Parcimonie et Analyse de Données en Astrophysique

J.-L. Starck

CEA, IRFU, Service d'Astrophysique, France

jstarck@cea.fr

http://jstarck.free.fr

jstarck@cea.fr

http://jstarck.free.fr

http://www.cosmostat.org

Sparsity Everywhere

Sparsity Tour

Sparsity and inverse problems

Sparsity and PLANCK

Sparsity and Euclid

What is Sparsity?

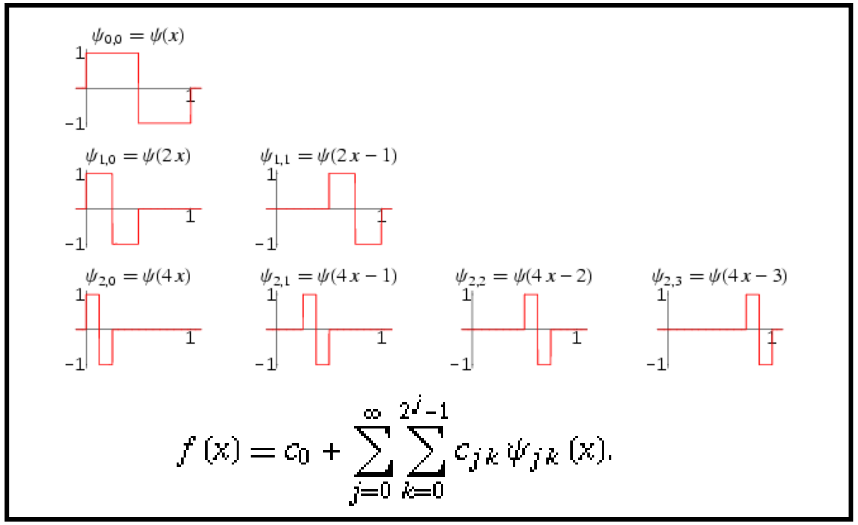
A signal s (n samples) can be represented as sum of weighted elements of a given dictionary

Dictionary (basis, frame)

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

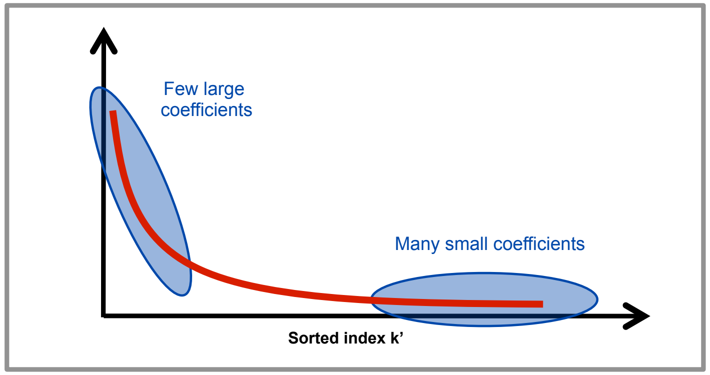
Atoms

Ex: Haar wavelet

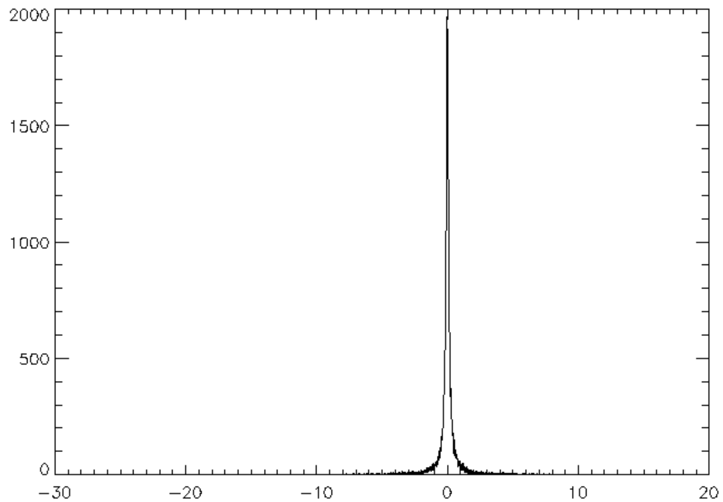


coefficients

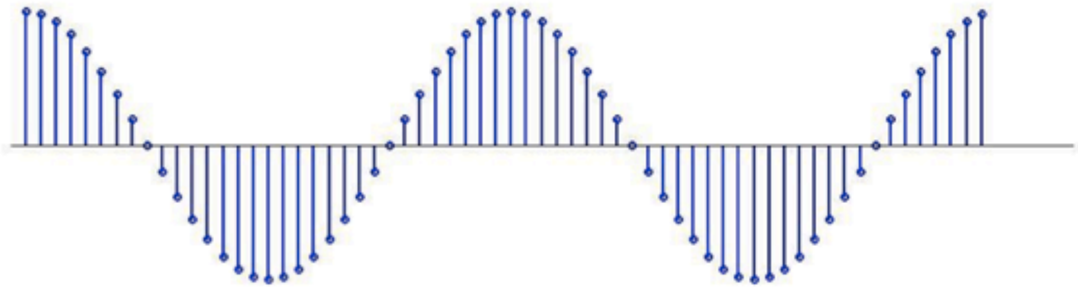
$$s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi\alpha$$



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



Strict Sparsity: k-sparse signals

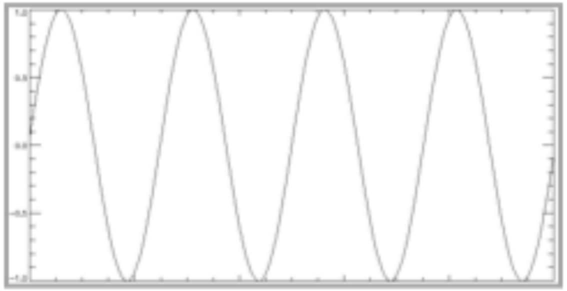


A sine wave in real space…

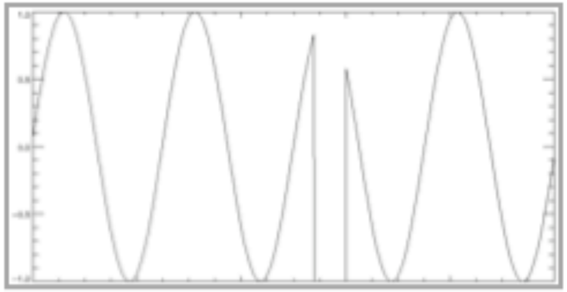
…can be a Dirac in Fourier space.

Sinusoids are sparse in the Fourier domain.

Minimizing the l_0 norm

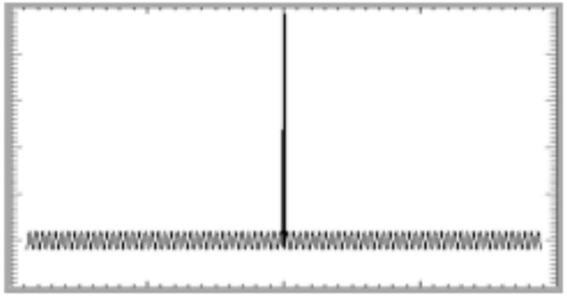


Sine curve



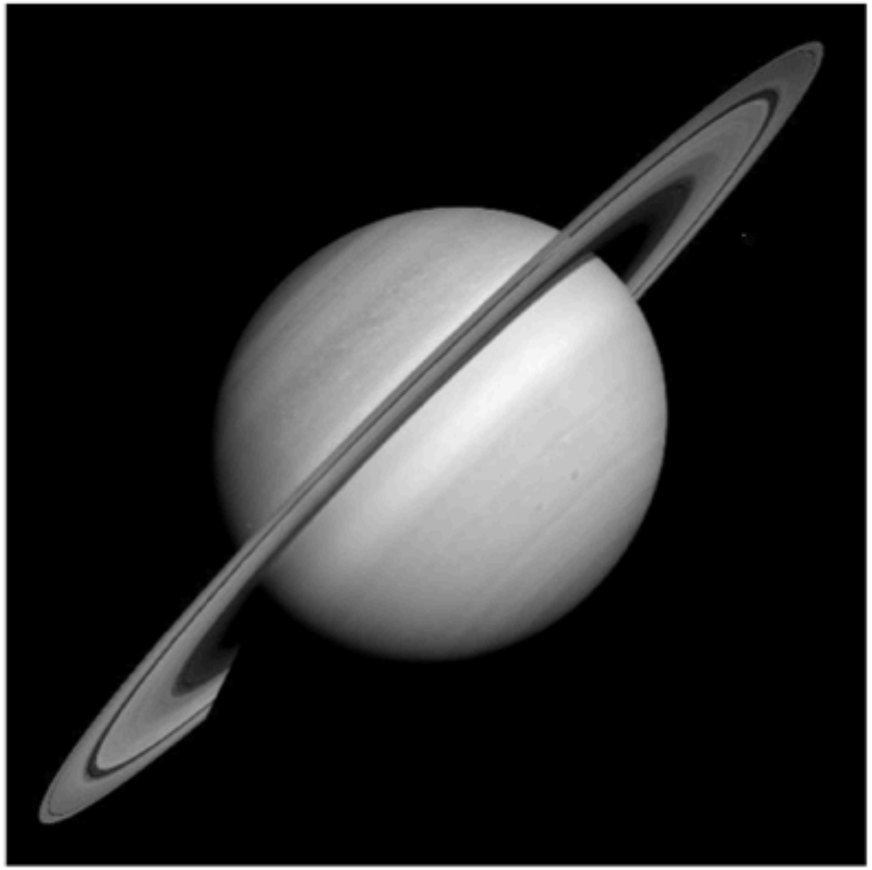
Truncated sine curve

TF of a sine curve



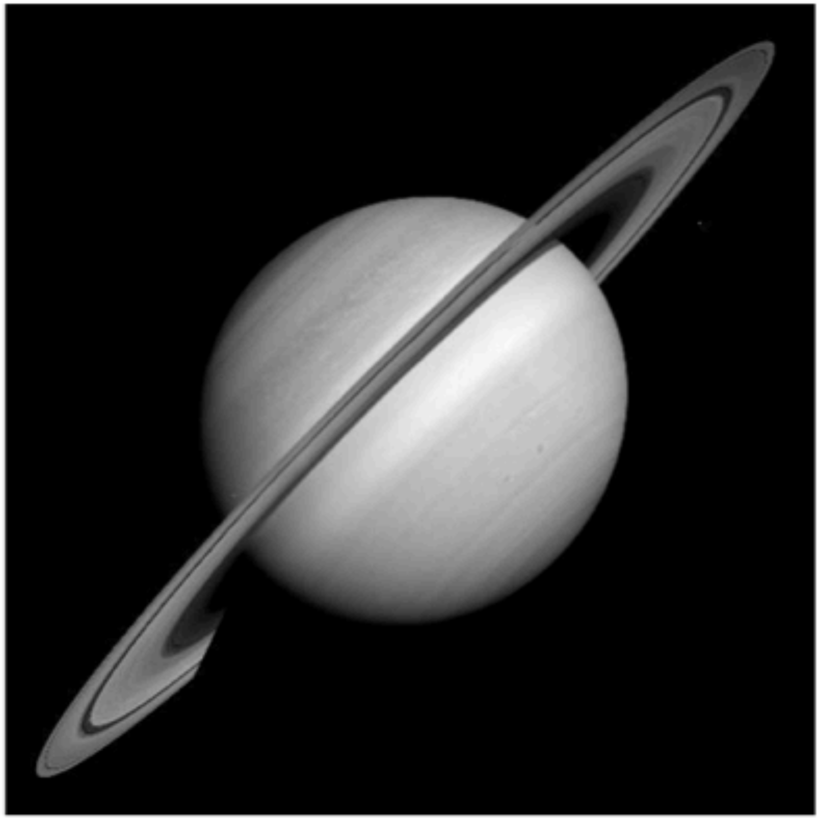
TF of a truncated sine curve

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



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

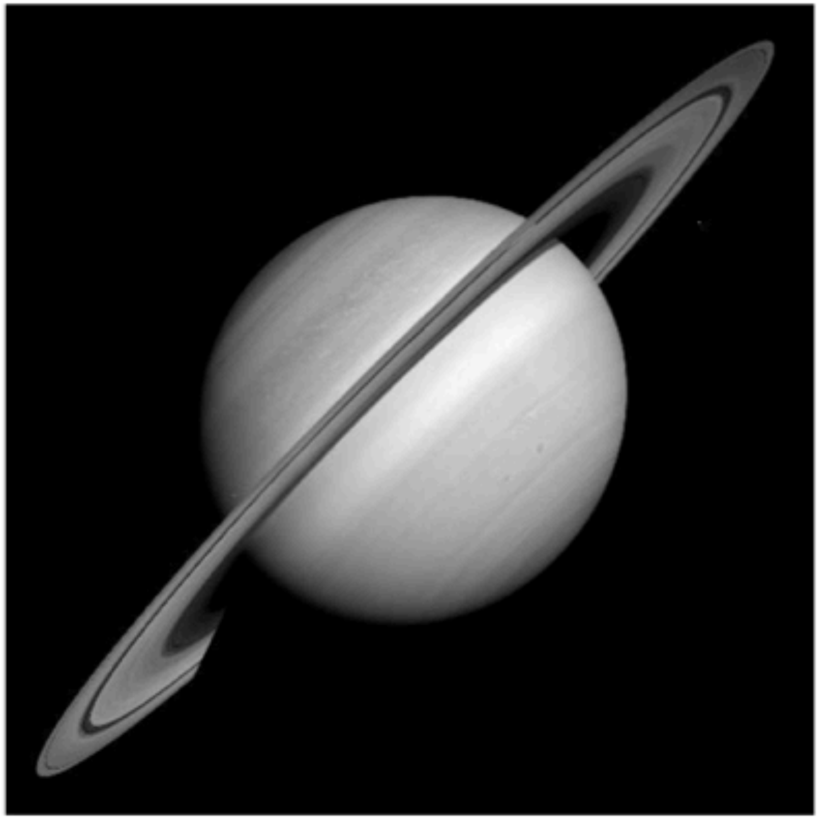
The top 1% of the coefficients concentrate only 8.66% of the energy. Not sparse…

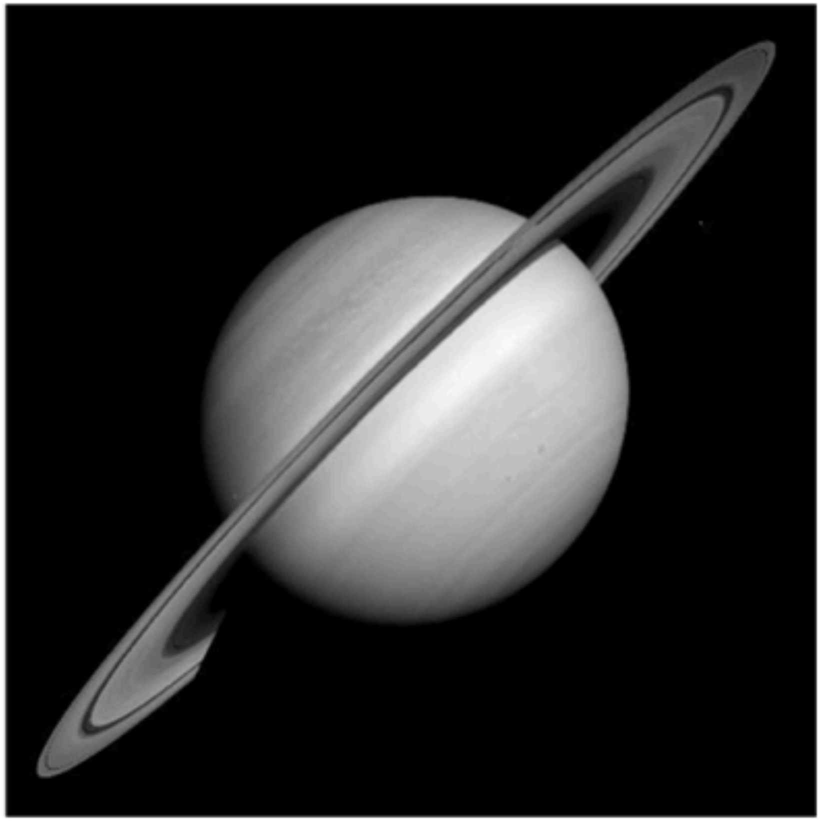


1% largest coefficients in wavelet space
(the others are set to 0)

The wavelet coefficients encode edges and large scale information.

Wavelet transform





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

Reconstruction, after throwing away 99% of the wavelet coefficients

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

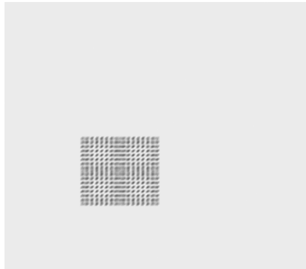
$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

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

Local DCT

Stationary textures

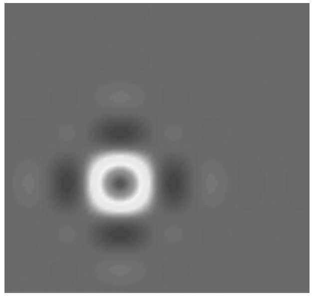
Locally oscillatory



Wavelet transform

Piecewise smooth

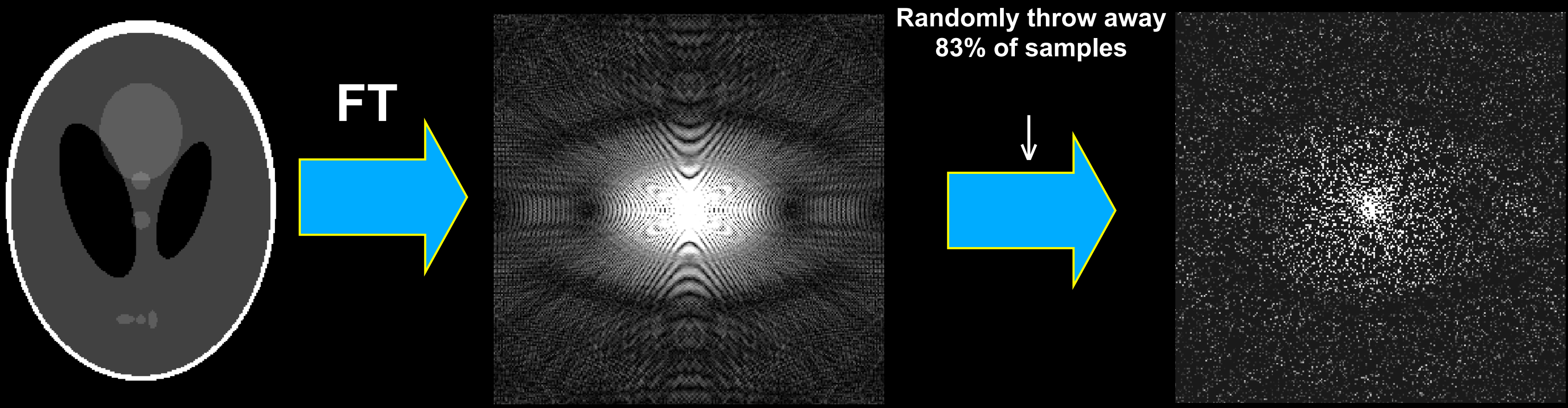
Isotropic structures



Curvelet transform

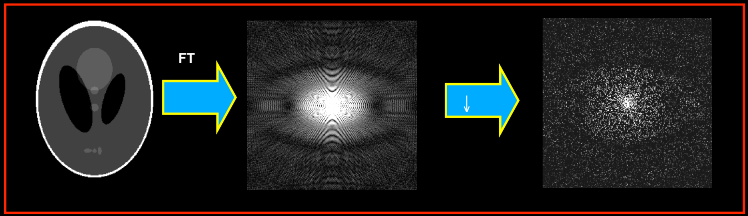
Piecewise smooth, edge

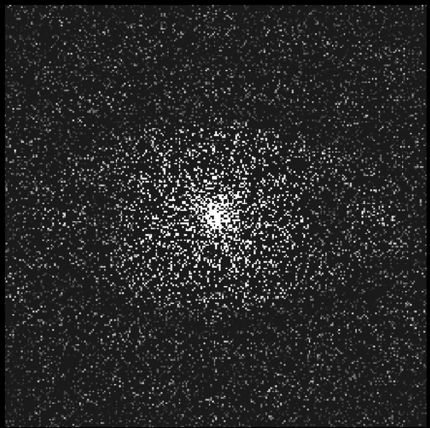
A Surprising Experiment*



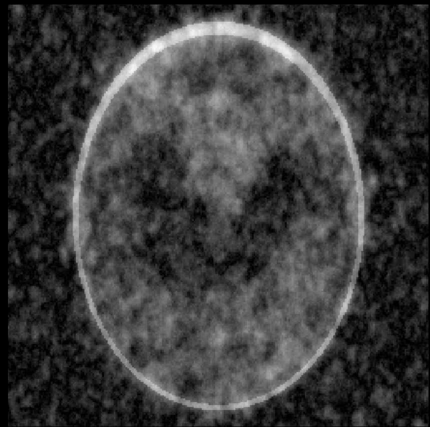
* E.J. Candes, J. Romberg and T. Tao.

