

Sparsity and Multiscale Methods in Astronomy

J.-L. Starck

Irfu, SEDI-SAP,

Service d'Astrophysique

CEA-Saclay, France.

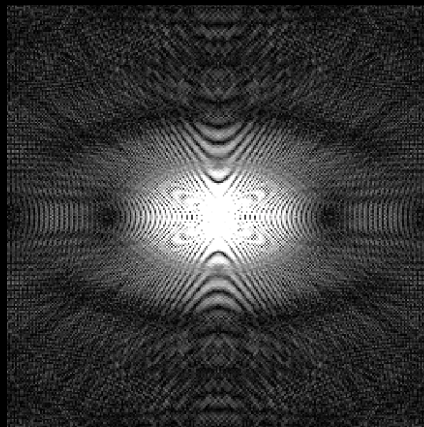
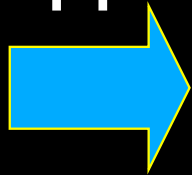
jstarck@cea.fr

<http://jstarck.free.fr>

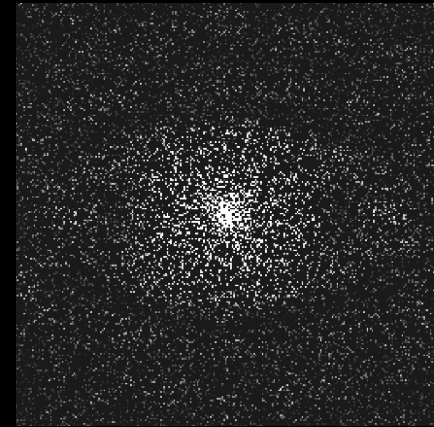
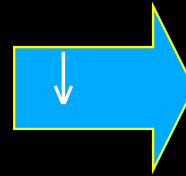
- **Introduction:**
 - ➔ **Compressed Sensing (CS)**
 - ➔ **Sparsity**
- **Compressed Sensing in Astronomy**
 - ➔ **CS Data**
 - ➔ **CS and the Herschel Satellite**
- **Sparsity in 2D**
- **Sparsity in 3D and the Spatial Distribution of galaxies**
- **Morphological Diversity**
 - ➔ **Component Separation**
 - ➔ **Inpainting**
 - ➔ **Multichannel Data**
- **Sparsity and the PLANCK project**



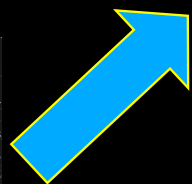
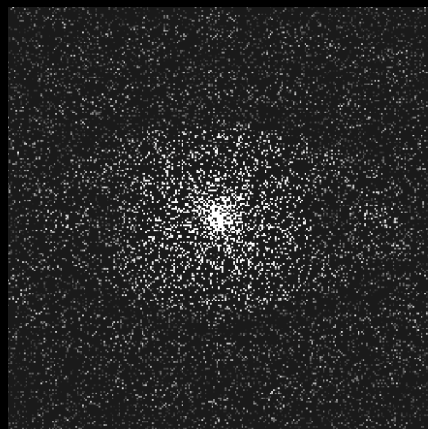
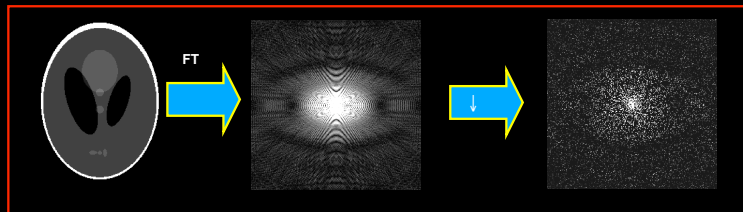
FT



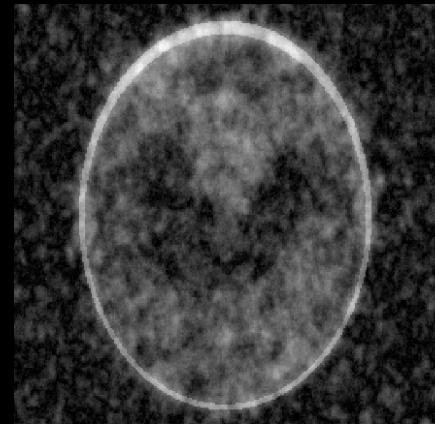
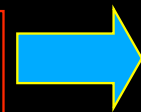
Randomly throw
away 83% of
samples



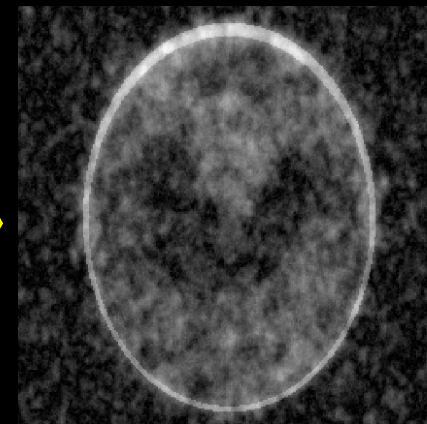
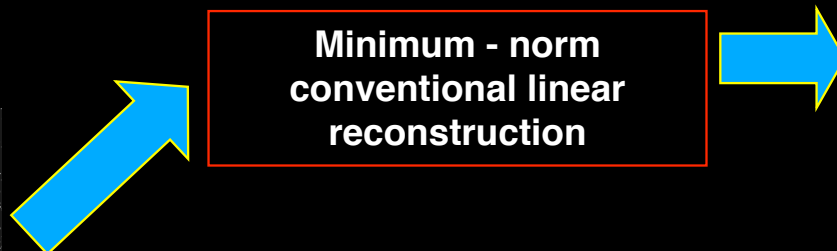
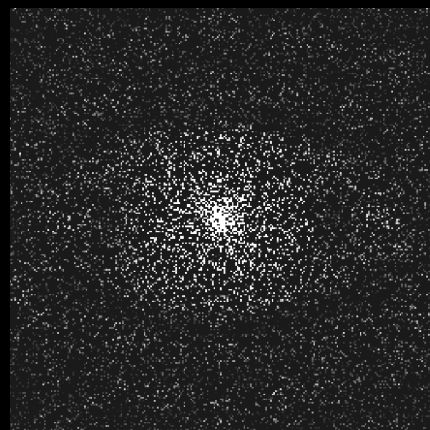
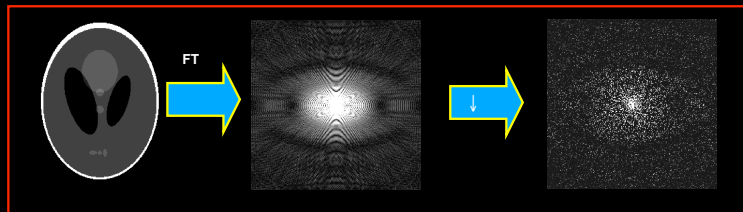
* E.J. Candes, J. Romberg and T. Tao.



**Minimum - norm
conventional linear
reconstruction**



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





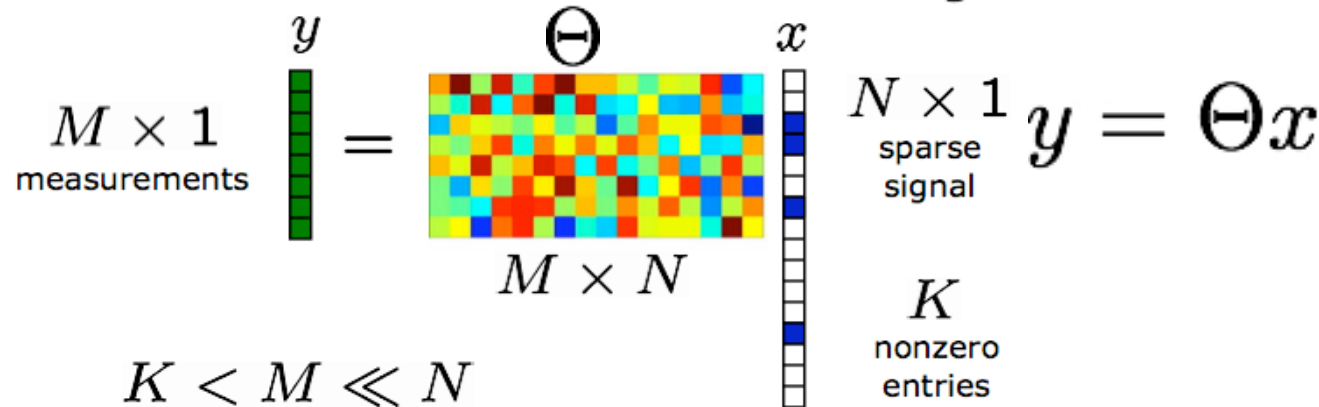
Compressed Sensing

- * 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 $\sim K \log N$ incoherent measurements”

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



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

⇒Application: Compression, tomography, ill posed inverse problem.

Compressed Sensing Reconstruction

Measurements: $y_k = \langle x, \theta_k \rangle$

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

In practice, x is sparse in a given **dictionary**: $x = \Phi\alpha$

and we need to solve: $\min_\alpha \|\alpha\|_1 \quad \text{s.t.} \quad y = \Theta_\Lambda \Phi\alpha$

The mutual incoherence is defined as $\mu_{\Theta, \Phi} = \max_{i,j} |\langle \theta_i, \phi_j \rangle|$

the number of required measurements is : $m \geq C \cdot \mu_{\Theta, \Phi}^2 \cdot S \cdot \log n$

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

$$(P0) \text{ Minimize } \|\alpha\|_0 \text{ subject to } \mathcal{S} = \phi\alpha$$

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

$$(P1) \text{ Minimize } \|\alpha\|_1 \text{ subject to } \mathcal{S} = \phi\alpha$$

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

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

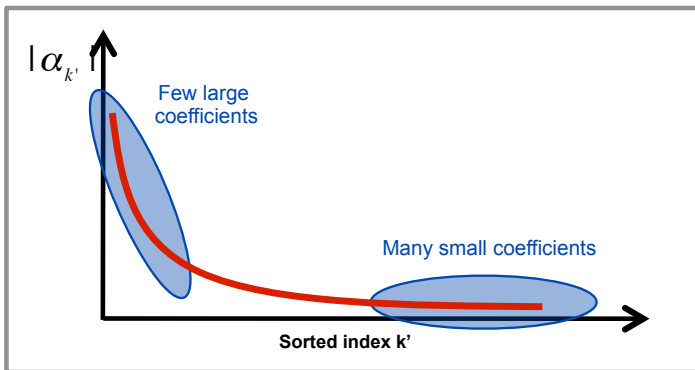
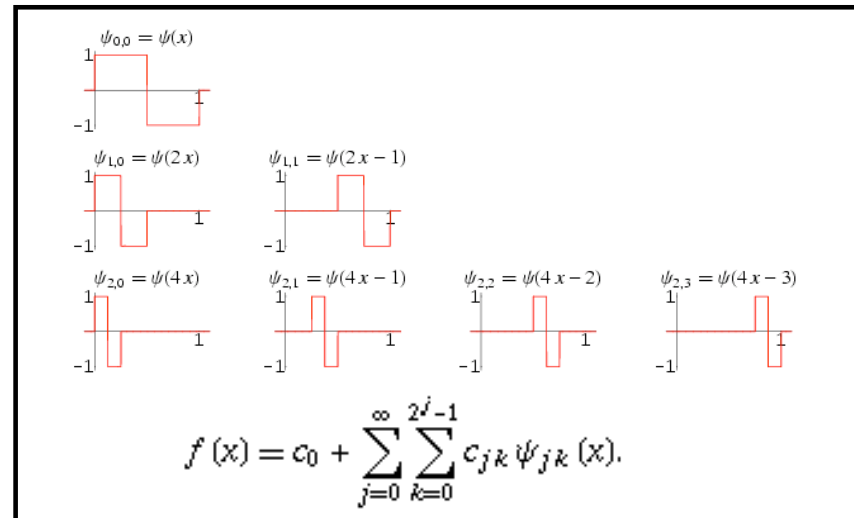
What is Sparsity ?

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

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

Ex: Haar wavelet

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



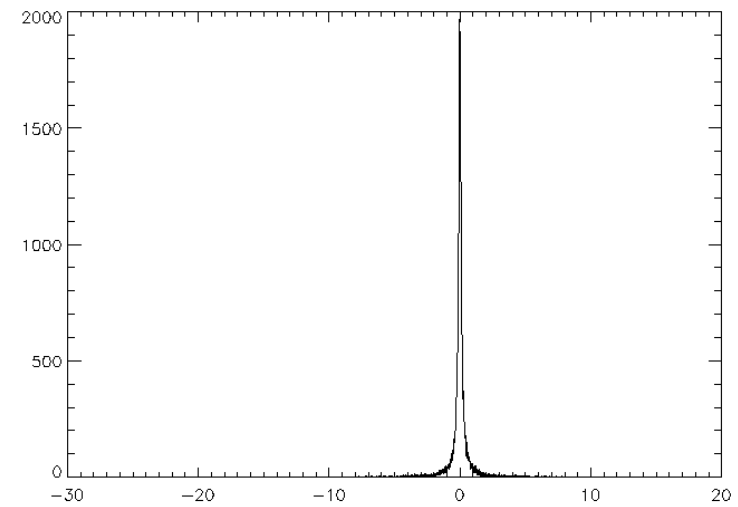
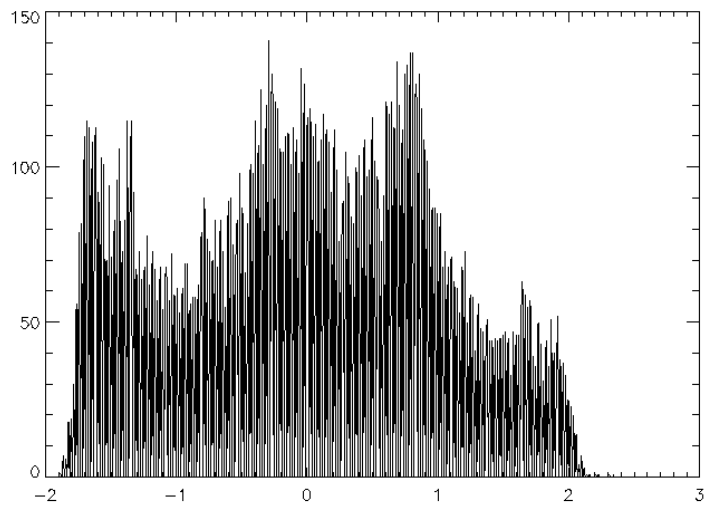
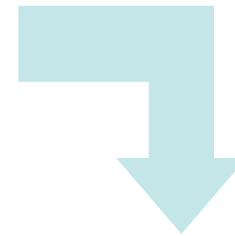
The new sampling theory, *Compressed Sensing*, is strongly related to Sparsity.

Representing Barbara

Direct Space



Curvelet Space



JPEG / JPEG2000

Original BMP
300x300x24
270056 bytes



JPEG 1:68
3983 bytes



JPEG2000 1:70
3876 bytes



Multiscale Transforms

Critical Sampling

(bi-) Orthogonal WT
Lifting scheme construction
Wavelet Packets
Mirror Basis

Redundant Transforms

Pyramidal decomposition (Burt and Adelson)
Undecimated Wavelet Transform
Isotropic Undecimated Wavelet Transform
Complex Wavelet Transform
Steerable Wavelet Transform
Dyadic Wavelet Transform
Nonlinear Pyramidal decomposition (Median)

New Sparse Representations

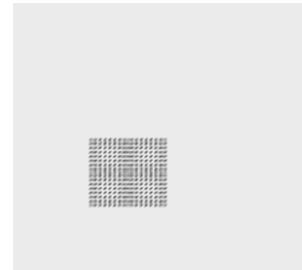
Contourlet
Bandelet
Finite Ridgelet Transform
Platelet
(W-)Edgelet
Adaptive Wavelet
KSVD
Grouplet

Ridgelet
Curvelet (Several implementations)
Wave Atom

Looking for adapted representations

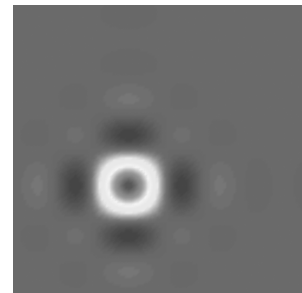
Local DCT

Stationary textures
Locally oscillatory



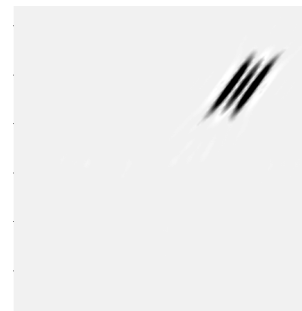
Wavelet transform

Piecewise smooth
Isotropic structures



Curvelet transform

Piecewise smooth,
edge



Compressive Sensing Resources

<http://www.dsp.ece.rice.edu/cs/>

More than 200 related papers already!

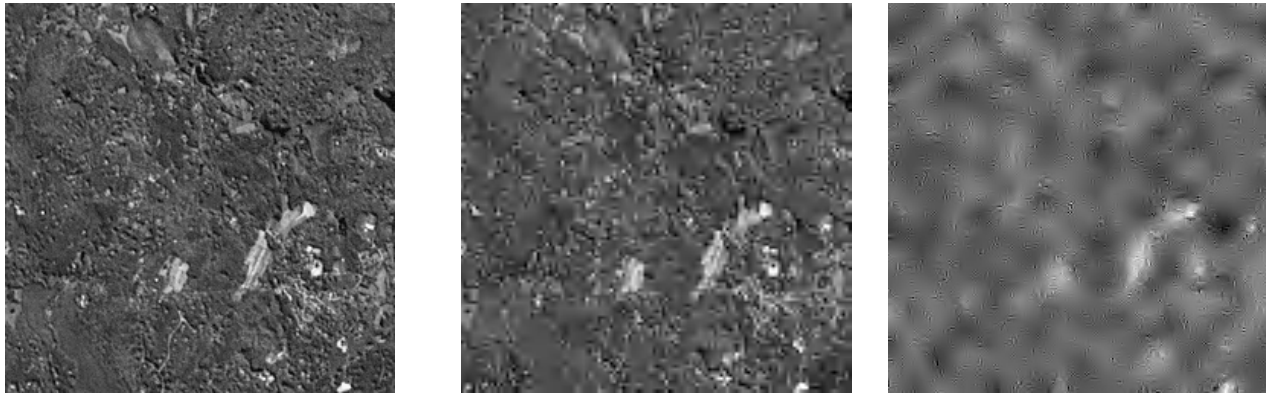
- Compressive Sensing
- Extensions of Compressive Sensing
- Multi-Sensor and Distributed Compressive Sensing
- Compressive Sensing Recovery Algorithms
- Foundations and Connections
- High-Dimensional Geometry
- ℓ_1 Norm Minimization
- Statistical Signal Processing
- Machine Learning
- Bayesian Methods
- Finite Rate of Innovation
- Multi-band Signals
- Data Stream Algorithms
- Compressive Imaging
- Medical Imaging
- Analog-to-Information Conversion
- Biosensing
- Geophysical Data Analysis
- Hyperspectral Imaging
- Compressive Radar
- Astronomy
- Communications

+ software available

Compressed Sensing For Data Compression

Compressed Sensing presents several interesting properties for data compress:

- Compression is **very fast** ==> good for on-board applications.
- Very **robust** to bit loss during the transfer.
- **Decoupling** between compression/decompression.
- **Data protection.**
- **Linear Compression.**



But clearly not as competitive to JPEG or JPEG2000 to compress an image.

Compressed Sensing Impact in astronomy

$$y = \Theta x$$

Typical Astronomical Data related to CS

- (radio-) Interferometry: $\Theta =$ Fourier transform
 $\Phi =$ Id (or Wavelet transform)
- Period detection in temporal series
 $\Theta =$ Id
 $\Phi =$ Fourier transform
- Gamma Ray Instruments (Integral) - Acquisition with coded masks

CS gives another point of view on some existing methods

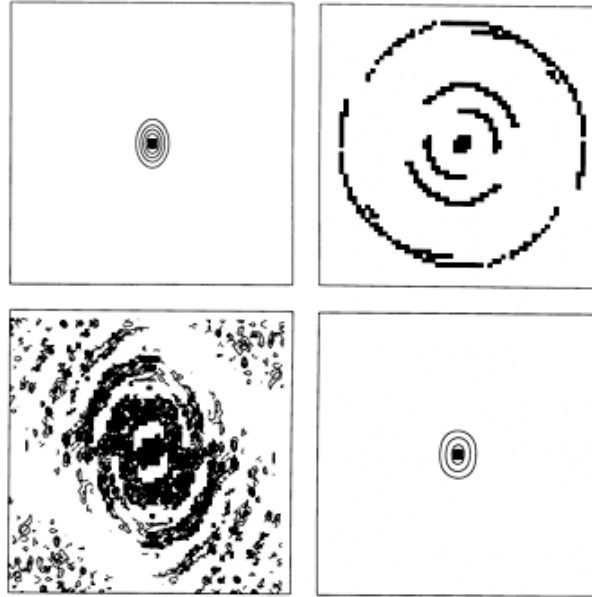
- Inpainting: $\Theta =$ Id

$$\min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad y = M\Phi\alpha = Mx$$

New problems that can be addressed by CS

==> Data compression: the case of Herschel satellite

(Radio-) Interferometry

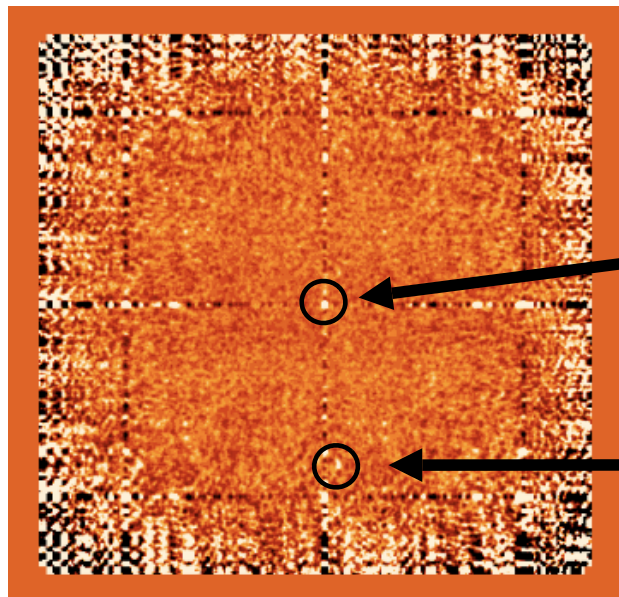
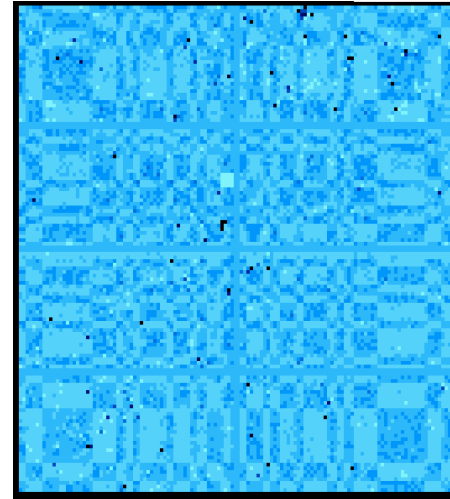
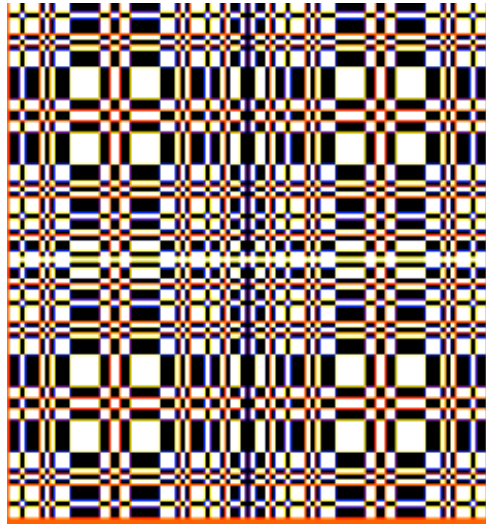


J.L. Starck, A. Bijaoui, B. Lopez, and C. Perrier, "Image Reconstruction by the Wavelet Transform Applied to Aperture Synthesis", *Astronomy and Astrophysics*, 283, 349--360, 1994.

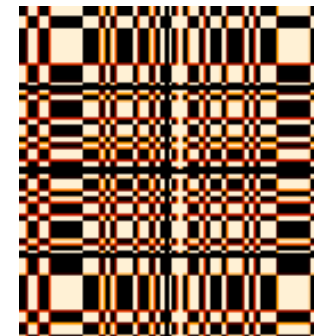
Wavelet - CLEAN minimizes well the l_0 norm

But recent l_0 - l_1 minimization algorithms would be clearly much faster.

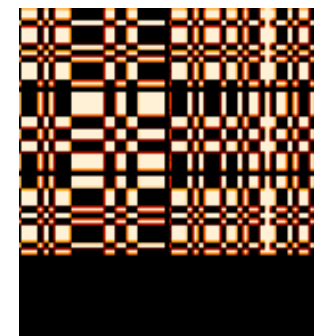
INTEGRAL/IBIS Coded Mask



Excess 1
Position (SPSF fit)
Identification
→ Modelling



Excess 2
Position (SPSF fit)
Identification
→ Modelling





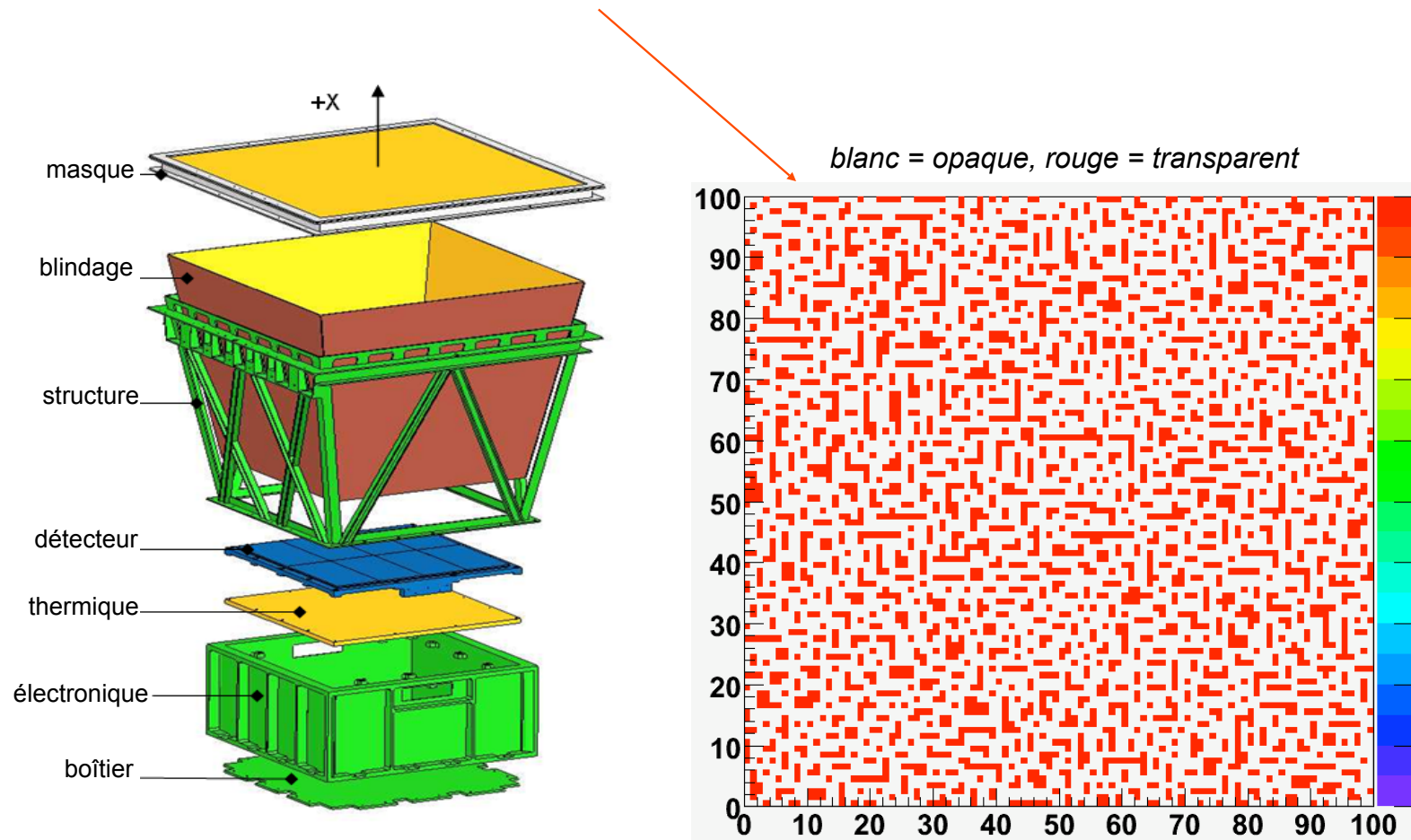
dapnia



saclay

ECLAIRs

- **ECLAIRs** france-chinese satellite 'SVOM' (launch in 2013)
Gamma-ray detection in energy range 4 - 120 keV
Coded mask imaging (at 460 mm of the detector plane)
- detector 1024 cm² of Cd Te (80 x 80 pixels)
- mask (100 x 100 pixels)



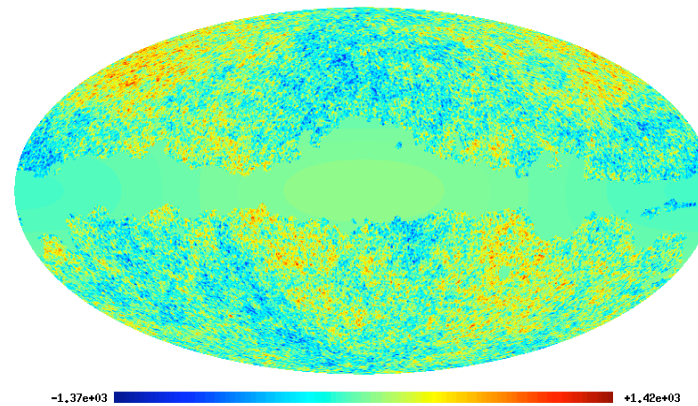
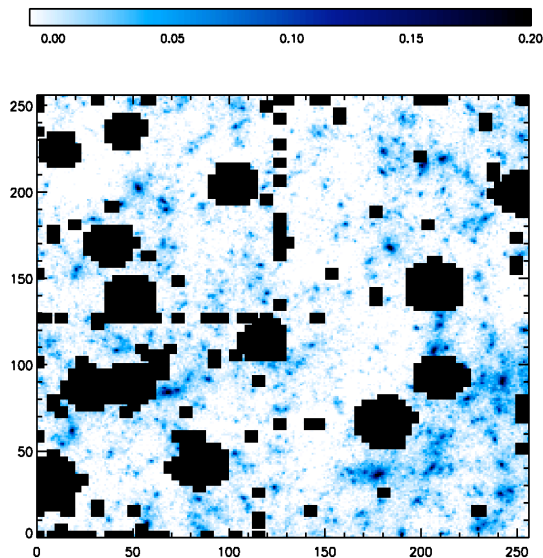
Interpolation of Missing Data: Inpainting

•M. Elad, J.-L. Starck, D.L. Donoho, P. Querre, "Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA)", *ACHA*, Vol. 19, pp. 340-358, 2005.

