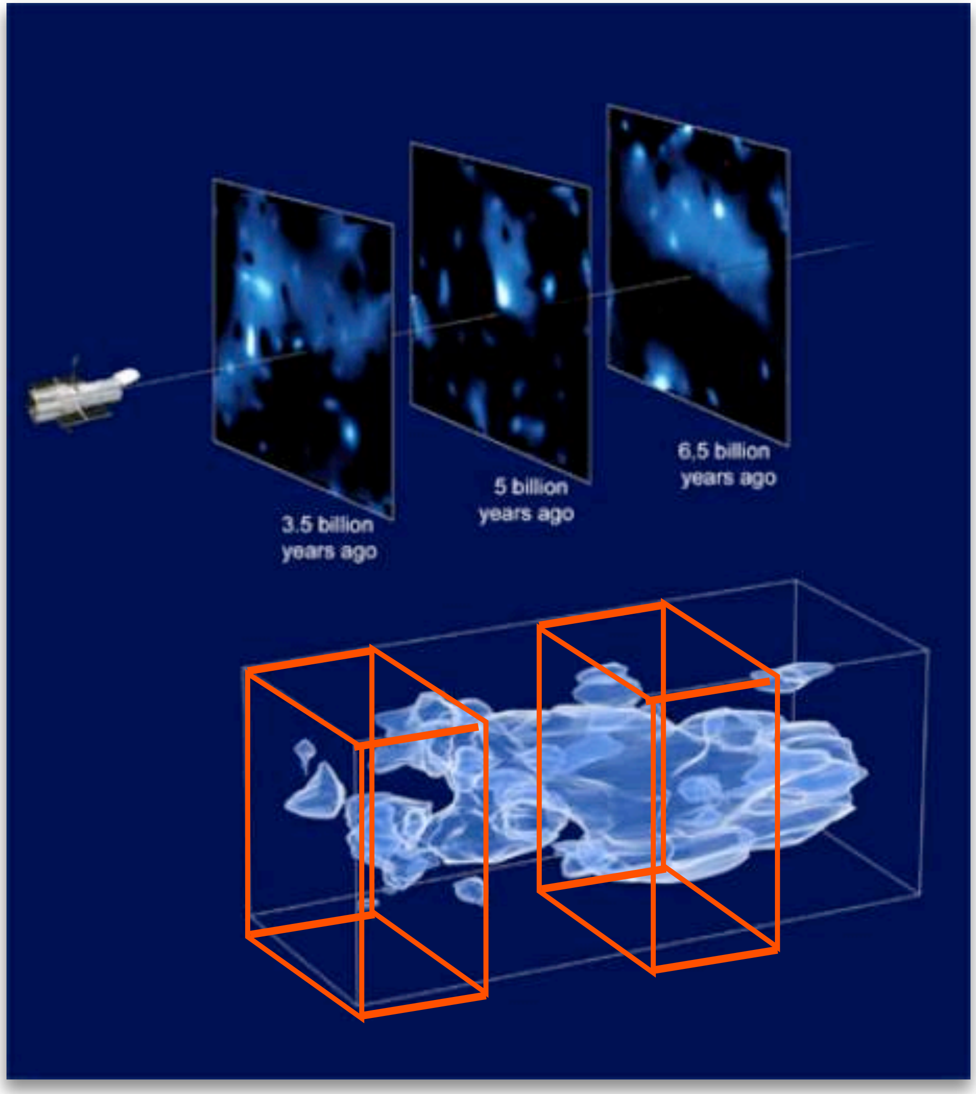
- ***ECLAIRs france-chinese satellite 'SVOM' (launch in 2014-2015)***
  Gamma-ray detection in energy range 4 - 120 keV
  Coded mask imaging (at 460 mm of the detector plane)

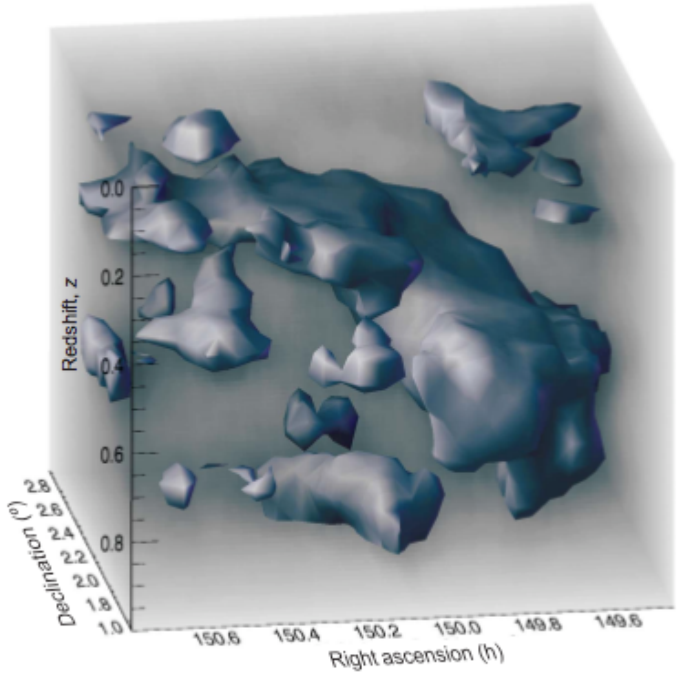
Physical mask pattern
(46 x 46 pixels of 11.7 mm)