A Surprising Result*



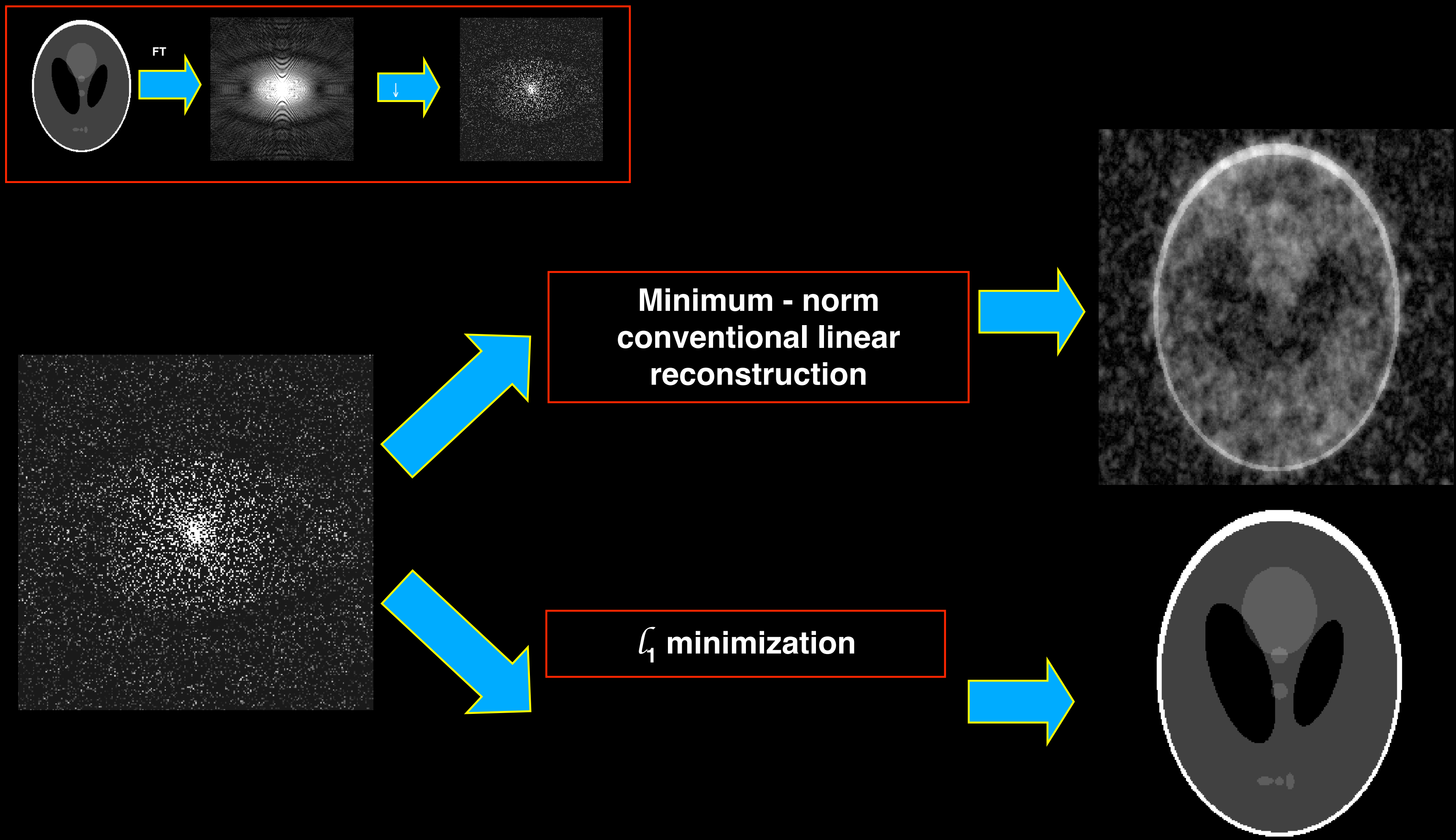


Minimum - norm conventional linear reconstruction



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

A Surprising Result*



E.J. Candes

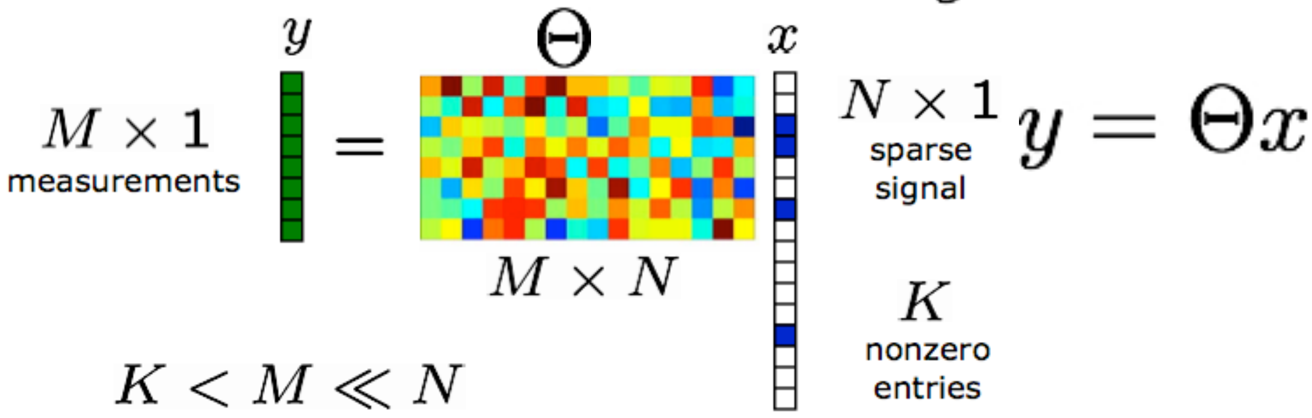
Compressed Sensing

* 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* $y = \Theta x$



$$K < M \ll N$$

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

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

Compressed Sensing Reconstruction

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

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**: $$x = \Phi\alpha$$

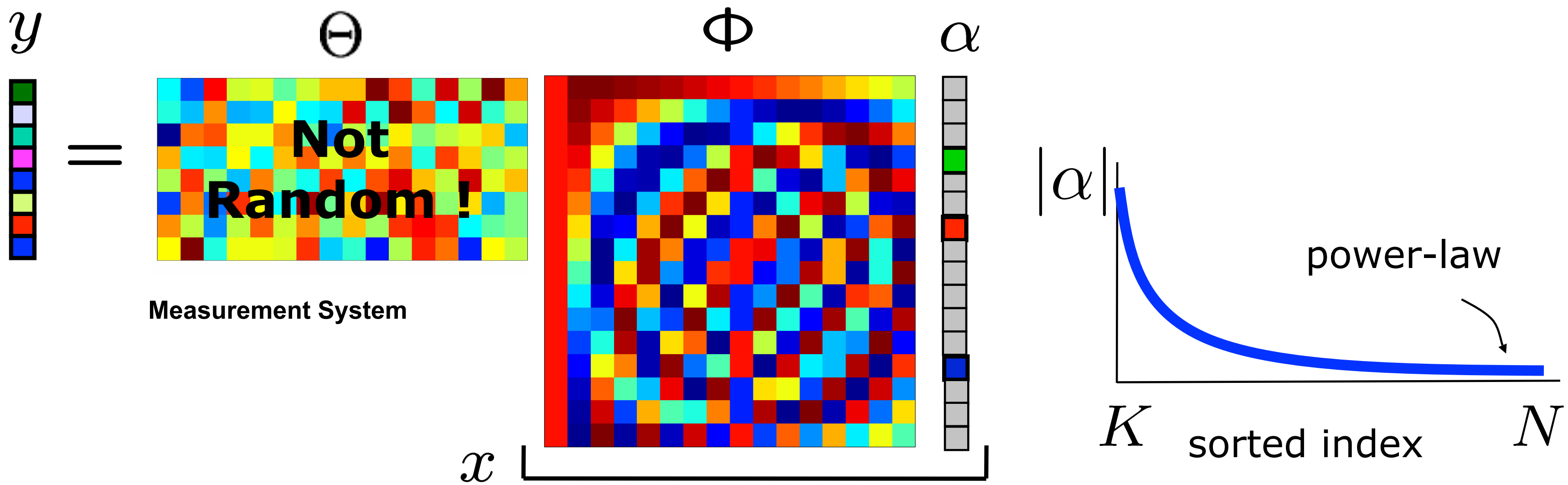
and we need to solve: $$\min_\alpha \|\alpha\|_1 \quad \text{s.t.} \quad y = \Theta_\Lambda \Phi\alpha$$

The mutual incoherence is defined as $$\mu_{\Theta,\Phi} = \sqrt{N} \max_{i,k} \left| \left\langle \phi_i, \theta_k \right\rangle \right|$$

the number of required measurements is : $$m \geq C\mu^2_{\Theta,\Phi} K \log n$$

Soft Compressed Sensing Definition

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



Prior: Data Representation System

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.

Reconstruction via non linear processing:

$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad y = \Theta \Phi \alpha$$

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).

==> Link the sparsity and the sampling through the Compressed Sensing.

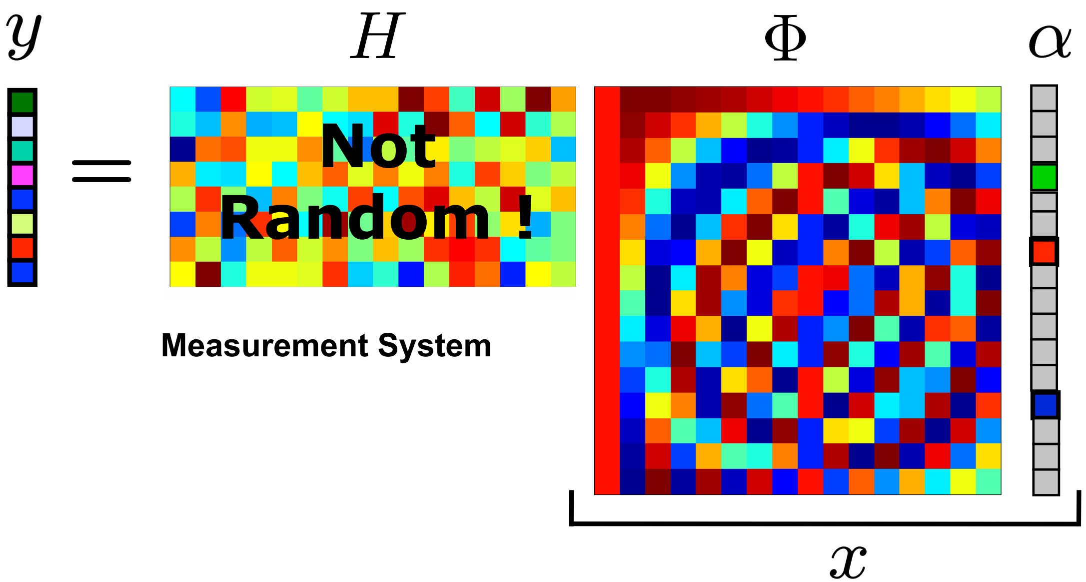
INVERSE PROBLEM TOUR and SPARSE RECOVERY

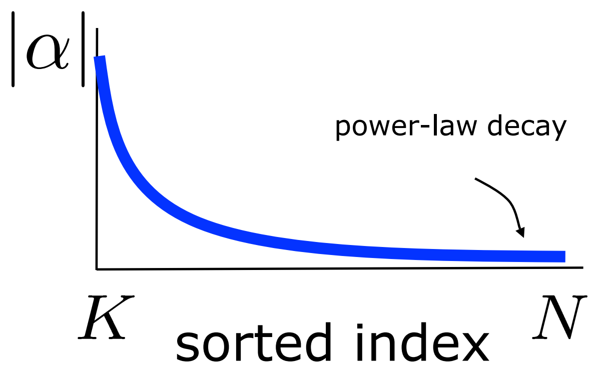
$$Y = HX + N$$

$X = \Phi\alpha$, and α is sparse

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

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





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

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

==> 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.

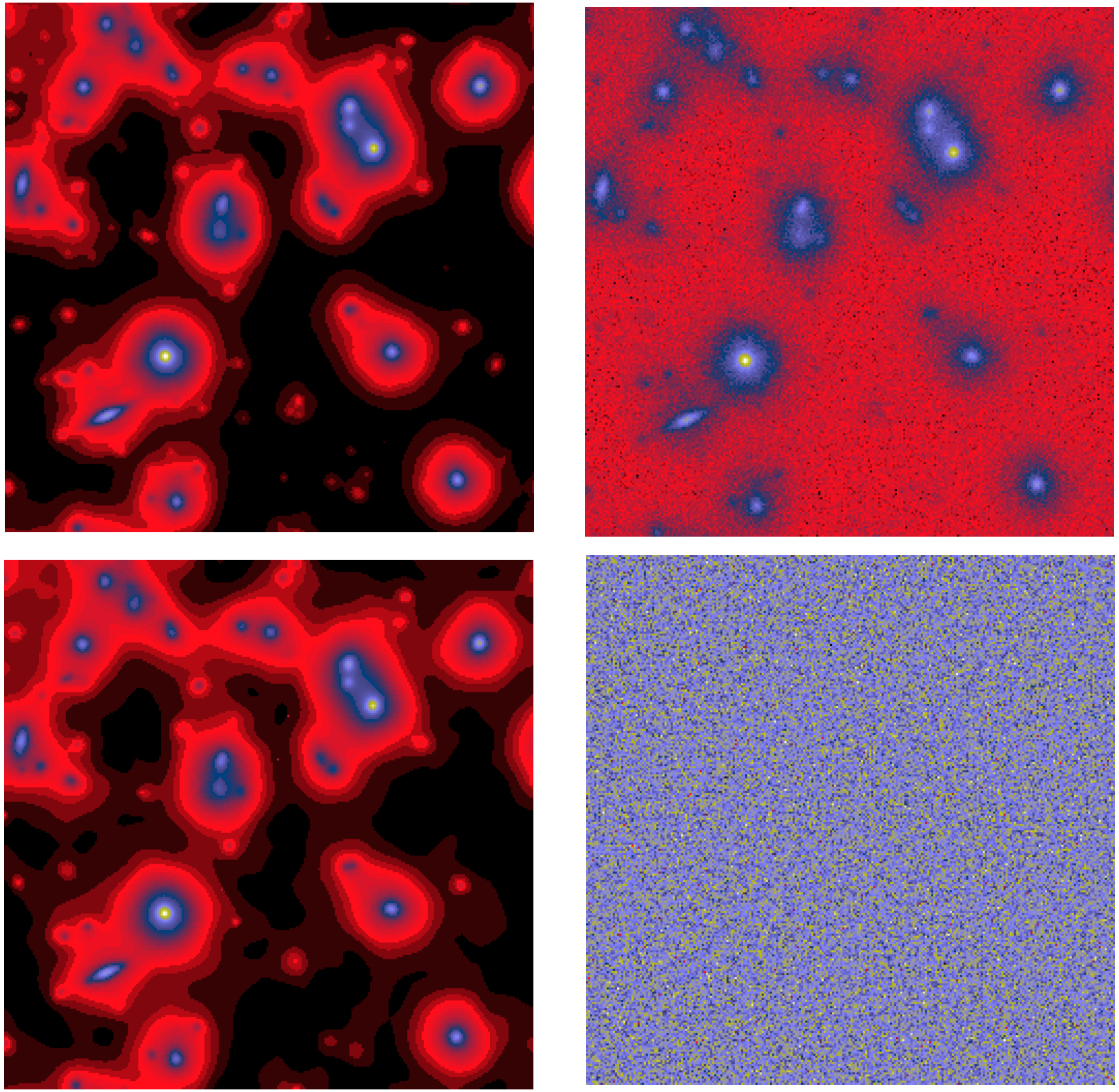
Inverse Problems and Iterative Thresholding Minimizing Algorithm

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 $\lambda^{(n)}$ ch 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 ; etc.



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
- Ell-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

Data Representation Tour

- Computational harmonic analysis seeks representations of a signal as linear combinations of basis, frame, dictionary, element :

$$s_i = \sum_{k=1}^{K} \alpha_k \phi_k$$

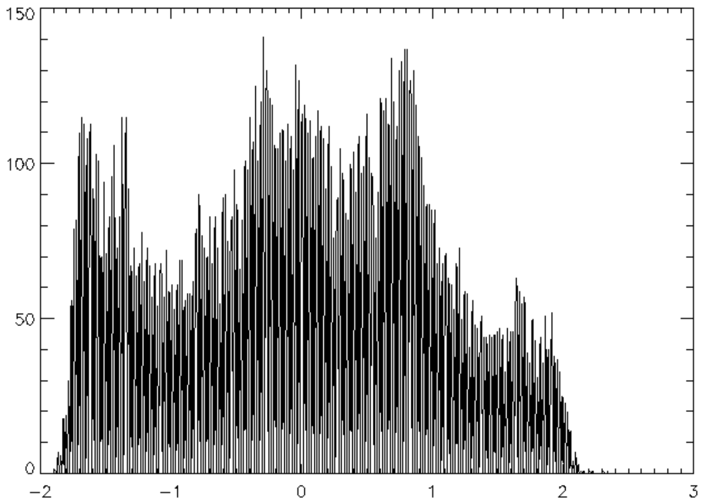
coefficients basis, frame

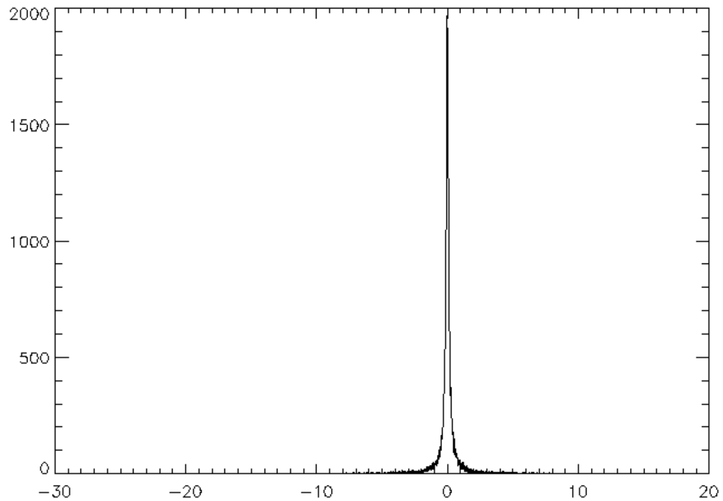
- Fast calculation of the coefficients α_k
- Analyze the signal through the statistical properties of the coefficients

Representing Barbara

Direct Space

Curvelet Space





The Great Father Fourier - Fourier Transforms

Any Periodic function can be expressed as linear combination of basic trigonometric functions

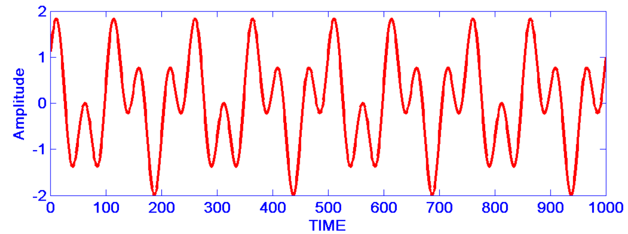
(Basis functions used are sine and cosine)

Jean-Baptiste-Joseph Fourier
(1768-1830)

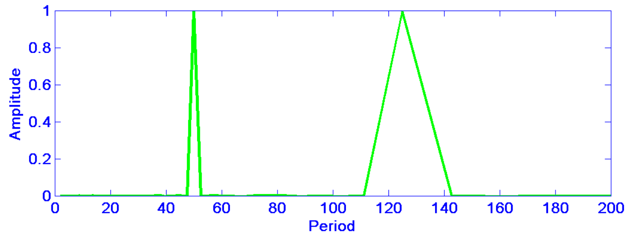
$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-2\pi ift}\,dt$$

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{2\pi ift}\,df$$

Time domain



Frequency domain

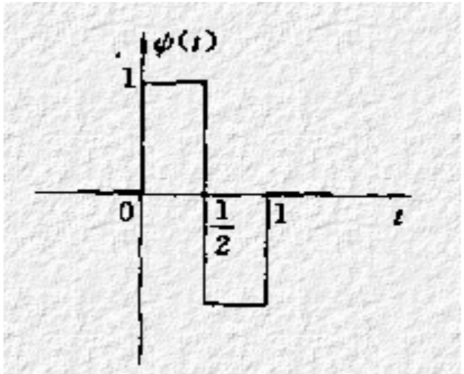


Alfred Haar Wavelet (1909):

The first mention of wavelets appeared in an appendix to the thesis of Haar

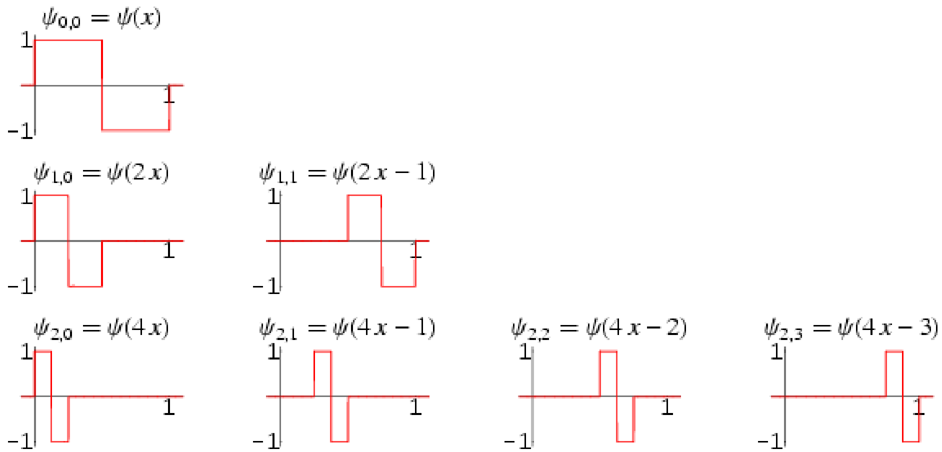
- **With *compact support,* vanishes outside of a finite interval**
- **Not continuously differentiable**
- **Wavelets are functions defined over a finite interval and having an average value of zero.**

$$f(x) = c_0 + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} c_{jk} \psi_{jk}(x).$$



$$\Psi(x) = \begin{cases} 1 & 0 \le x \le \frac{1}{2} \\ -1 & \frac{1}{2} < x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Haar wavelet



==> What kind of $\psi(t)$ could be useful?

. Impulse Function (Haar): Best time resolution

. Sinusoids (Fourier): Best frequency resolution

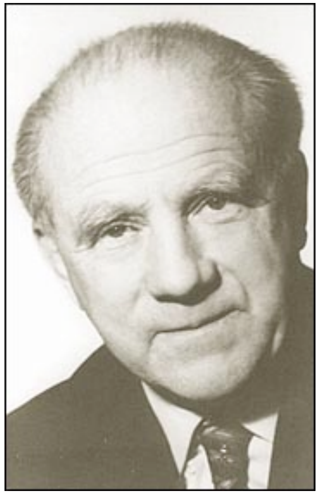
==> We want both of the best resolutions

==> Heisenberg, 1930

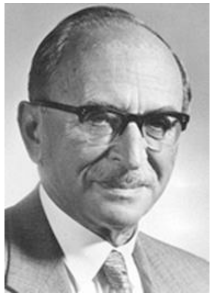
Uncertainty Principle

There is a lower bound for

$$\Delta t \cdot \Delta\omega$$



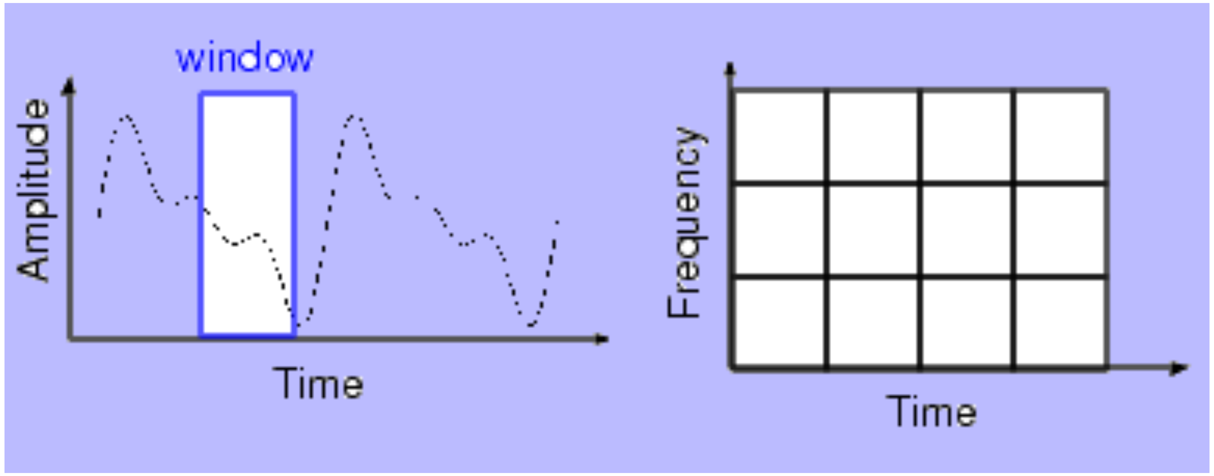
SFORT TIME FOURIER TRANSFORM (STFT)



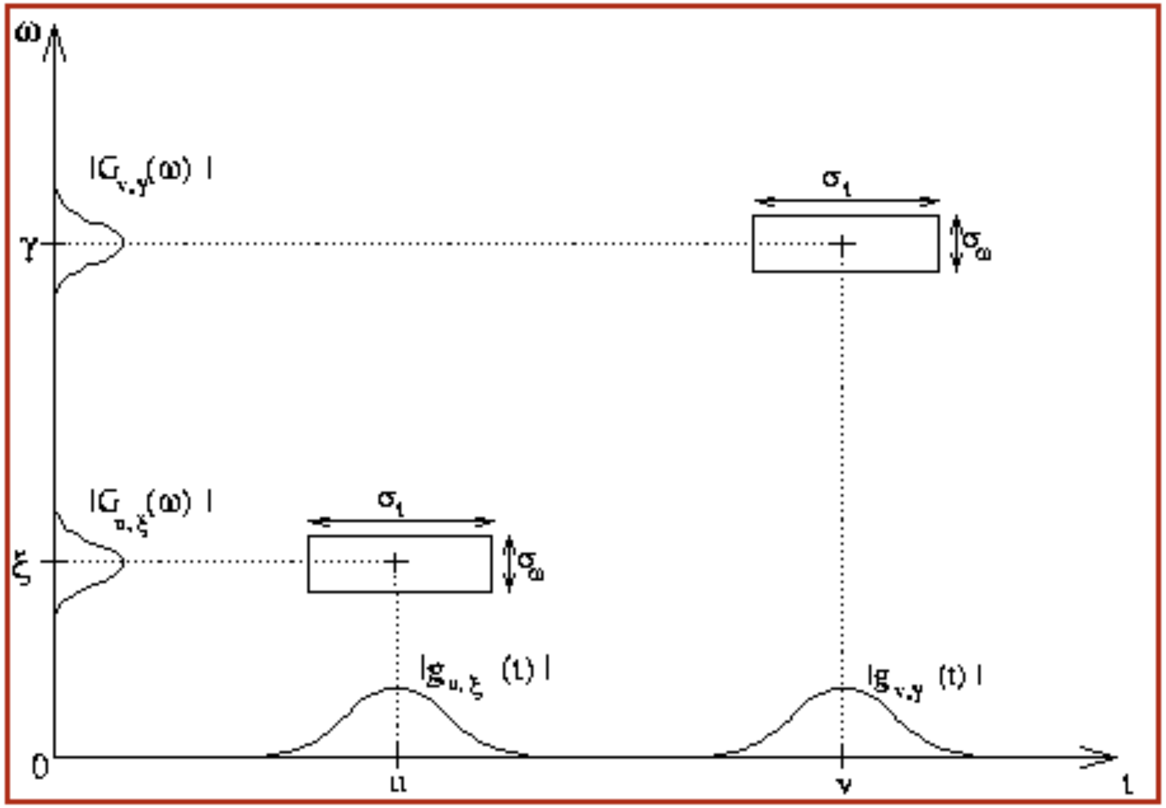
- **Dennis Gabor (1946) Used STF**

To analyze only a small section of the signal at a time -- a technique called *Windowing the Signal.*