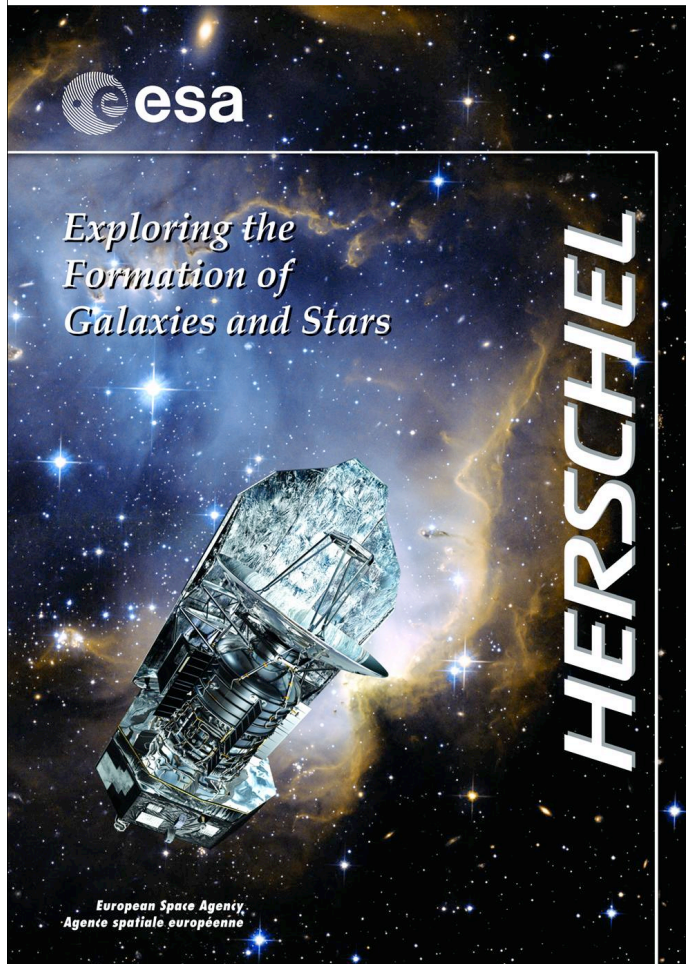
•M.J. Fadili, J.-L. Starck and F. Murtagh, "Inpainting and Zooming using Sparse Representations", in press.

$$\Theta_{\Lambda} = \text{Id}_{\Lambda} \quad \min_{\alpha} \|\alpha\|_0 \quad \text{s.t.} \quad y = M\Phi\alpha = Mx$$

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



HERSCHEL



This space telescope has been designed to observe in the far-infrared and sub-millimeter wavelength range.

Its launch is scheduled for the beginning of 2009. The shortest wavelength band, 57-210 microns, is covered by PACS (Photodetector Array Camera and Spectrometer).

Herschel data transfer problem:

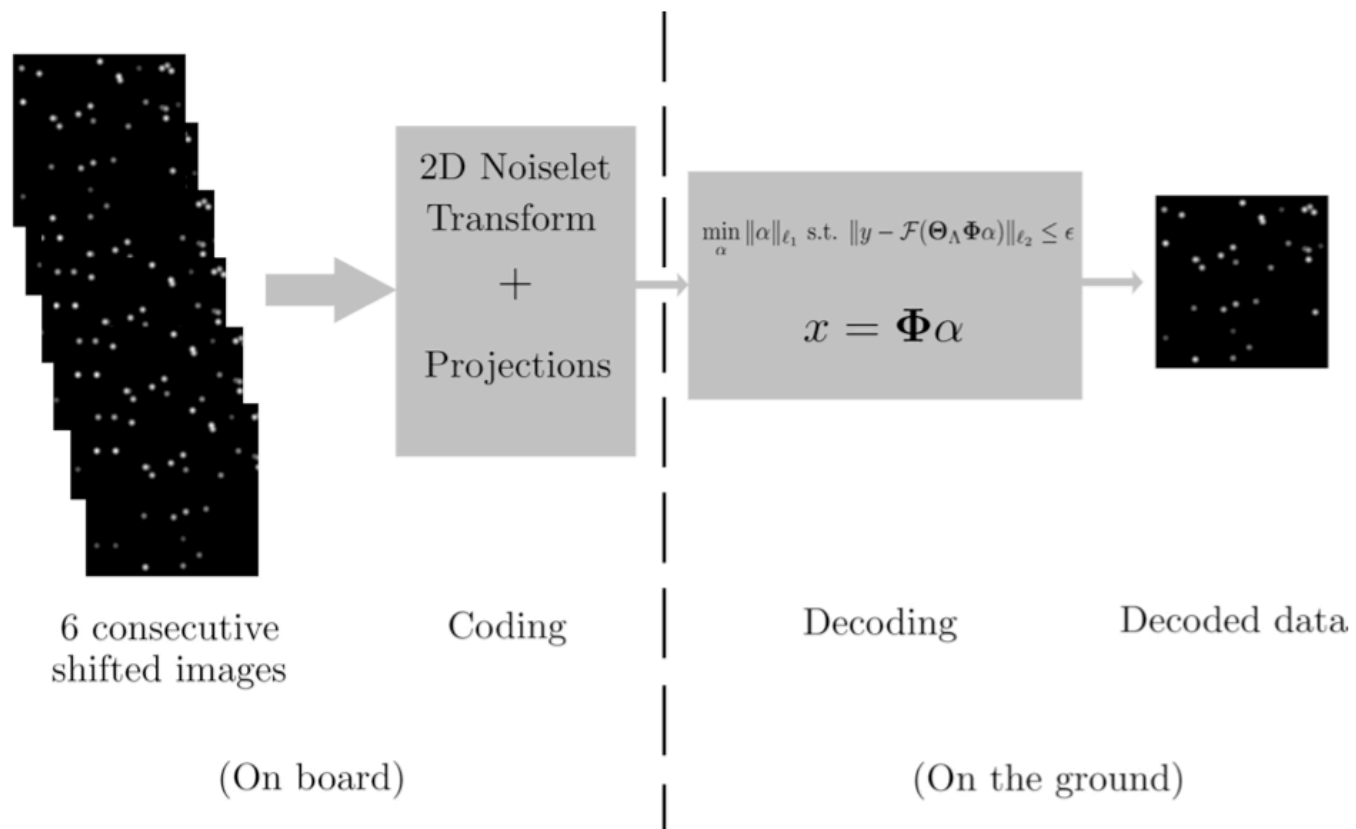
- no time to do sophisticated data compression on board.

- a compression ratio of 6 must be achieved.

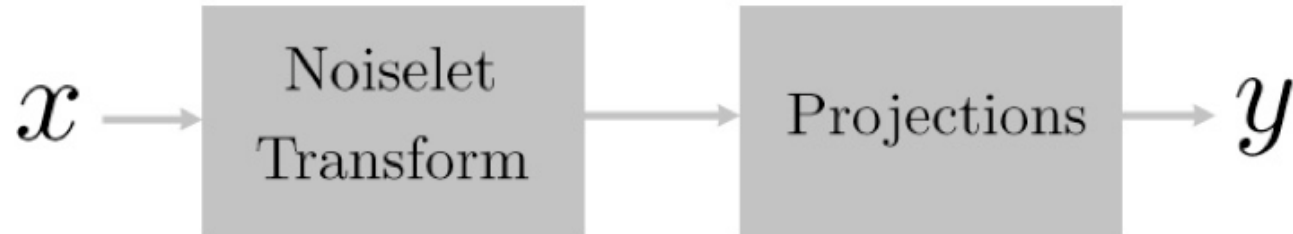
==> solution: averaging of six successive images on board

CS may offer another alternative.

The proposed Herschel compression scheme



The coding scheme



Good measurements must be incoherent with the basis in which the data are assumed to be sparse.

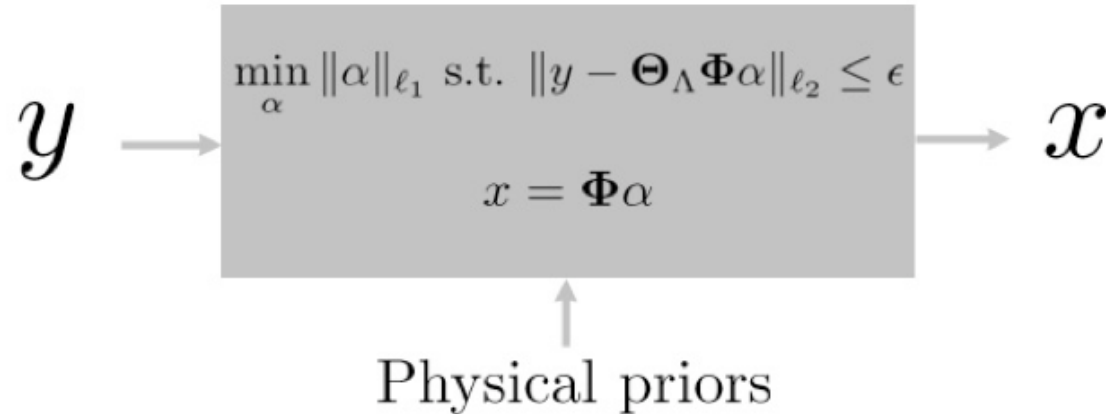
Noiselets (Coifman, Geshwind and Meyer, 2001) are an orthogonal basis that is shown to be highly incoherent with a wide range of practical sparse representations (wavelets, Fourier, etc).

Advantages:

Low computational cost ($O(n)$)

Most astronomical data are sparsely represented in a wide range of wavelet bases

The decoding scheme



$$\alpha^{(h)} = \mathcal{S}_{\gamma} \left\{ \alpha^{(h-1)} + \Phi^T \Theta^T \left(y^{\#} - \mathbf{I}_{\Lambda} \Theta \Phi \alpha^{(h-1)} \right) \right\}$$

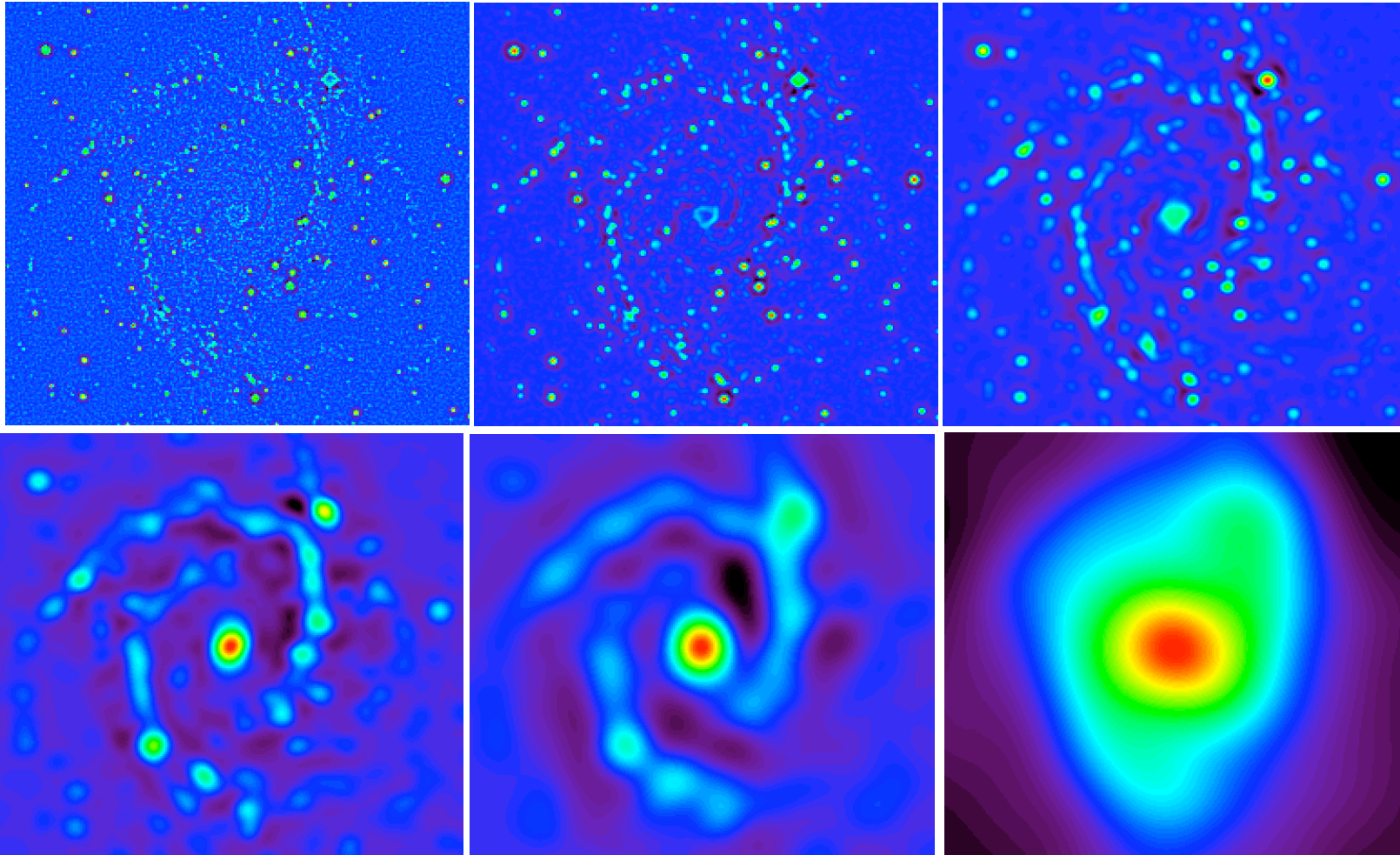
1. Set the number of iterations I_{\max} and threshold $\gamma^{(0)} = \|\Phi^T \Theta^T y^{\#}\|_{\infty}$. $x^{(0)}$ is set to zero.
2. While $\gamma^{(h)}$ is higher than a given lower bound γ_{\min} :
 - Compute the measurement projection of $x^{(h-1)}$:
 $y^{(h)} = \mathbf{I}_{\Lambda} \Theta x^{(h-1)}$.
 - Estimate the current coefficients $\alpha^{(h)}$:
 $\alpha^{(h)} = \mathcal{S}_{\gamma^{(h)}} \left\{ \alpha^{(h-1)} + \Phi^T \Theta^T [y^{\#} - y^{(h)}] \right\}$.
 - Get the new estimate of x by reconstructing from the selected coefficients $\alpha^{(h)}$:
 $x^{(h)} = \Phi \alpha^{(h)}$.
3. Decrease the threshold $\gamma^{(h)}$ following a given strategy.

Isotropic Undecimated Wavelet Transform (a trous algorithm)

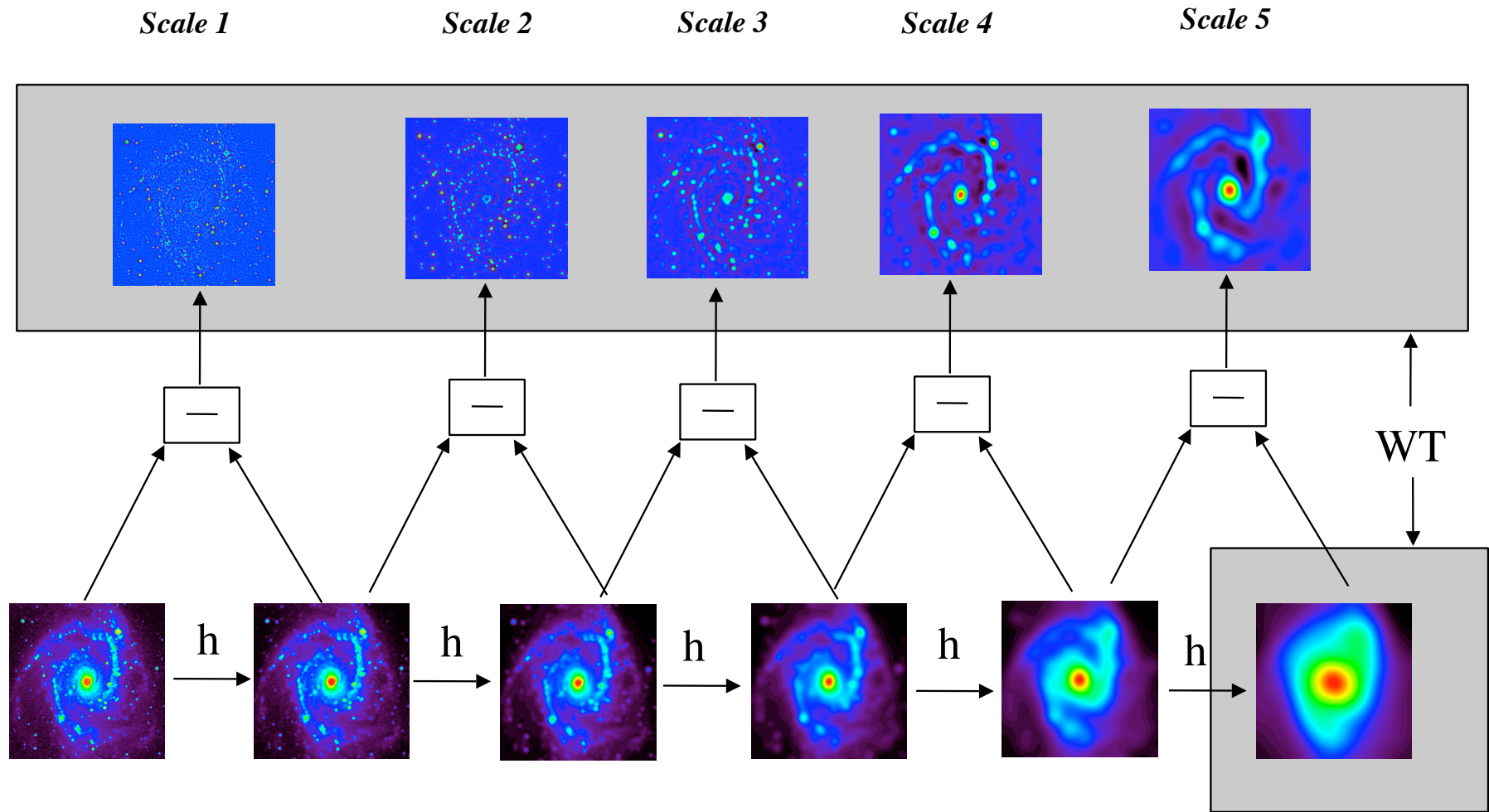
$$\varphi = B_3 - \text{spline}, \quad \frac{1}{2}\psi\left(\frac{x}{2}\right) = \frac{1}{2}\varphi\left(\frac{x}{2}\right) - \varphi(x)$$

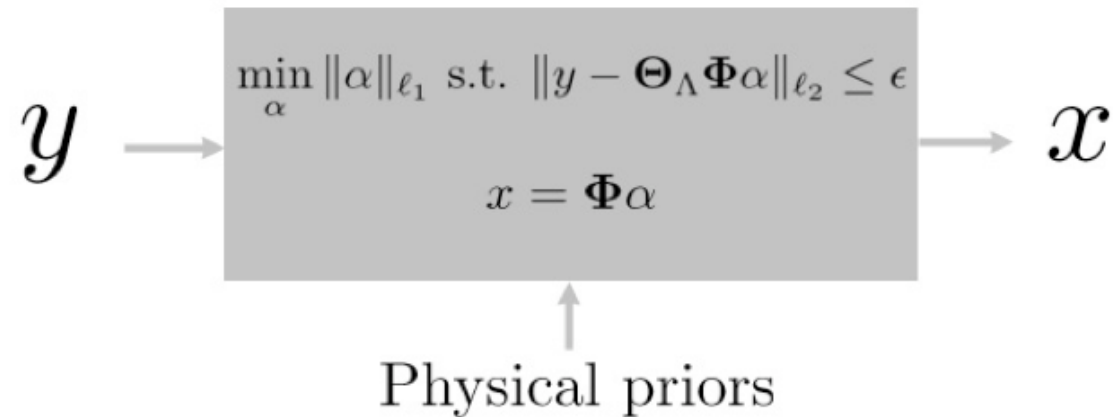
$$h = [1, 4, 6, 4, 1]/16, \quad g = \delta - h, \quad \tilde{h} = \tilde{g} = \delta$$

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



ISOTROPIC UNDECIMATED WAVELET TRANSFORM

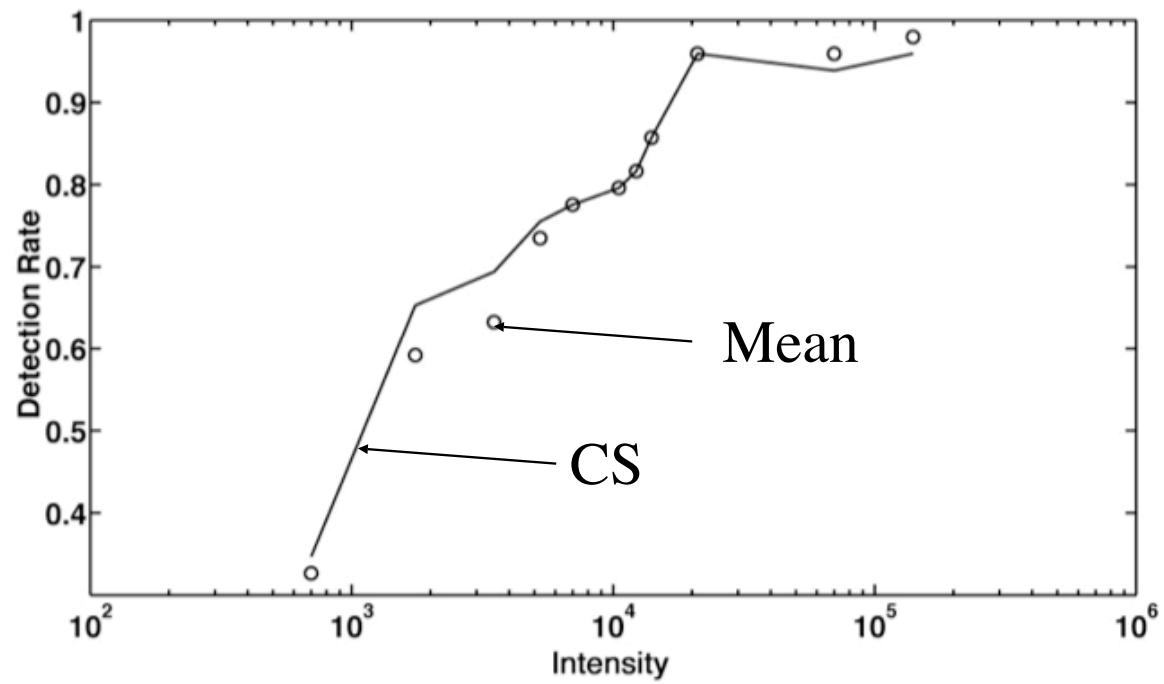




Six consecutive observations of the same field can be decompressed together:

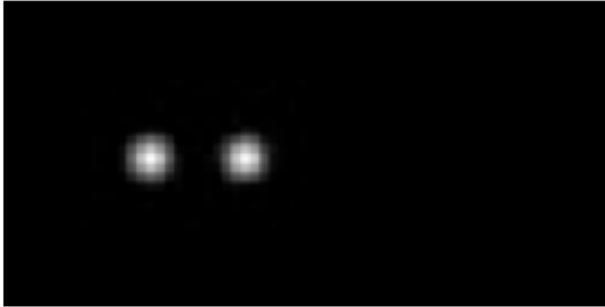
$$x^{(n+1)} = \frac{1}{6} \mathcal{S}_{\Phi, \lambda^{(n)}} \left\{ x^{(n)} + \sum_{i=1, \dots, 6} \Theta^T (y - \Theta x^{(n)}) \right\}$$

Sensitivity: CS versus mean of 6 images

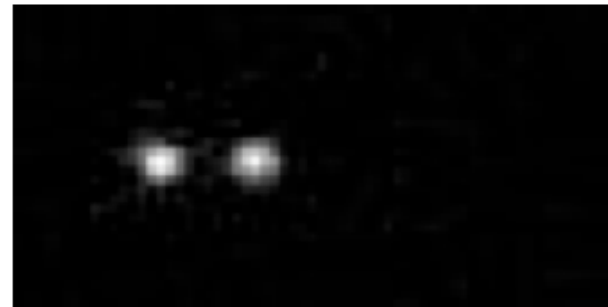
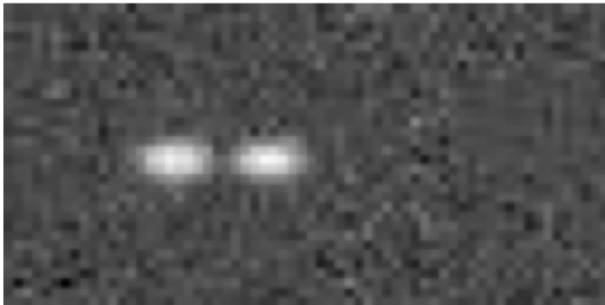
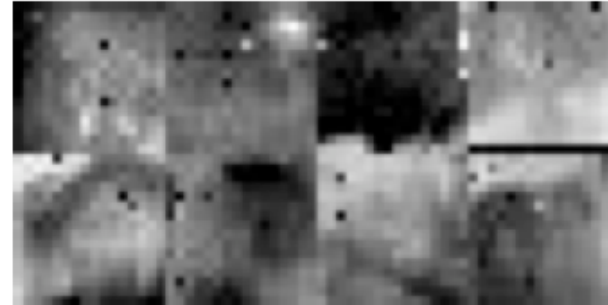


Resolution: CS versus Mean

Simulated image



Simulated noisy image with flat and dark



Mean of six images

Compressed sensing reconstructed images

Resolution limit versus SNR

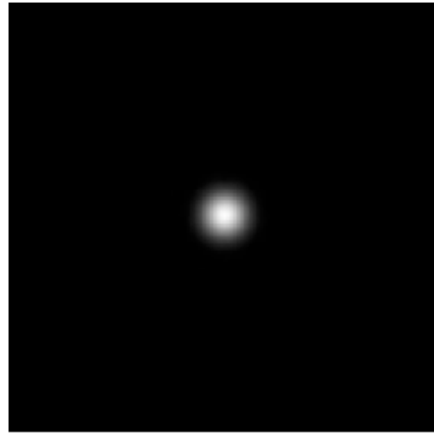
SNR	-17.3	-9.35	-3.3	0.21	2.7	4.7	6.2	7.6	8.7
Intensity	900	2250	4500	6750	9000	11250	13500	15750	18000
MO6	3	3	3	3	3	3	3	3	3
CS	2.33	2.33	2	2	2	2	2	2	2

THE CS-BASED COMPRESSION ENTAILS A RESOLUTION GAIN EQUAL TO A 30% OF THE SPATIAL RESOLUTION PROVIDED BY MO6.

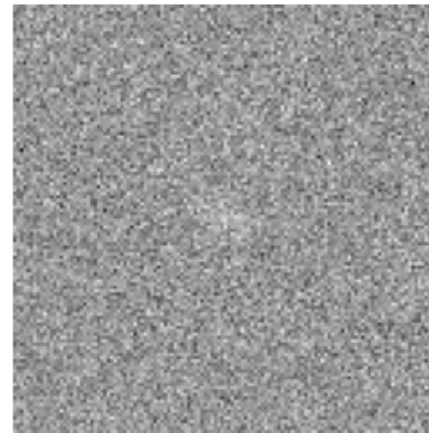
CS and Herschel Status

- CS compression is implemented in the Herschel on-board software (as an option).
- CS Tests in flight will be done.
- Software developments required for an efficient decompression (taking into account dark, flatp-field, PSF, etc).
- The CS decompression is fully integrated in the data processing pipeline.

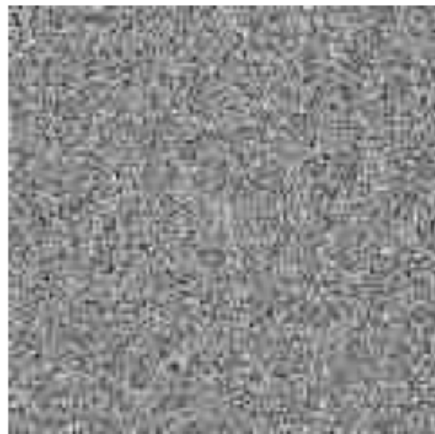
Data Fusion: JPEG versus Compressed Sensing



Simulated source



One of the 10 observations



Averaged of the 10 JPEG compressed images (CR=4)

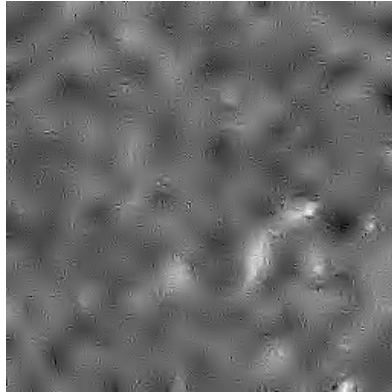


Reconstruction from the 10 compressed sensing images (CR=4)

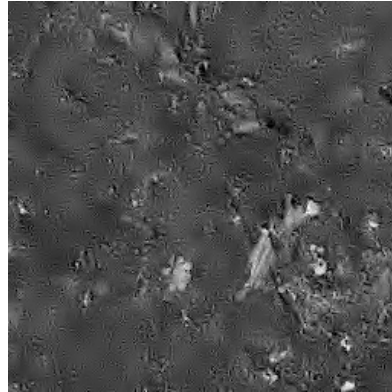
JPEG2000 Versus Compressed Sensing

Compression Rate: 25

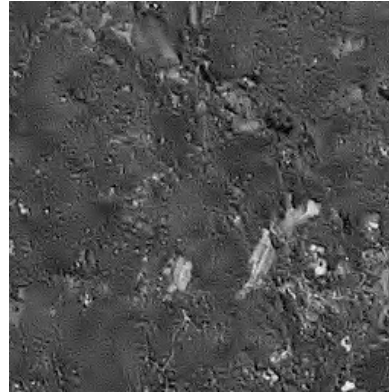
One observation



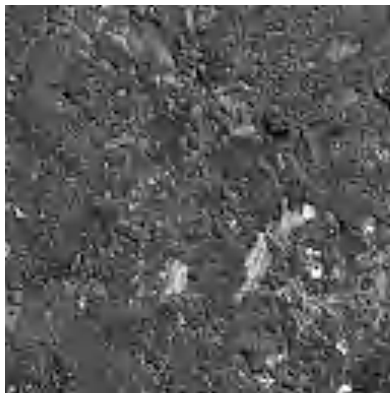
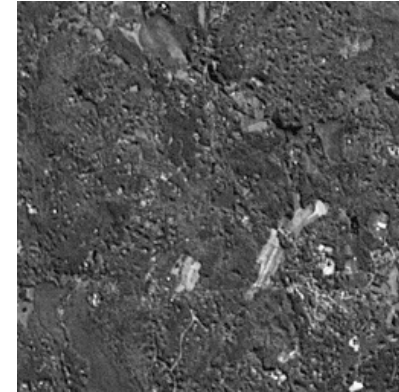
10 observations



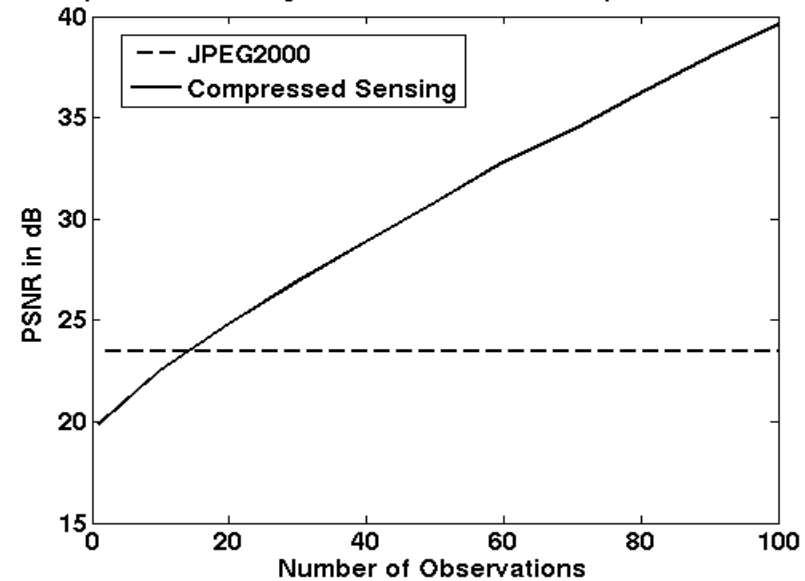
20 observations



100 observations



Compressed Sensing versus JPEG2000 - Compression Rate : 4%



Conclusions on CS

Compressed Sensing gives us a clear direction for:

- (radio-) interferometric data reconstruction
- periodic signals with sampled irregularly
- gamma-ray image reconstruction

CS provides an interesting framework and a good theoretical support for our inpainting work.