ECLAIR could become the first CS-Designed Astronomical Instrument

Pseudo-3D Weak Lensing





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):

$$\boxed{\kappa = Q\delta + N}$$

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

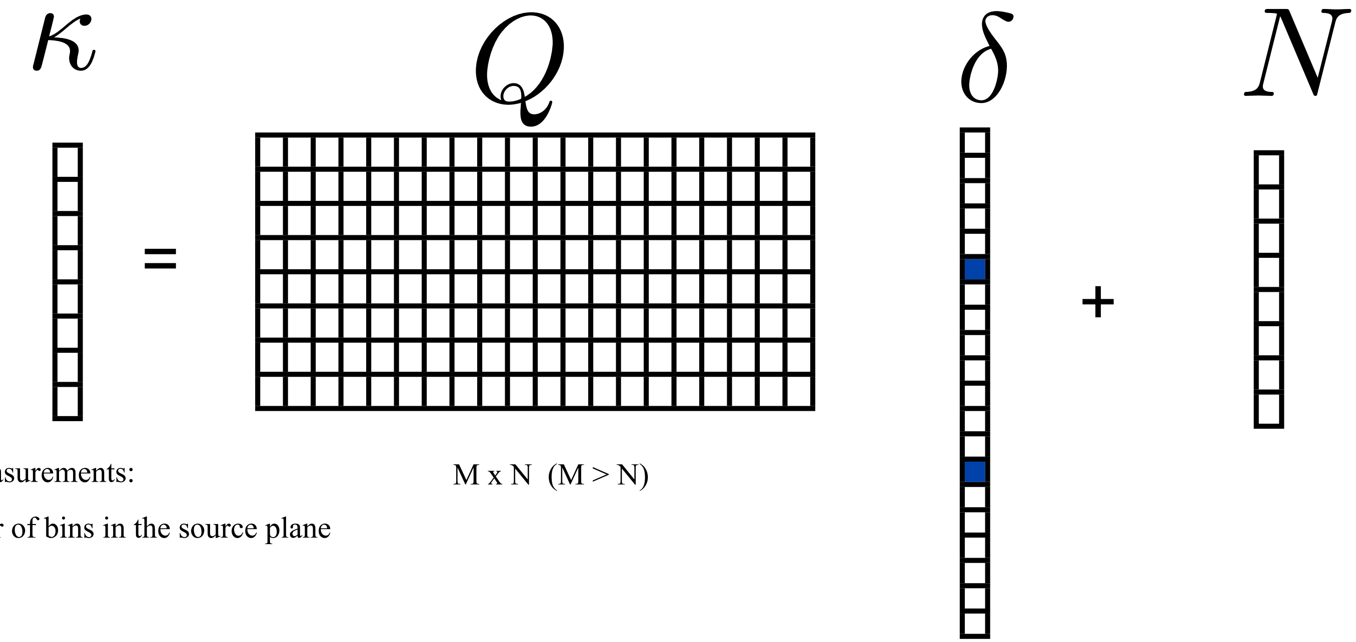
$$Q_{i\ell} = \frac{3H_0^2\Omega_M}{2c^2}\int_{w_\ell}^{w_{\ell+1}} dw \frac{\overline{W}^{(i)}(w) f_K(w)}{a(w)}\,, \quad \overline{W}^{(i)}(w) = \int_0^{w^{(i)}} dw' \frac{f_K(w-w')}{f_K(w')}\left(p(z)\frac{dz}{dw}\right)_{z=z(w')}$$

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.

3D Weak Lensing



M measurements:

number of bins in the source plane

M x N (M > N)

N redshift bin for the density contrast

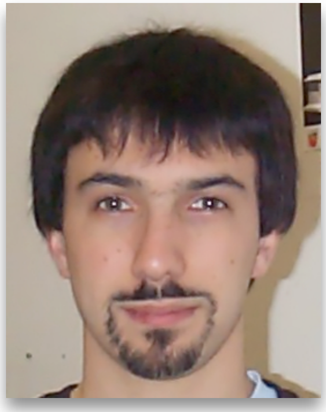
δ is sparse.