- **The Segment of Signal is Assumed *Stationary***



Heisenberg Box



$$\sigma_t^2 \, \sigma_\omega^2 \geq \frac{1}{4}.$$

Yves Meyer

A Major Breakthrough

Daubechies, 1988 and Mallat, 1989

Daubechies:

Compactly Supported Orthogonal and Bi-Orthogonal Wavelets

Mallat:

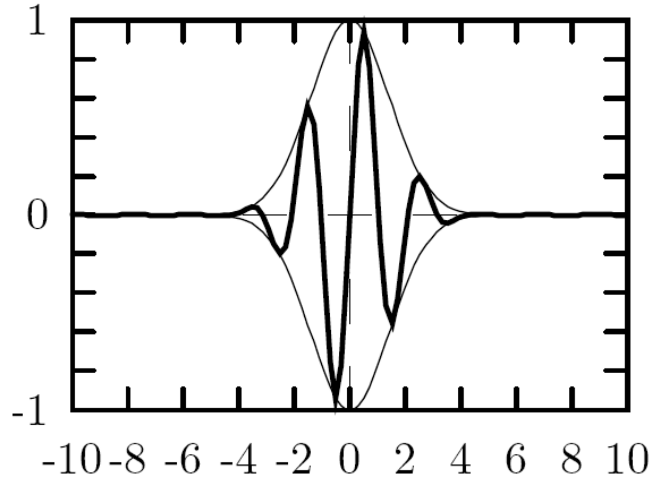
Theory of Multiresolution Signal Decomposition

Fast Algorithm for the Computation of Wavelet Transform Coefficients using Filter Banks

Candidate analyzing functions for piecewise smooth signals

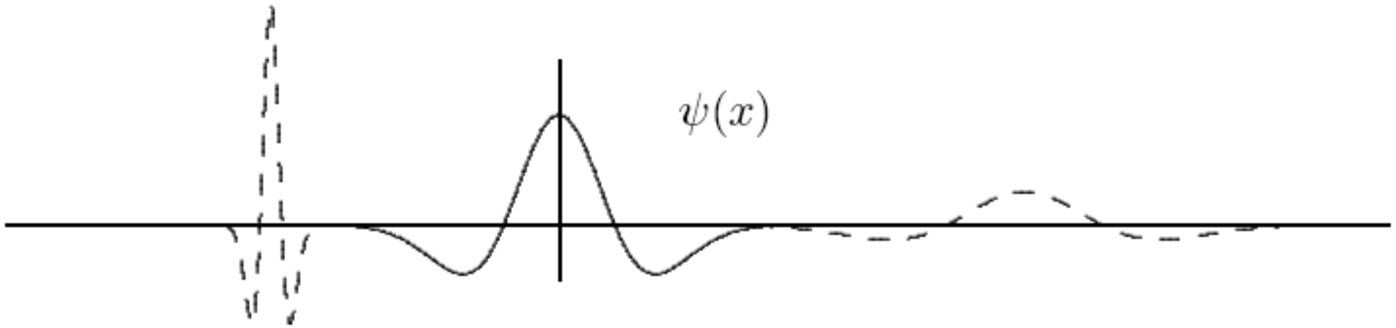
- Windowed fourier transform or Gaborlets :

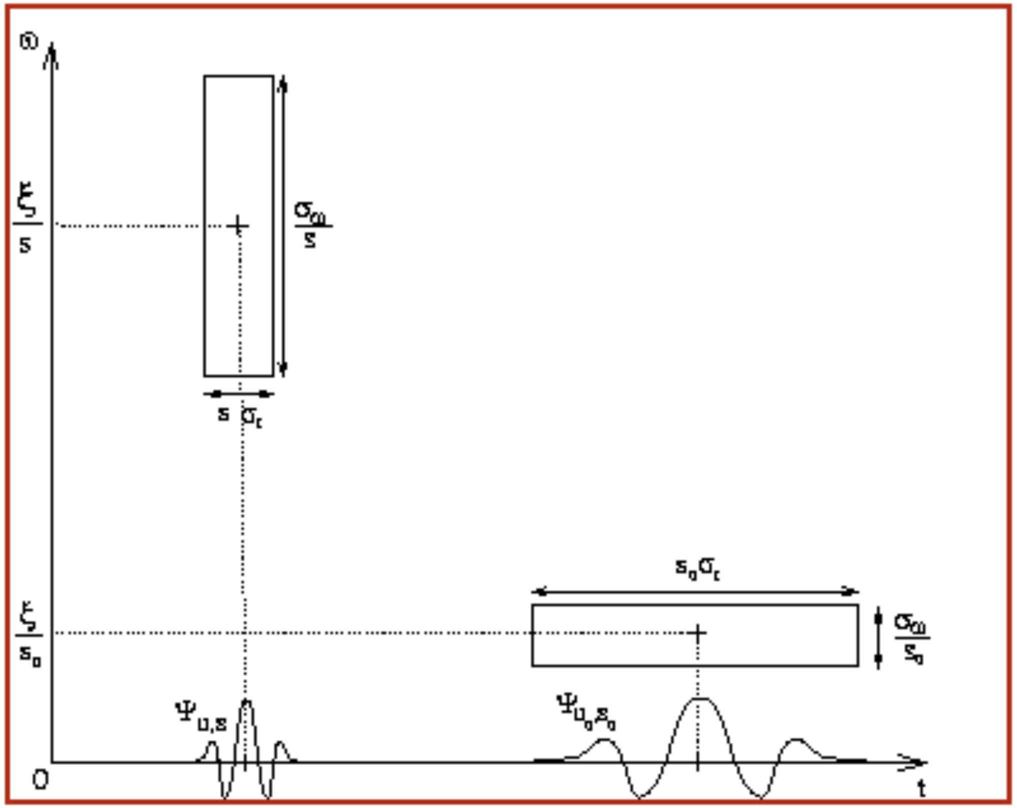
$$\psi_{\omega,b}(t) = g(t-b)e^{i\omega t}$$



- Wavelets :

$$\psi_{a,b}(t) = \frac{1}{\sqrt{b}}\psi(\frac{t-a}{b})$$



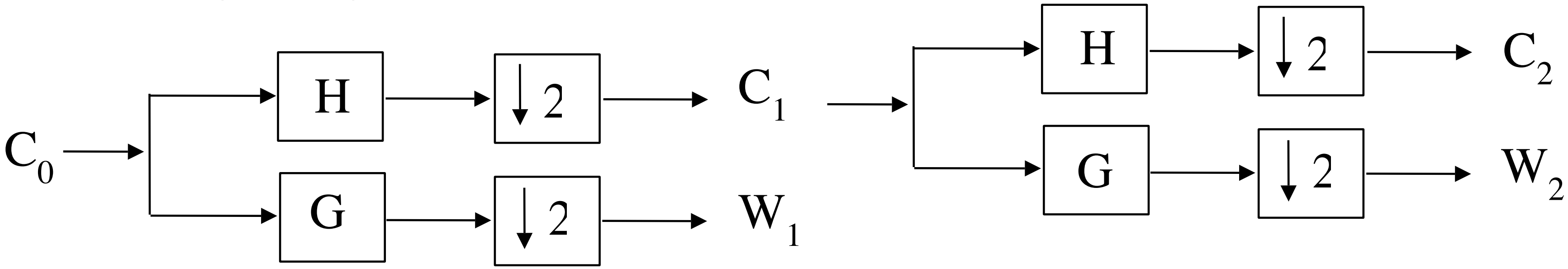


ω
ξ/s
σ_ω/s
s σ_t
s_0σ_t
ξ/s_0
σ_ω/s_0
Ψ_u,s
Ψ_u_0,s_0
O
t

The Orthogonal Wavelet Transform (OWT)

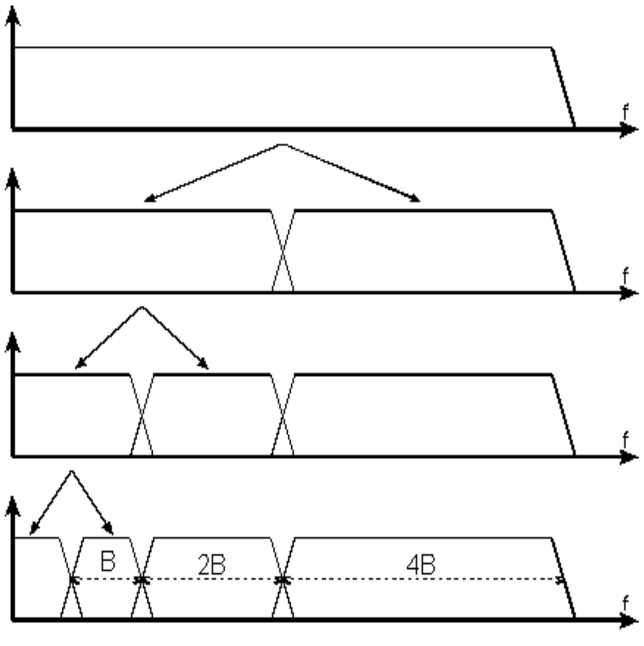
$$s_l = \sum_k c_{J,k}\phi_{J,l}(k) + \sum_k \sum_{j=1}^{J} \psi_{j,l}(k) w_{j,k}$$

Transformation



$$c_{j+1,l} = \sum_h h_{k-2l} c_{j,k} = (\bar{h} * c_j)_{2l}$$

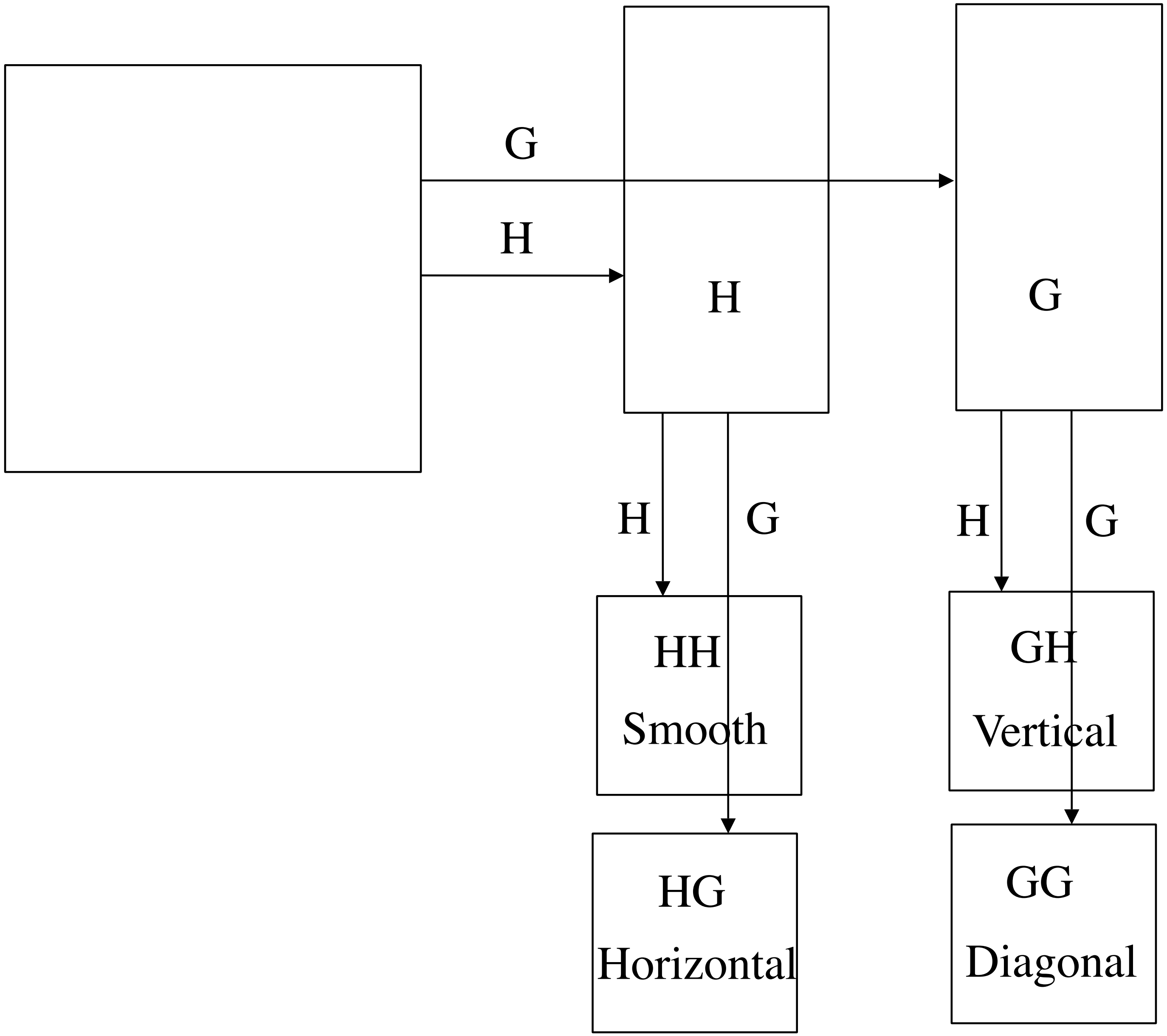
$$w_{j+1,l} = \sum_h g_{k-2l} c_{j,k} = (\bar{g} * c_j)_{2l}$$

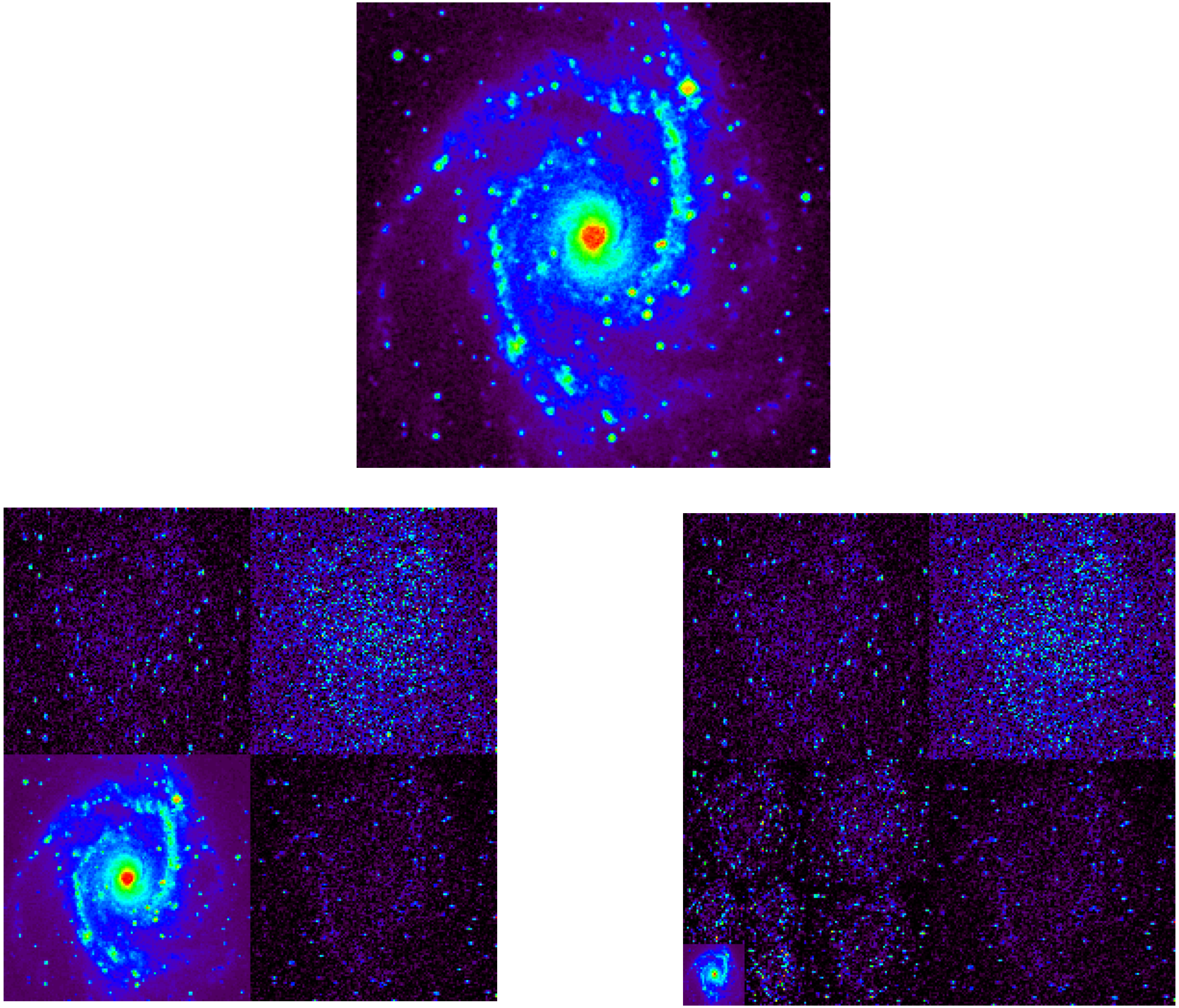


Reconstruction:

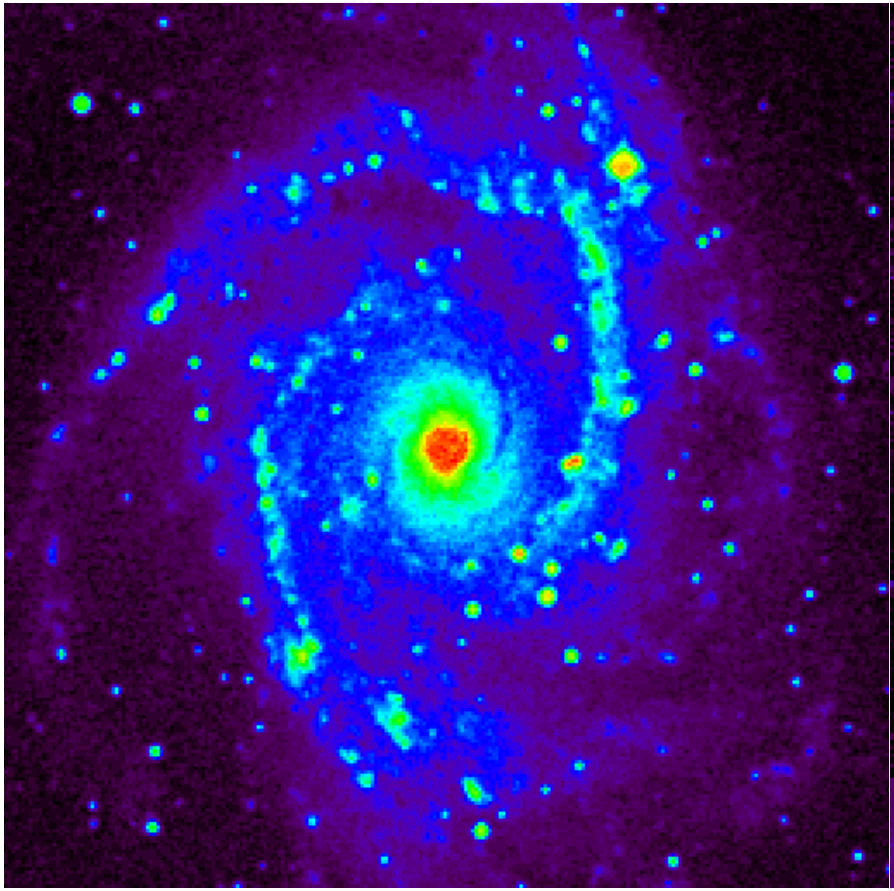
$$c_{j,l} = \sum_k \tilde{h}_{k+2l} c_{j+1,k} + \tilde{g}_{k+2l} w_{j+1,k} = \tilde{h} * \breve{c}_{j+1} + \tilde{g} * \breve{w}_{j+1}$$

$$\breve{x} = (x_1, 0, x_2, 0, x_3, \ldots, 0, x_j, 0, \ldots, x_{n-1}, 0, x_n)$$