CS can be a good solution for on board data compression.

CS is a highly competitive solution for compressed data fusion.

Perspective

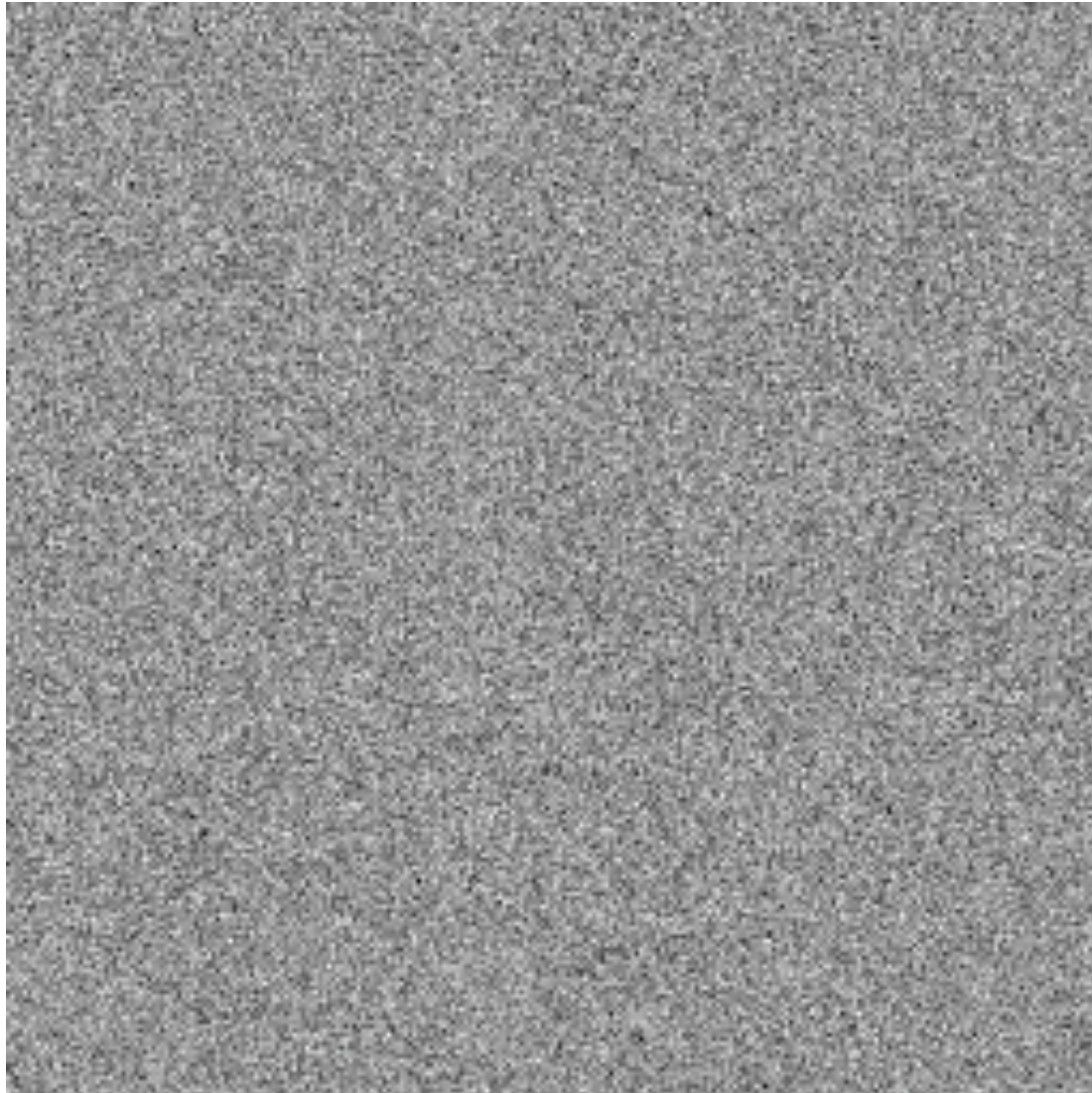
Design of CS astronomical instruments ?

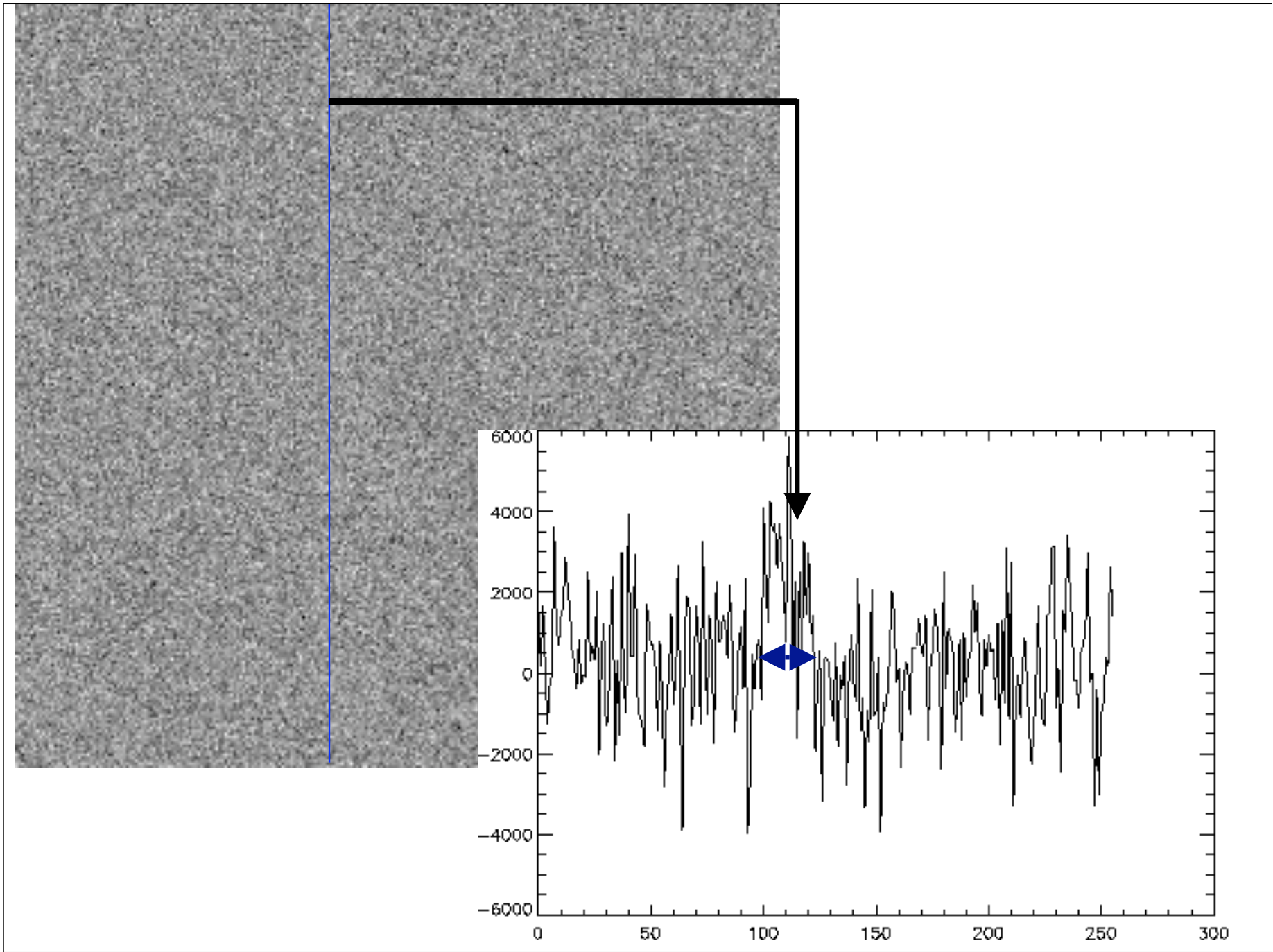
PREPRINT: <http://fr.arxiv.org/abs/0802.0131>

Problems related to the WT

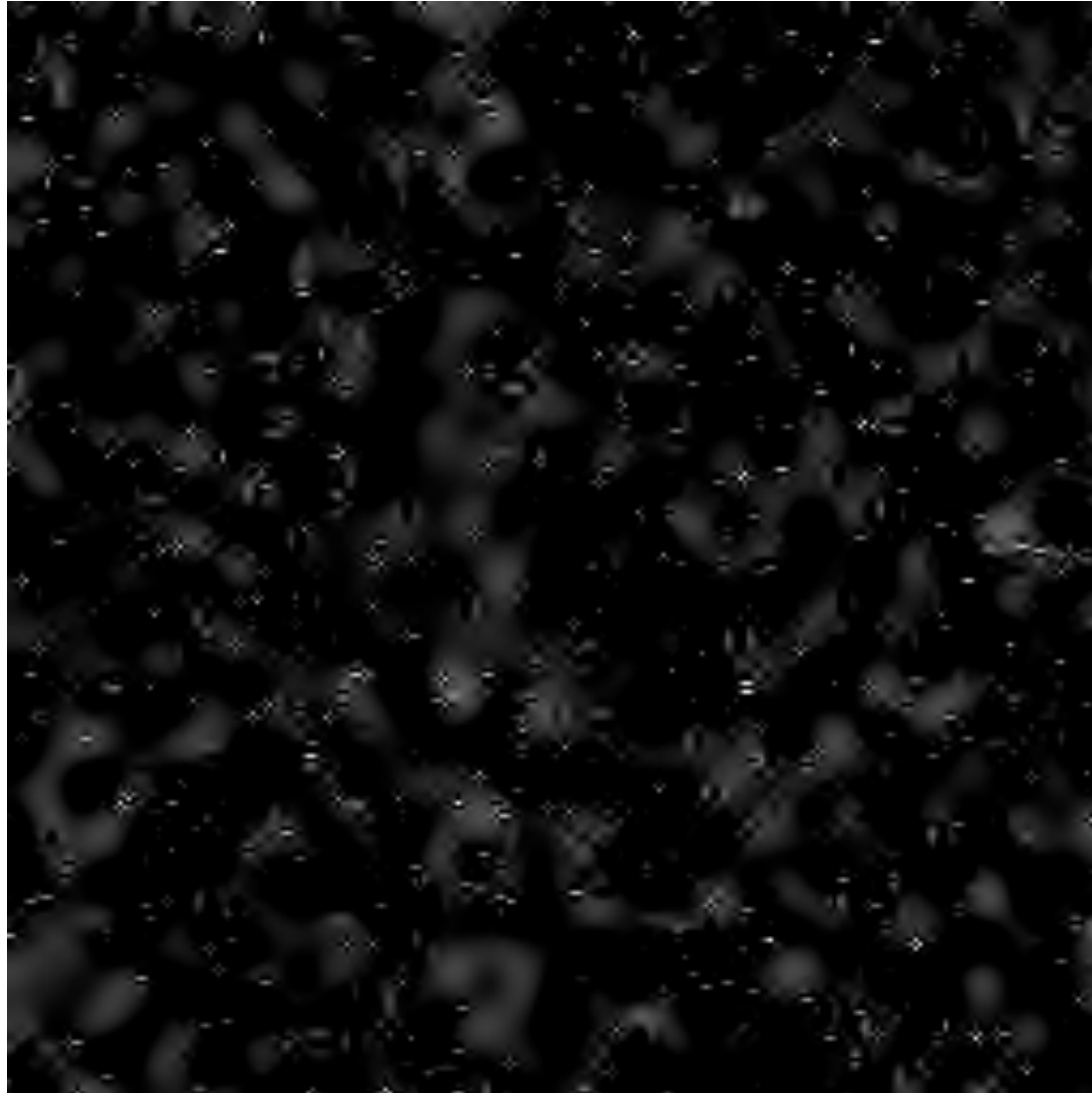
- 1) Edges representation:
if the WT performs better than the FFT to represent edges in an image, it is still not optimal.
- 2) There is only a fixed number of directional elements independent of scales.
- 3) Limitation of existing scale concepts:
there is no highly anisotropic elements.

SNR = 0.1

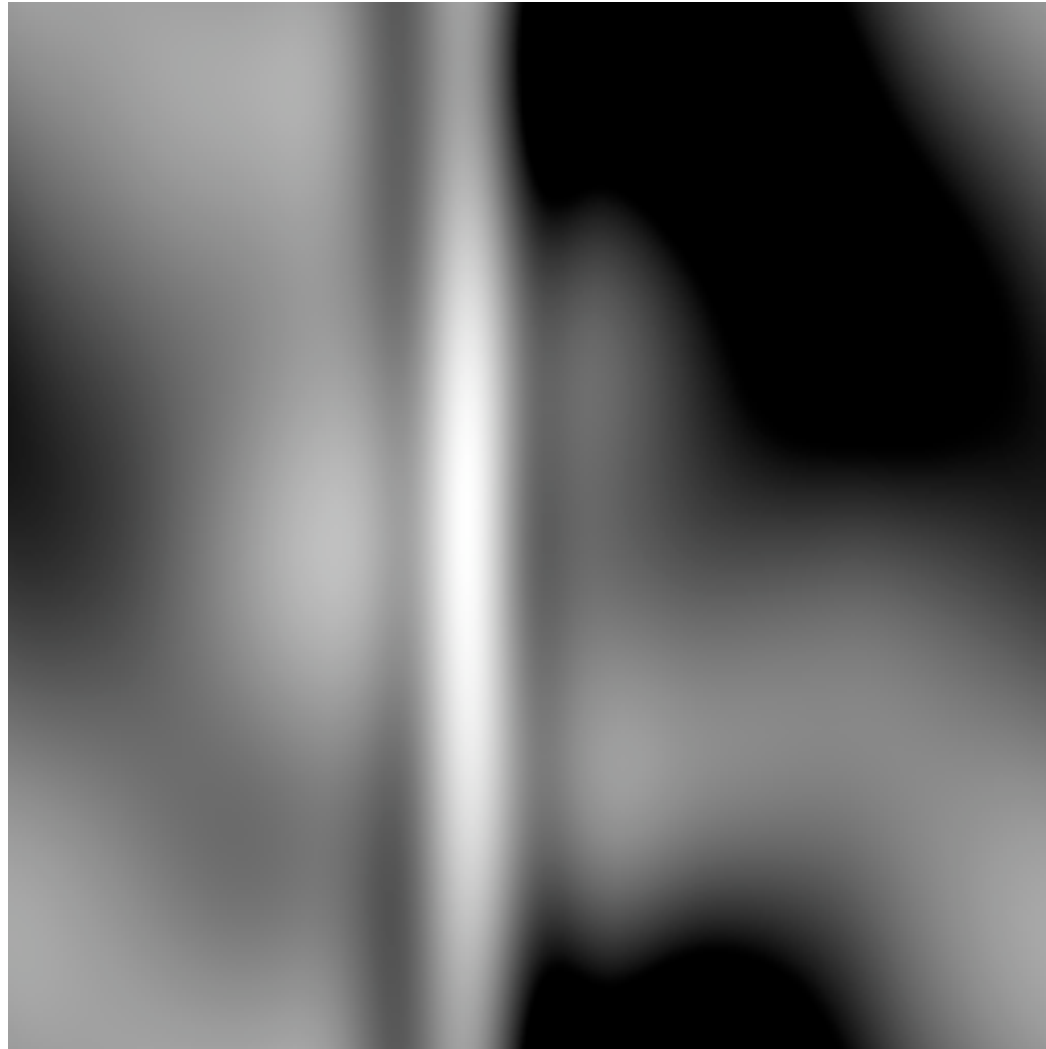




Undecimated Wavelet Filtering (3 sigma)

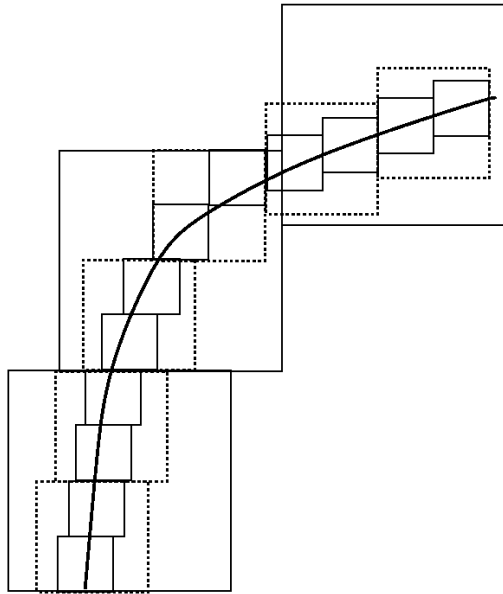


Ridgelet Filtering (5sigma)

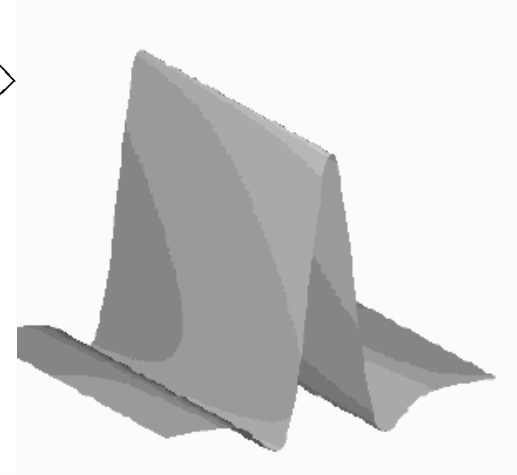
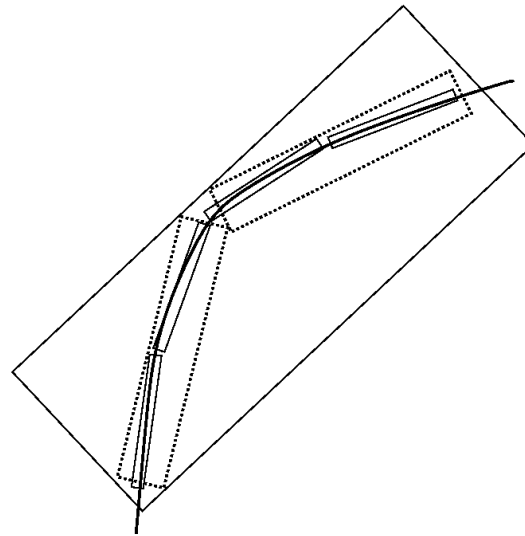


Wavelets and edges

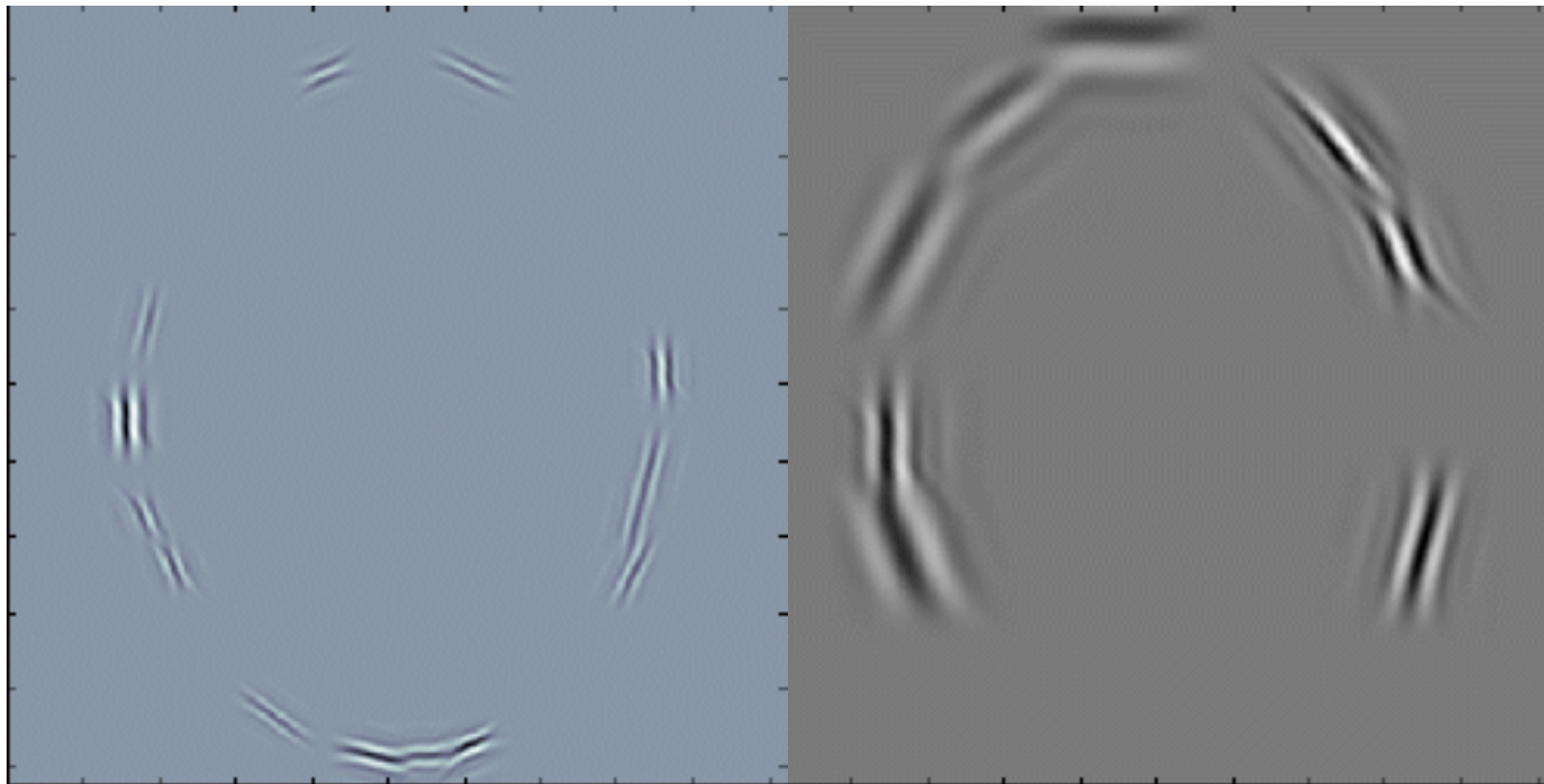
- many wavelet coefficients are needed to account for edges i.e. singularities along lines or curves :



- need dictionaries of strongly anisotropic atoms :



ridgelets, curvelets, contourlets, bandelettes, etc.

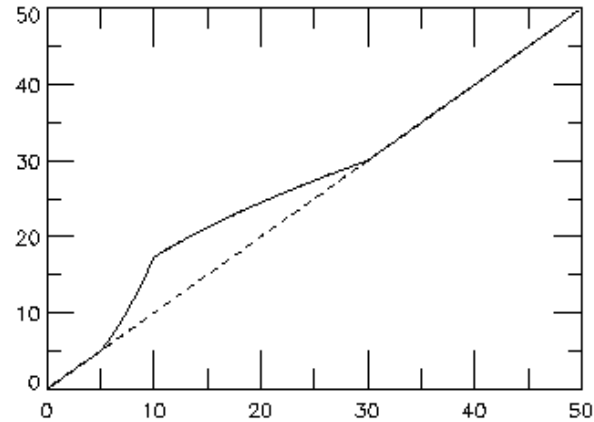


- J.-L. Starck, E. Candes, and D.L. Donoho, "**The Curvelet Transform for Image Denoising**", IEEE Transactions on Image Processing , 11, 6, pp 670 -684, 2002.
- J.-L. Starck, M.K. Nguyen and F. Murtagh, "**Wavelets and Curvelets for Image Deconvolution: a Combined Approach**", Signal Processing, 83, 10, pp 2279-2283, 2003.
- J.-L. Starck, E. Candes, and D.L. Donoho, "**Astronomical Image Representation by the Curvelet Transform**" , Astronomy and Astrophysics, 398, 785--800, 2003.
- J.-L. Starck, F. Murtagh, E. Candes, and D.L. Donoho, "**Gray and Color Image Contrast Enhancement by the Curvelet Transform**", IEEE Transaction on Image Processing, 12, 6, pp 706--717, 2003.

CONTRAST ENHANCEMENT

Gray and Color Image Contrast Enhancement by the Curvelet Transform,
IEEE Transaction on Image Processing, 12, 6, pp 706--717, 2003.

*Modified
curvelet
coefficient*



Curvelet coefficient

$$\tilde{I} = C_R(y_c(C_T I)) \quad \left\{ \begin{array}{ll} y_c(x, \sigma) = 1 & x < c\sigma \\ y_c(x, \sigma) = \frac{x - c\sigma}{c\sigma} \left(\frac{m}{c\sigma}\right)^p + \frac{2c\sigma - x}{c\sigma} & c\sigma \leq x < 2c\sigma \\ y_c(x, \sigma) = \left(\frac{m}{x}\right)^p & 2c\sigma \leq x < m \\ y_c(x, \sigma) = \left(\frac{m}{x}\right)^s & x > m \end{array} \right.$$

CONTRAST ENHANCEMENT USING THE CURVELET TRANSFORM

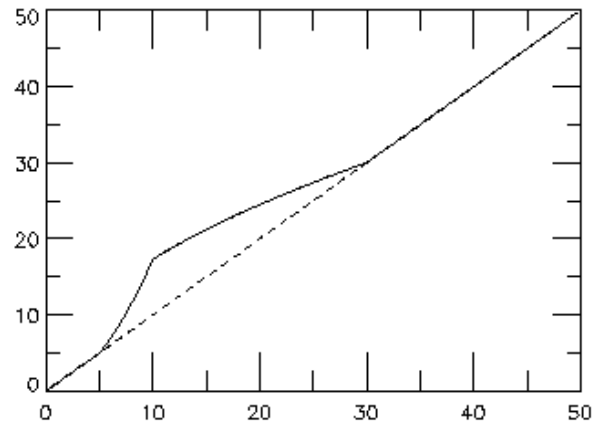
J.-L. Starck, F. Murtagh, E. Candes and D.L. Donoho, "Gray and Color Image Contrast Enhancement by the Curvelet Transform",

IEEE Transaction on Image Processing, 12, 6, 2003.

$$\tilde{I} = C_R(y_c(C_T I))$$

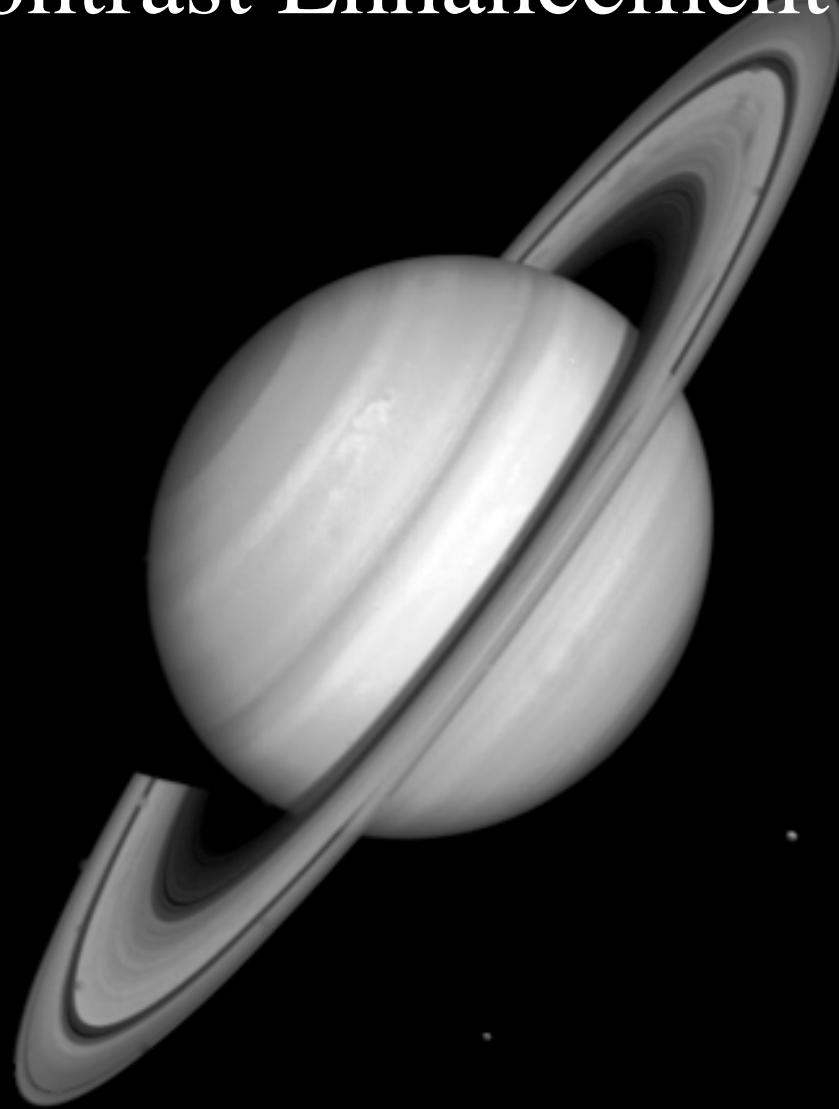
$$y_c(x, \sigma) = \begin{cases} 1 & \text{if } x < c\sigma \\ \frac{x - c\sigma}{c\sigma} \left(\frac{m}{c\sigma}\right)^p + \frac{2c\sigma - x}{c\sigma} & \text{if } x < 2c\sigma \\ \left(\frac{m}{x}\right)^p & \text{if } 2c\sigma \leq x < m \\ \left(\frac{m}{x}\right)^s & \text{if } x > m \end{cases}$$