Q spreads out the information in δ along κ bins.

More unkown than measurements

3D Weak Lensing

Matter in the Universe as a Natural Compressed Sensing Operator

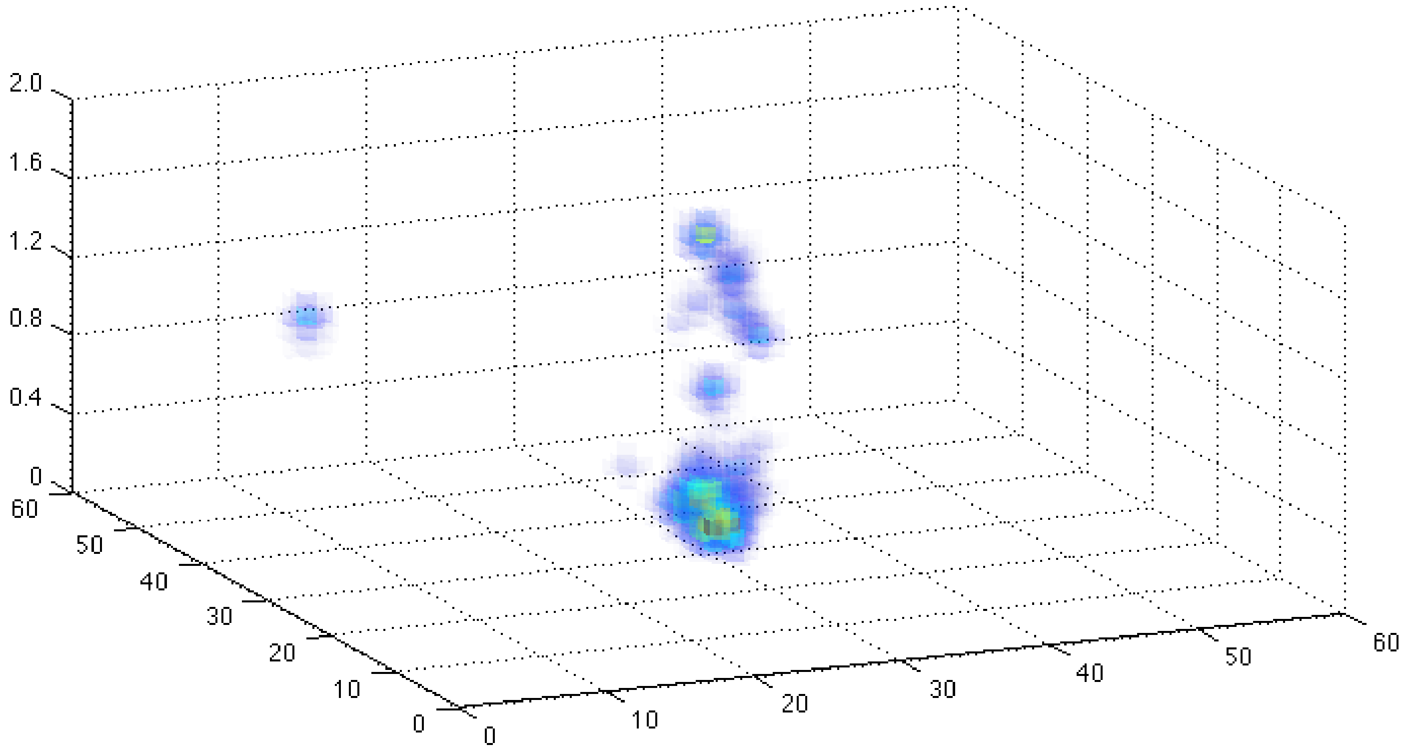


A. Leonard, F.-X. Dupe, J.-L. Starck, "A compressed sensing approach to 3D weak lensing", Astronomy and Astrophysics, arXiv:1111.6478, A&A, in press.

$$\min_{\delta} \| \delta \|_1 \quad s.t. \ \frac{1}{2} \| \kappa - Q\delta \|^2_{\mathbf{\Sigma}^{-1}} \leq \epsilon$$

Recent optimization method, based on proximal theory, such as Chambolle & Pock (2010) can be used to find the solution.

3D Weak Lensing



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).

Sparsity/CS in Astrophysics

- Sparsity is very efficient for
  - Inverse problems (denoising, deconvolution, etc).
  - Inpainting
  - Component Separation.
- Compressed Sensing
  - New algorithms to process radio-interferommetric data
  - New approaches to analyze data.
  - New instruments design.
- Perspectives
  - PLANCK component separation.
  - Euclid