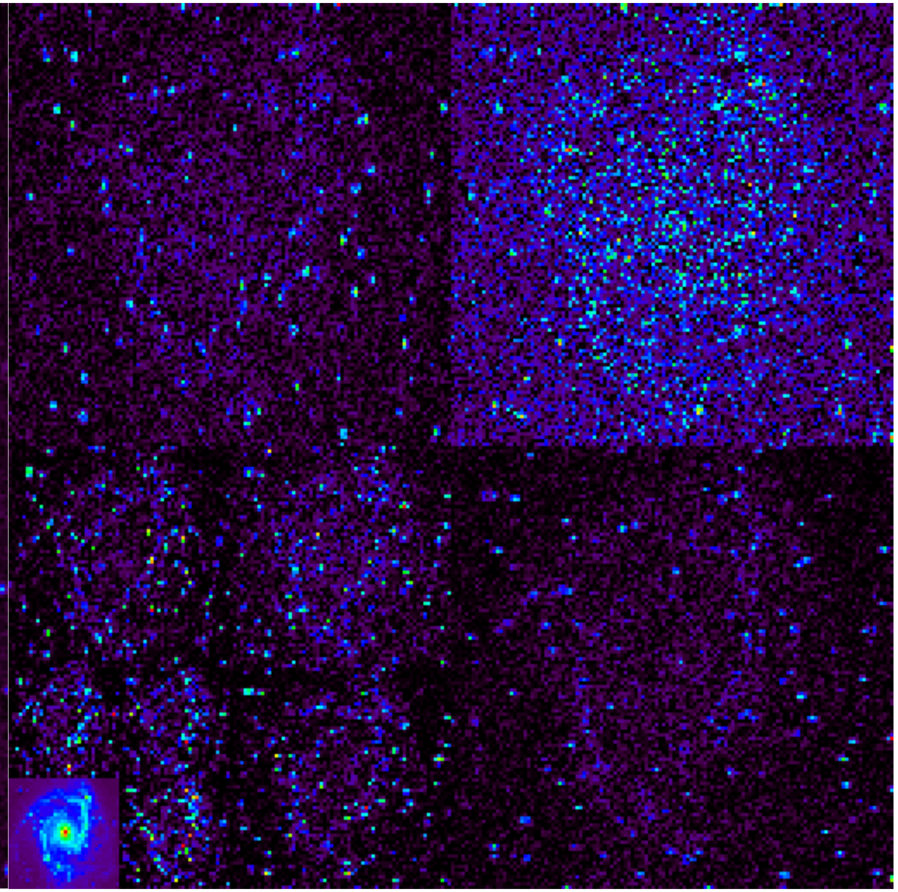

G
H
H
G
H
G
H
G
HH
Smooth
GH
Vertical
HG
Horizontal
GG
Diagonal



NGC2997



NGC2997 WT



JPEG/JPEG 2000

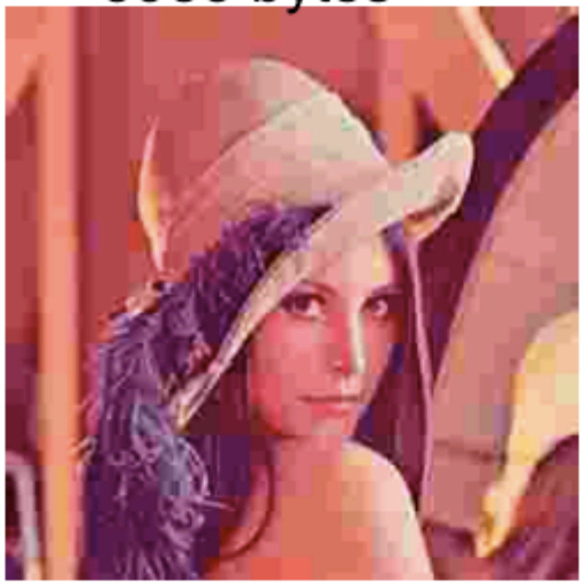
Original BMP

300x300x24

270056 bytes

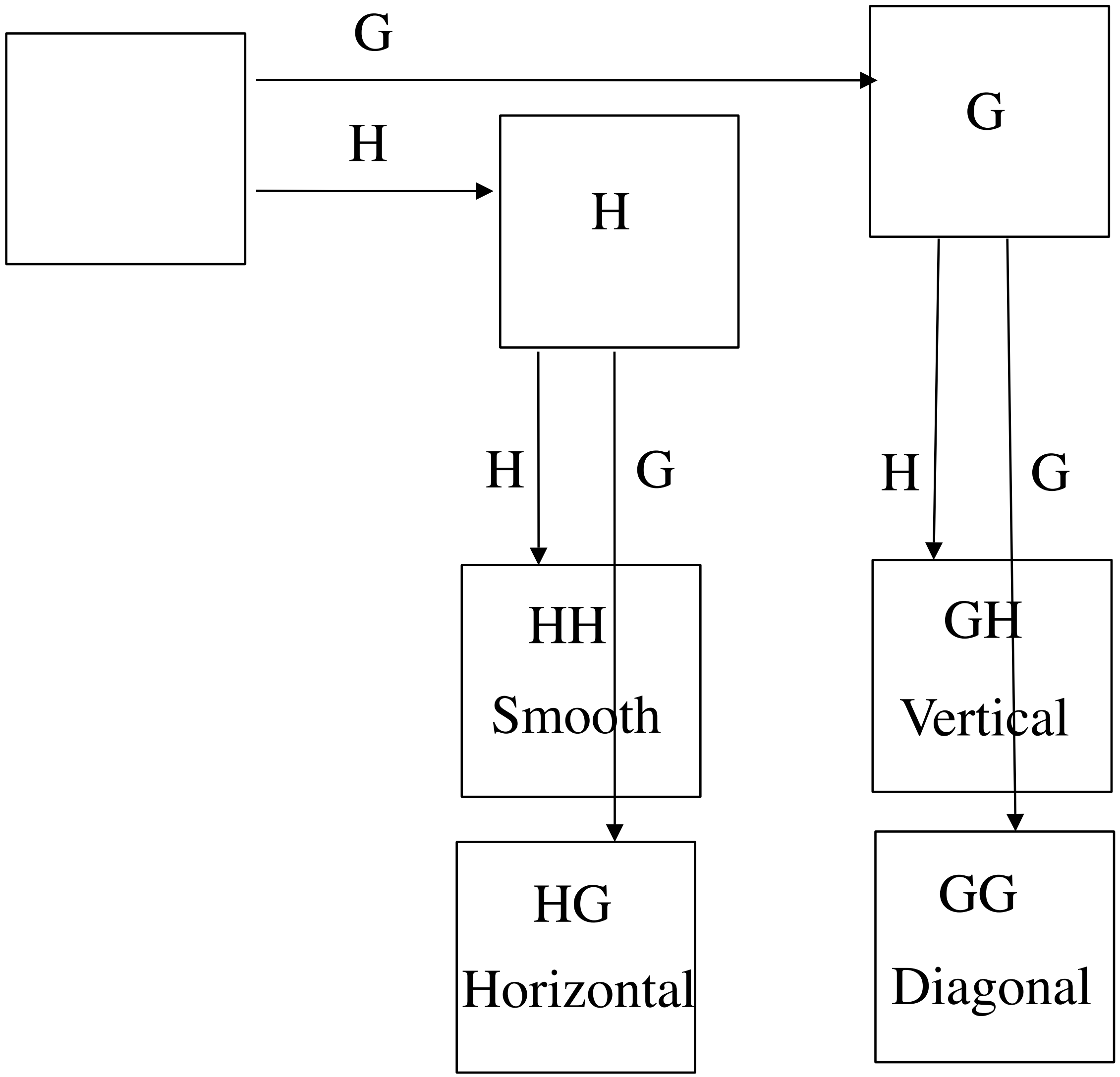
JPEG 1:68

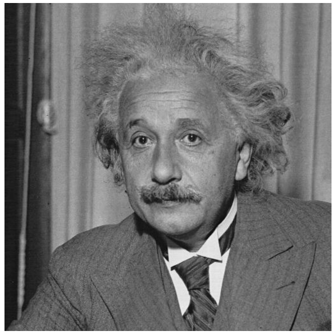
3983 bytes



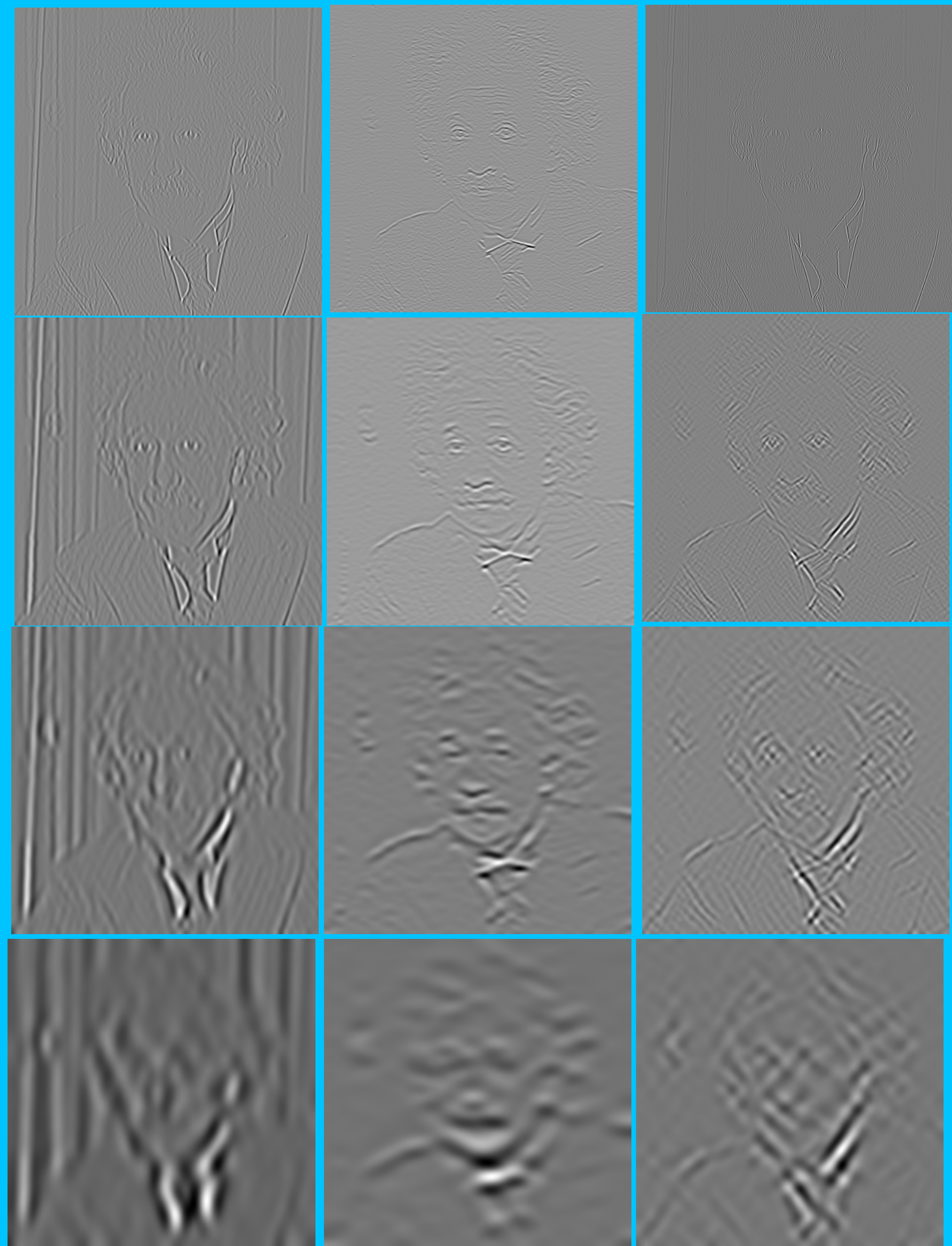
JPEG2000 1:70

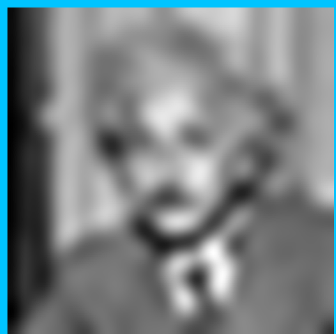
3876 bytes

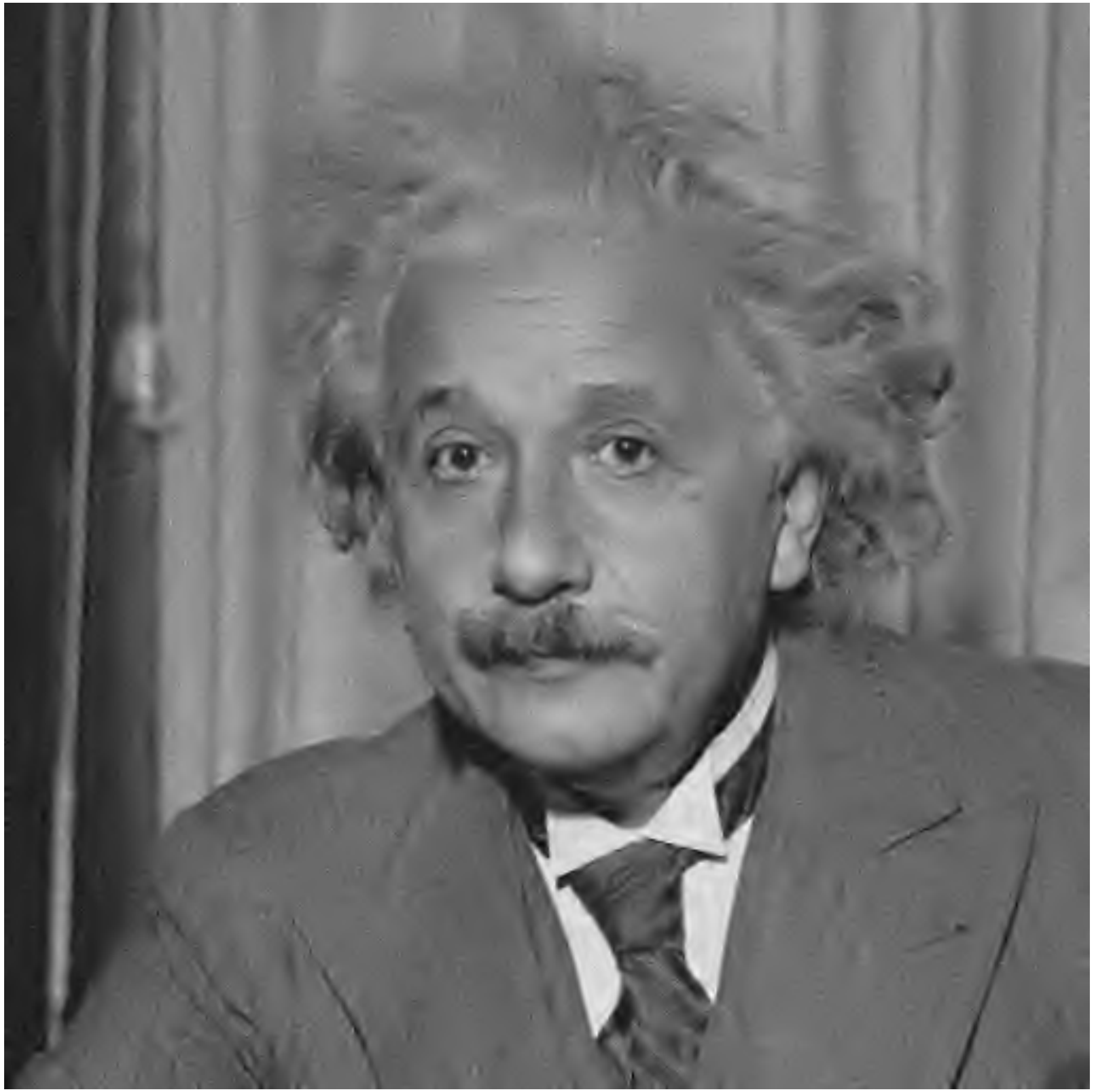

G
H
G
H
H
G
H
G
HH
Smooth
GH
Vertical
HG
Horizontal
GG
Diagonal



Undecimated Wavelet Transform







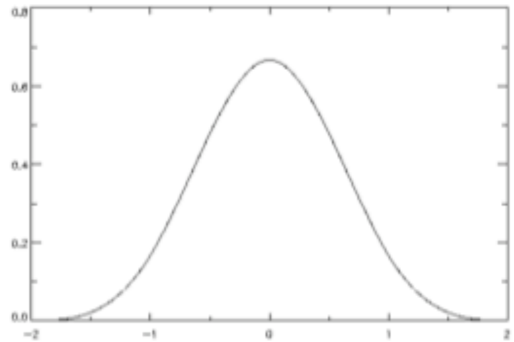
Wavelet Transform in Astronomy

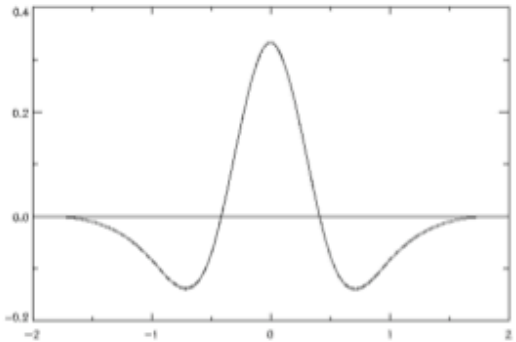
The Isotropic Wavelet and Scaling Functions

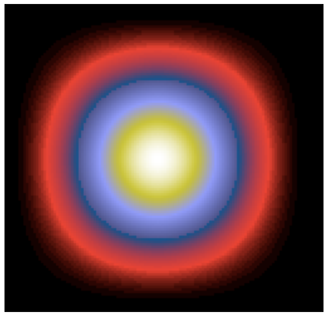
$$B_3(x) = \frac{1}{12}(\mid x-2\mid^3 - 4\mid x-1\mid^3 + 6\mid x\mid^3 - 4\mid x+1\mid^3 + \mid x+2\mid^3)$$

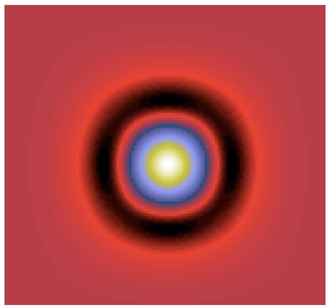
$$\psi(x,y) = B_3(x)B_3(y)$$

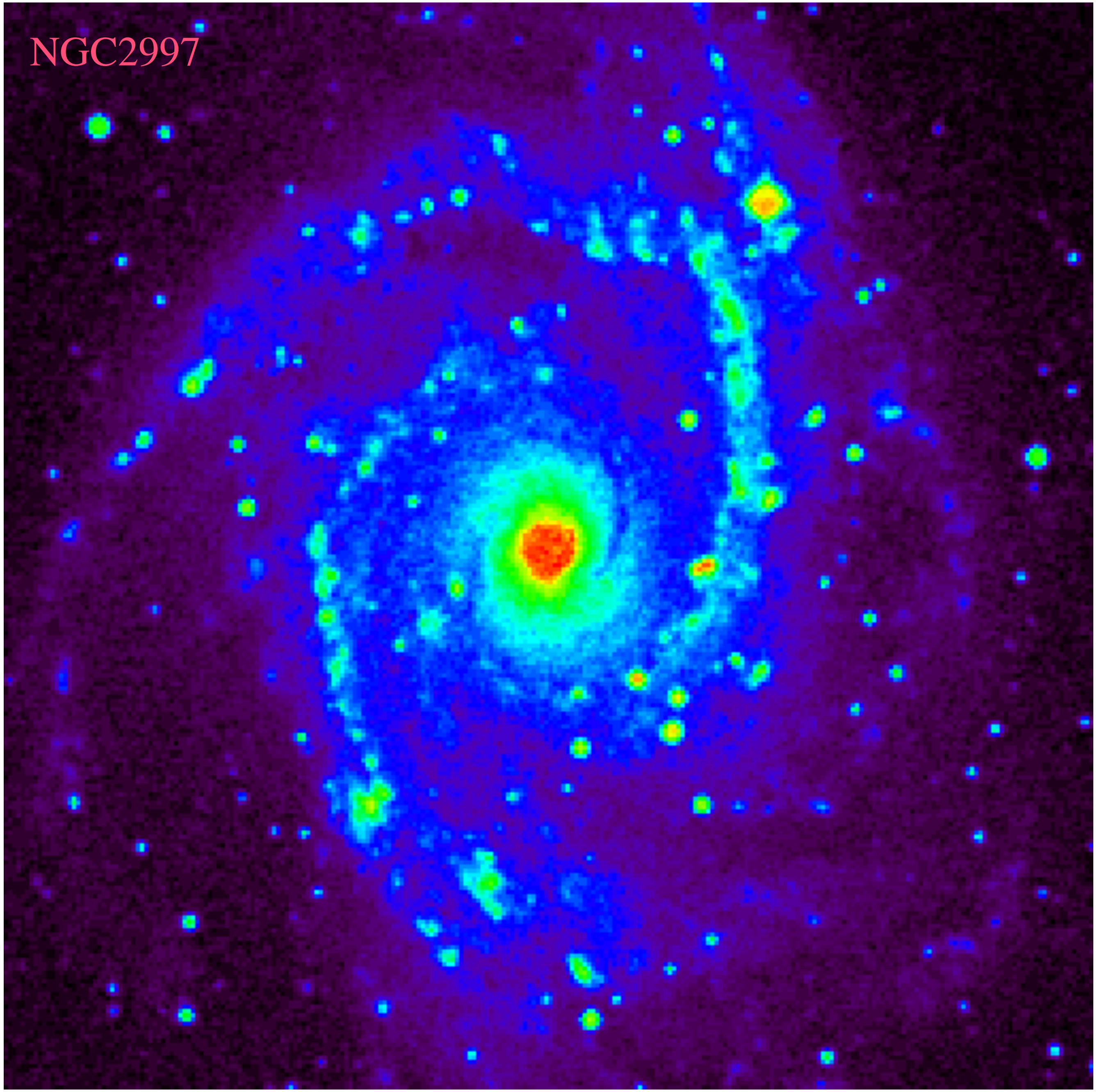
$$\frac{1}{4}\psi(\frac{x}{2},\frac{y}{2}) = \phi(x,y) - \frac{1}{4}\phi(\frac{x}{2},\frac{y}{2})$$



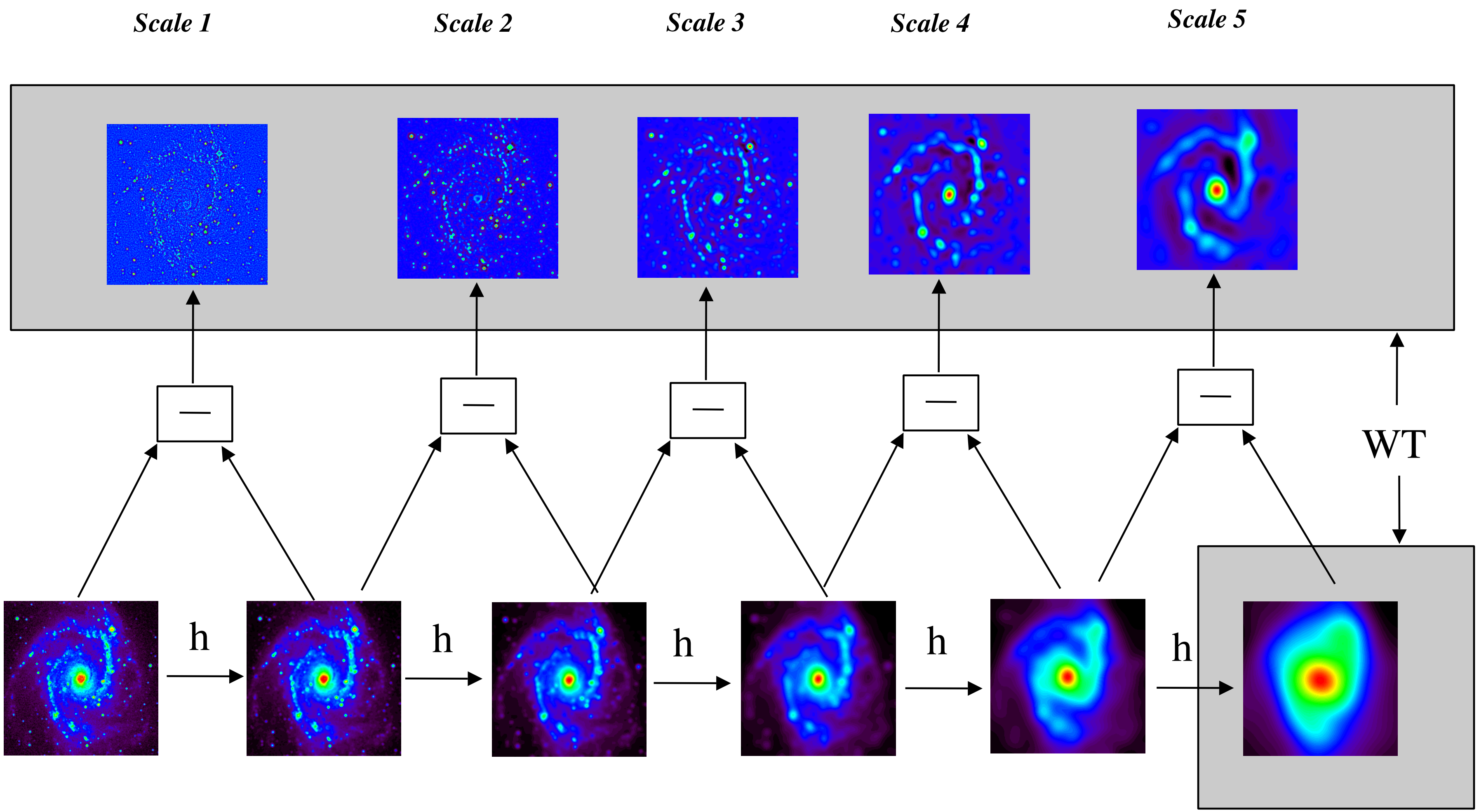







NGC2997

ISOTROPIC UNDECIMATED WAVELET TRANSFORM

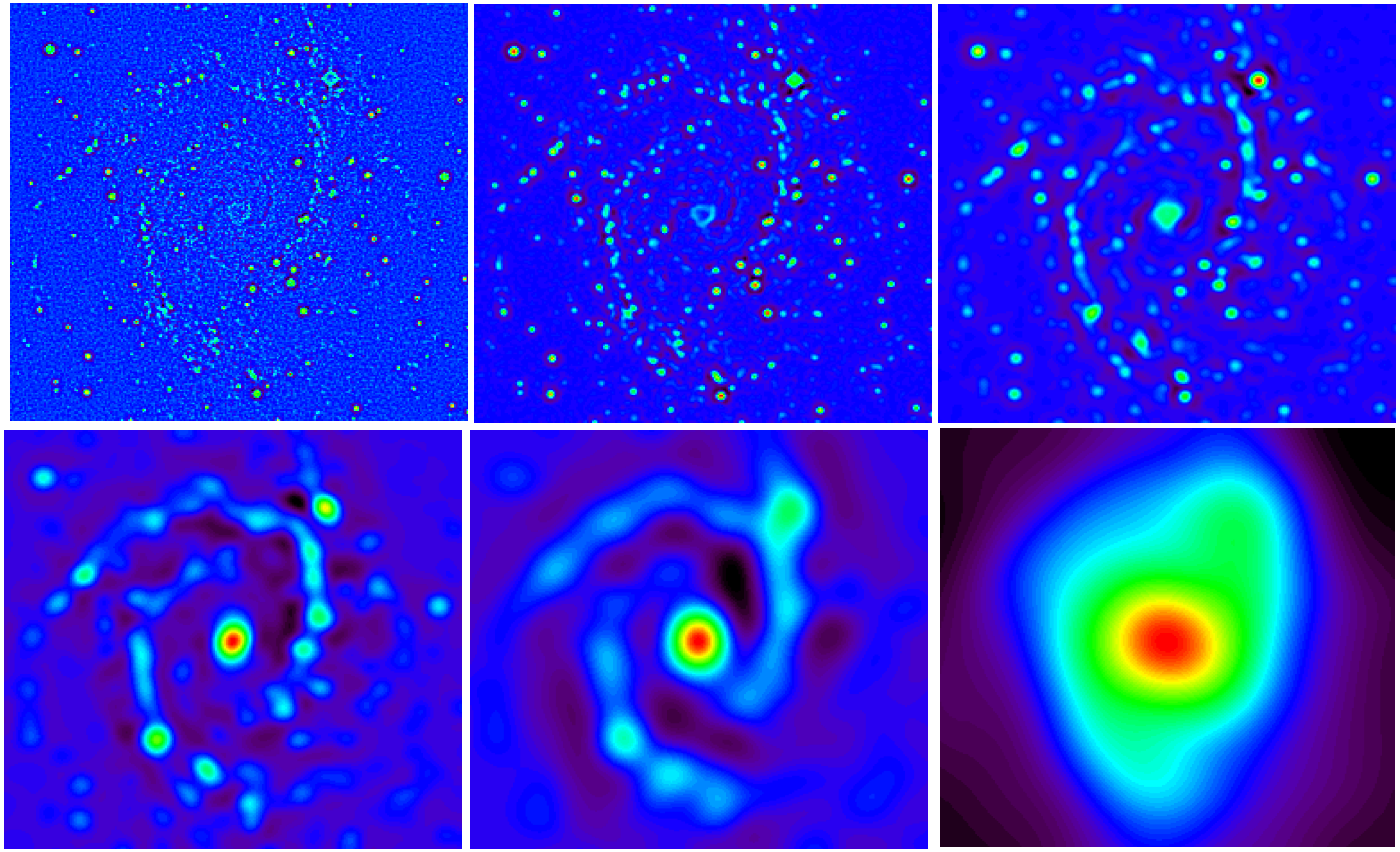


The STARLET Transform
Isotropic Undecimated Wavelet Transform (a trous algorithm)