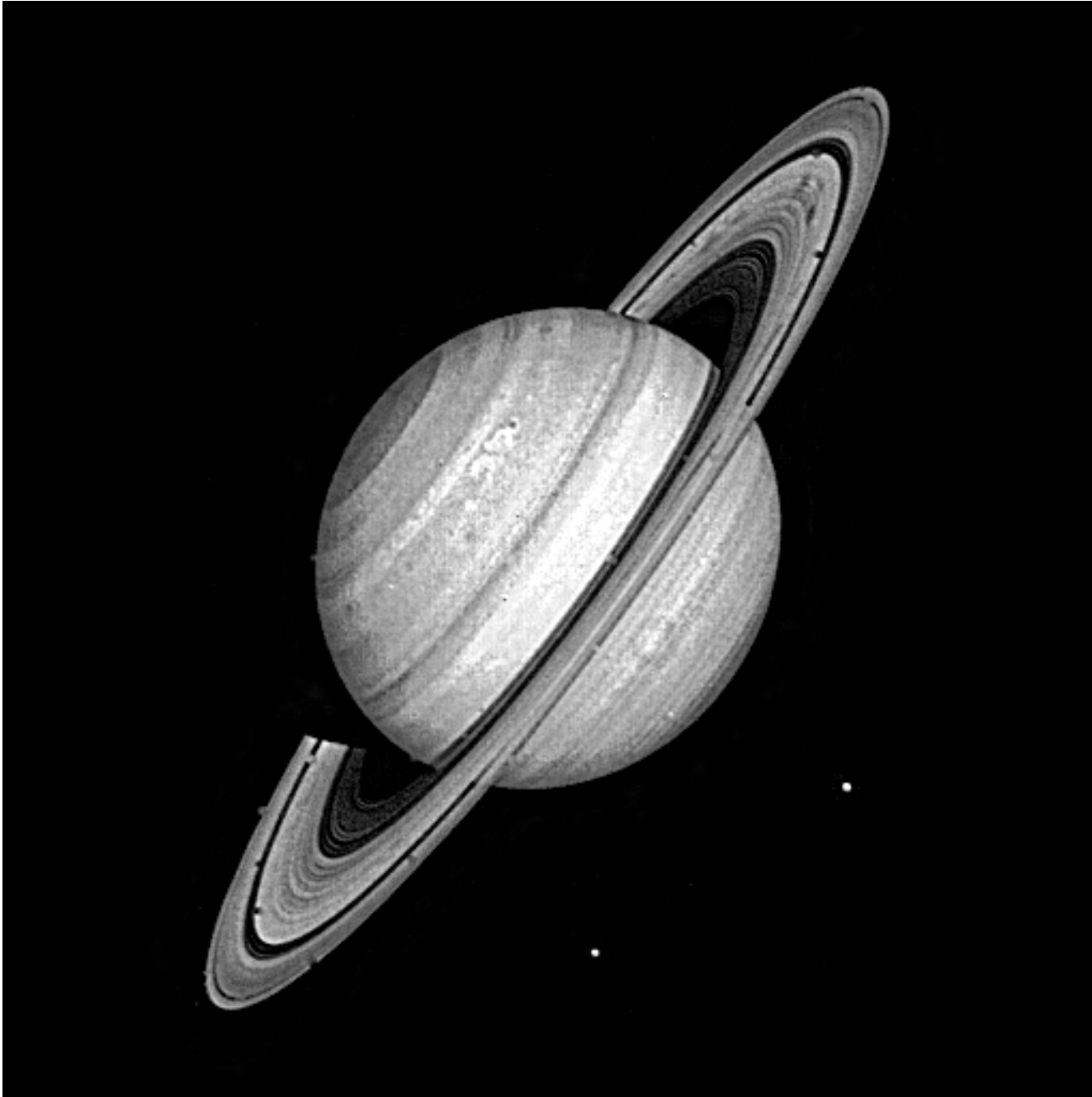
*Modified
curvelet
coefficient*

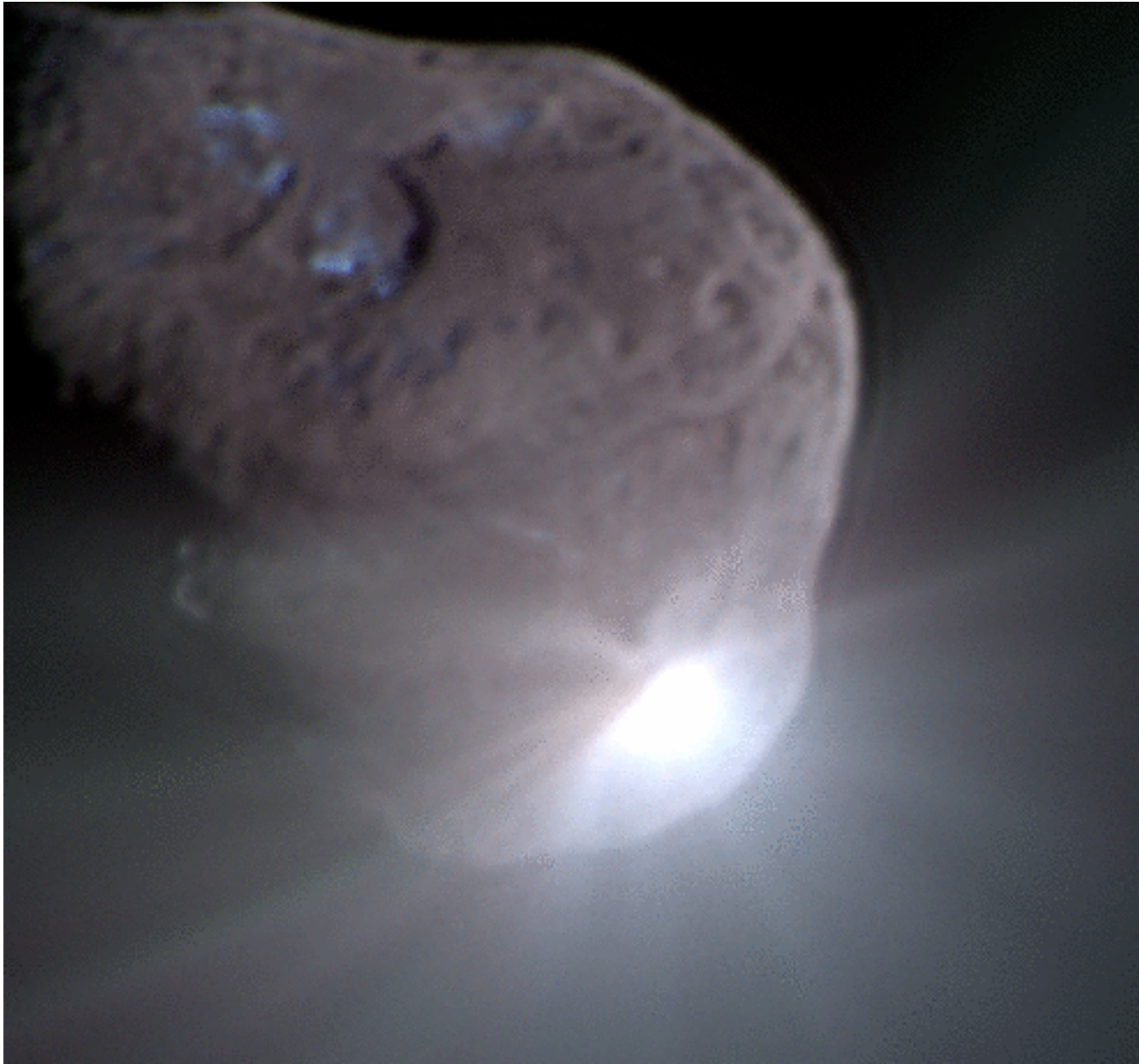


Curvelet coefficient

Contrast Enhancement







CURVELET FILTERING

NOISE MODELING

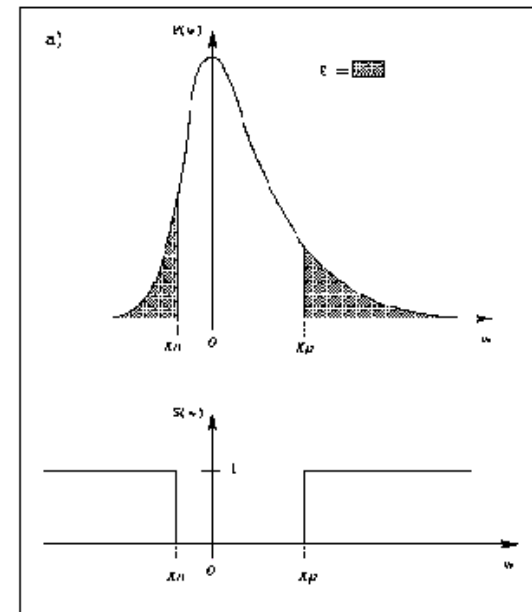
For a positive coefficient: $P = Prob\{W \dots w\}$

For a negative coefficient $P = Prob\{W \dots w\}$

Given a threshold t :

if $P > t$, the coefficient could be due to the noise.

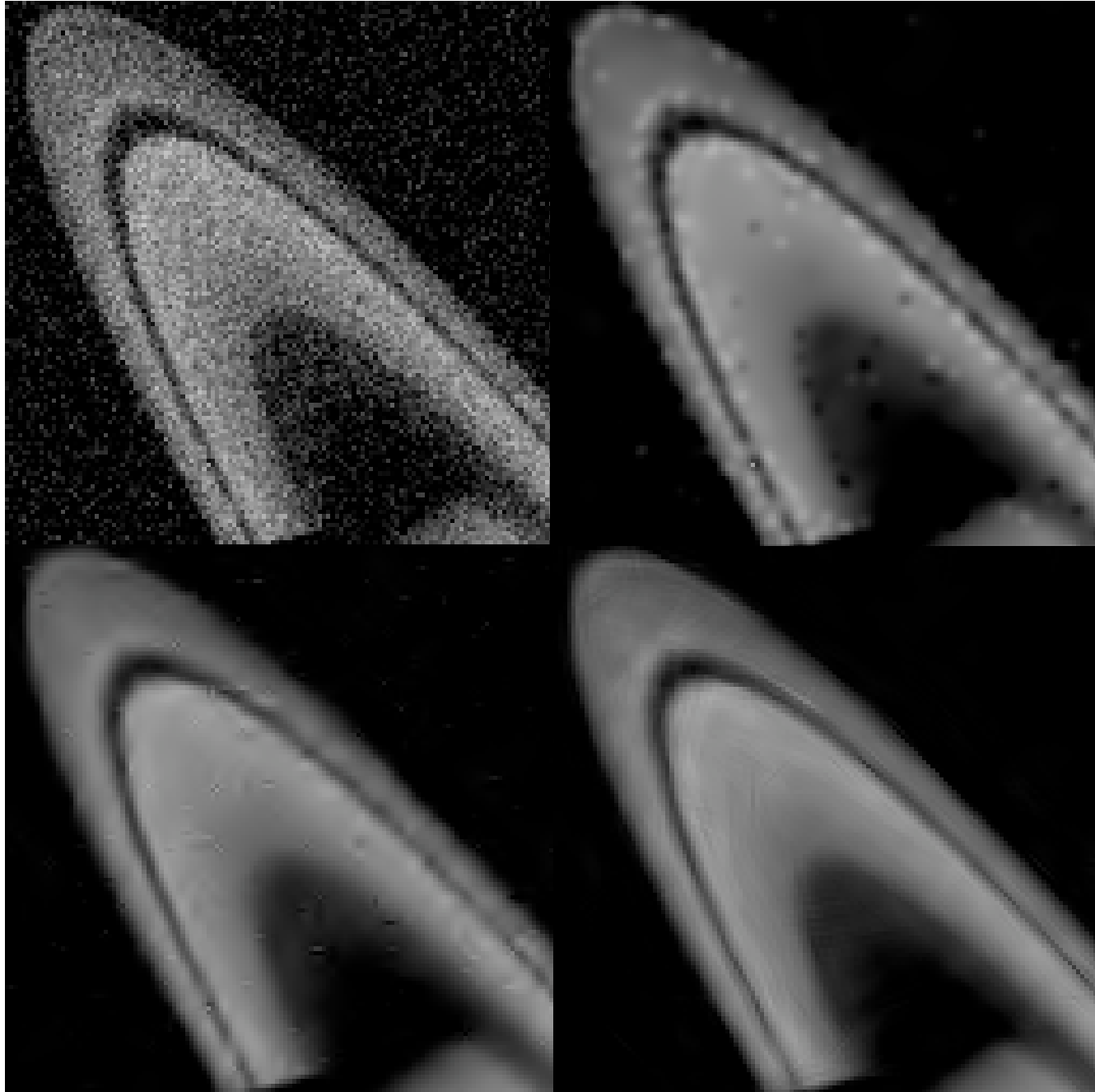
if $P < t$, the coefficient cannot be due to the noise,
and a **significant coefficient** is detected.



$$\tilde{y} = C_R [\delta(C_T y)]$$

Hard Thresholding:

$$\begin{aligned} \delta(c) &= c && \text{if } |c| \geq t \\ &= 0 && \text{if } |c| < t \end{aligned}$$



INVERSE PROBLEMS

$$Y = HX + N$$

PB 1: find X knowing Y, H and the statistical properties of the noise N

Ex: Astronomical image deconvolution

Weak lensing

PB 2: find X and H knowing Y and the statistical properties of the noise N

Ex: Blind deconvolution

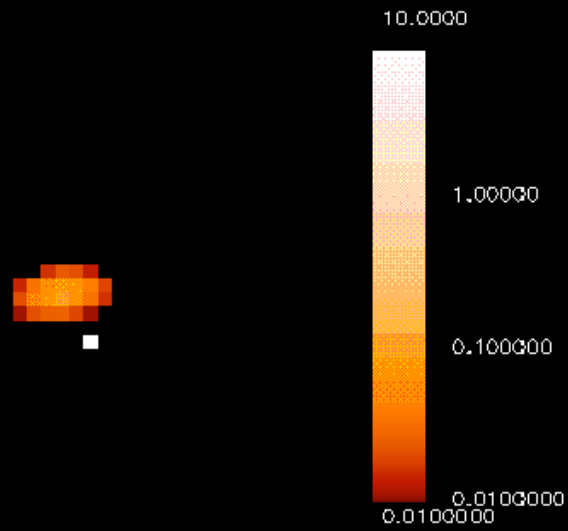
Multichannel Data (PCA, ICA, etc)

Ill posed problem, i.e. not an unique and stable solution \implies Regularization

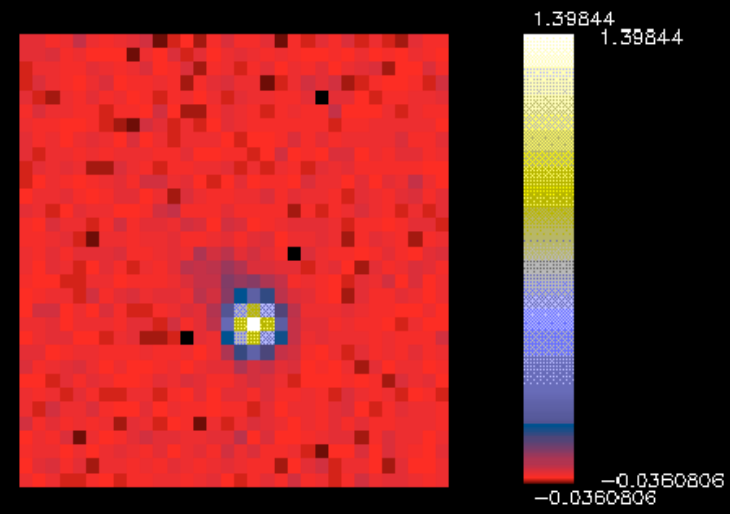
$$\|Y - HX\|^2 \quad \text{with some constraints on } X$$

\implies Sparsity constraint (i.e. $\|X\|_0$)

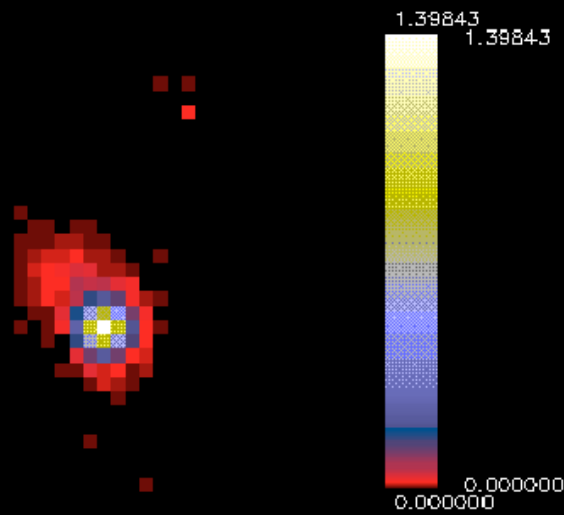
Simulation : faint galaxy nearby a bright star : original



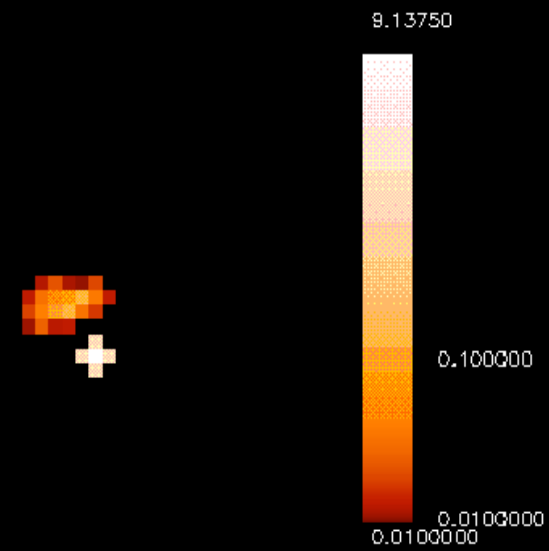
Simulation:weak galax. neara bright *, convolv. with ISOCAM Psf,noise



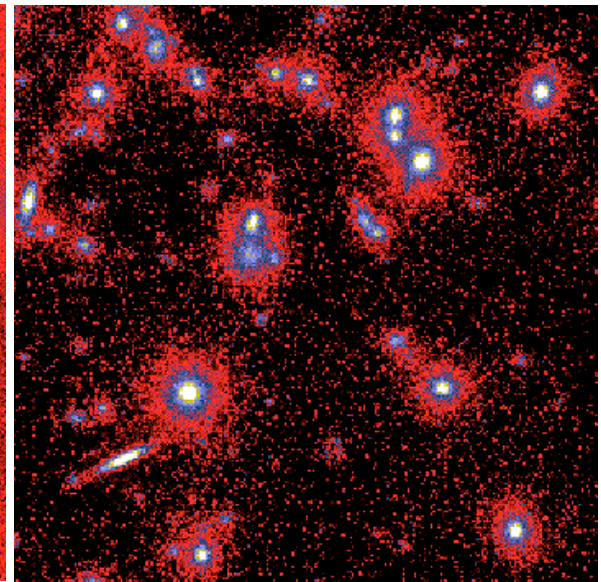
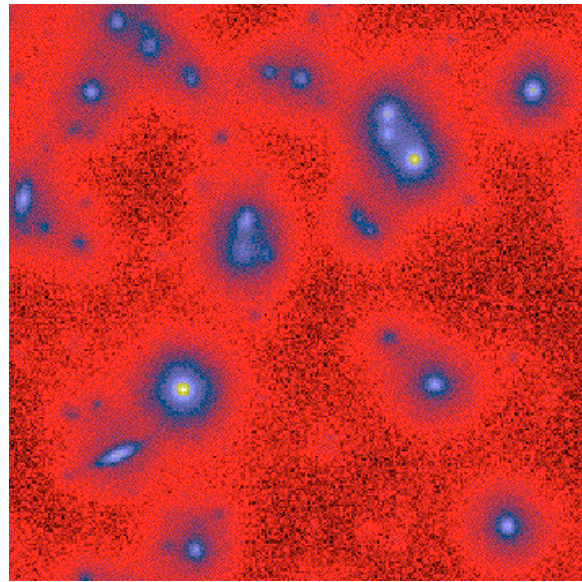
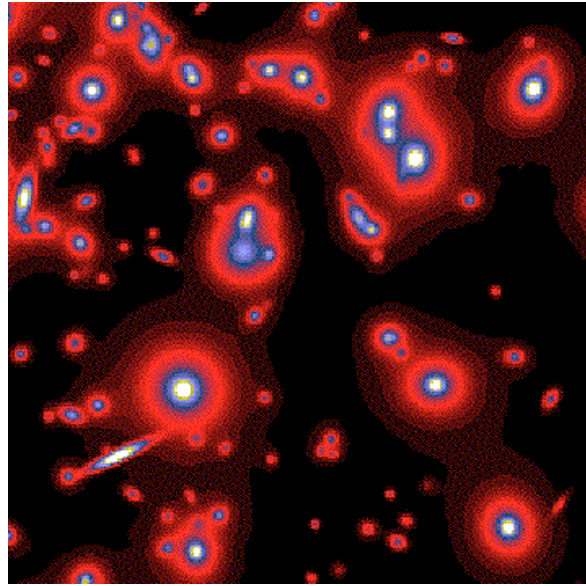
Simulation:weak galax. near a bright * : after filtering



Simulation : faint galaxy nearby a bright star : after deconvolution



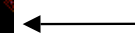
DECONVOLUTION SIMULATION



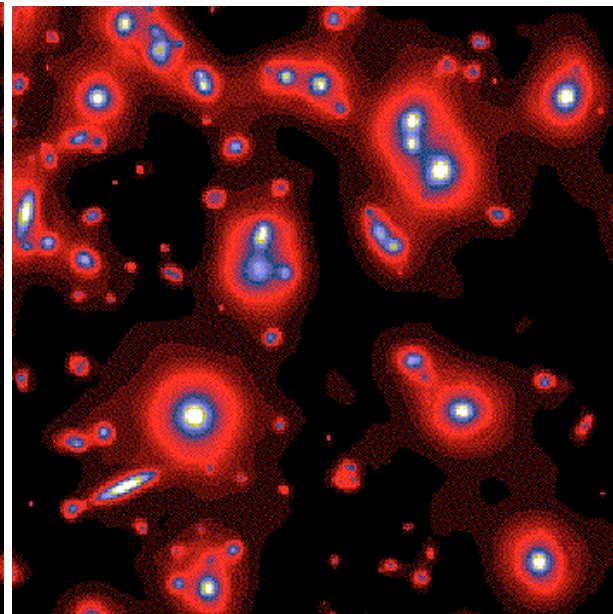
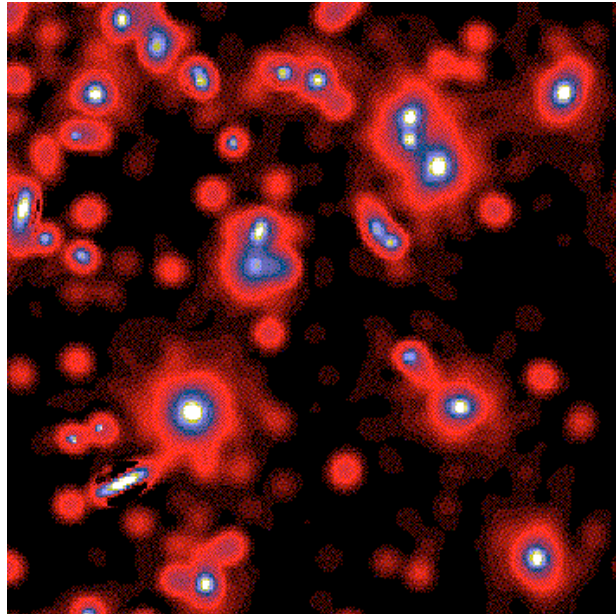
LUCY



Wavelet

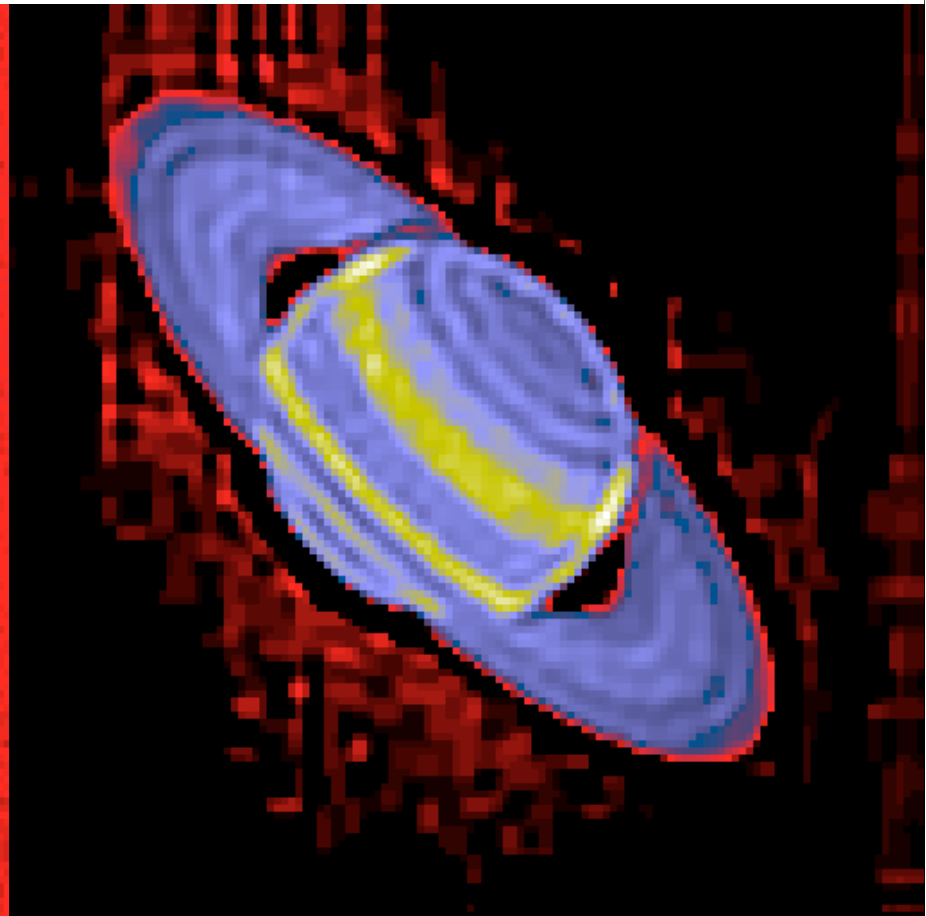
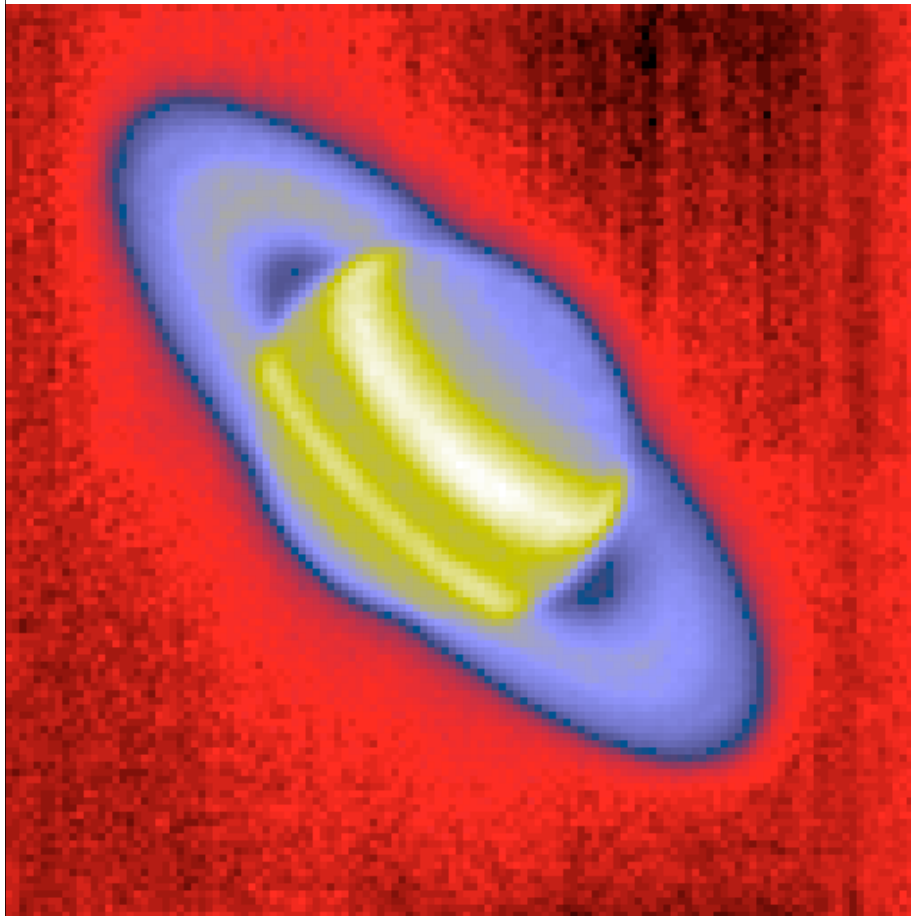
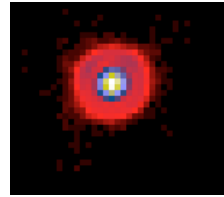


PIXON



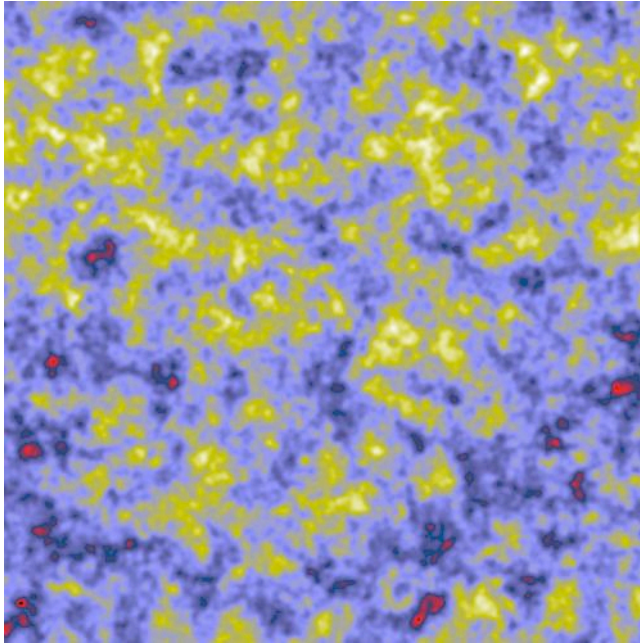


E. Pantin, J.-L. Starck, and F. Murtagh, "Deconvolution and Blind Deconvolution in Astronomy",
in *Blind image deconvolution: theory and applications*,
K. Egiazarian and P. Campisi (Eds), CRC Press, pp 277--317, 2007.