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

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

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

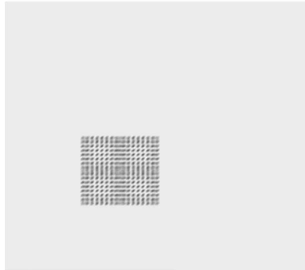


Looking for adapted representations

Local DCT

Stationary textures

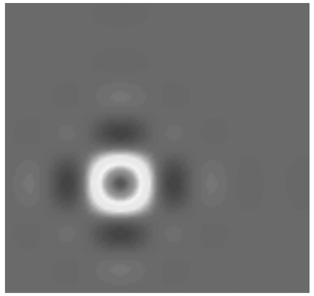
Locally oscillatory



Wavelet transform

Piecewise smooth

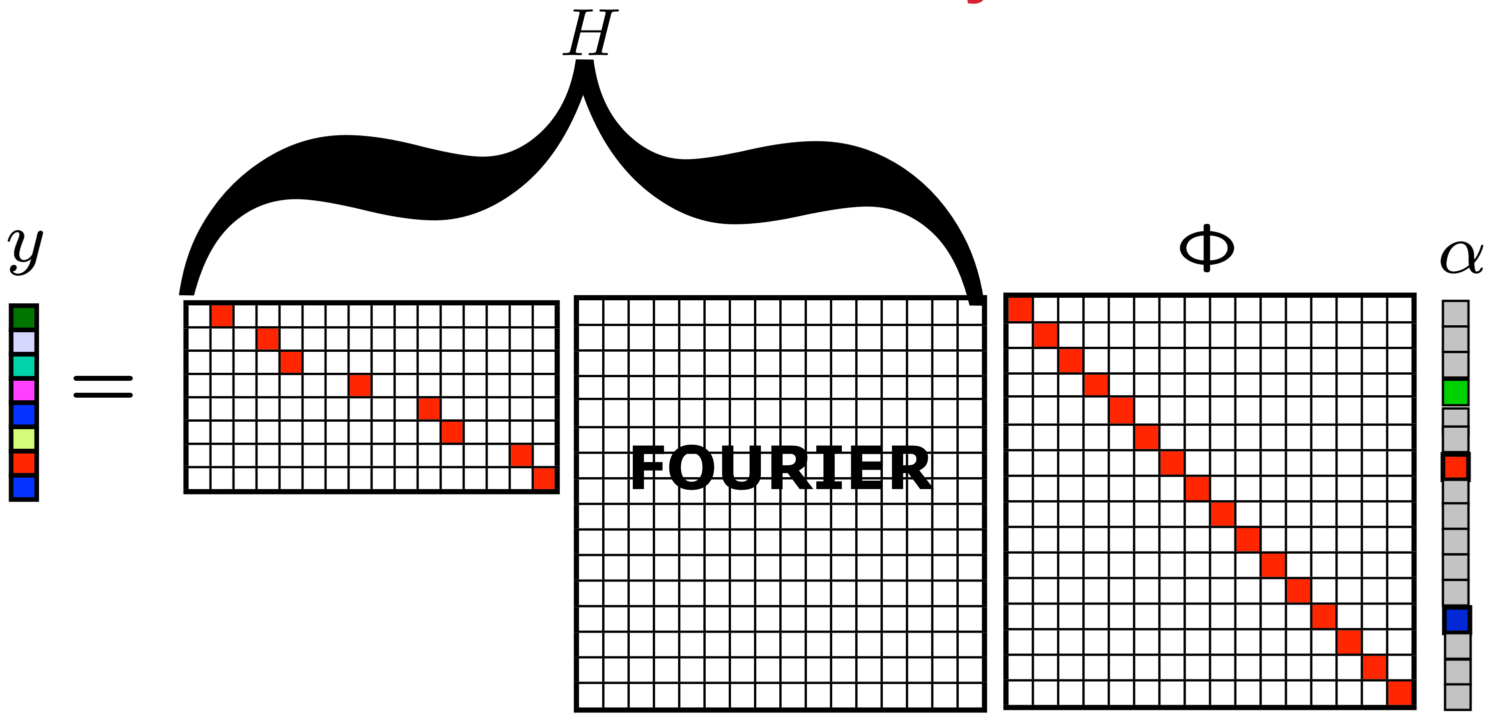
Isotropic structures



Curvelet transform

Piecewise smooth,
edge

Radio-Interferometry

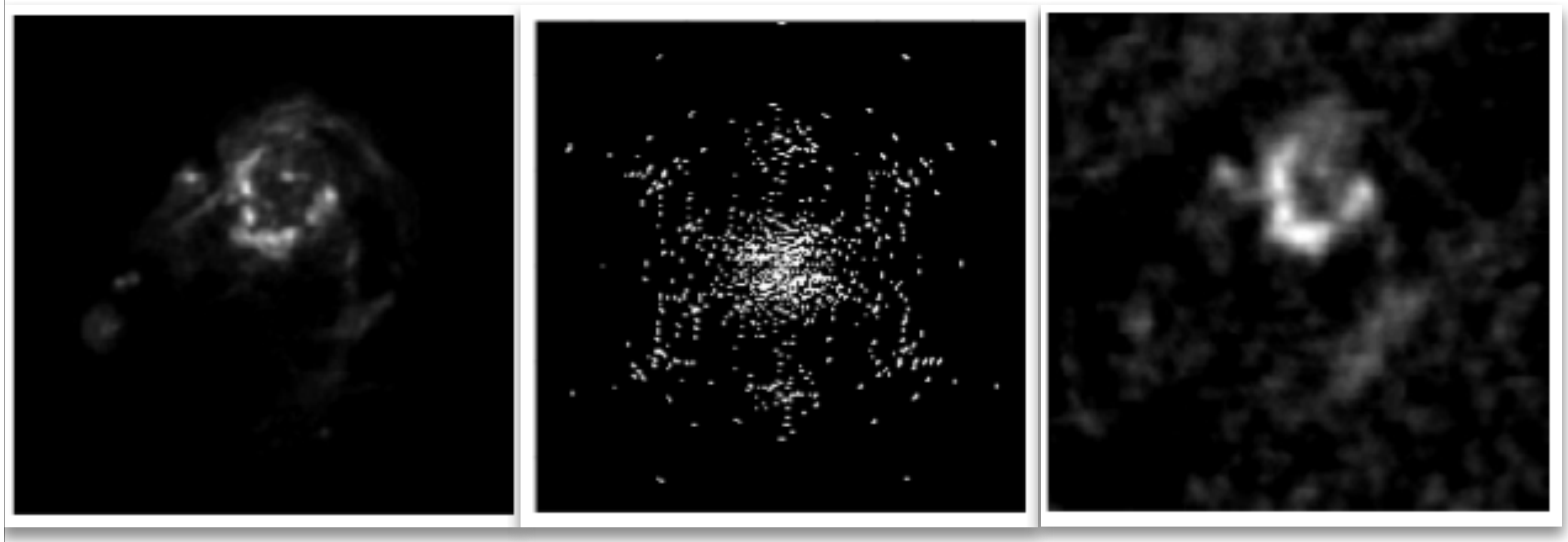


Measurement System

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

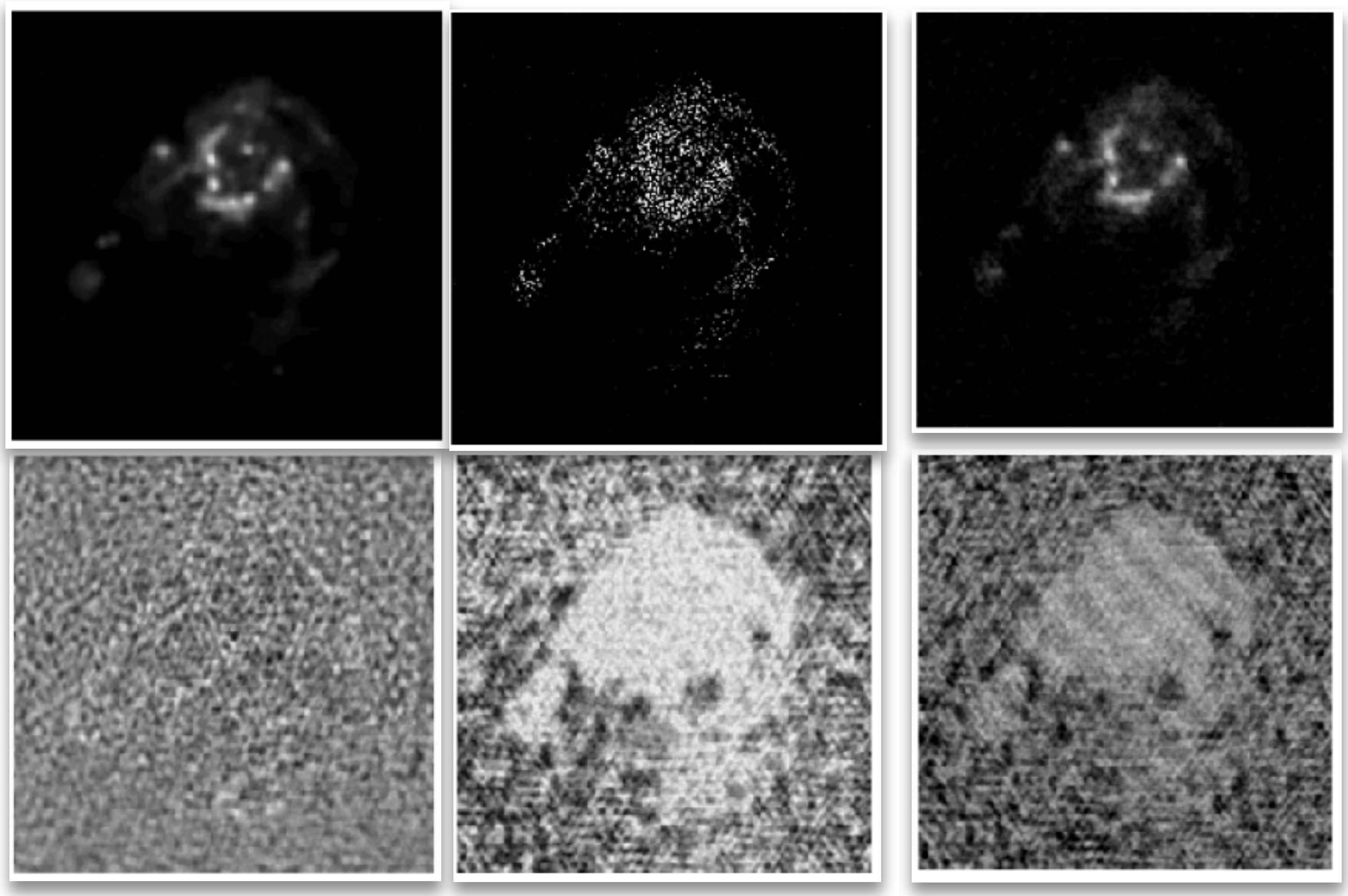
CS-Radio Astronomy

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



Australian Square Kilometer Array Pathfinder (ASKAP) radio telescope.

CS-Radio Astronomy

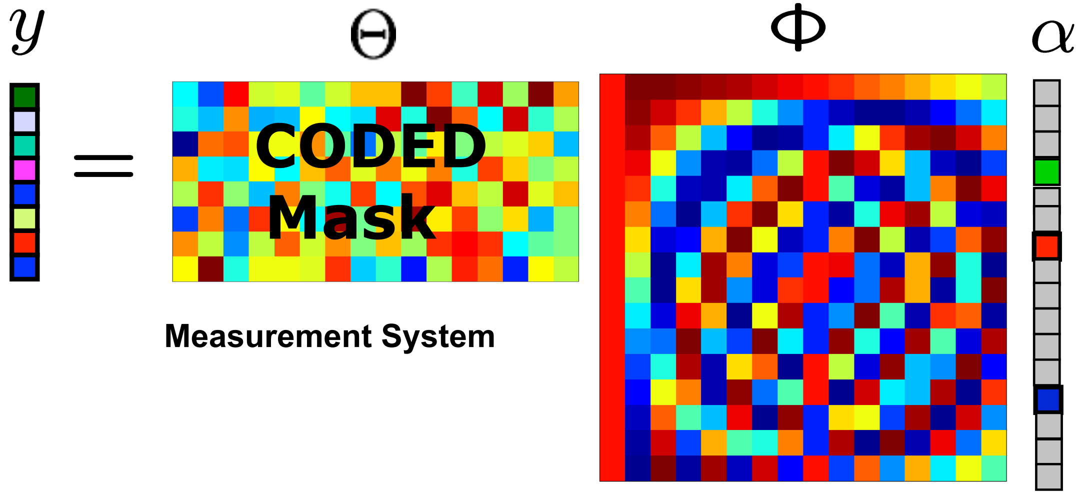


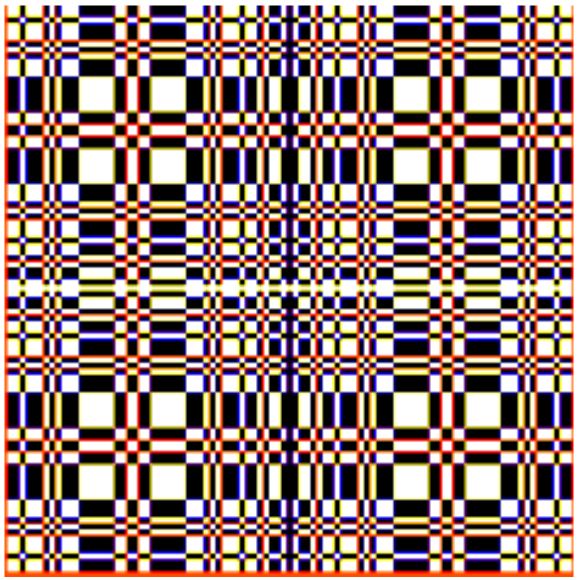
Hogbom CLEAN

MEM residual

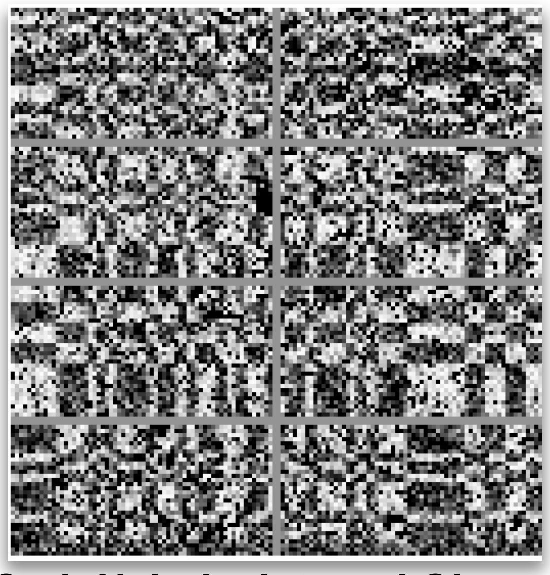
==> McEwen & Wiaux presentation in the ICIP Astronomy and Cosmology Session.

Gamma Ray Instruments (Integral) - Acquisition with coded masks





INTEGRAL/IBIS Coded Mask



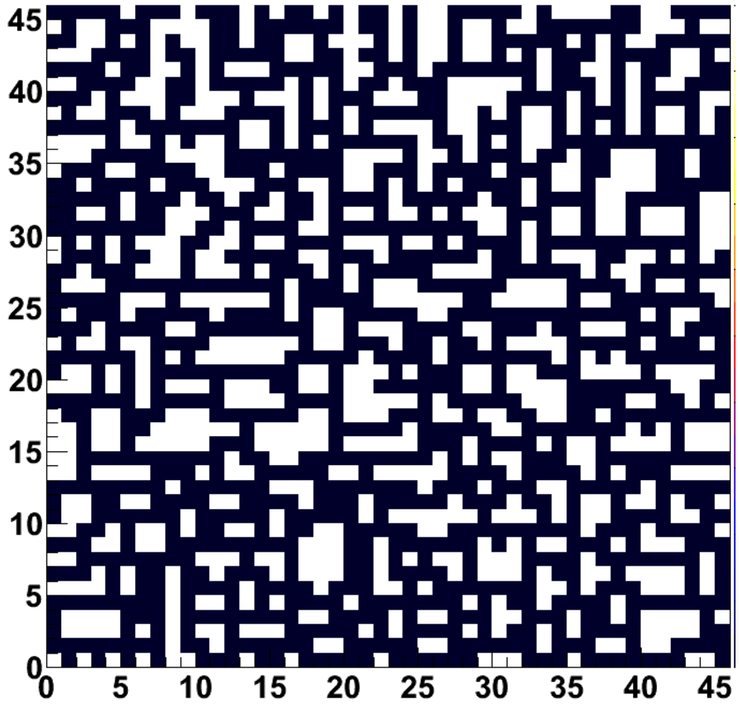
Crab Nebula Integral Observation

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

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.

Multiscale Transforms

<u>Critical Sampling</u>

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

<u>Redundant Transforms</u>

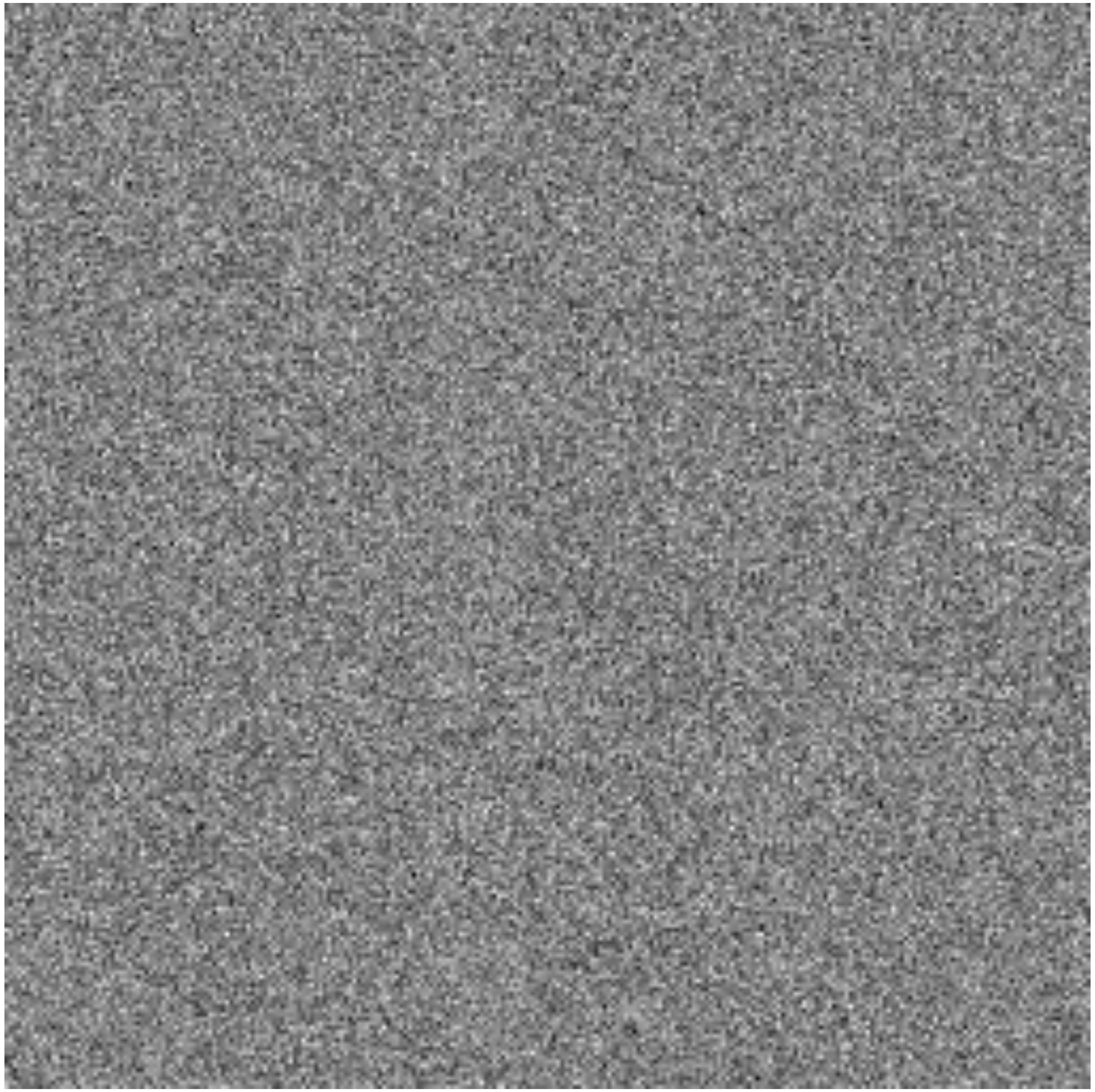
Pyramidal decomposition (Burt and Adelson)
Undecimated Wavelet Transform
Isotropic Undecimated Wavelet Transform
Complex Wavelet Transform
Steerable Wavelet Transform
Dyadic Wavelet Transform
Nonlinear Pyramidal decomposition (Median)