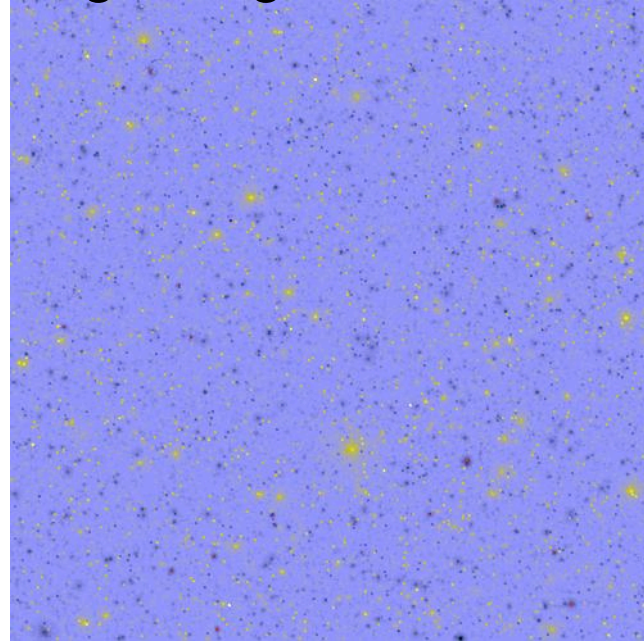


Detection of non-Gaussian Cosmological Signatures

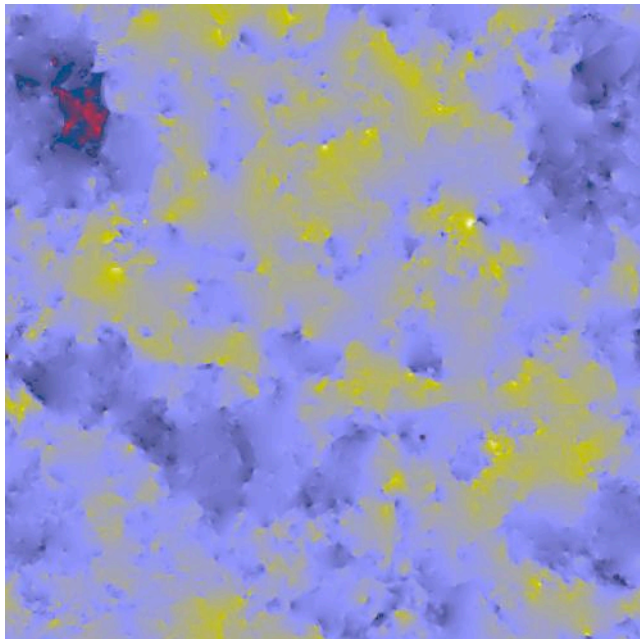
CMB



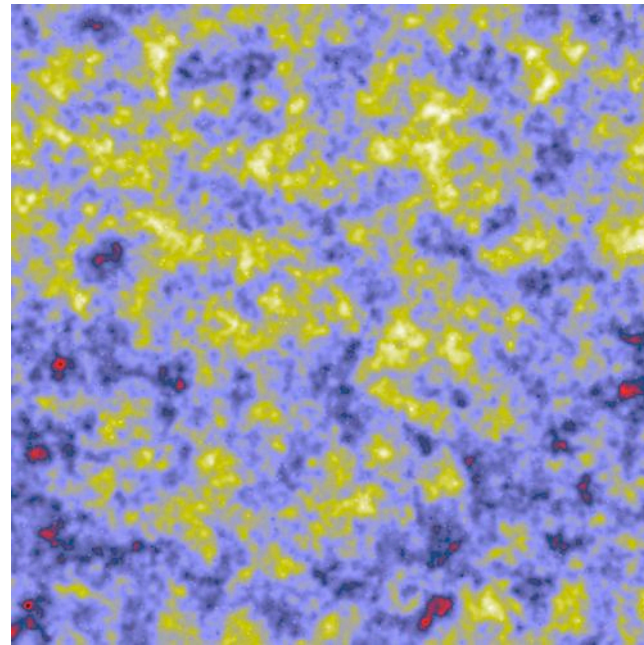
SZ

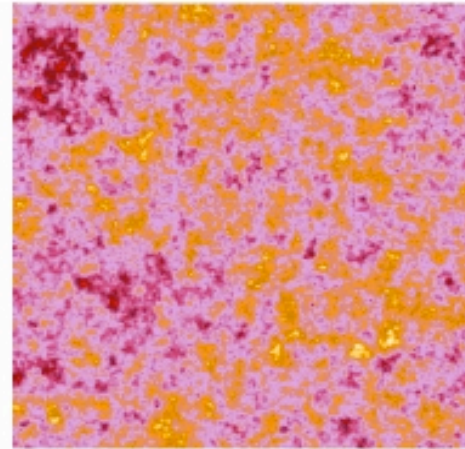
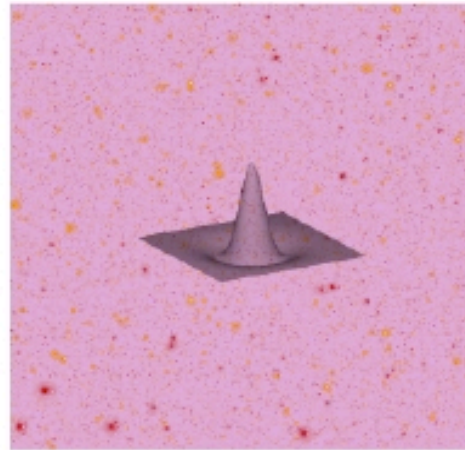
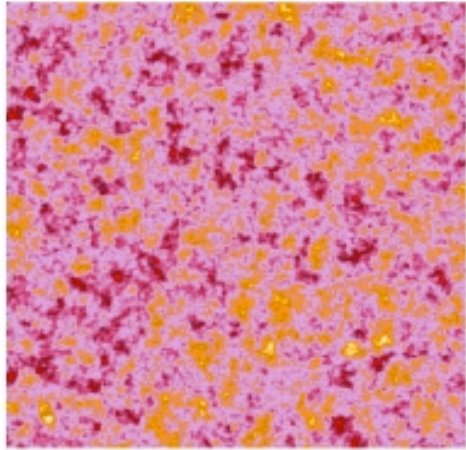


CS



Total





Results



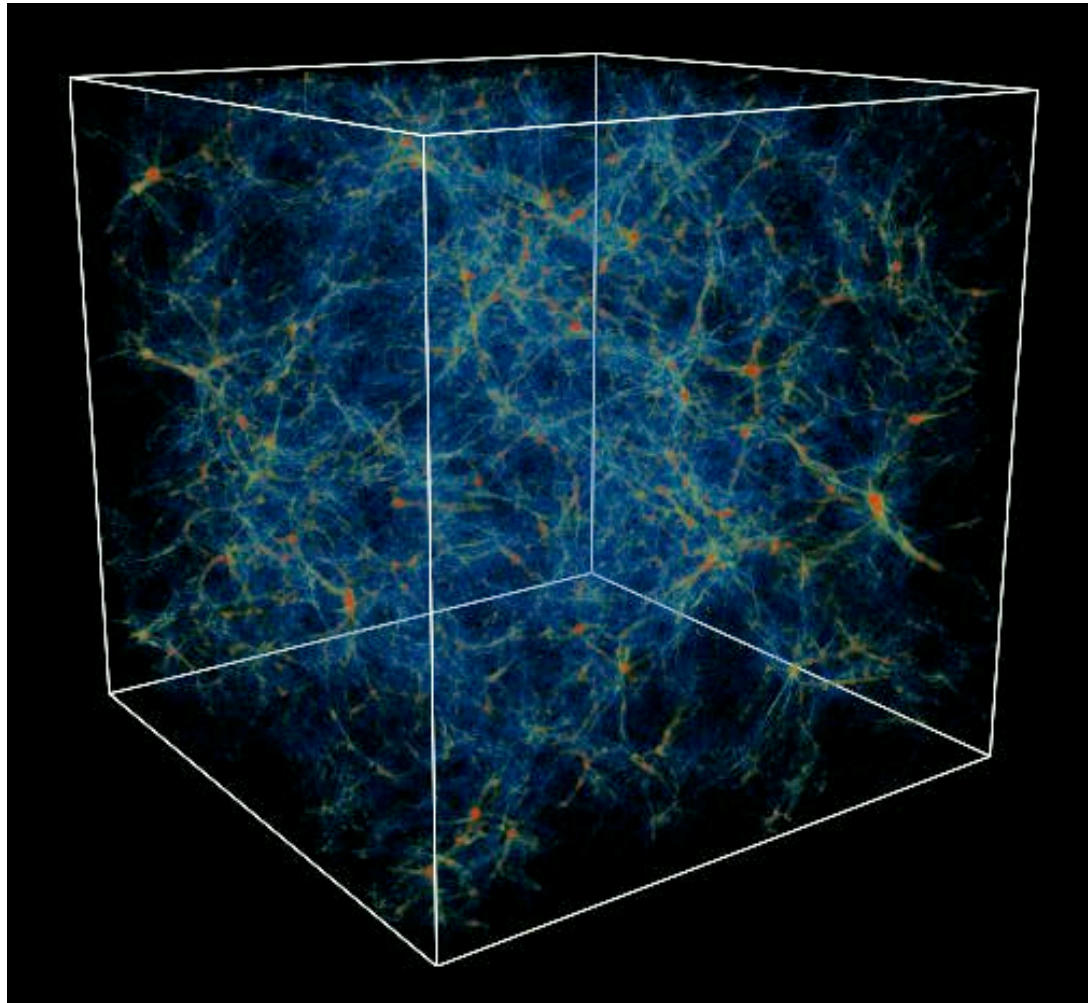
- **Curvelets are NOT sensitive to KSZ but are sensitive to cosmic strings**

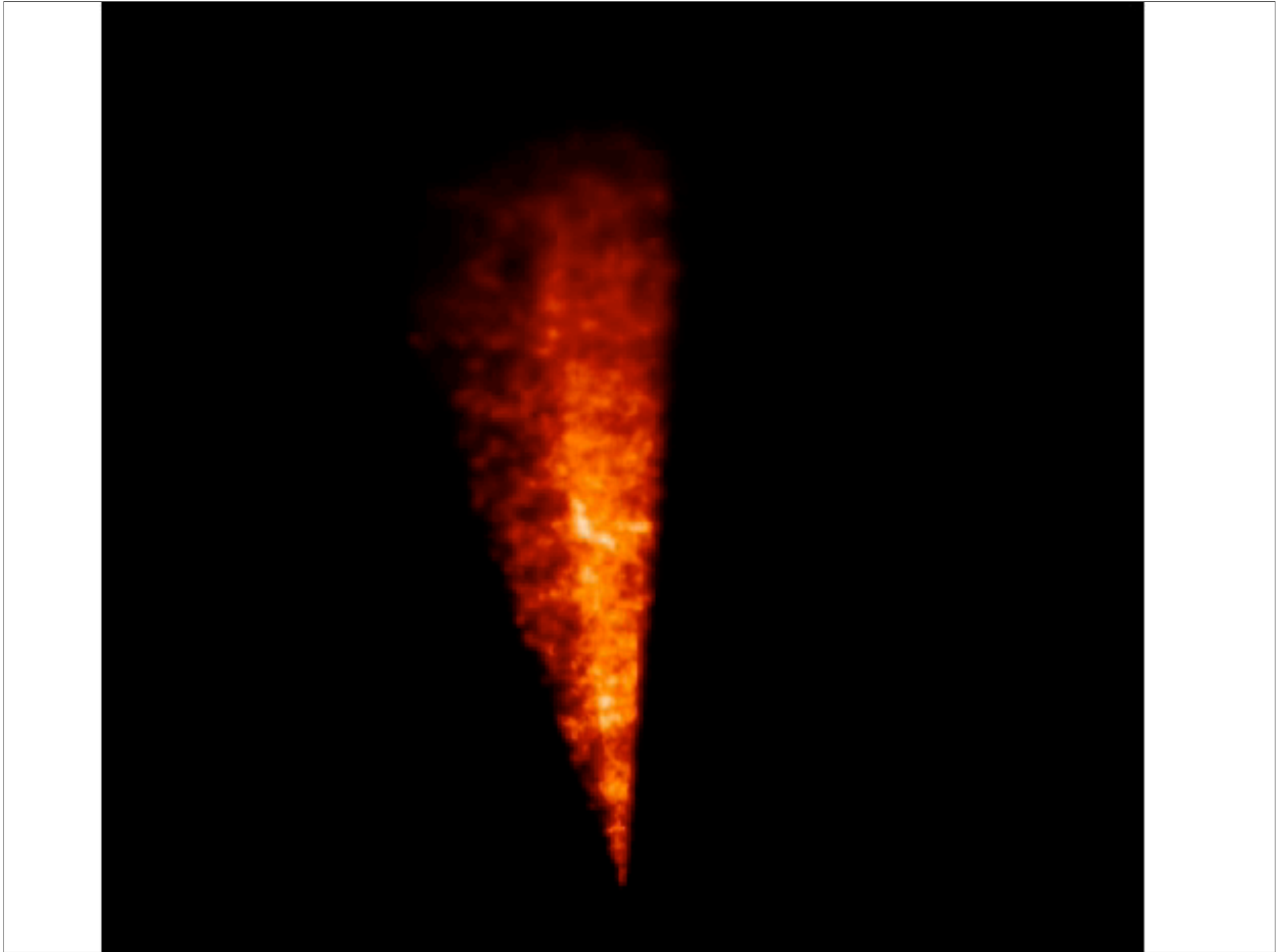
	Bi-orthogonal WT	Ridgelet	Curvelet
CMB+KSZ	1106.	0.1	10.12
CMB+CS	1813.	5.7	198.
CMB+CS+KSZ	1040.	5.9	165.

J.-L. Starck, N. Aghanim and O. Forni, **Detecting cosmological non-Gaussian signatures by multi-scale methods**, Astron. and Astrophys., 416, 9--17, 2004 .

J. Jin, J.-L. Starck, D.L. Donoho, N. Aghanim and O. Forni, **Cosmological Non-Gaussian Signatures Detection: Comparison of Statistical Tests**, Eurasip Journal on Applied Signal Processing, 15 pp 2470-2485, 2005.

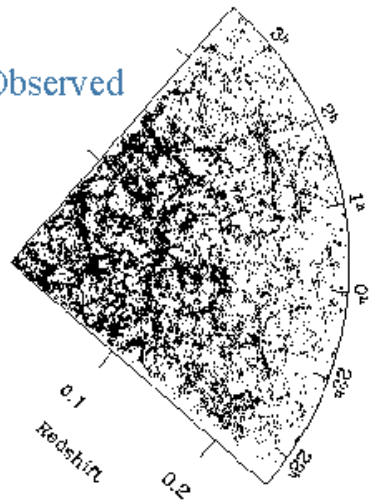
Spatial distribution of the galaxies



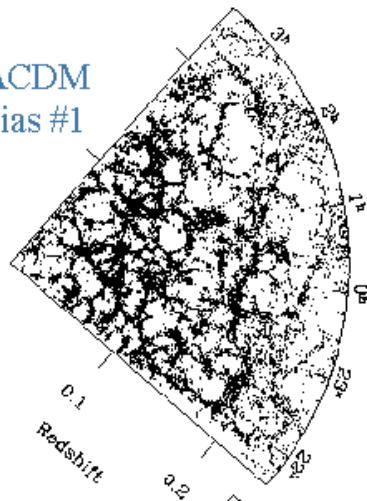


Models vs observations

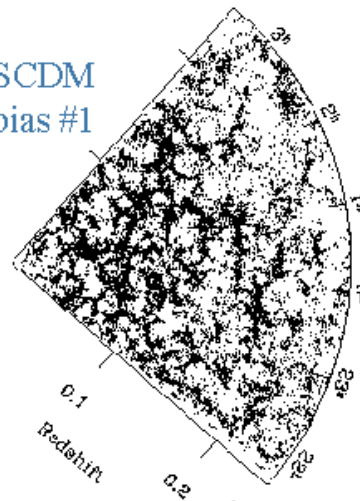
Observed



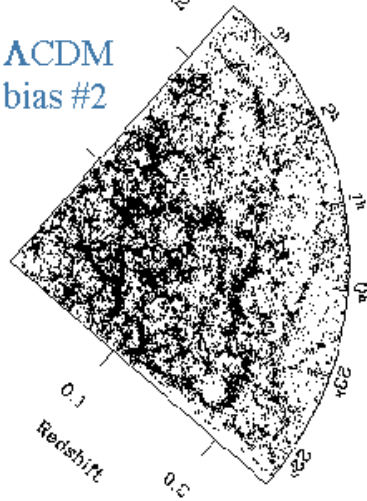
Λ CDM
bias #1



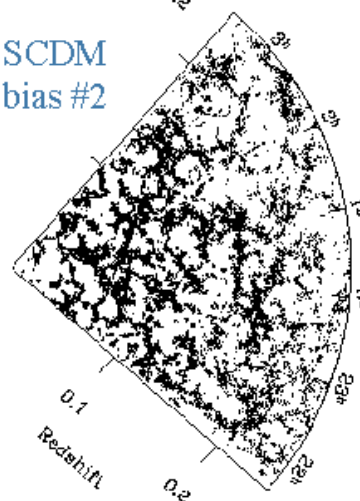
SCDM
bias #1



Λ CDM
bias #2



SCDM
bias #2



The Two-Point Correlation Function

A measure of the deviation from randomness:

$$\xi(r) = \frac{n_{DD}(r)}{n_{RR}(r)} - 1$$

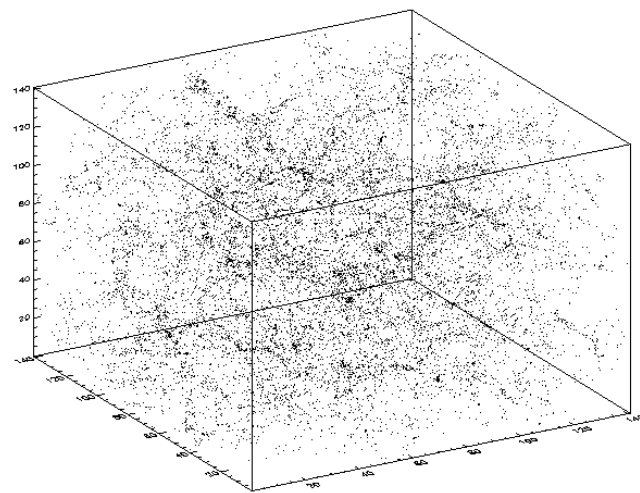
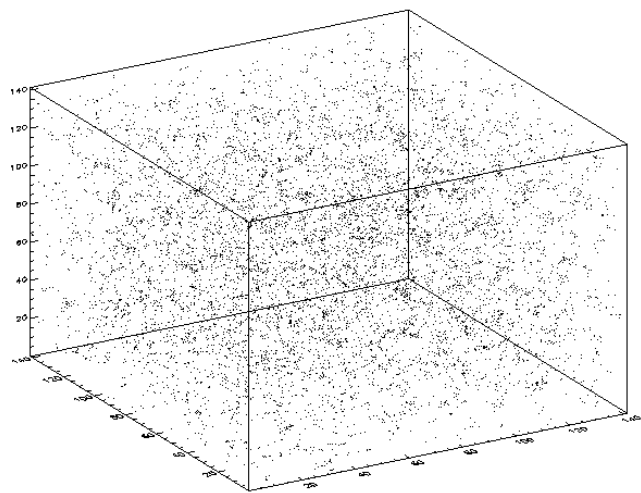
$n_{DD}(r)$ = number of pairs with a separation of r in the data

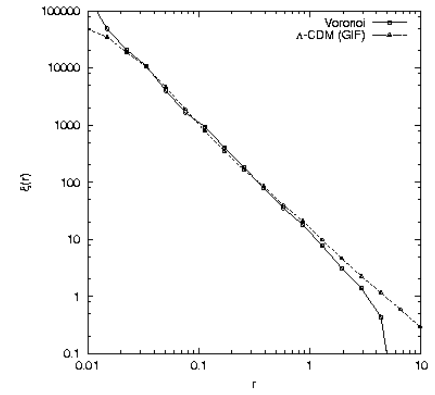
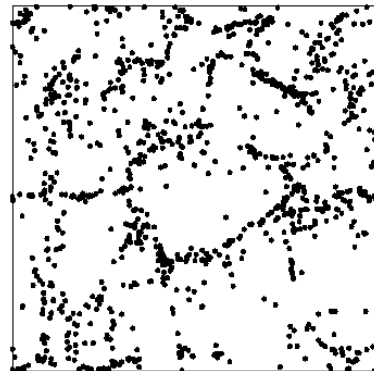
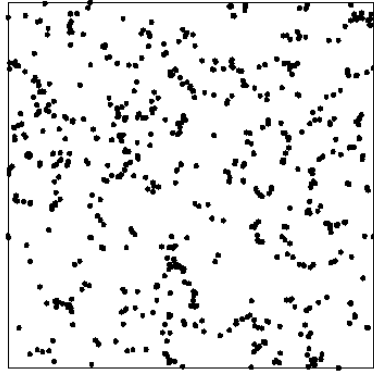
$n_{RR}(r)$ = number of pairs with a separation of r for a randomly distributed data set

Methods

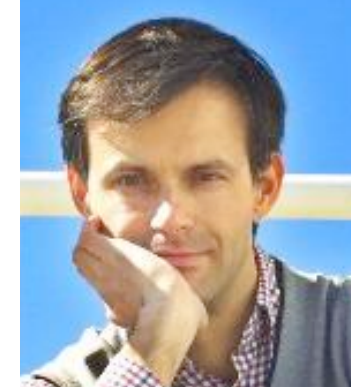
- . Two or three point correlation function**
- . Genus curve**
- . Voronoi Tessellation**
- . Minimal spanning trees**
- . Power spectrum**
- . Fractals**

Simulations





3D MULTISCALE TRANSFORMS



- 1) **3D WAVELET TRANSFORM: Isotropic Structures**
- 2) **3D RIDGELET TRANSFORM: Sheet like Structures**
- 3) **3D BEAMLET TRANSFORM: Filaments**

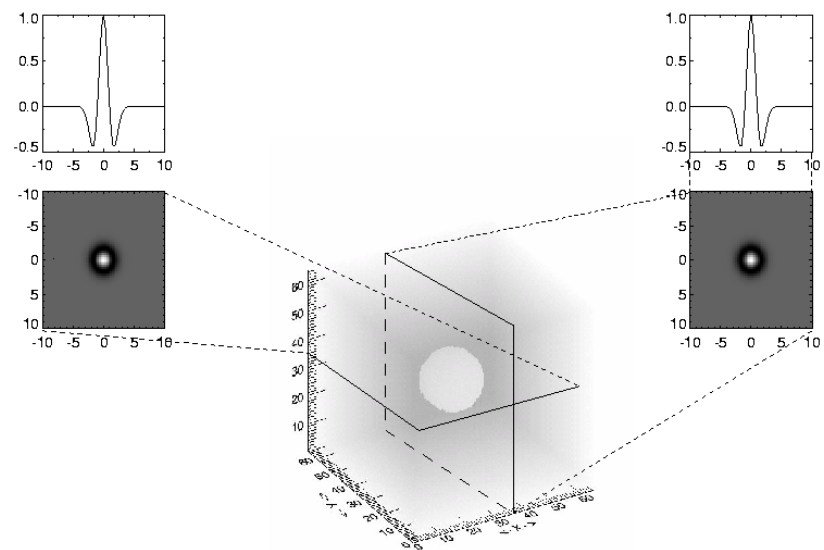
⇒ **Statistical information extraction from all transforms**

◦J.-L. Starck, V.J. Martinez, D.L. Donoho, O. Levi, P. Querre, E. Saar, **Analysis of the spatial distribution of galaxies by multiscale methods**, Eurasip Journal on Applied Signal Processing, 15, pp 2455-2469, 2005.

◦V. Martinez, J.-L. Starck, E. Saar, D.L. Donoho, P. de la Cruz, S. Paredes and S. Reynolds, **"Morphology of the Galaxy Distribution from Wavelet Denoising"**, ApJ, 634, pp 744--755, 2005.

◦E. Saar, V. martinez, J.-L. Starck and D. Donoho, **"Multi-scale morphology of the galaxy distribution"**, MNRAS , 374, 1030-1044, 2007.

3D Wavelet Function



3D Ridgelet Transform

The three-dimensional continuous ridgelet transform of a function $f \in L^2(\mathbf{R}^3)$ is given by:

$$\mathcal{R}_f : \mathbf{R}^4 \rightarrow \mathbf{R}$$

$$\mathcal{R}_f(a, b, \theta_1, \theta_2) = \int \overline{\psi}_{a,b,\theta_1,\theta_2}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

where $a > 0$, $b \in \mathbf{R}$, $\theta_1 \in [0, 2\pi[$ and $\theta_2 \in [0, \pi[$.

The *ridgelet* function is defined by:

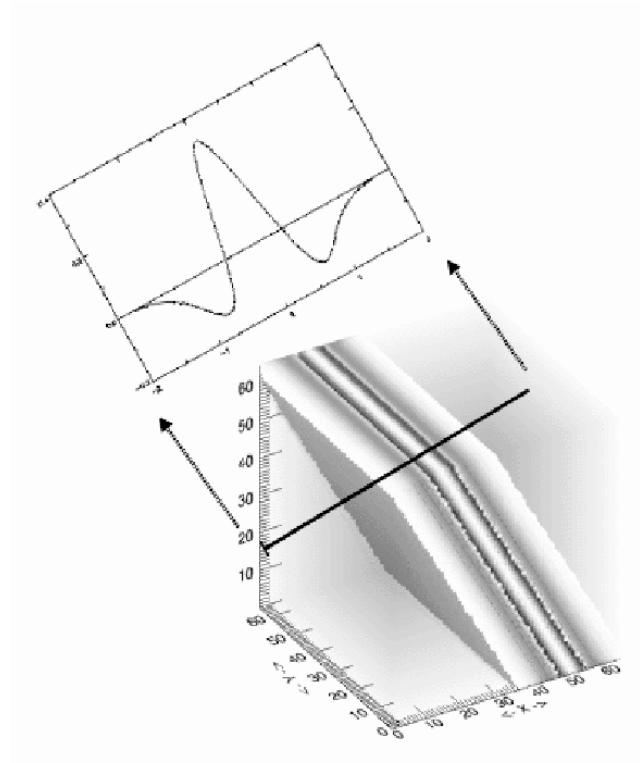
$$\psi_{a,b,\theta_1,\theta_2} :$$

$$\mathbf{R}^3 \rightarrow \mathbf{R}$$

$$\psi_{a,b,\theta_1,\theta_2}(x_1, x_2, x_3) =$$

$$\psi((x_1 \cos \theta_1 \cos \theta_2 + a^{-1/2} x_2 \sin \theta_1 \cos \theta_2 + x_3 \sin \theta_2 - b)/a);$$

3D Ridgelet Function



3D Beamlet Transform

the three-dimensional continuous beamlet transform of a function $f \in L^2(\mathbf{R}^3)$ is given by:

$$\mathcal{B}_f : \mathbf{R}^5 \rightarrow \mathbf{R}$$

$$\mathcal{B}_f(a, b_1, b_2, \theta_1, \theta_2) = \int \overline{\psi}_{a,b,\theta_1,\theta_2}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}.$$

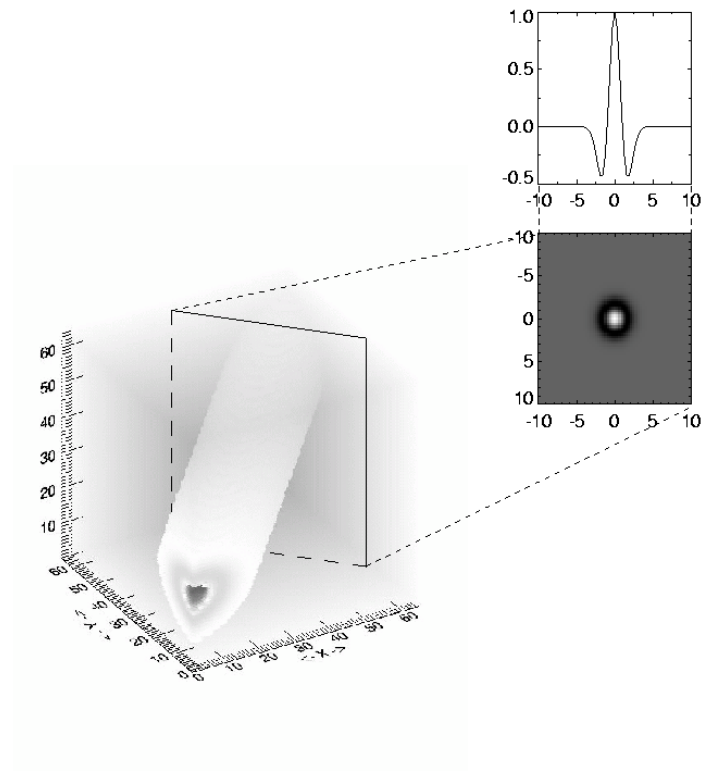
where $a > 0$, $b_1, b_2 \in \mathbf{R}$, $\theta_1 \in [0, 2\pi[$ and $\theta_2 \in [0, \pi[$.

The *beamlet* function is defined by:

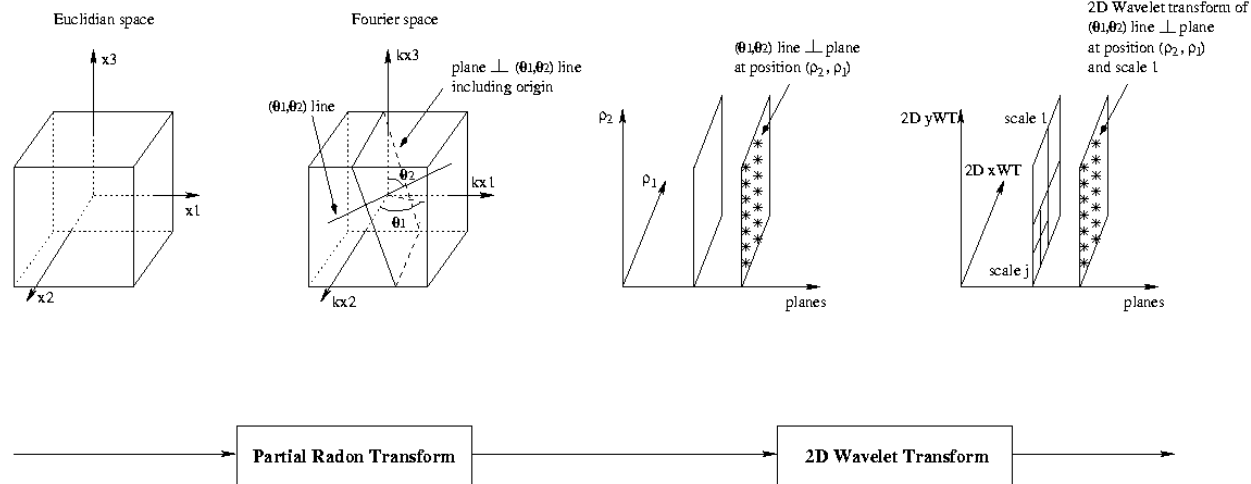
$$\psi_{a,b_1,b_2,\theta_1,\theta_2} : \mathbf{R}^3 \rightarrow \mathbf{R}$$

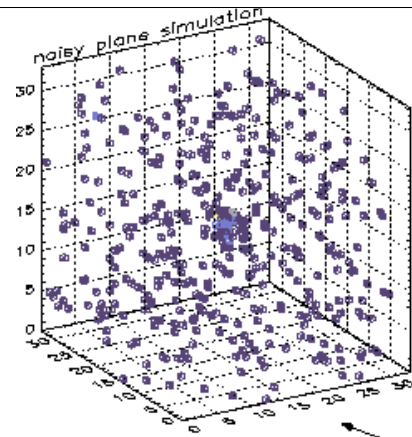
$$\psi_{a,b_1,b_2,\theta_1,\theta_2}(x_1, x_2, x_3) = a^{-1/2} \psi\left(\frac{-x_1 \sin \theta_1 + x_2 \cos \theta_1 + b_1}{a}, \frac{x_1 \cos \theta_1 \cos \theta_2 + x_2 \sin \theta_1 \cos \theta_2 - x_3 \sin \theta_2 + b_2}{a}\right);$$

3D Beamlet Function

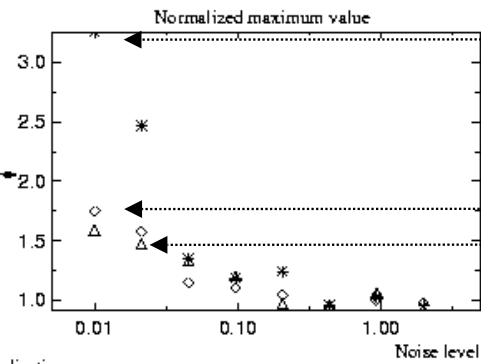


3D beamlet transform flowgraph



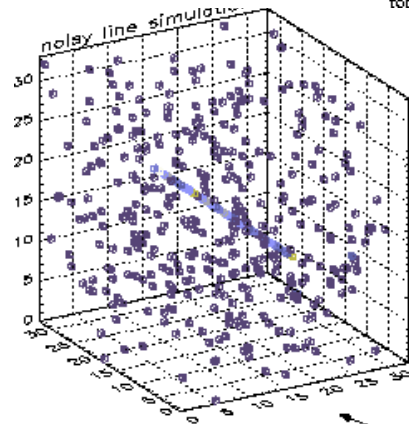


Poisson realisation for a low noise level

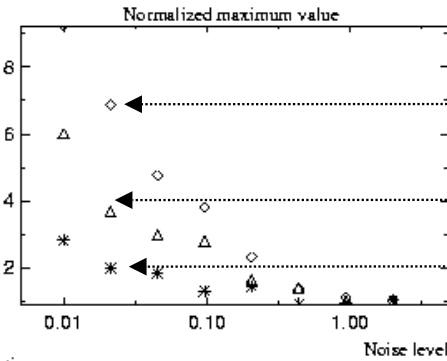


Wavelet

Beamlet
Ridgelet

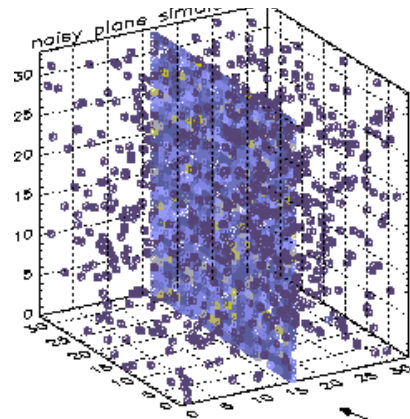


Poisson realisation for a low noise level

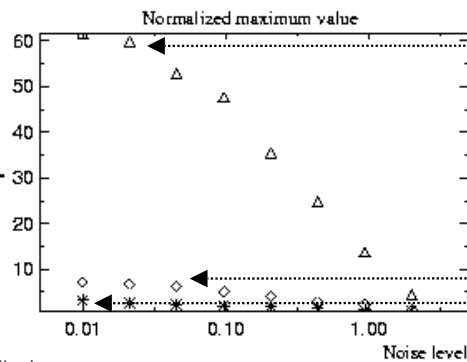


Beamlet

Ridgelet
Wavelet



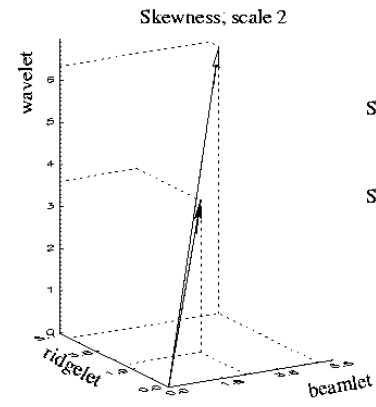
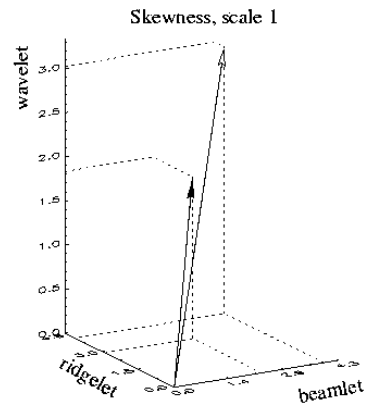
Poisson realisation for a low noise level



Ridgelet

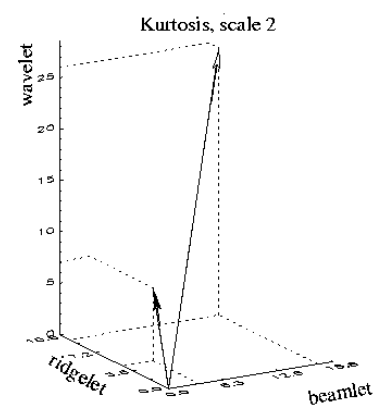
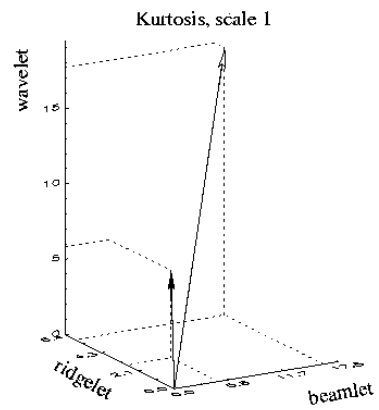
Beamlet
Wavelet

Skewness and Kurtosis



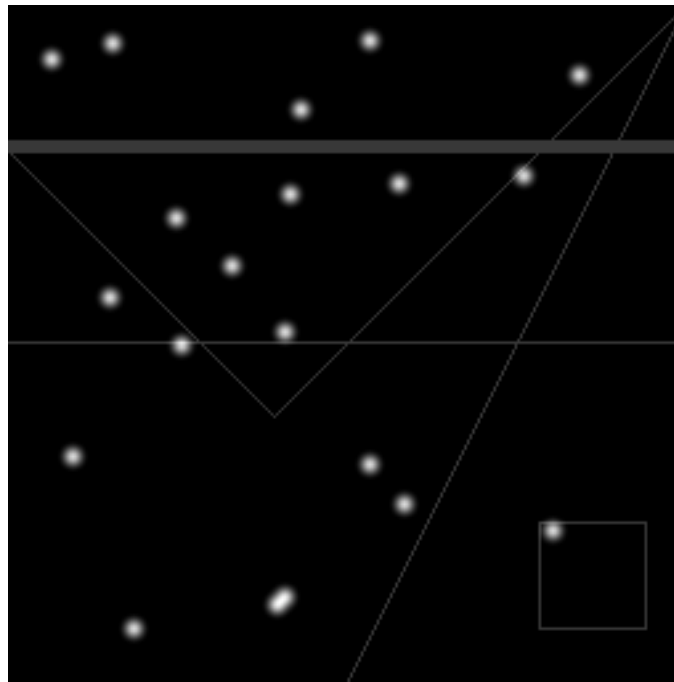
Simulated file 1 ↑

Simulated file 2 ↑

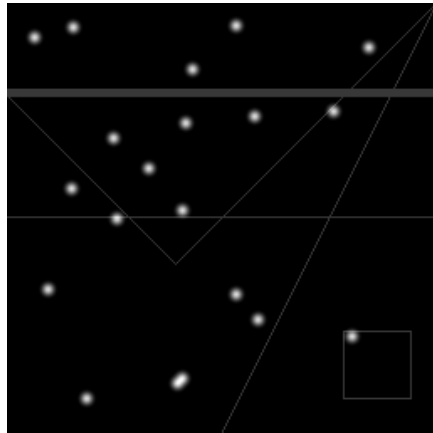


A difficult issue

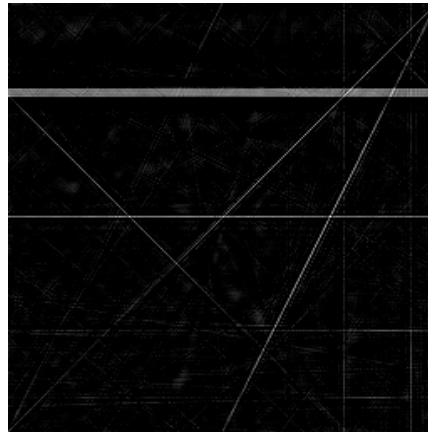
Is there any representation that well represents the following image ?



Going further

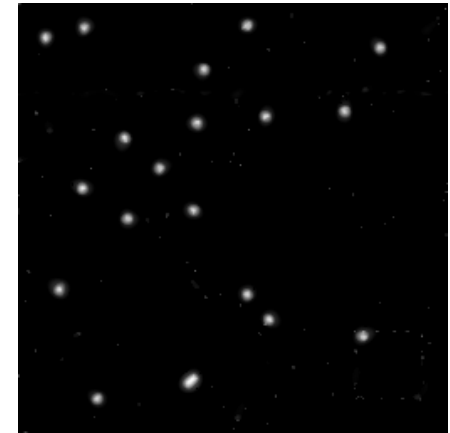


=



Lines

+



Gaussians



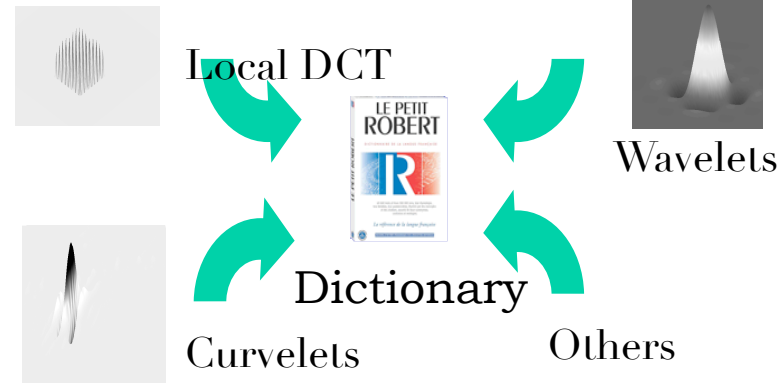
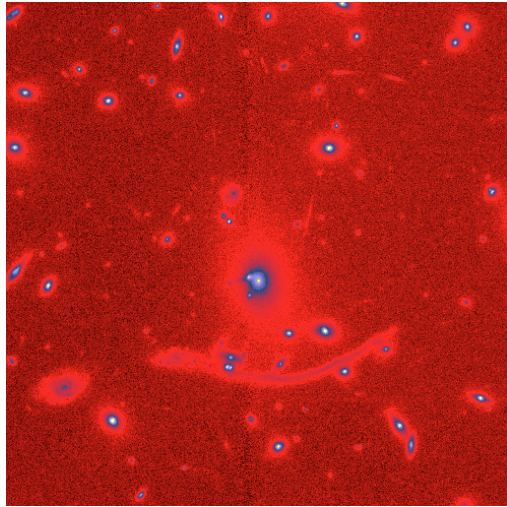
Curvelets



Wavelets

REDUNDANT REPRESENTATIONS

Morphological Diversity



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

Model:

$$s = \sum_{k=1}^L s_k + n$$

and s_k ($s_k = \phi_k \alpha_k$) is sparse in ϕ_k .



New Perspectives



Morphological Component Analysis (MCA)

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

$$J(s_1, \dots, s_L) = \left\| s - \sum_{k=1}^L s_k \right\|_2^2 + \lambda \sum_{k=1}^L \|T_k s_k\|_p$$

Morphological Component Analysis (MCA)

$$J(s_1, \dots, s_L) = \left\| s - \sum_{k=1}^L s_k \right\|_2^2 + \lambda \sum_{k=1}^L \|T_k s_k\|_p$$

- Initialize all s_k to zero
- Iterate $j=1, \dots, Niter$
 - Iterate $k=1, \dots, L$

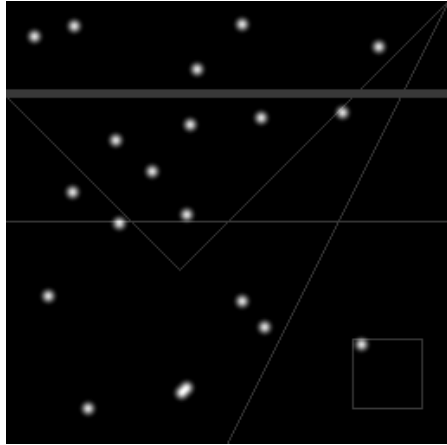
Update the k th part of the current solution by fixing all other parts and minimizing:

$$J(s_k) = \left\| s - \sum_{i=1, i \neq k}^L s_i - s_k \right\|_2^2 + \lambda^{(j)} \|T_k s_k\|_p$$

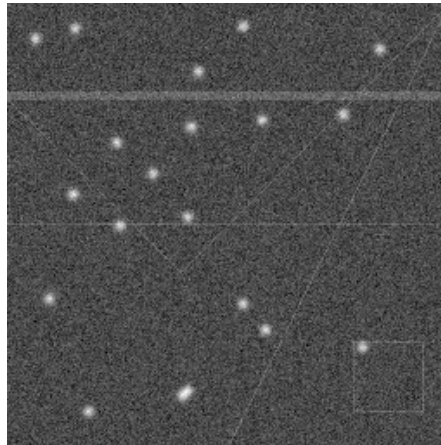
Which is obtained by a simple **hard**/soft thresholding of: $s_r = s - \sum_{i=1, i \neq k}^L s_i$

- Decrease the threshold $\lambda^{(j)}$

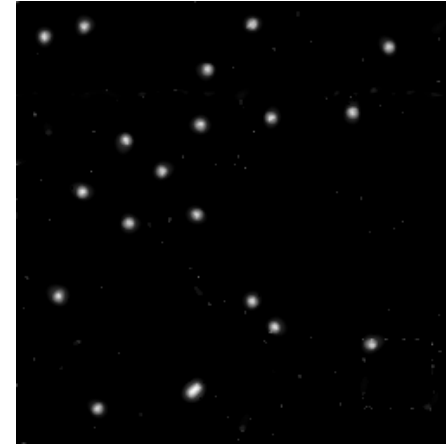
$$\text{MIN}_{s_1, s_2} (\|Ws_1\|_p + \|Cs_2\|_p) \text{ subject to } \|s - (s_1 + s_2)\|_2^2 < \varepsilon$$



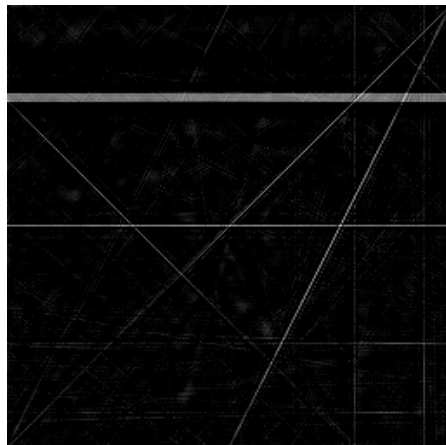
a) Simulated image (gaussians+lines)



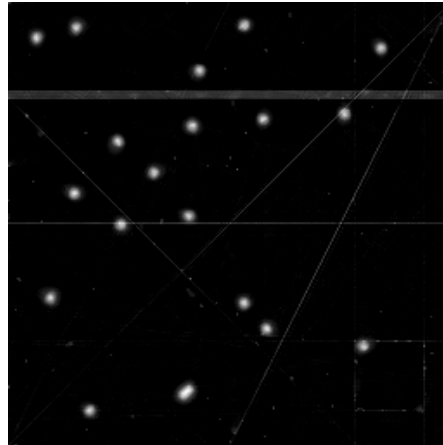
b) Simulated image + noise



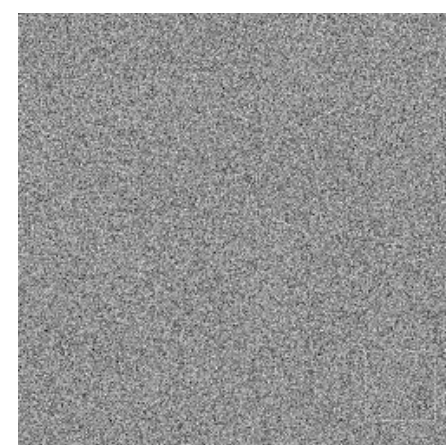
c) A trous algorithm



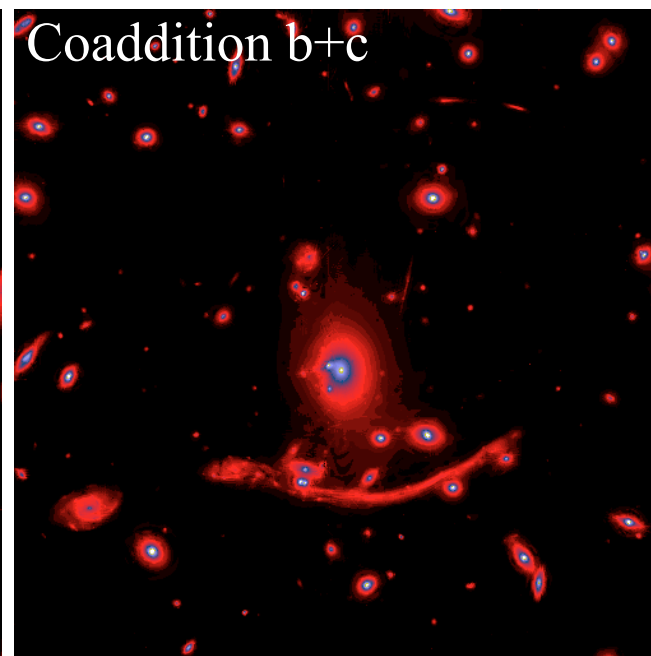
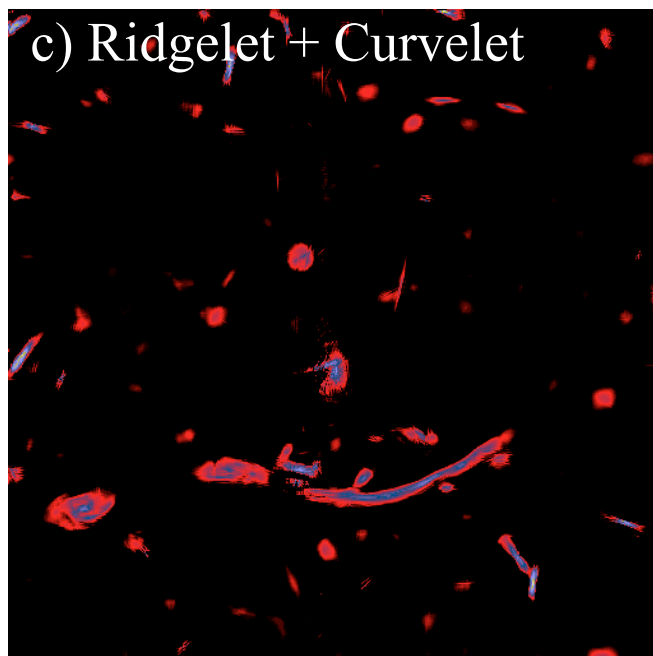
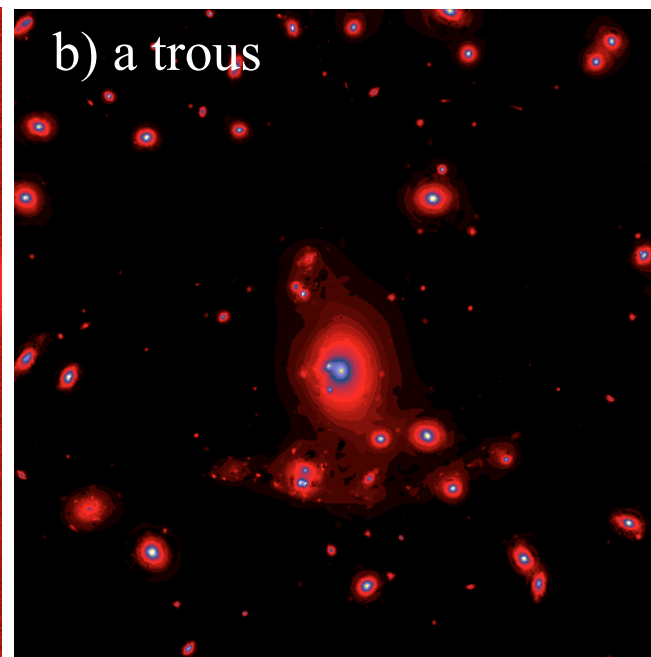
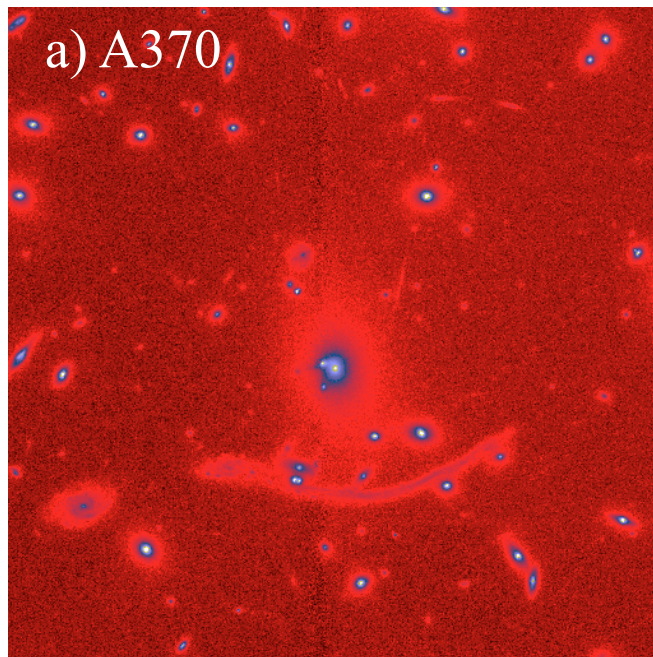
d) Curvelet transform

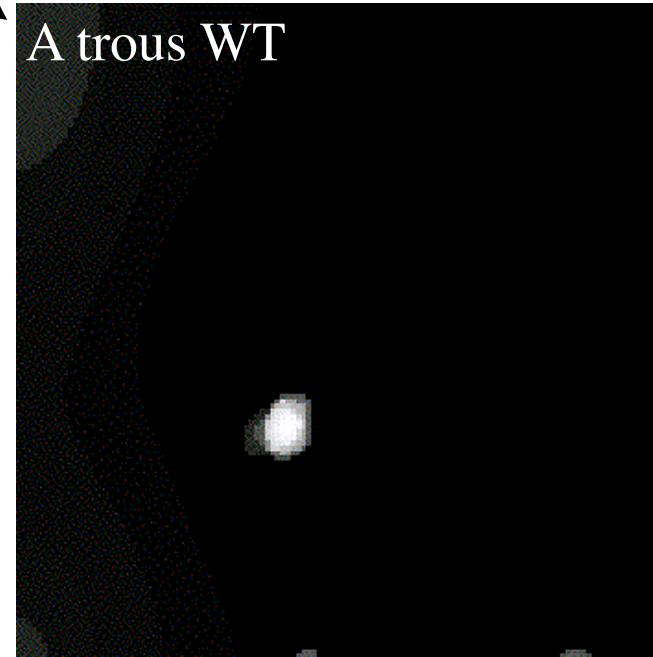
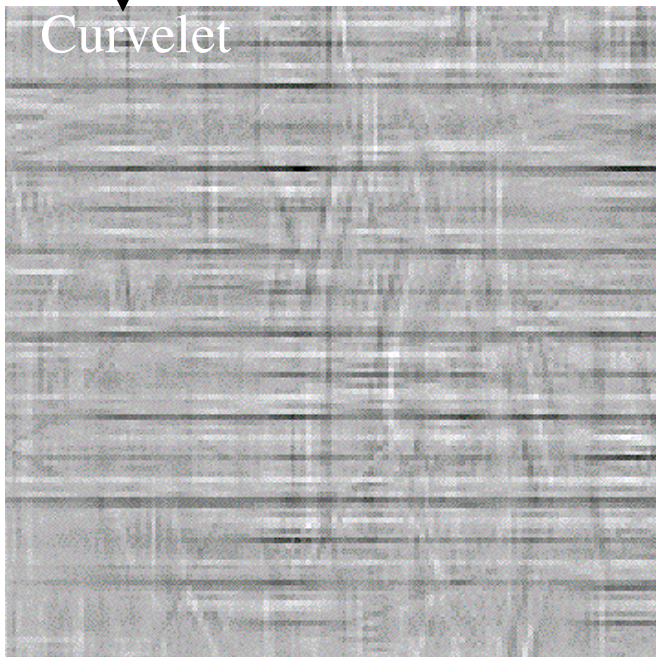


e) coaddition c+d



f) residual = e-b





Galaxy SBS 0335-052
10 micron
GEMINI-OSCIR



Interpolation of Missing Data

$$J(s_1, \dots, s_L) = \left\| M \left(s - \sum_{k=1}^L s_k \right) \right\|_2^2 + \lambda \sum_{k=1}^L \| T_k s_k \|_p$$

Where M is the mask: $M(i,j) = 0 \implies$ missing data

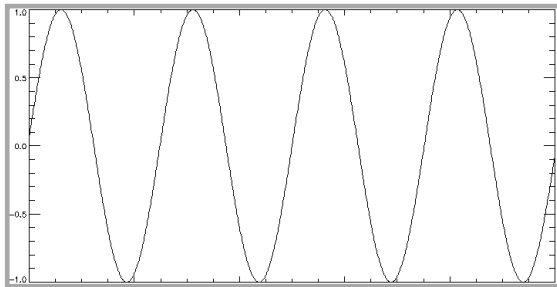
$M(i,j) = 1 \implies$ good data

If the data are composed of a piecewise smooth component + texture

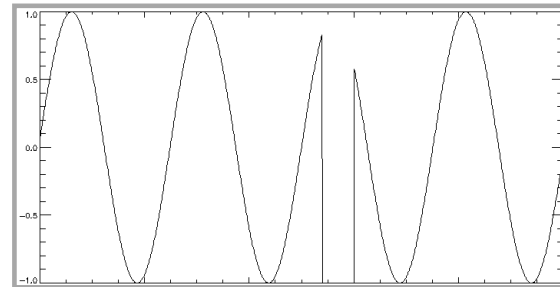
$$J(X_t, X_n) = \left\| M(X - X_t - X_n) \right\|_2^2 + \lambda (\| \mathbf{C} X_n \|_1 + \| \mathbf{D} X_t \|_1) + \gamma \text{TV}(X_n)$$

•M. Elad, J.-L. Starck, D.L. Donoho, P. Querre, "Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA)", *ACHA*, Vol. 19, pp. 340-358, November 2005.

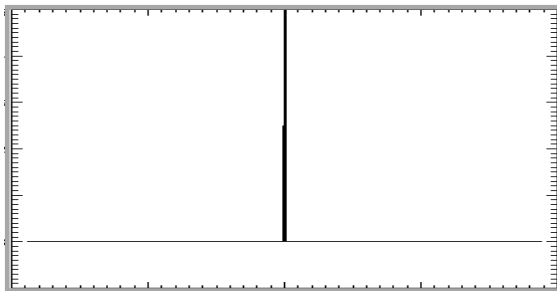
•M.J. Fadili, J.-L. Starck and F. Murtagh, "Inpainting and Zooming using Sparse Representations", *in press*.



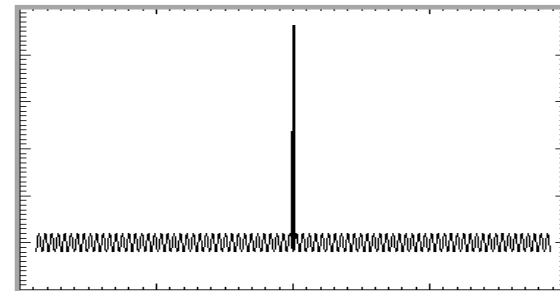
Sine curve



Truncated sine curve



TF of a sine curve



TF of a truncated sine curve

. Initialize all s_k to zero

. Iterate $j=1, \dots, \text{Niter}$

- Iterate $k=1, \dots, L$

- Update the k th part of the current solution by fixing all other parts and minimizing:

$$J(s_k) = \left\| M\left(s - \sum_{i=1, i \neq k}^L s_i - s_k\right) \right\|_2^2 + \lambda \|T_k s_k\|_1$$

Which is obtained by a simple soft thresholding of :

$$s_r = M\left(s - \sum_{i=1, i \neq k}^L s_i\right)$$

20%



50%



80%



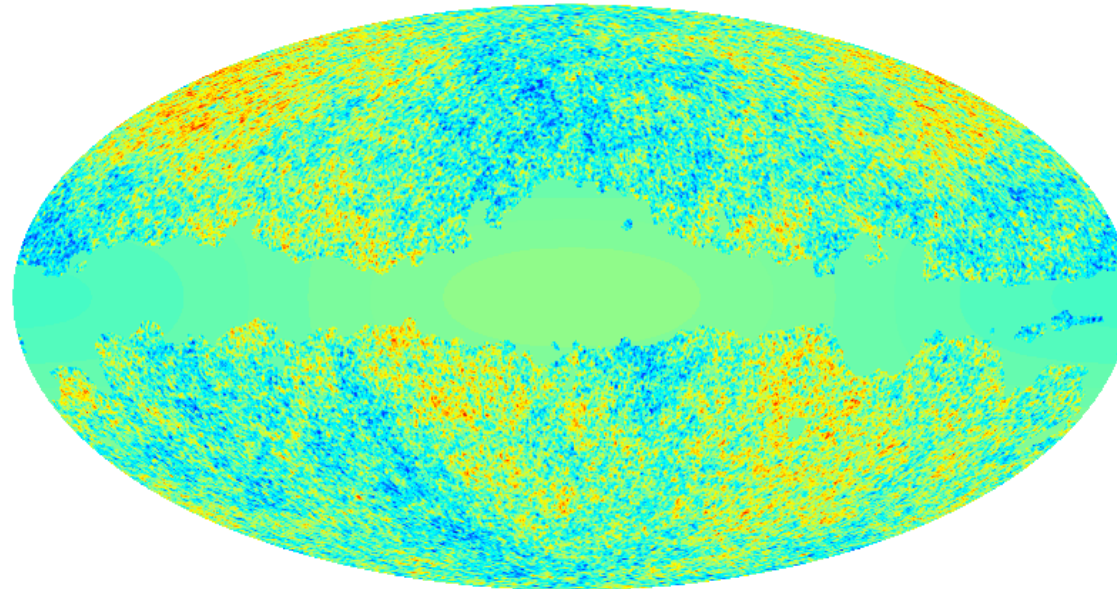


Inpainted with the curvelet dictionary (80% data missing)



Jalal Fadili's web page (<http://www.greyc.ensicaen.fr/~jfadili>).

WHY INPAINTING IS USEFUL FOR THE CMB ?