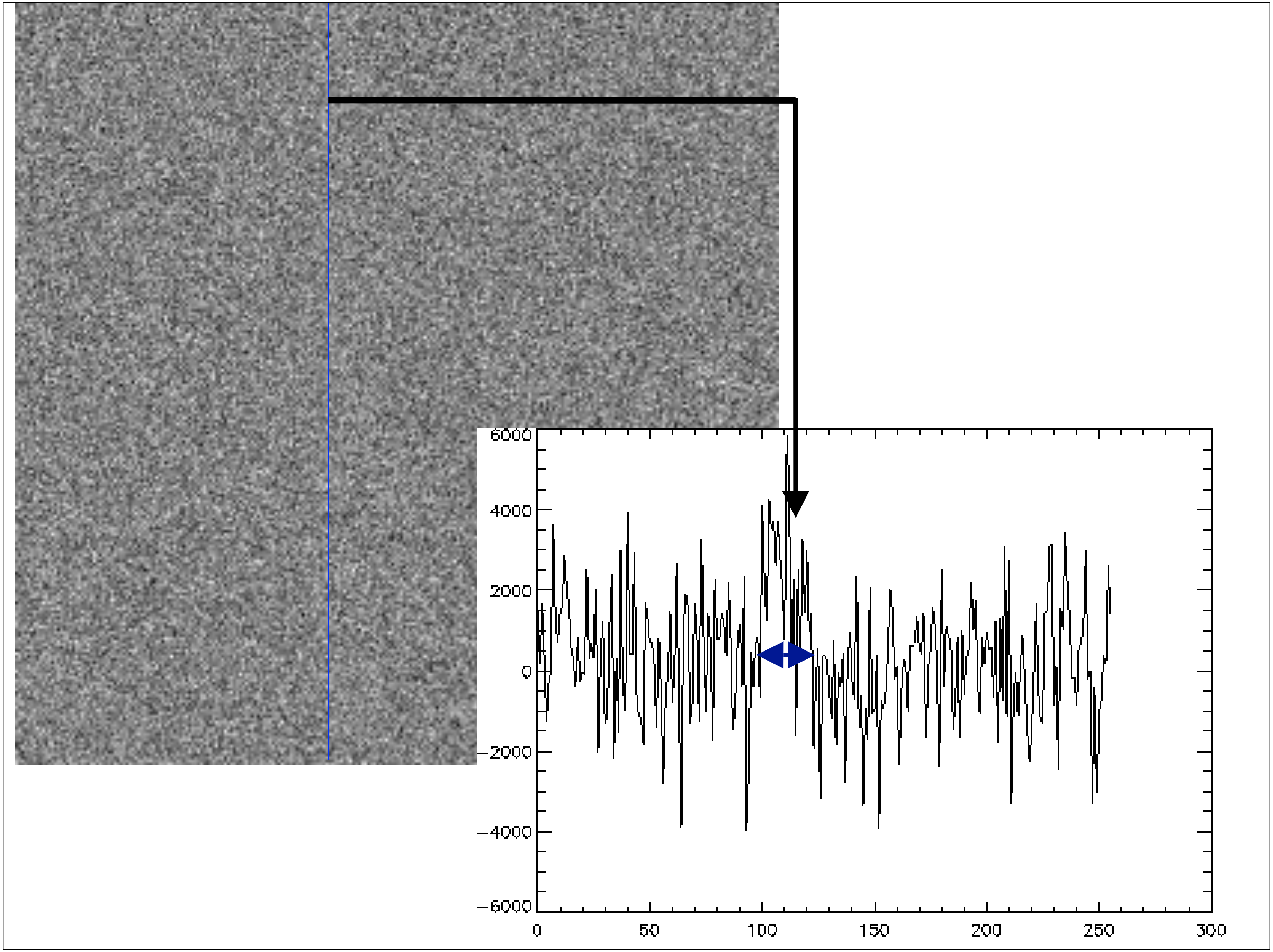
New Multiscale Construction

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

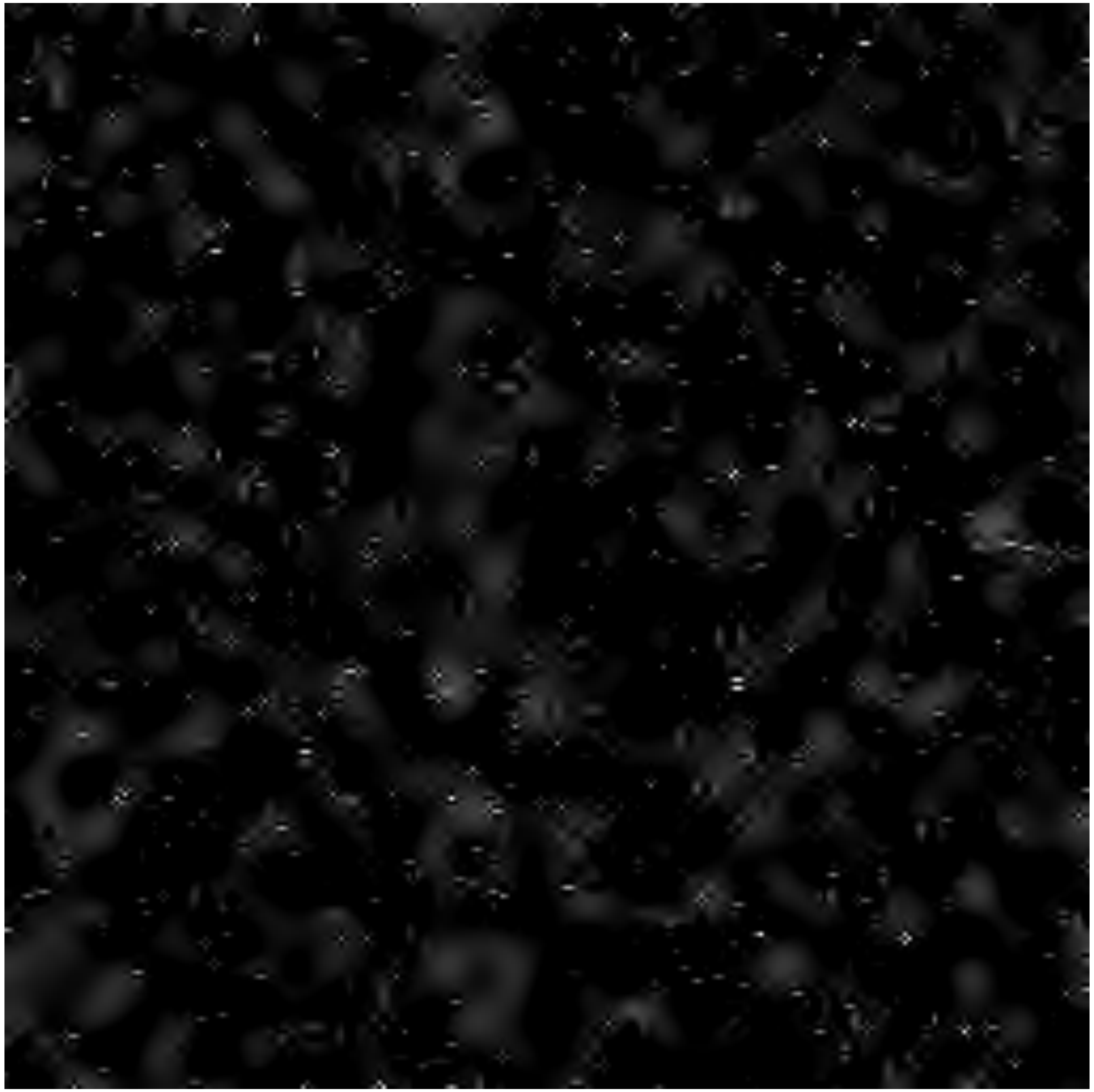
Ridgelet
Curvelet (Several implementations)
Wave Atom

SNR = 0.1

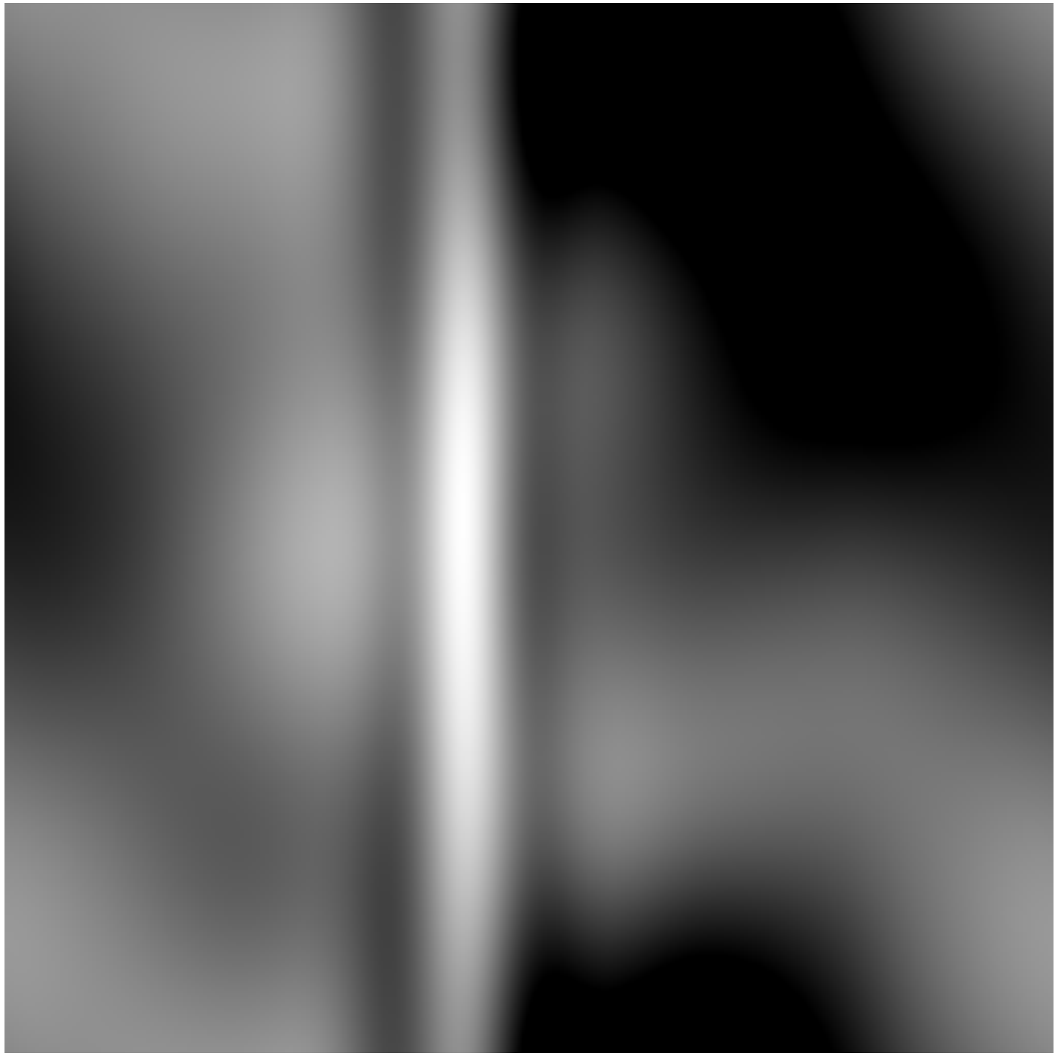



6000
4000
2000
0
−2000
−4000
−6000
0
50
100
150
200
250
300

Undecimated Wavelet Filtering (3 sigma)



Ridgelet Filtering (5sigma)

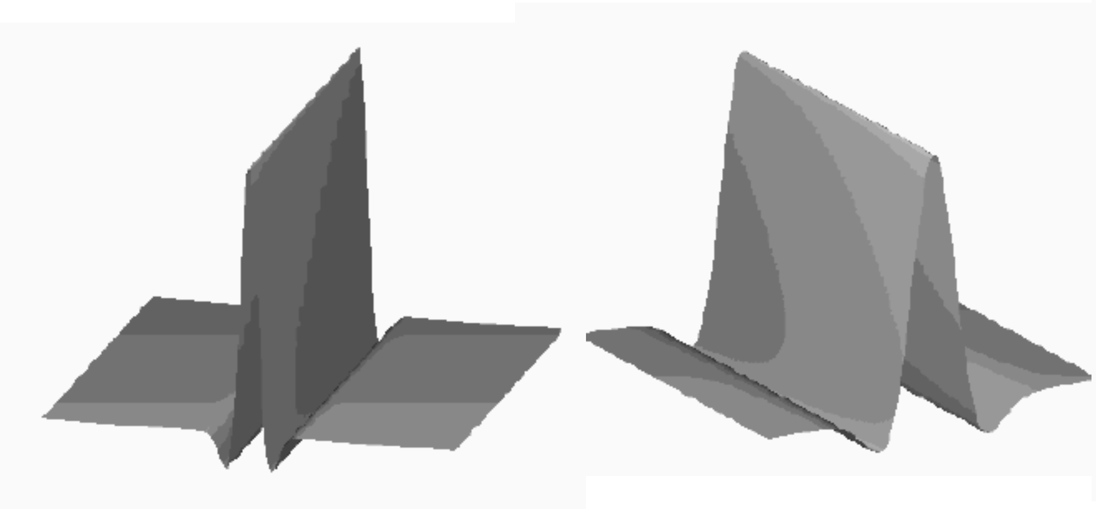


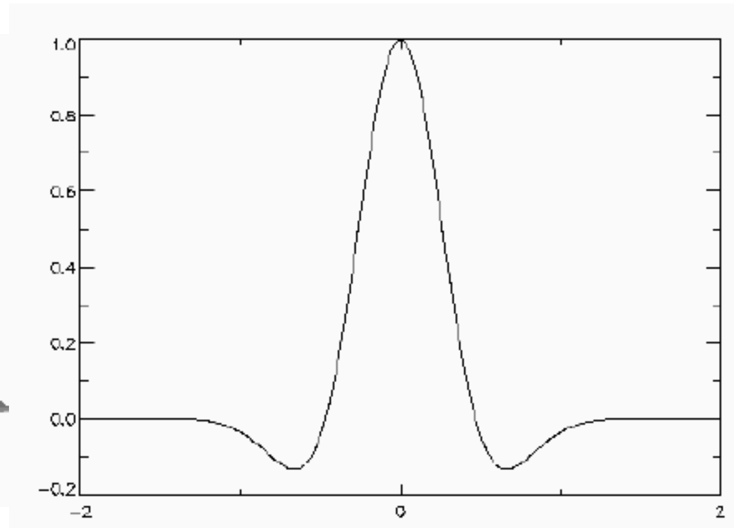
Continuous Ridgelet Transform

Ridgelet Transform (Candes, 1998): $R_f(a,b,\theta) = \int \psi_{a,b,\theta}(x) f(x) dx$

Ridgelet function: $\psi_{a,b,\theta}(x) = a^{\frac{1}{2}} \psi\left(\frac{x_1 \cos(\theta) + x_2 \sin(\theta) - b}{a}\right)$

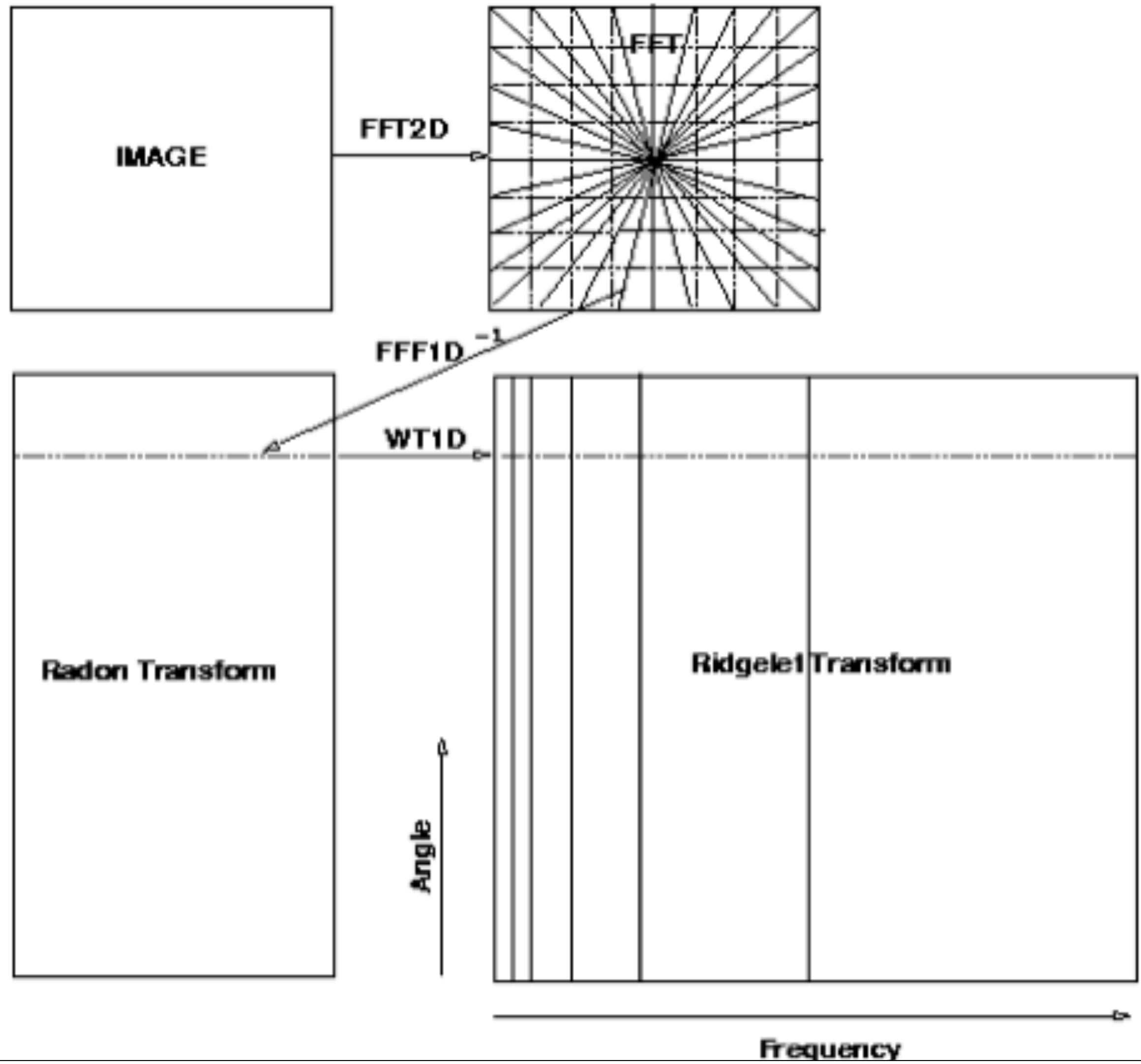
The function is constant along lines. Transverse to these ridges, it is a wavelet.





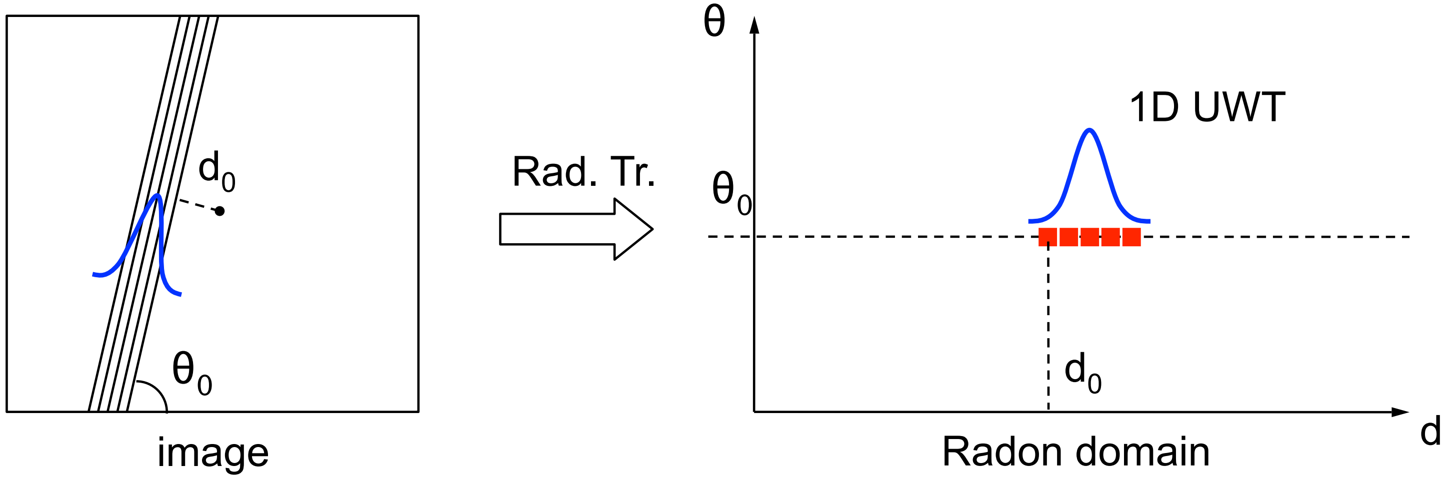
The ridgelet coefficients of an object f are given by analysis of the Radon transform via:

$$R_f(a,b,\theta) = \int Rf(\theta,t)\psi\left(\frac{t-b}{a}\right)dt$$



Ridgelet Denoising

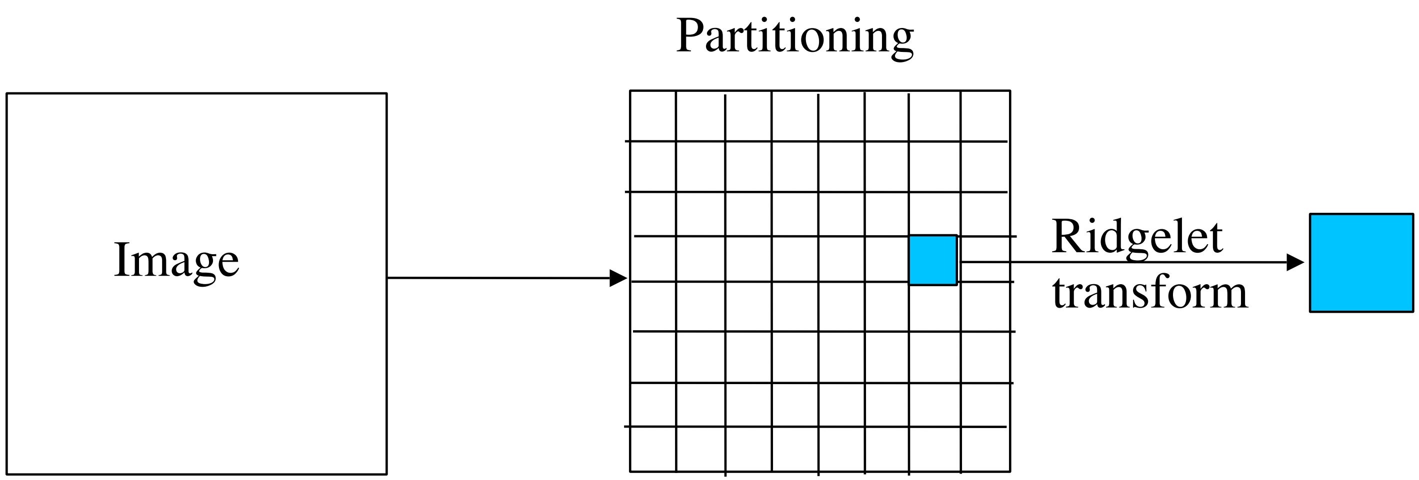
- Ridgelet transform: Radon + 1D Wavelet



1. Rad. Tr.
2. For each line, apply the same denoising scheme as before
3. Rad. Tr.$^{-1}$

Local Ridgelet Transform

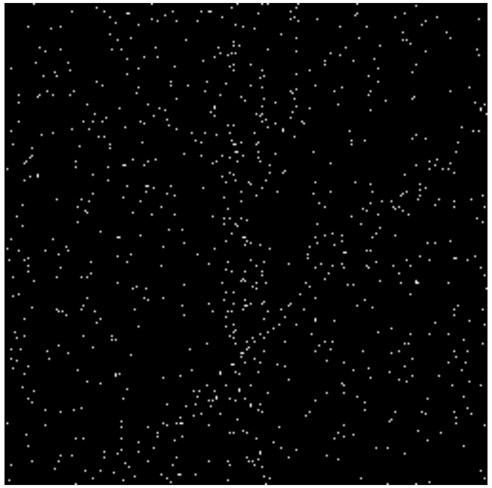
The ridgelet transform is optimal to find only lines of the size of the image.
To detect line segments, a partitioning must be introduced. The image is decomposed into blocks, and the ridgelet transform is applied on each block.