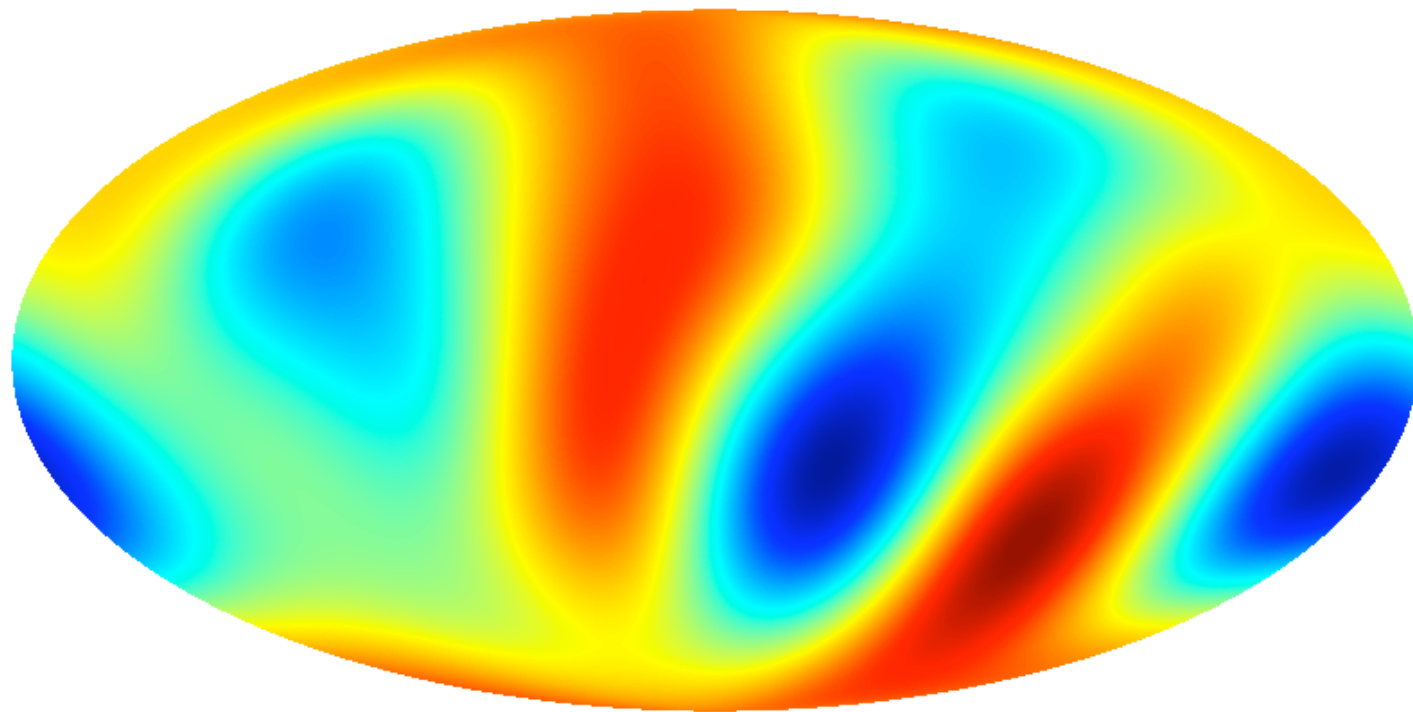
-1.37e+03  +1.42e+03

- Point sources removal for preprocessing
- Gaussianity test.
- Power estimation with the minimum of correlation.
- Any analysis where the mask is a problem.

▪ P. Abrial, Y. Moudden, J.L. Starck, M.J. Fadili, J. Delabrouille, and M. Nguyen, *CMB Data Analysis and Sparsity*, Statistical Methodology, in press

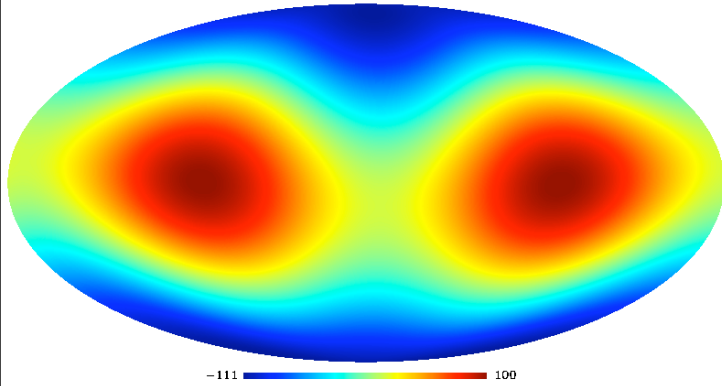
▪ P. Abrial, Y. Moudden, J.L. Starck, et al., "*Morphological Component Analysis and inpainting on the Sphere: Application in Physics and Astrophysics*", Journal of Fourier Analysis and Applications (JFAA), 13, 6, pp 729-748, 2007.

WMAP inpainting Scale 7

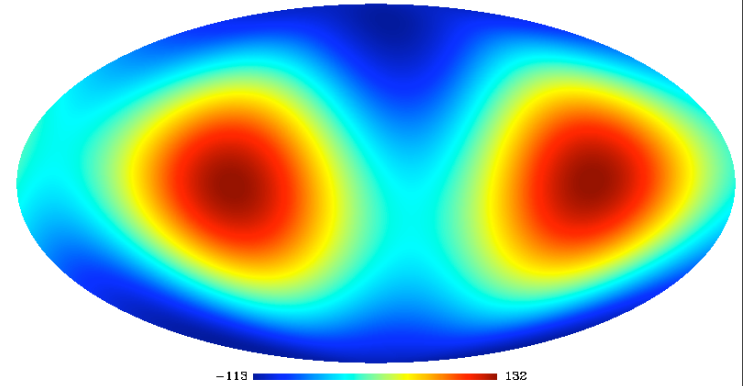


-0.022 0.017

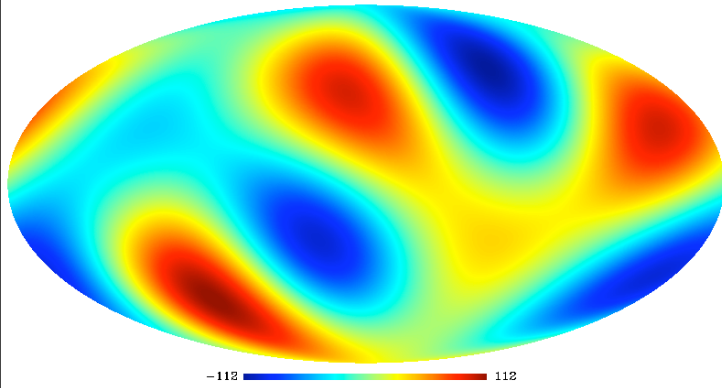
Simulated Data: l=2



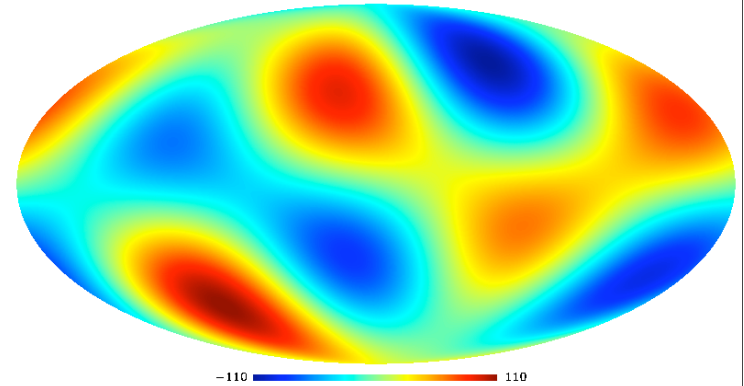
Simulated data (inpainting): l=2



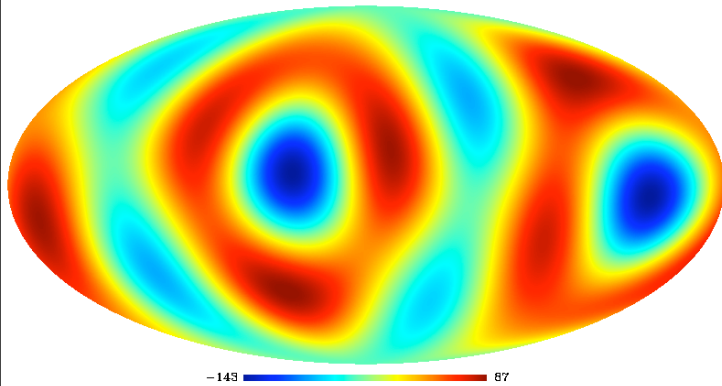
Simulated Data: l=3



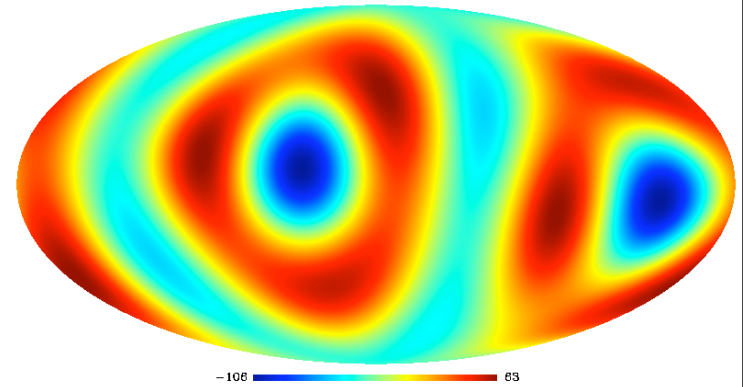
Simulated data (inpainting): l=3



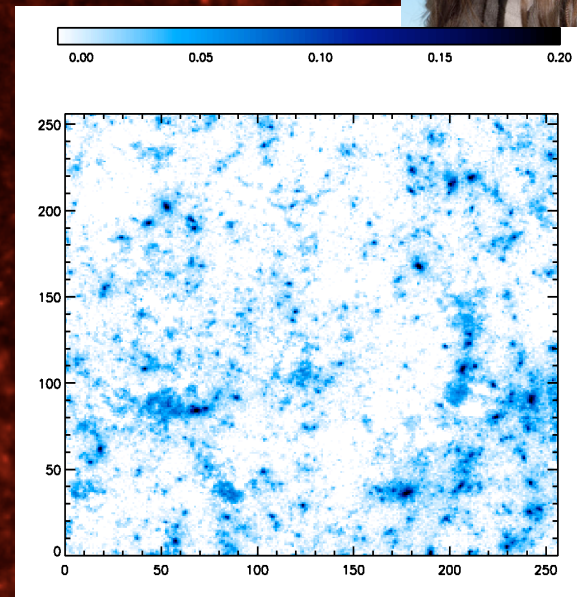
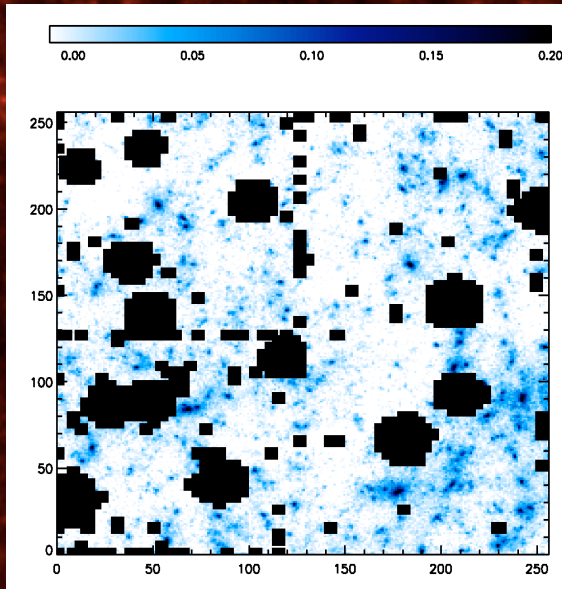
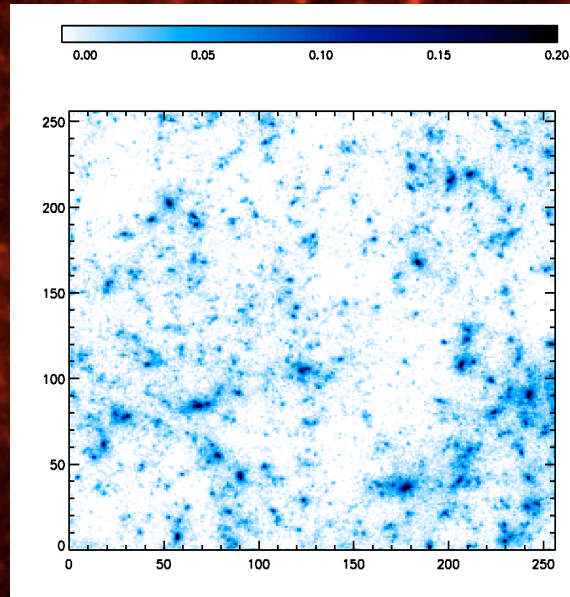
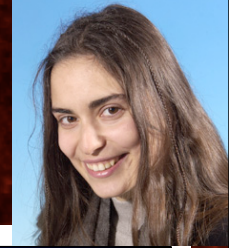
Simulated Data: l=4



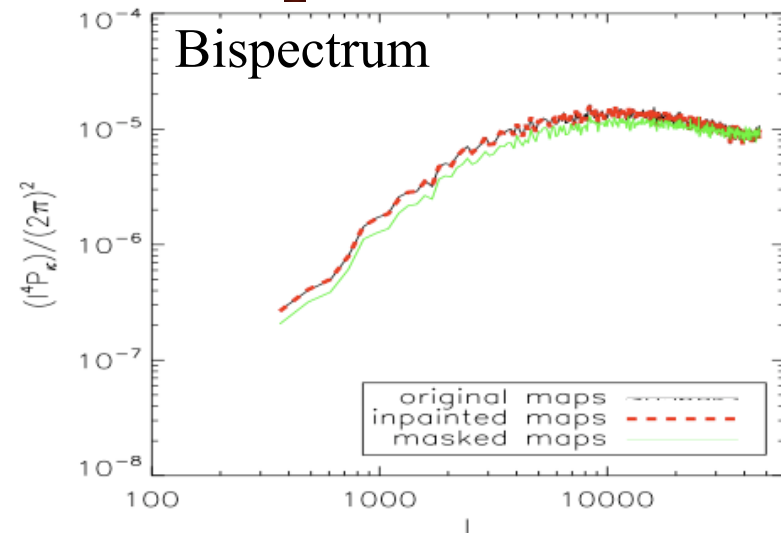
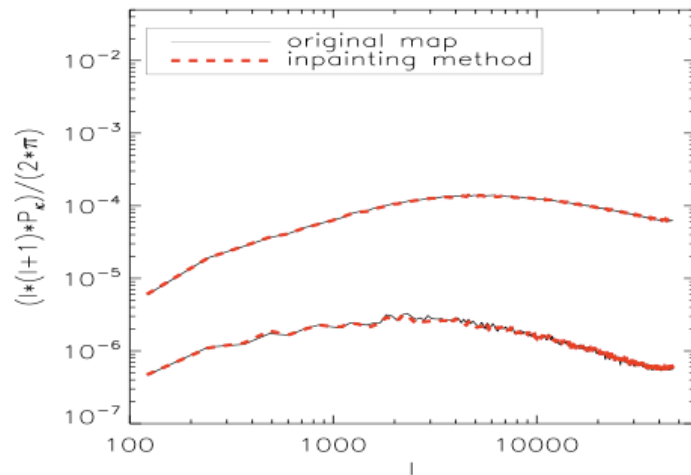
Simulated data (inpainting): l=4



Inpainting :



Power spectrum

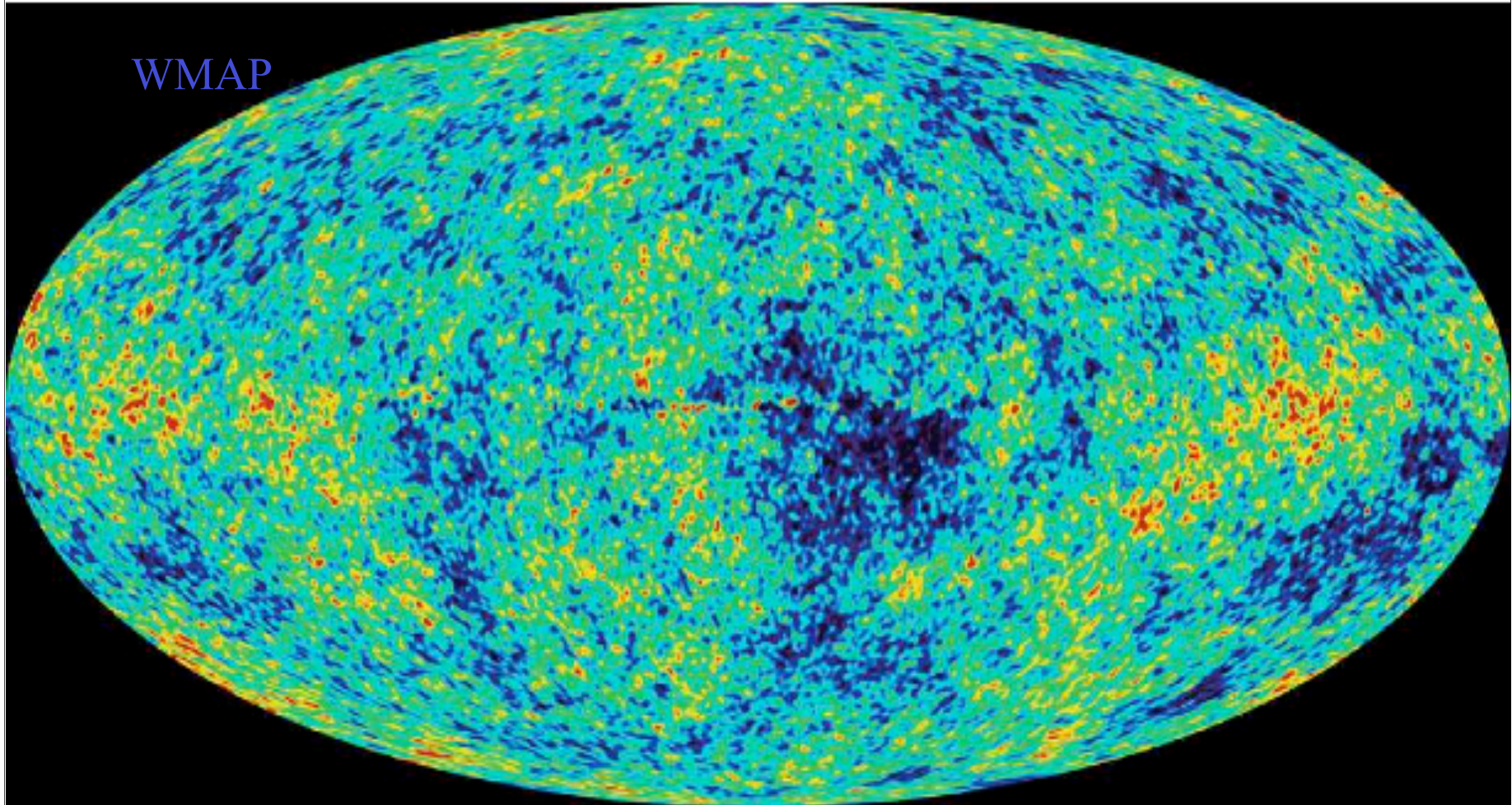


PLANCK PROJECT



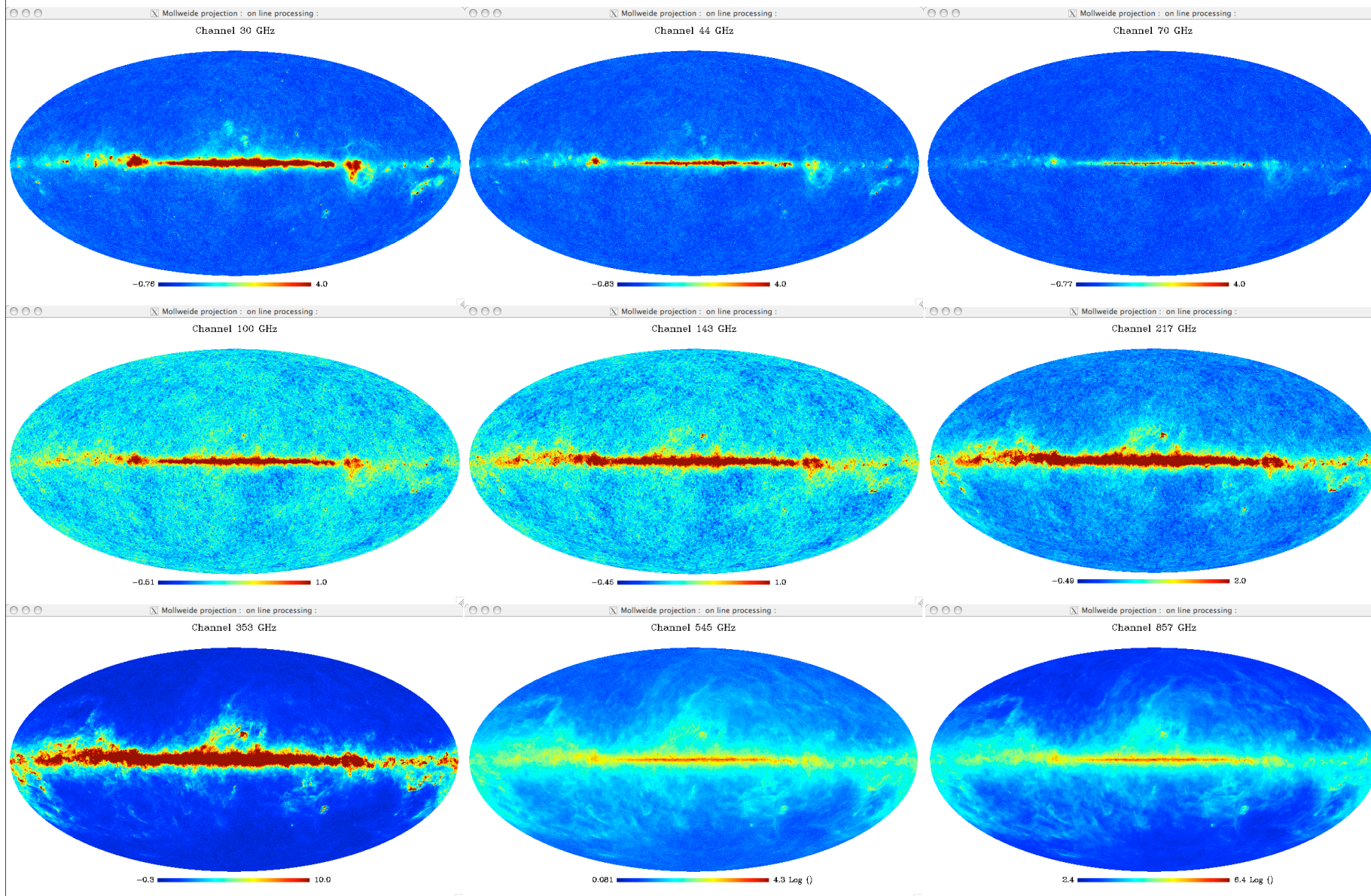
- **Successor of WMAP (better resolution, better sensitivity, more channels)**
- **Launch in 2008**
- **Two instruments LFI and HFI**
- **Nine maps at 30,44,70,100,143,217,353,545,857 GHz**
- **Angular resolutions: 33', 24', 14', 10', 7.1', 5', 5', 5', 5'**
- **Size of each map = 9 x 12 x 2048²**

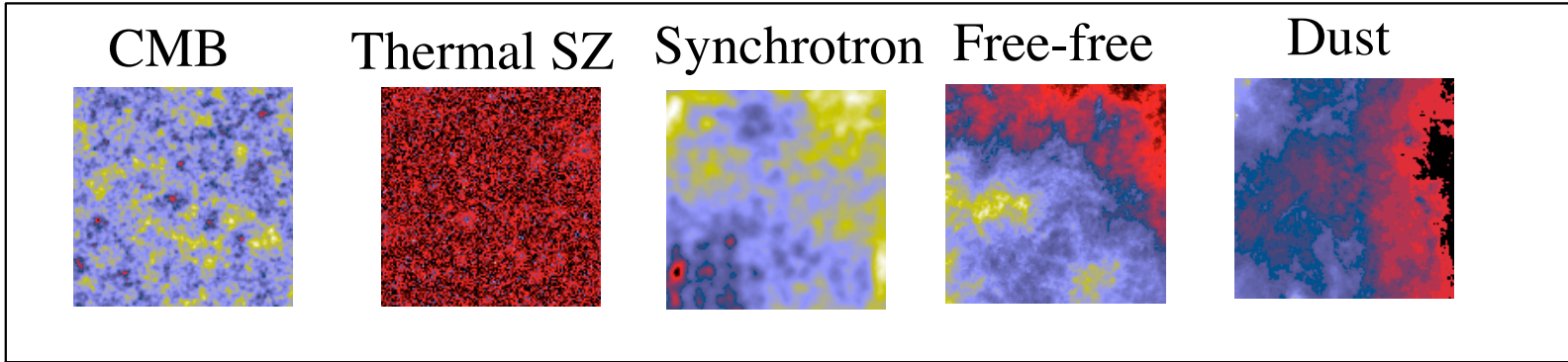
The CMB exhibits Fluctuations



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 370000 years old.

PLANCK Simulated Data

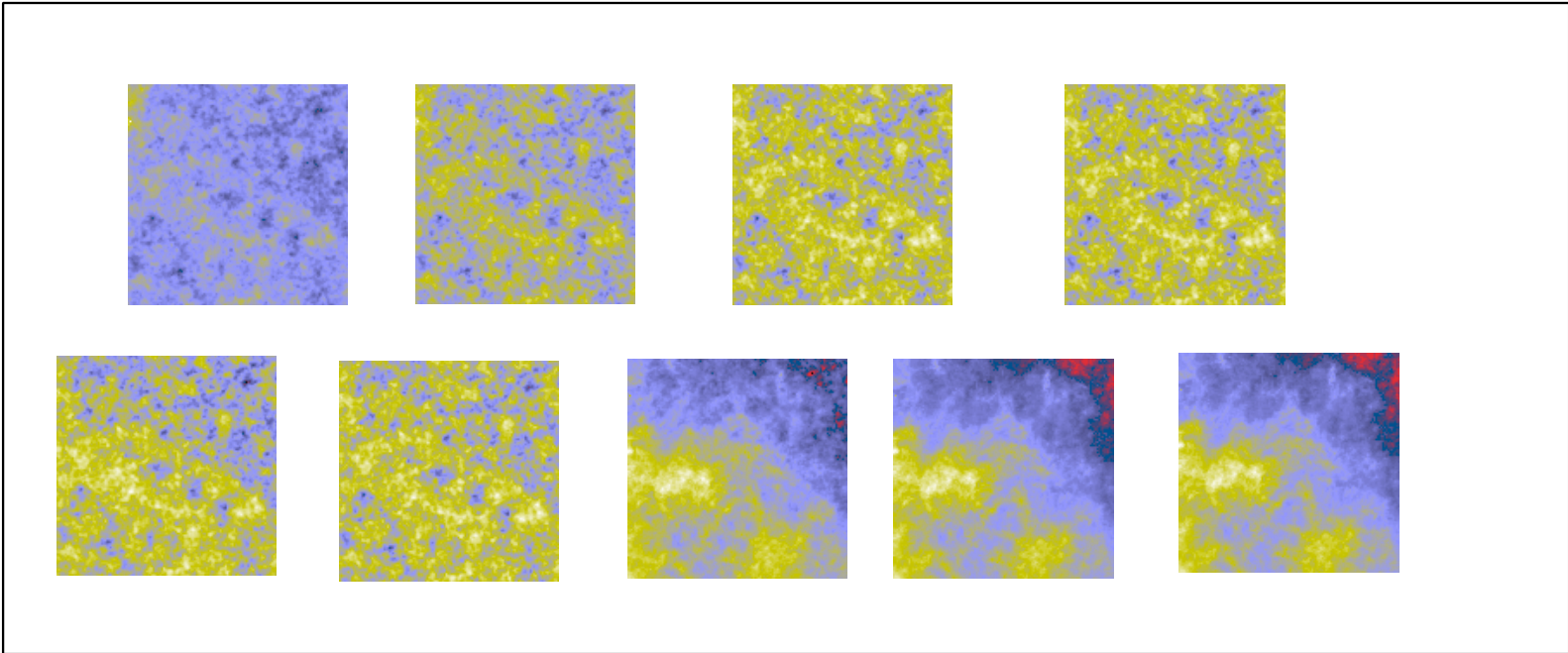




Sky components

Linear combination + PSF + Noise

Observations



METHODS

- Template Fitting
- Spectral Index Fitting
- Independent Component Analysis: JADE, FastICA, ...
- SMICA (Delabrouille et al, 1998) (assume all components are Gaussian)

$$Y = A X$$

$$Y = A X + N$$

==> Another approach: Sparsity

Morpho-Spectral Diversity

Data: $X = [x_1, \dots, x_m]$

Source: $S = [s_1, \dots, s_n]$

$$X = [x_1, \dots, x_m] = AS$$

$$x_l = \sum_{i=1}^n a_{i,l} s_i$$

$$\min_{\alpha} \|\alpha\|_p \text{ s.t. } \mathbf{X} = \sum_{\gamma \in \Gamma} \alpha_{\gamma} \psi_{\gamma}$$

$$\Phi_{\mathbf{A}} = [\Phi_{\mathbf{A},1}, \Phi_{\mathbf{A},2}]$$

$$\Phi_{\mathbf{S}}$$

Spatial Dictionary with
Spectral Dictionary

$$\Psi = [\Phi_{\mathbf{A},1} \otimes \Phi_{\mathbf{S}}, \Phi_{\mathbf{A},2} \otimes \Phi_{\mathbf{S}}]$$

Generalized MCA (GMCA)

• J. Bobin, Y. Moudden, J.-L. Starck and M. Elad, "Morphological Diversity and Source Separation", IEEE Transaction on Signal Processing, Vol 13, 7, pp 409--412, 2006.

• J. Bobin, J.-L. Starck, M.J. Fadili, and Y. Moudden, IEEE Transaction on Image Processing, "Sparsity, Morphological Diversity and Blind Source Separation", Vol 16, No 11, pp 2675--2681, 2007.



$$\text{Source: } S = [s_1, \dots, s_n] \quad \text{Data: } X = [x_1, \dots, x_m] = AS$$

We now assume that the sources are linear combinations of morphological components :

$$s_i = \sum_{k=1}^K c_{i,k} \quad \text{such that } \alpha_{i,k} = T_{i,k} c_{i,k} \text{ sparse}$$

$$\implies X_l = \sum_{i=1}^n A_{i,l} s_i = \sum_{i=1}^n A_{i,l} \sum_{k=1}^K c_{i,k}$$

$$\phi = \left[[\phi_{1,1}, \dots, \phi_{1,K}], \dots, [\phi_{n,1}, \dots, \phi_{n,K}] \right], \quad \alpha = S\phi^t = \left[[\alpha_{1,1}, \dots, \alpha_{1,K}], \dots, [\alpha_{n,1}, \dots, \alpha_{n,K}] \right]$$

GMCA aims at solving the following minimization:

$$\min_{A, c_{1,1}, \dots, c_{1,K}, \dots, c_{n,1}, \dots, c_{n,K}} = \sum_{l=1}^m \left\| X_l - \sum_{i=1}^n A_{i,l} \sum_{k=1}^K c_{i,k} \right\|_2^2 + \lambda \sum_{i=1}^n \sum_{k=1}^K \|T_{i,k} c_{i,k}\|_p$$

The GMCA Algorithm

• Initialize all C_k to zero, $\lambda_1 = \max(\alpha), \delta = \max(\alpha)/\text{Niter}$

• Iterate $t=1, \dots, \text{Niter}$

- Iterate $i=1, \dots, \text{NbrSource}$

Defining a multichannel residual D_i :
$$D_i = X - \sum_{i' \neq i} a^{i'} s_{i'}$$

Iterate $k=1, \dots, K_k$

- Least square estimate of $c_{i,k}$:
$$l_{i,k} = \frac{1}{a^{i^T} a^i} a^{i^T} (D_i - a^i \sum_{k' \neq k} c_{i,k'})$$

- Minimize:
$$J(\tilde{l}_{i,k}) = \left\| l_{i,k} - \tilde{l}_{i,k} \right\|_2^2 + \lambda_t \left\| T_{i,k} \tilde{l}_{i,k} \right\|_1$$

which is obtained by a simple hard/soft thresholding of $l_{i,k}$

$$s_k = \sum_i l_{k,i}$$

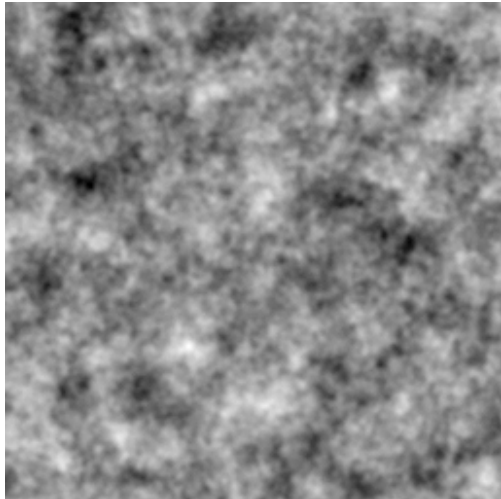
- $S = [s_1, \dots, s_K]^t$

- Estimation of the matrix A:
$$A = XS^t (SS^t)^{-1}$$

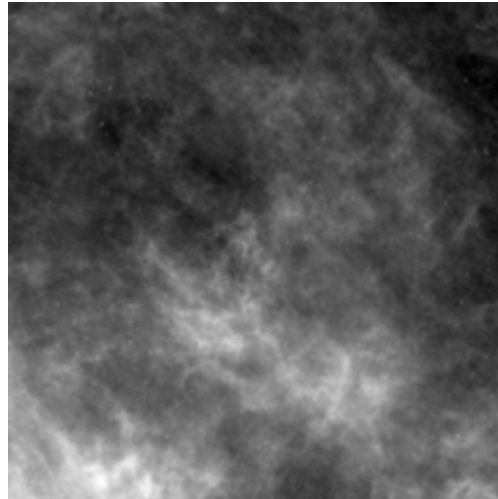
- Decrease $\lambda_{t+1} = \lambda_t - \delta$

The source images: 300x300 pixels corresponding to a field of 12,5x12,5 degrees.