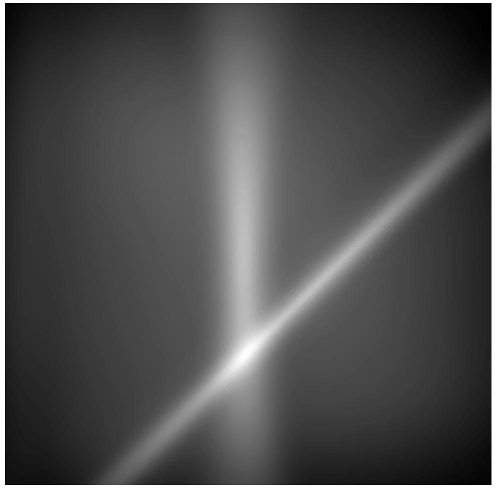
Poisson Noise and Line-Like Sources Restoration (MS-VST + Ridgelet)

B. Zhang, M.J. Fadili and J.-L. Starck, "Wavelets, Ridgelets and Curvelets for Poisson Noise Removal" ,ITIP, 2008.

underlying intensity image



simulated image of counts

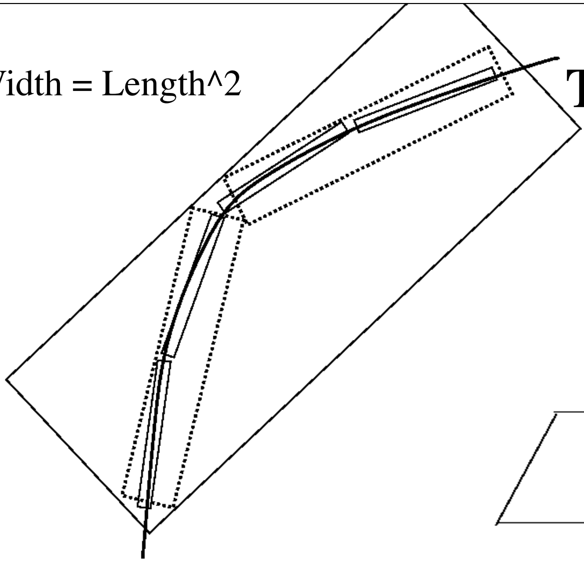


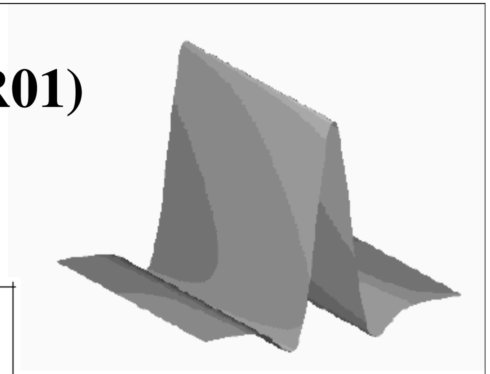
restored image
from the left image of counts

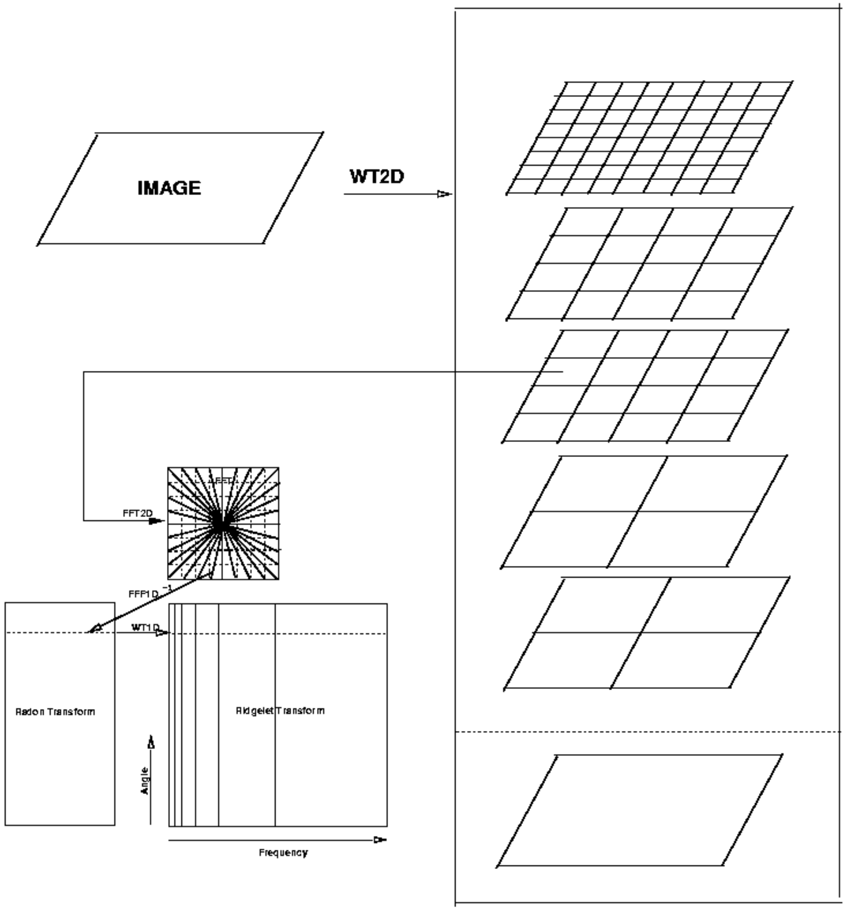
Max Intensity
background = 0.01
vertical bar = 0.03
inclined bar = 0.04

The Curvelet Transform (CUR01)

Width = Length^2



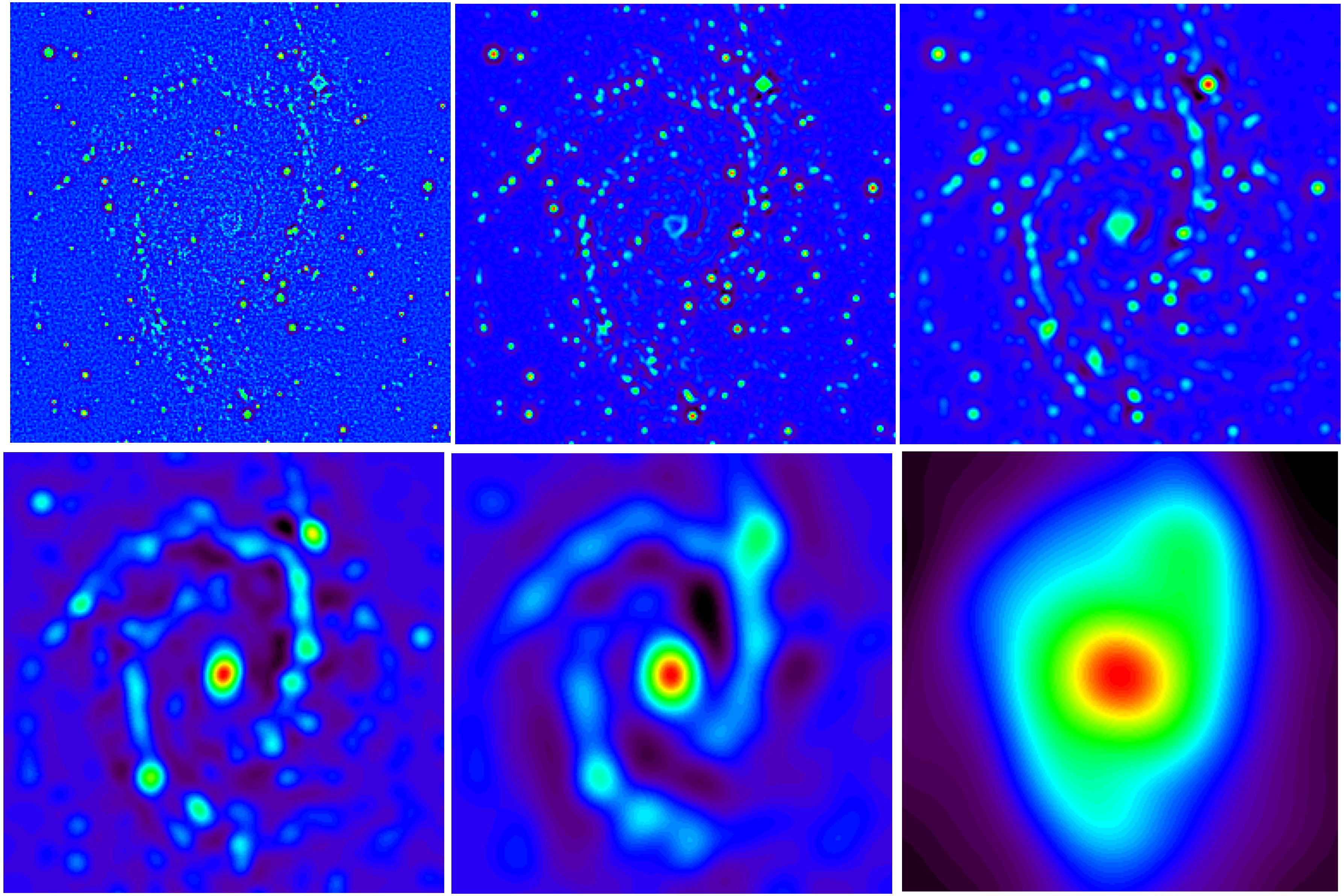




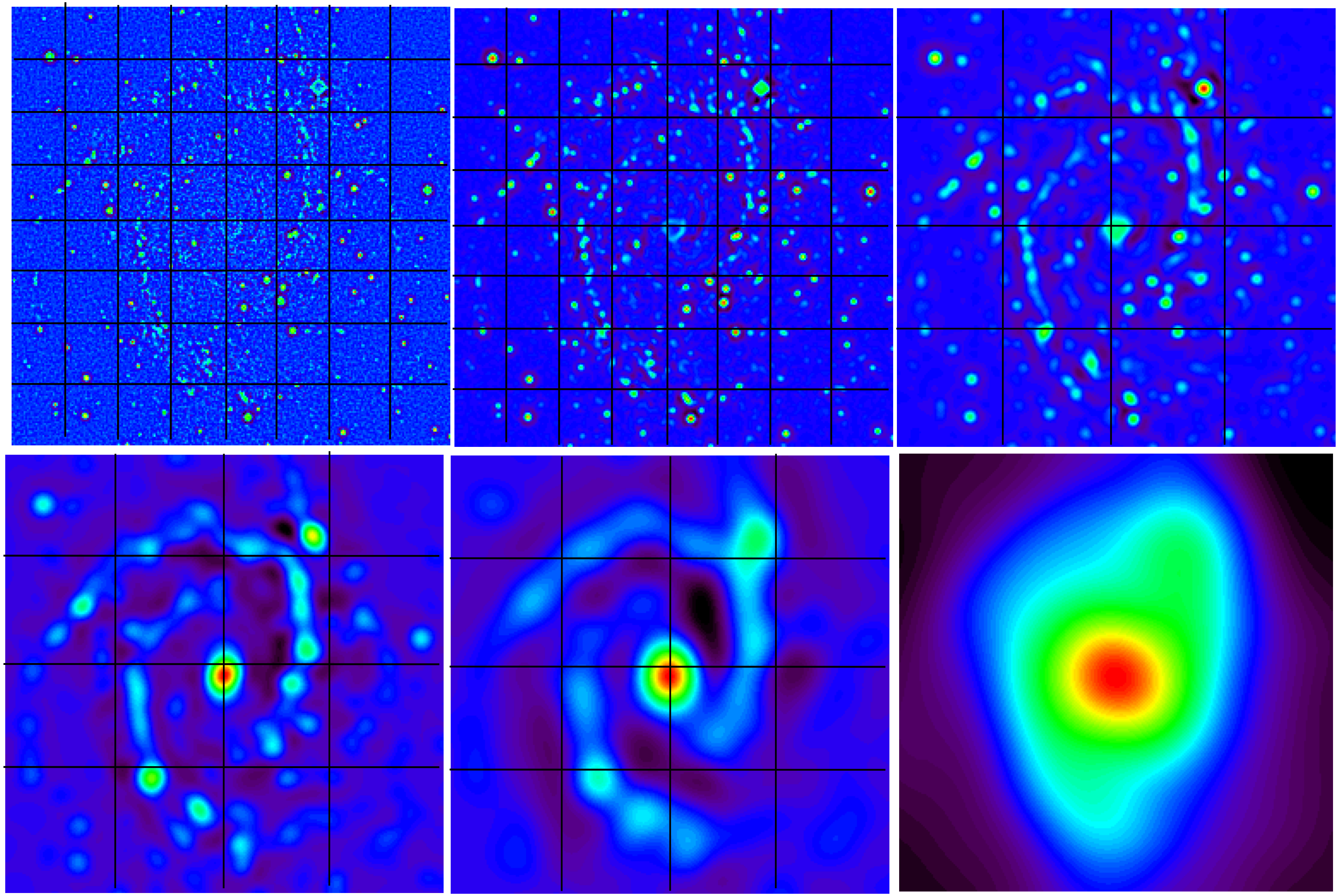
J.-L. Starck, E. Candes, D.L. Donoho ***The Curvelet Transform for Image Denoising****, IEEE Transaction on Image Processing, 11, 6, 2002.*

Undecimated Isotropic WT:

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



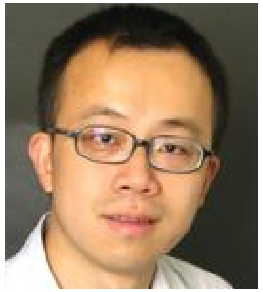
PARTITIONING

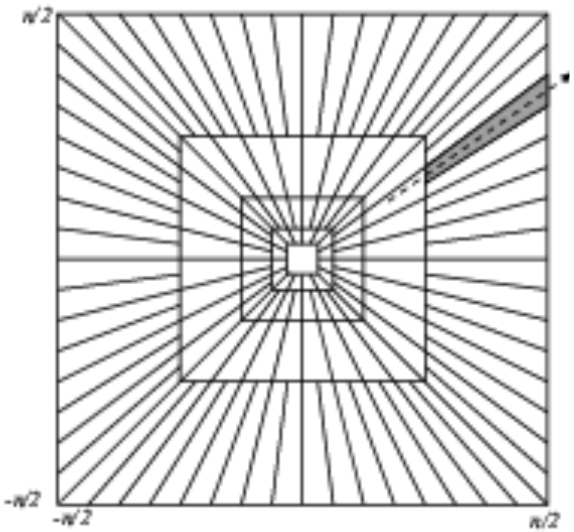


The Fast Curvelet Transform, Candes et al, 2005

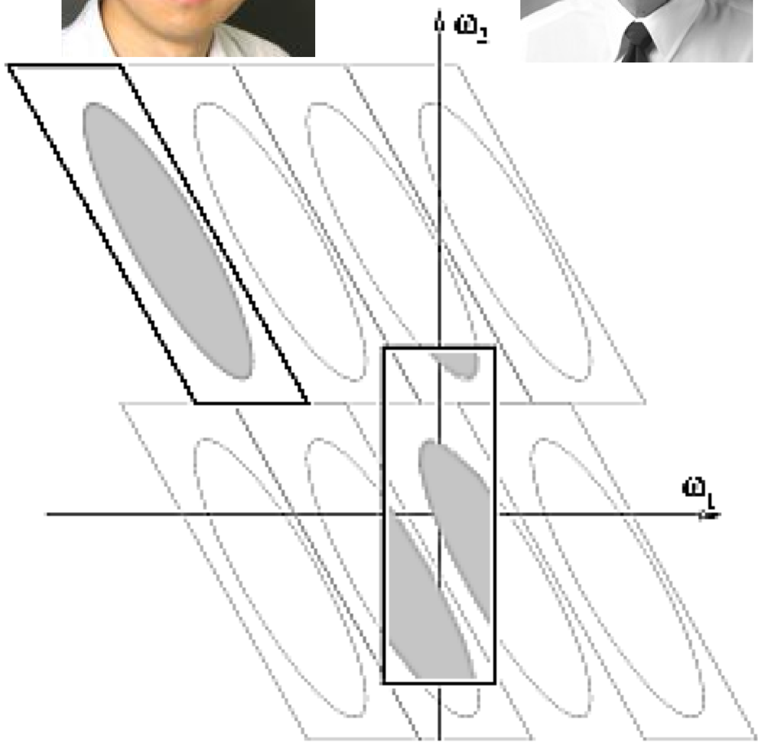
CUR03 - Fast Curvelet Transform using the USFFT
CUR04 - Fast Curvelet Transform using the Wrapping and 2DFFT





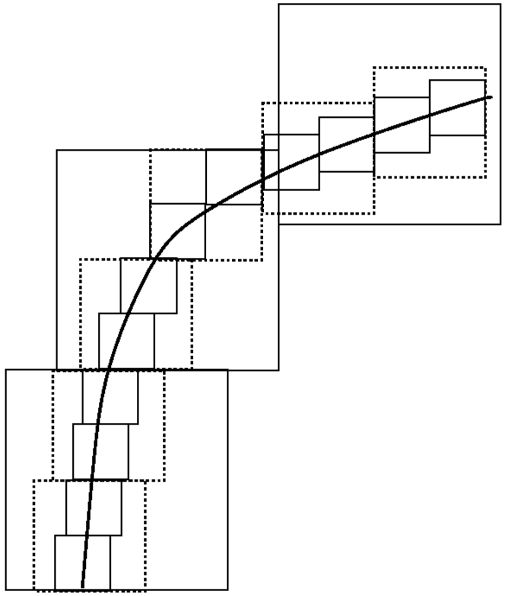



(a)

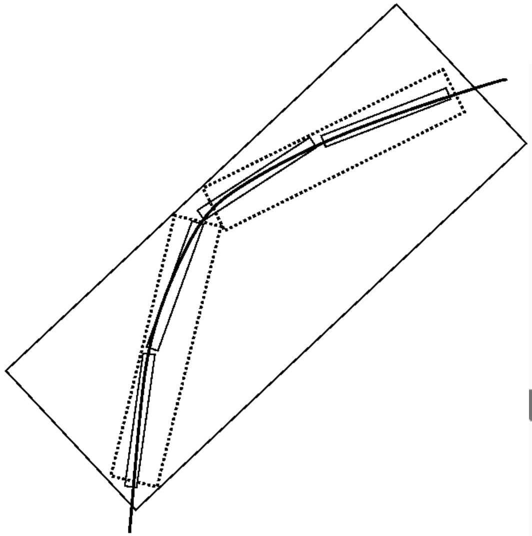


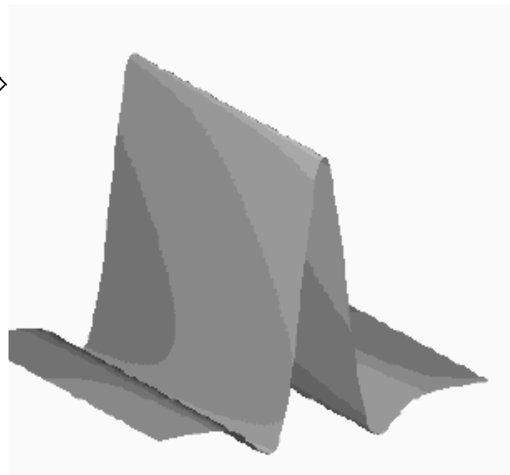
Wavelets and edges

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

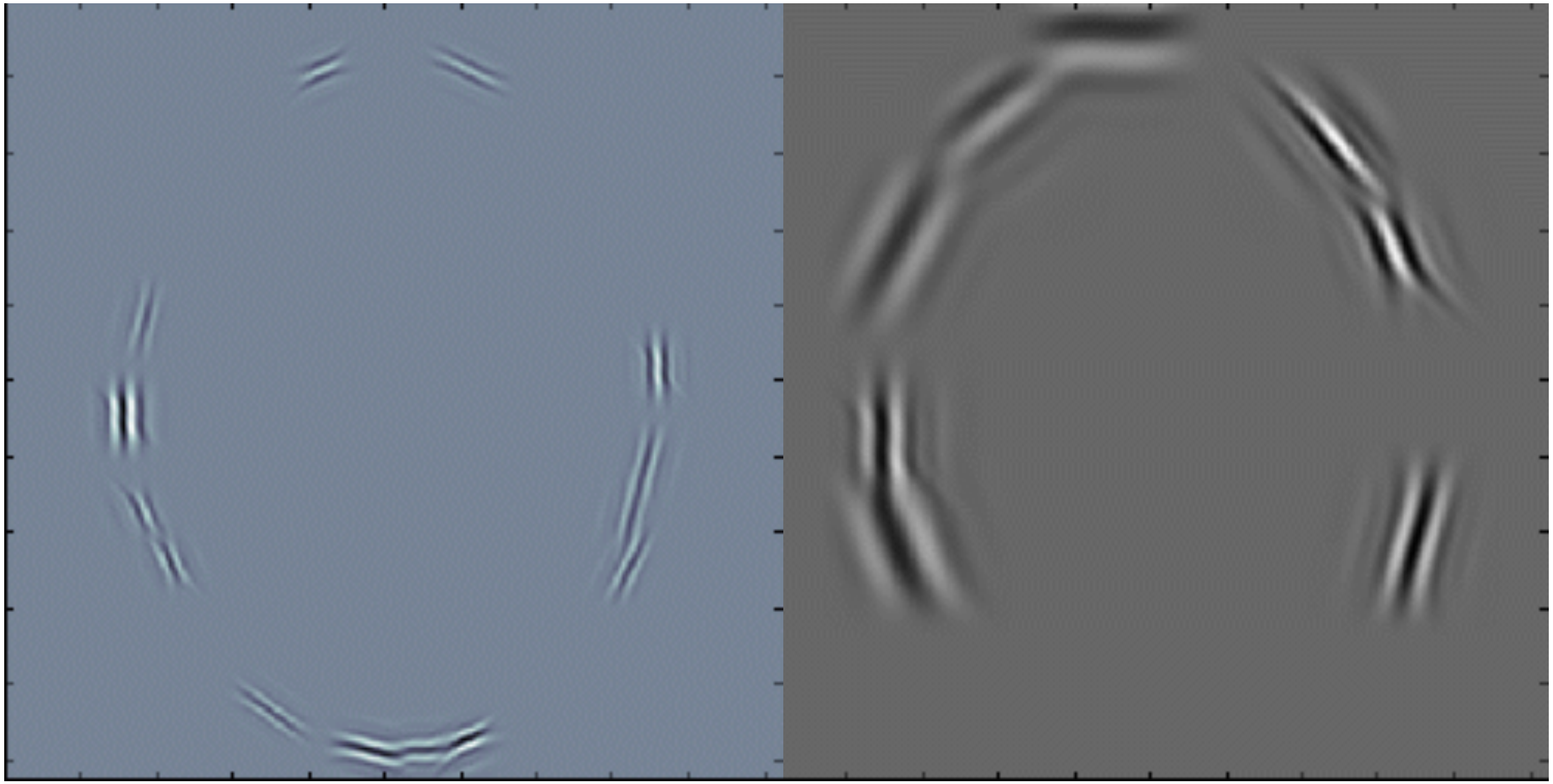


• need dictionaries of strongly anisotropic atoms :





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 Tansform**" , 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 USING THE CURVELET TRANSFORM

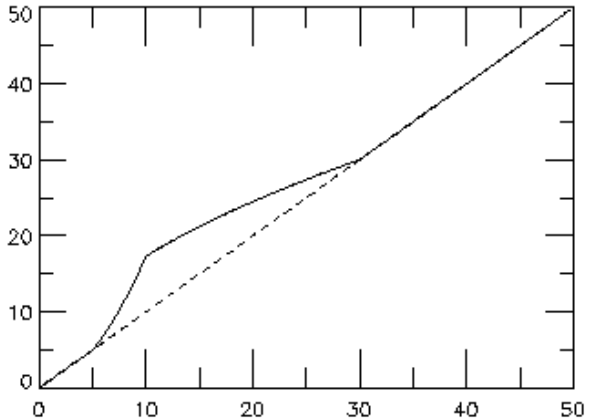
J.-L Starck, F. Murtagh, E. Candes and D.L. Donoho, ***"Gray and Color Image Contrast Enhancement by the Curvelet Transform"***,

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

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

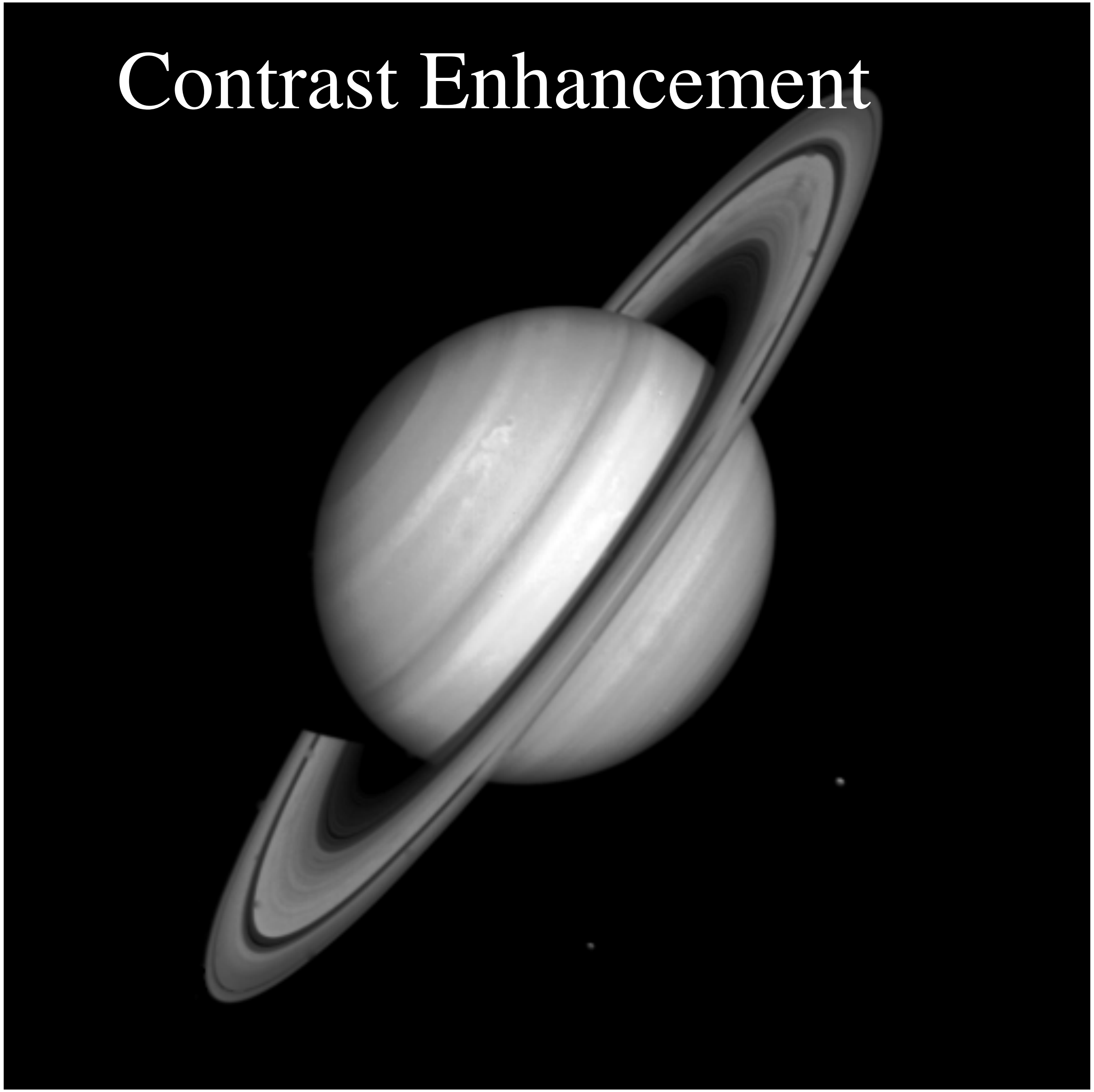
$$\begin{cases} y_c(x,\sigma) = 1 & \text{if } x < c\sigma \\ y_c(x,\sigma) = \dfrac{x - c\sigma}{c\sigma}\left(\dfrac{m}{c\sigma}\right)^p + \dfrac{2c\sigma - x}{c\sigma} & \text{if } x < 2c\sigma \\ y_c(x,\sigma) = \left(\dfrac{m}{x}\right)^p & \text{if } 2c\sigma \le x < m \\ y_c(x,\sigma) = \left(\dfrac{m}{x}\right)^s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