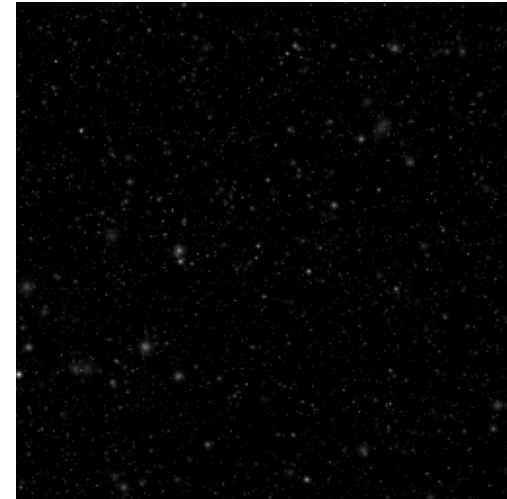
CMB



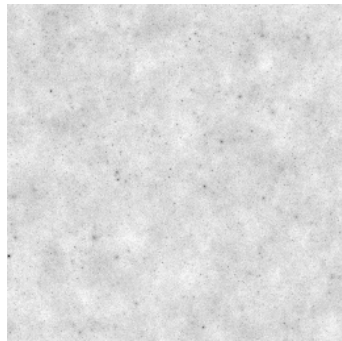
DUST



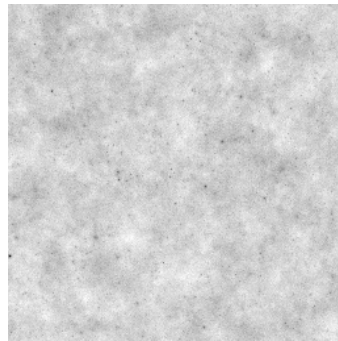
SZ



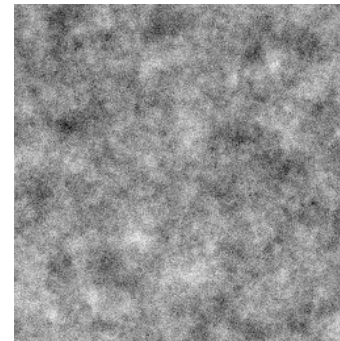
The six simulated HFI Channels
(100, 143, 217, 353, 545 and 857 GHz)



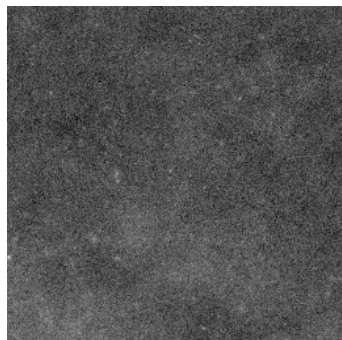
3.6 dB



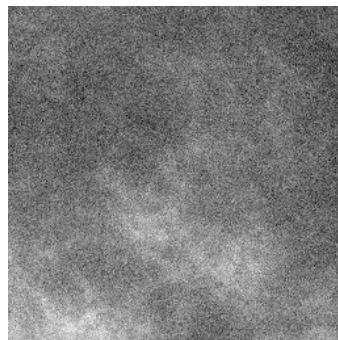
4.3 dB



1.4 dB



-3.7 dB

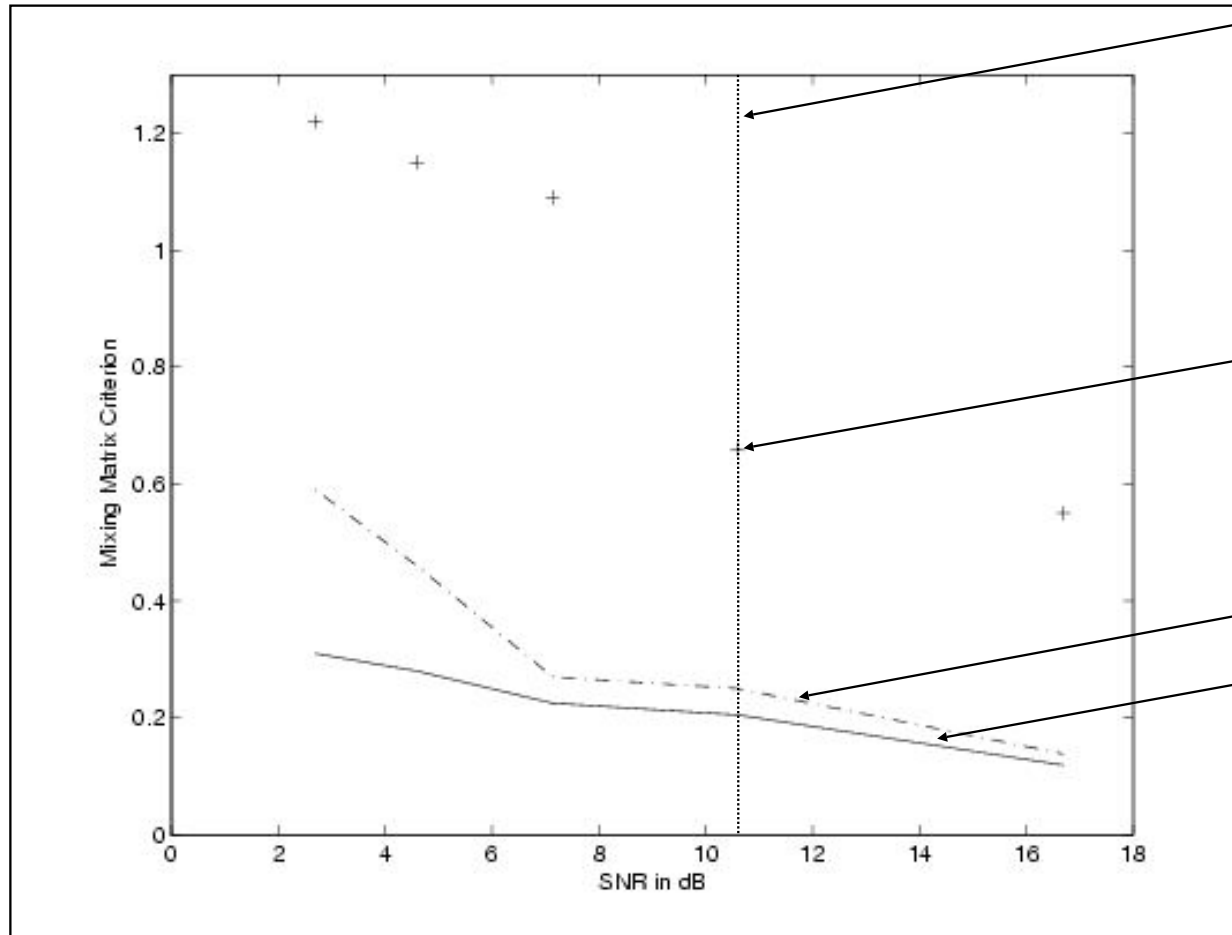


1.25 dB



9.35 dB

Mixing Matrix Estimation Error



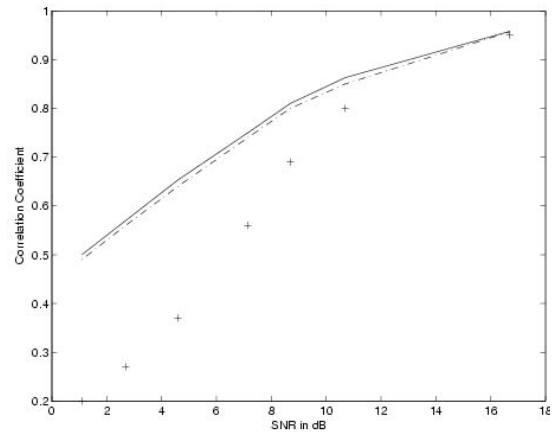
Planck
noise level

w-JADE

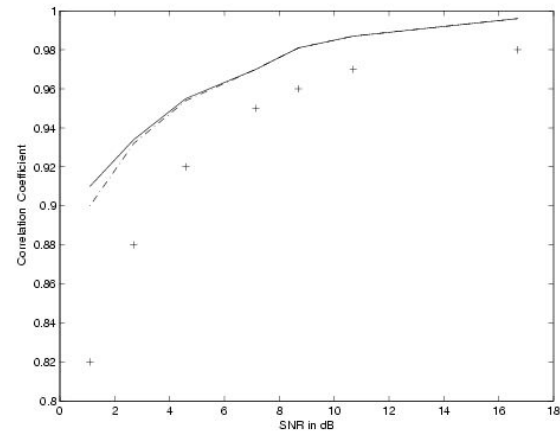
SMICA

GMCA

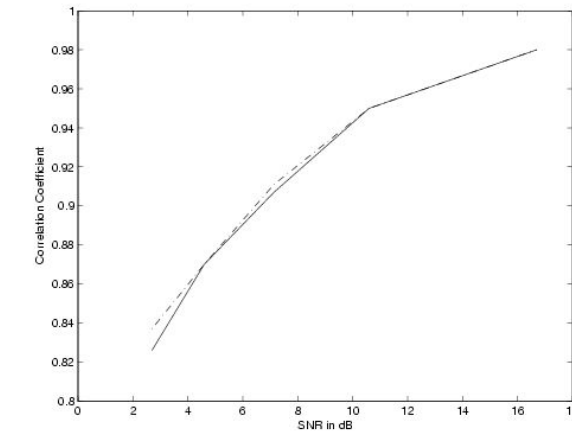
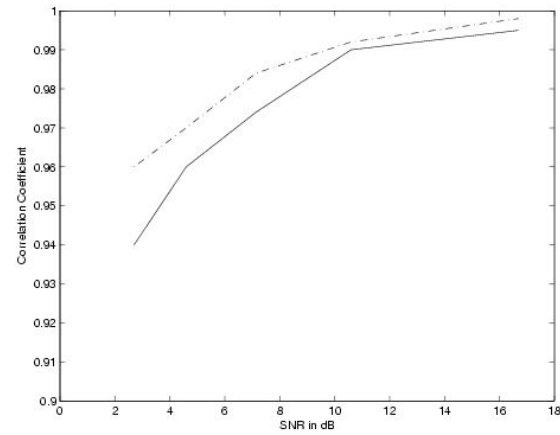
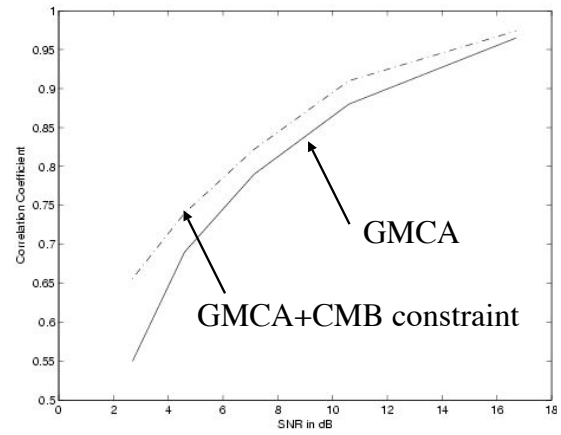
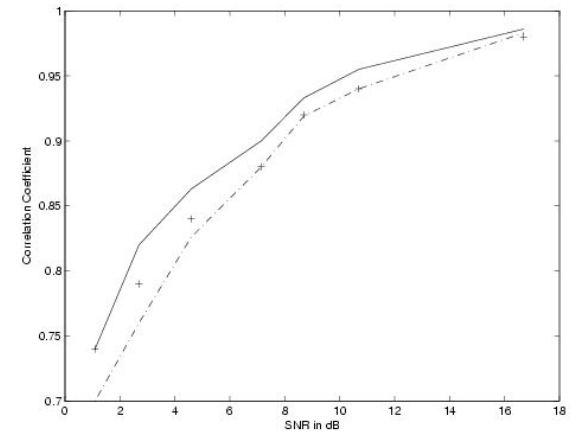
CMB



DUST



SZ



BSS experiment : Noiseless case

Original Sources



2 of 4 Mixtures

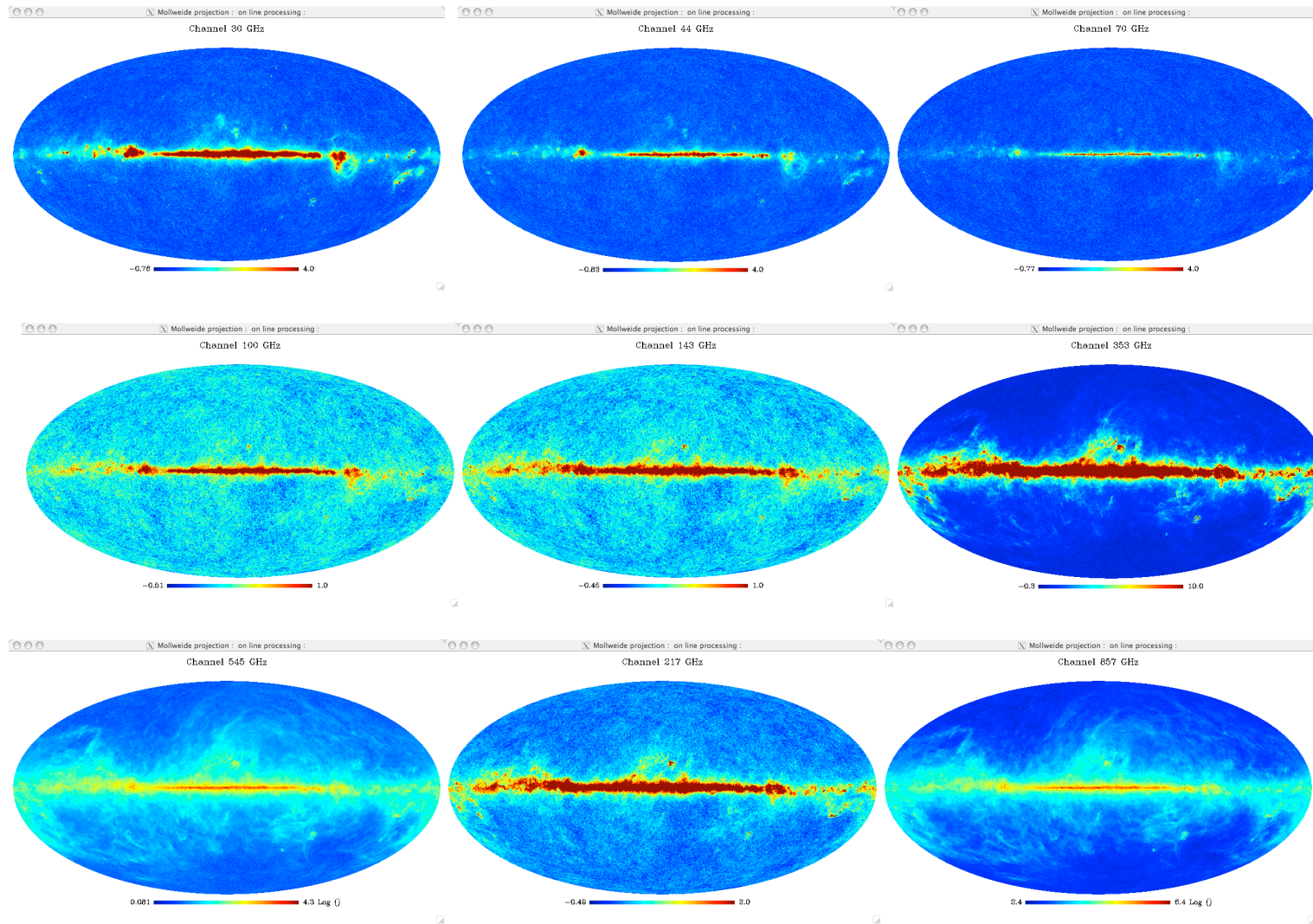


Noiseless experiment, 4 random mixtures, 4 sources

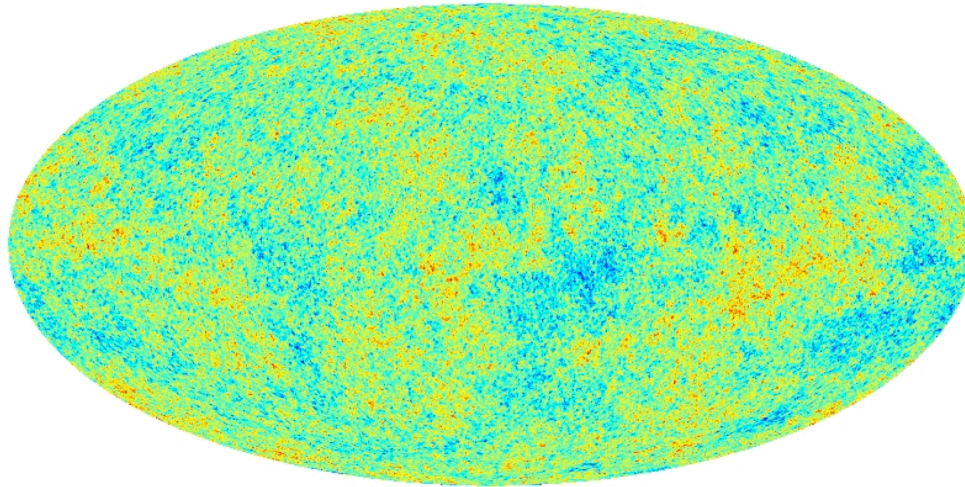
Results



We happlied GMCA to these nine components

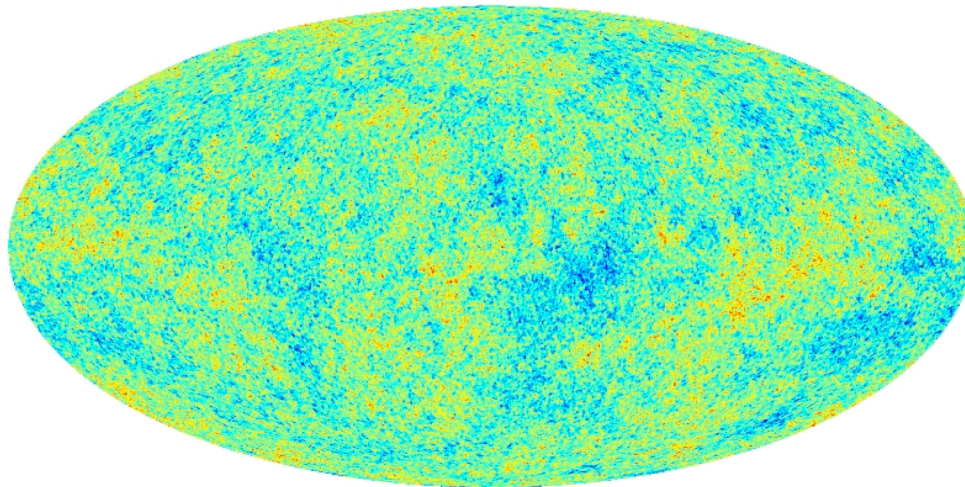


Planck



-0.50 0.50

Input CMB map (mK)

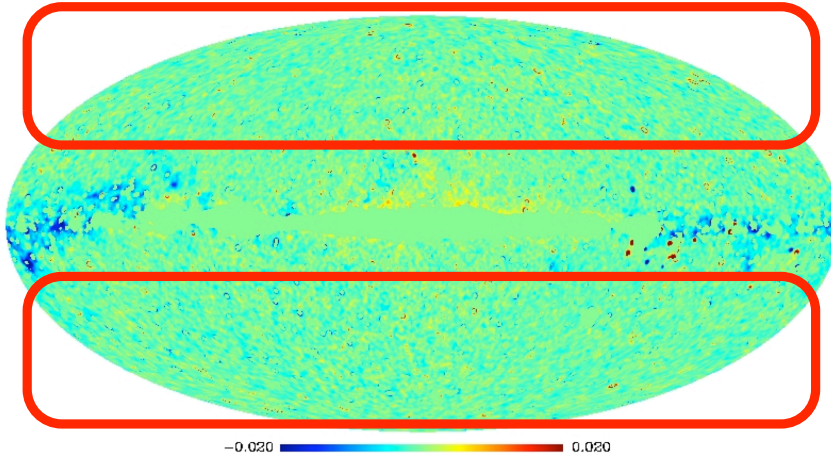


-0.50 0.50

Estimated map with GMCA

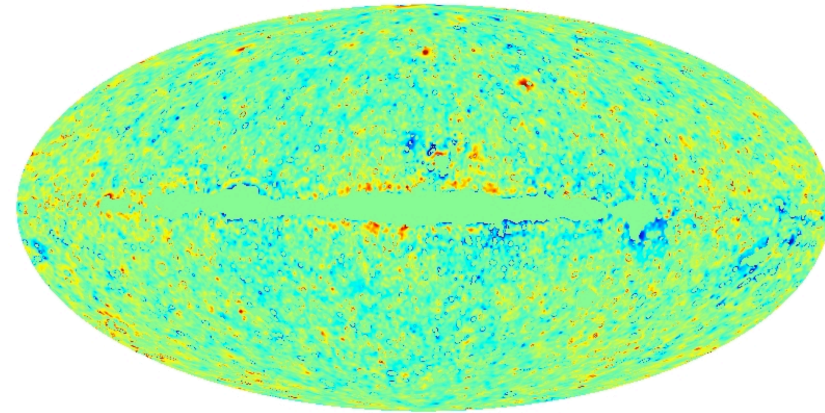
Planck

comparaisons - résidual maps



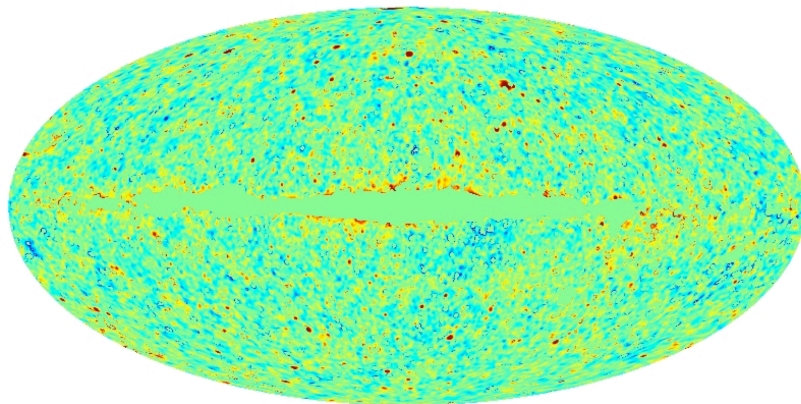
-0.020 0.020

GMCA



-0.020 0.020

SMICA

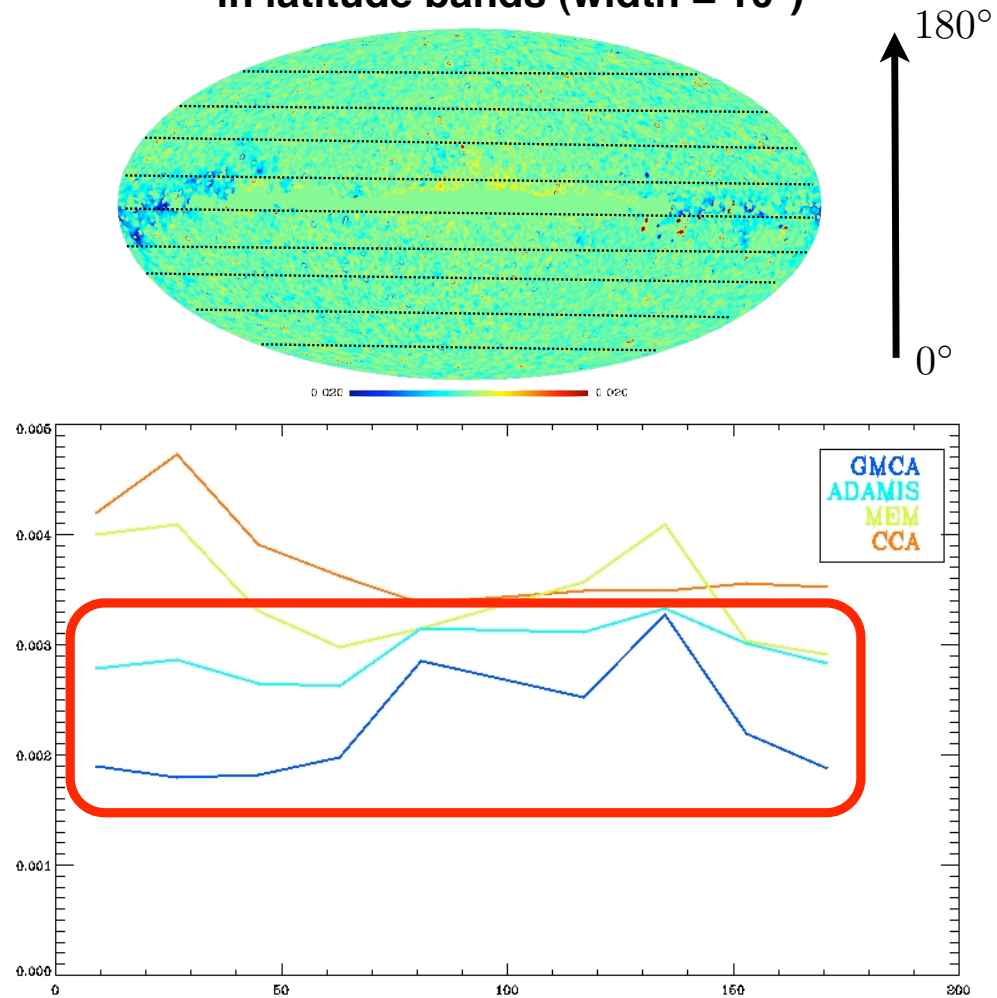


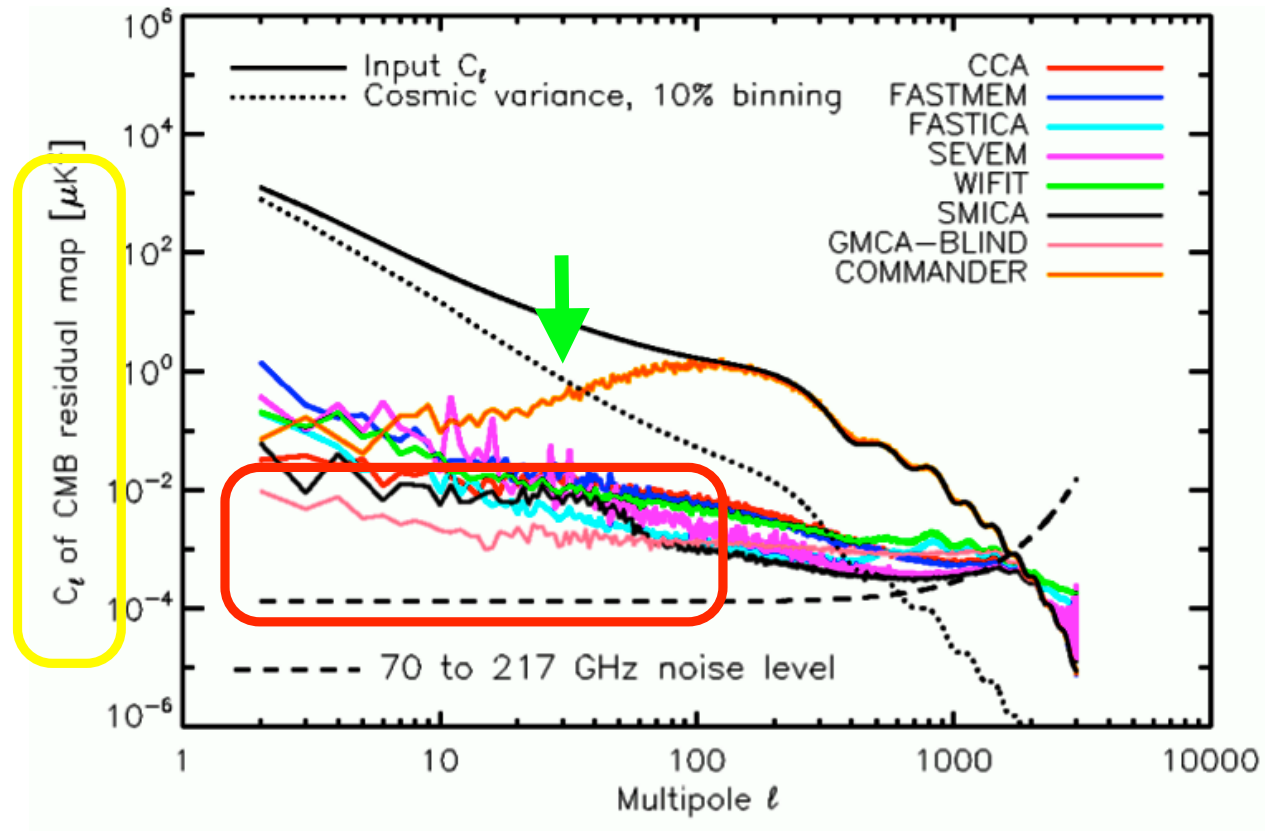
-0.020 0.020

CCA

**Residual maps convolved at
60 arcmin**

residual standard deviation (mK)
in latitude bands (width = 10°)





- S.M. Leach *et al.*, *Component separation methods for the Planck mission*, Astronomy and Astrophysics, 2008.

- Sparsity is very efficient for
 - Inverse problems (denoising, deconvolution, etc).
 - Inpainting
 - Component Separation.

- Perspectives
 - CS and Herschel.
 - PLANCK component separation.

If you want to know more...

Second edition available since September 2006