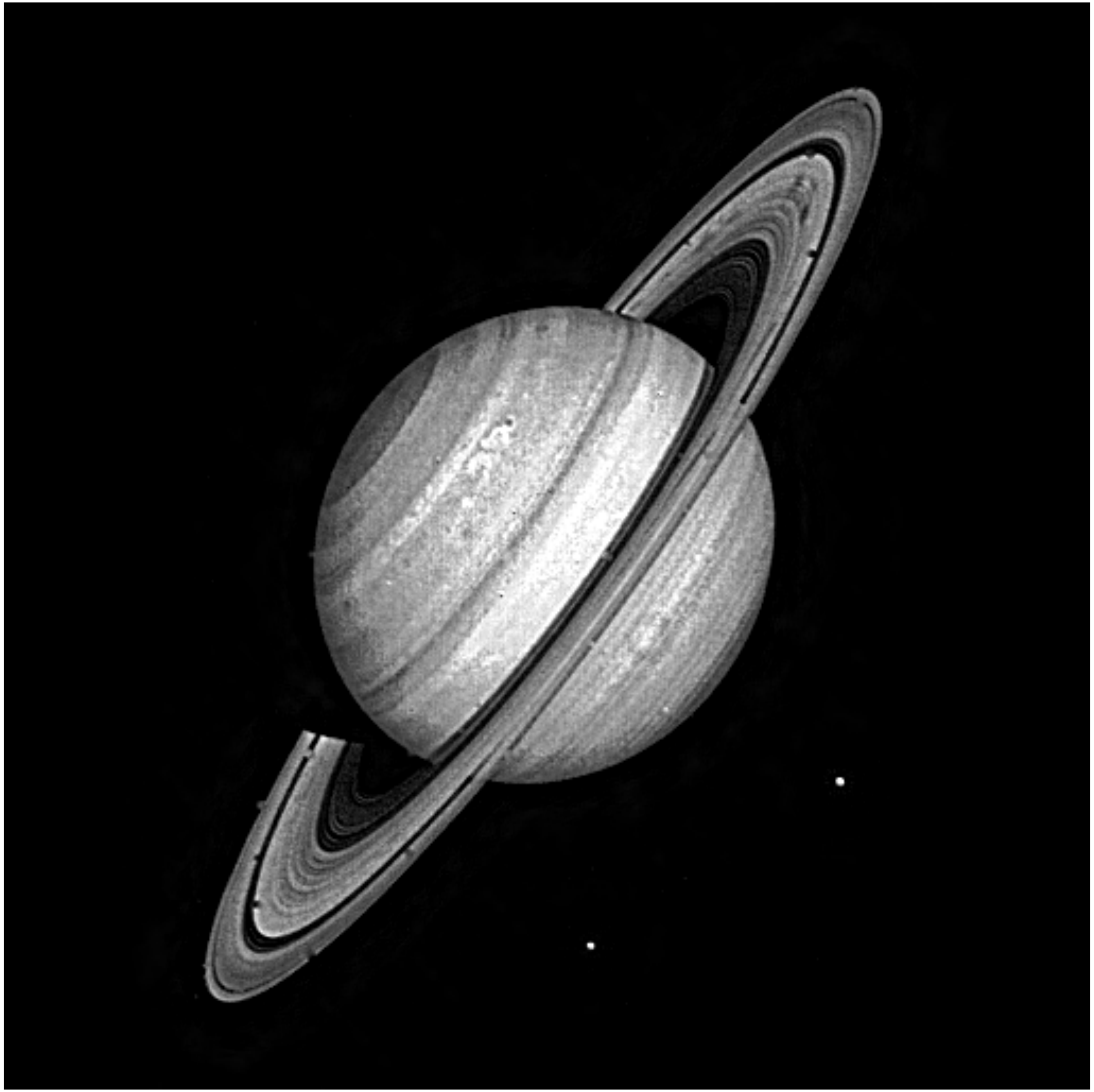
Modified curvelet coefficient



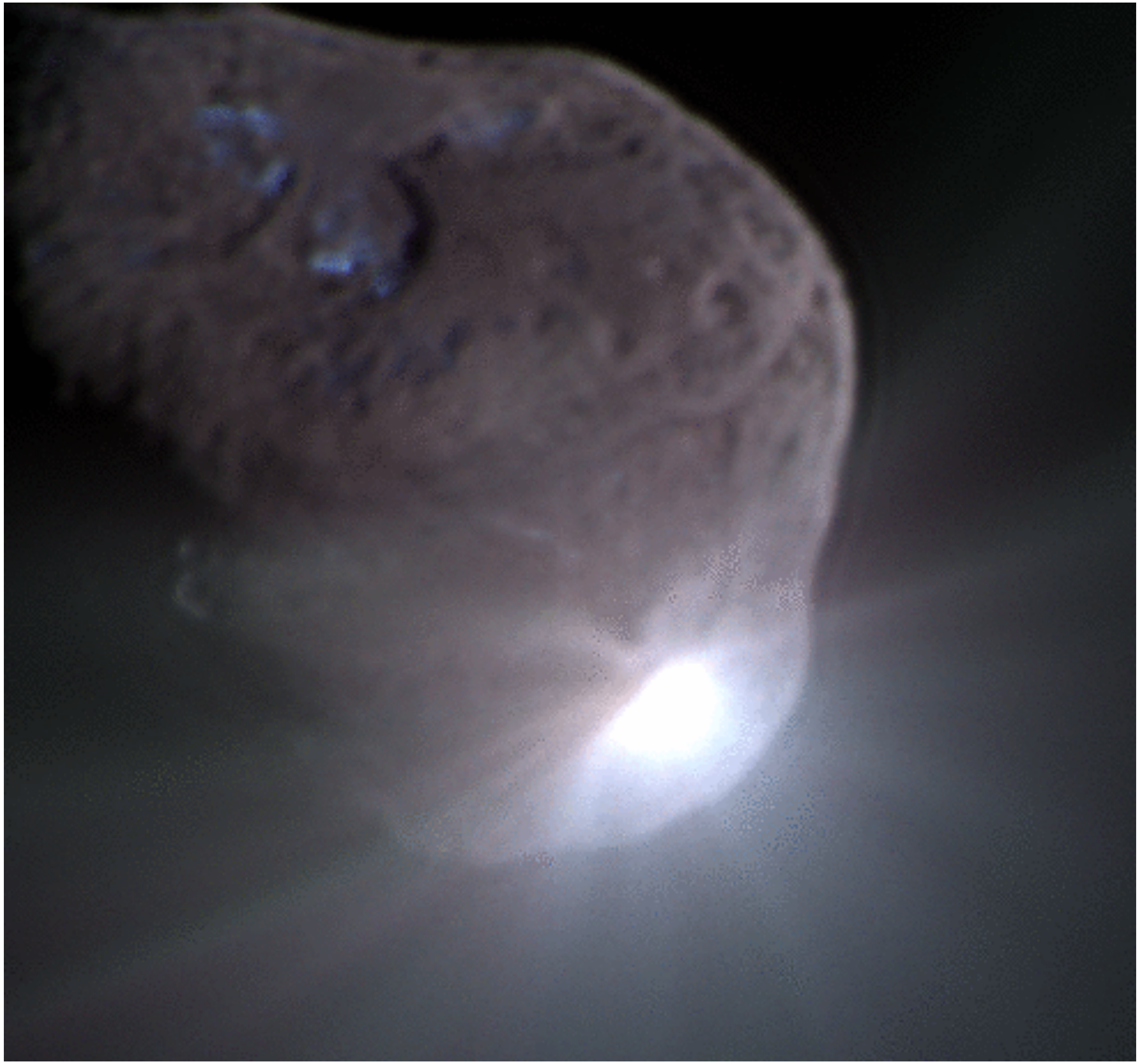
Curvelet coefficient

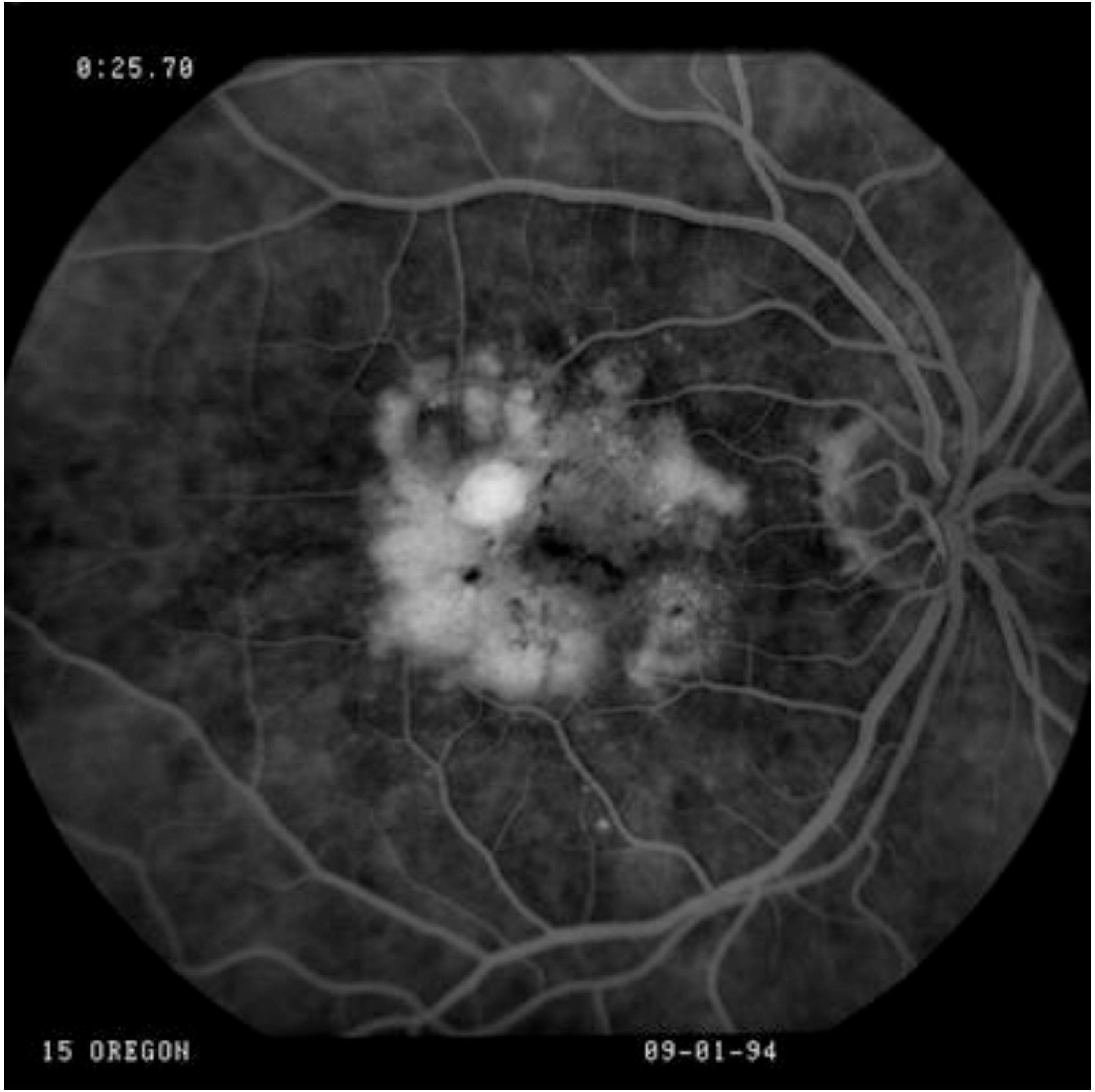
Contrast Enhancement

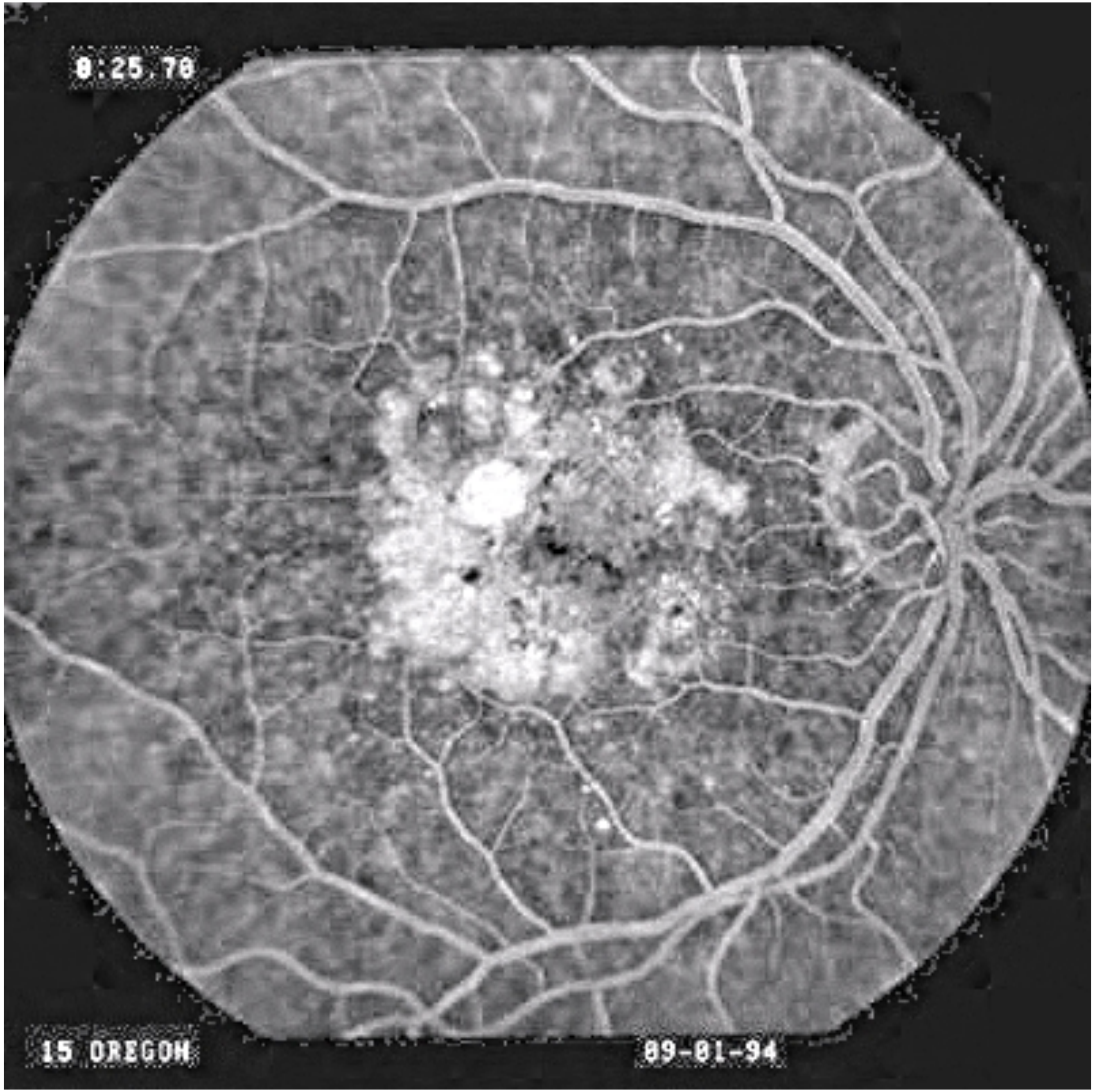


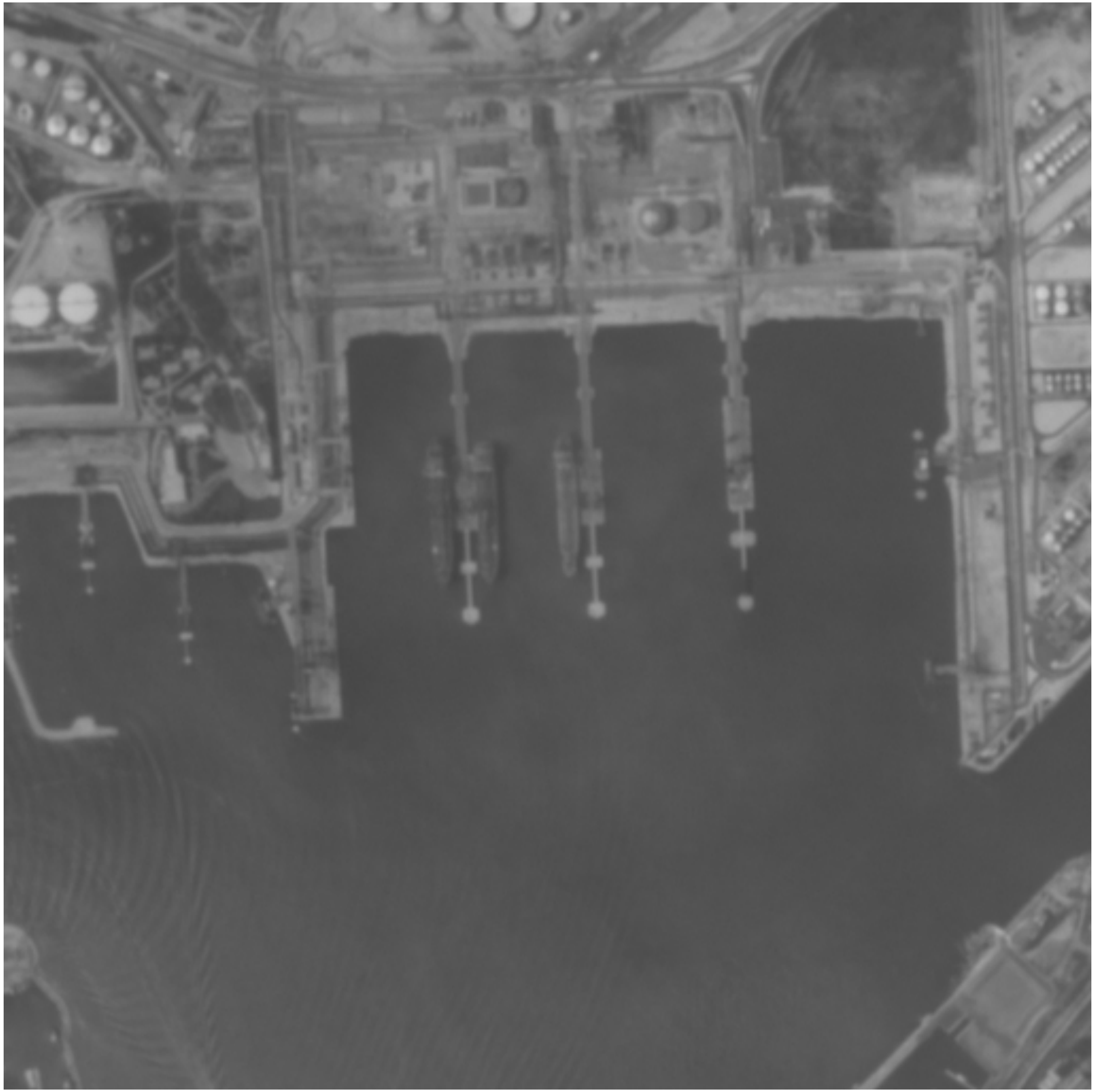


comet Tempel 1 on July 4, 2005




0:25.70
15 OREGON
09-01-94


0:25.70
15 OREGON
09-01-94



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 ==> Regularization

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

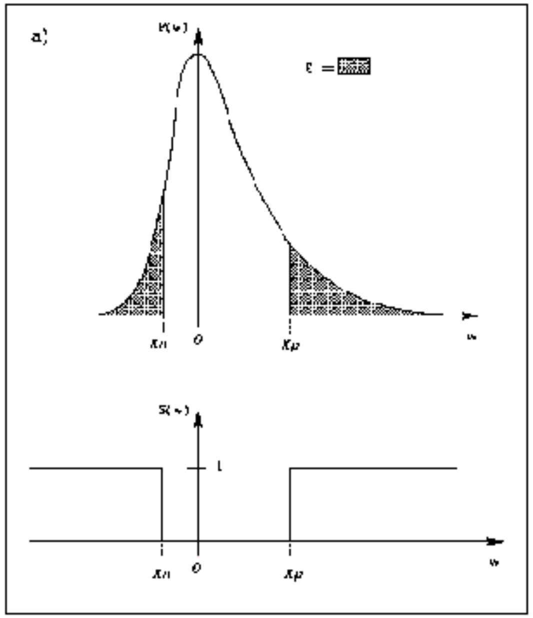
==> Sparsity constraint (i.e. $||X||_0$)

DENOISING

NOISE MODELING

For a positive coefficient: $P = \mathrm{Pr}\,ob(w > w_{j,x,y})$

For a negative coefficient: $P = \mathrm{Pr}\,ob(w < w_{j,x,y})$



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.

Hard Thresholding:
$$\delta(c) = c \quad \text{if } |c| \geq t$$
$$= 0 \quad \text{if } |c| < t$$

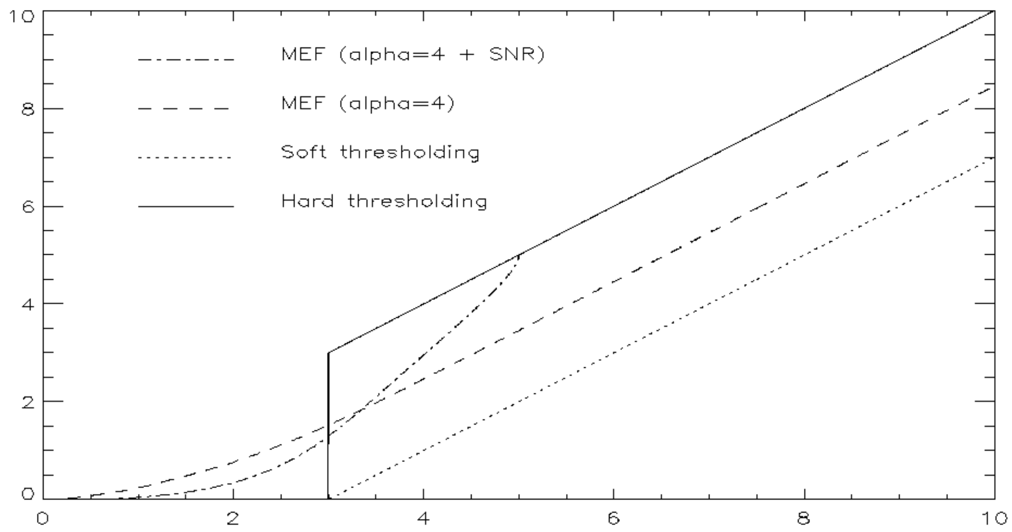
Soft Thresholding:
$$\delta(c) = \mathrm{sgn}(c)\left(|c| - t\right)_+$$

DENOISING ALGORITHM

- Take the wavelet transform of the data.
- For each wavelet scale j
    - Set to zero all coefficients with an absolute value lower than T_j (T_j is derived from the noise modeling).
- Apply the inverse wavelet transform to the thresholded coefficients.

$$\tilde{y} = W_R[\delta(W_T y)]$$

Filtered wavelet coefficients versus wavelet coefficients



Threshold estimation: Gaussian case

1. k-sigma: $T_j = k\sigma_j$

2. Universal Threshold: $T_j = \sqrt{2\log n}\sigma_j$

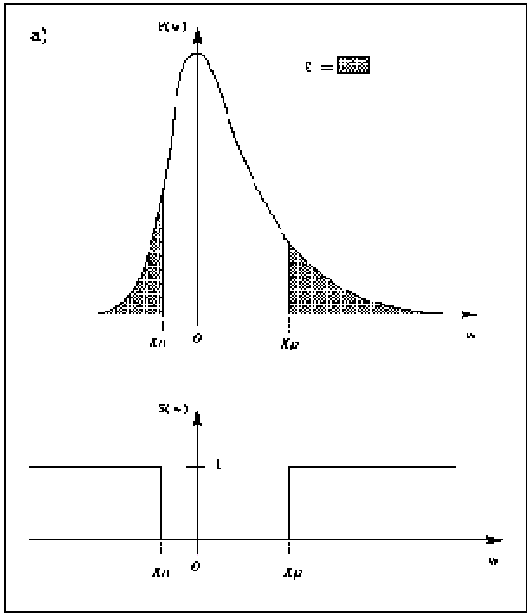
3. False Discovery Rate (FDR): compute the p-values for each wavelet coefficient $w_{j,l}$ at scale j and position l using the noise level σ_j . The user parameter α determines **the number of false detections as a percentage of the number of true detections. The FDR fixes the threshold.**

CURVELET FILTERING

NOISE MODELING

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

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

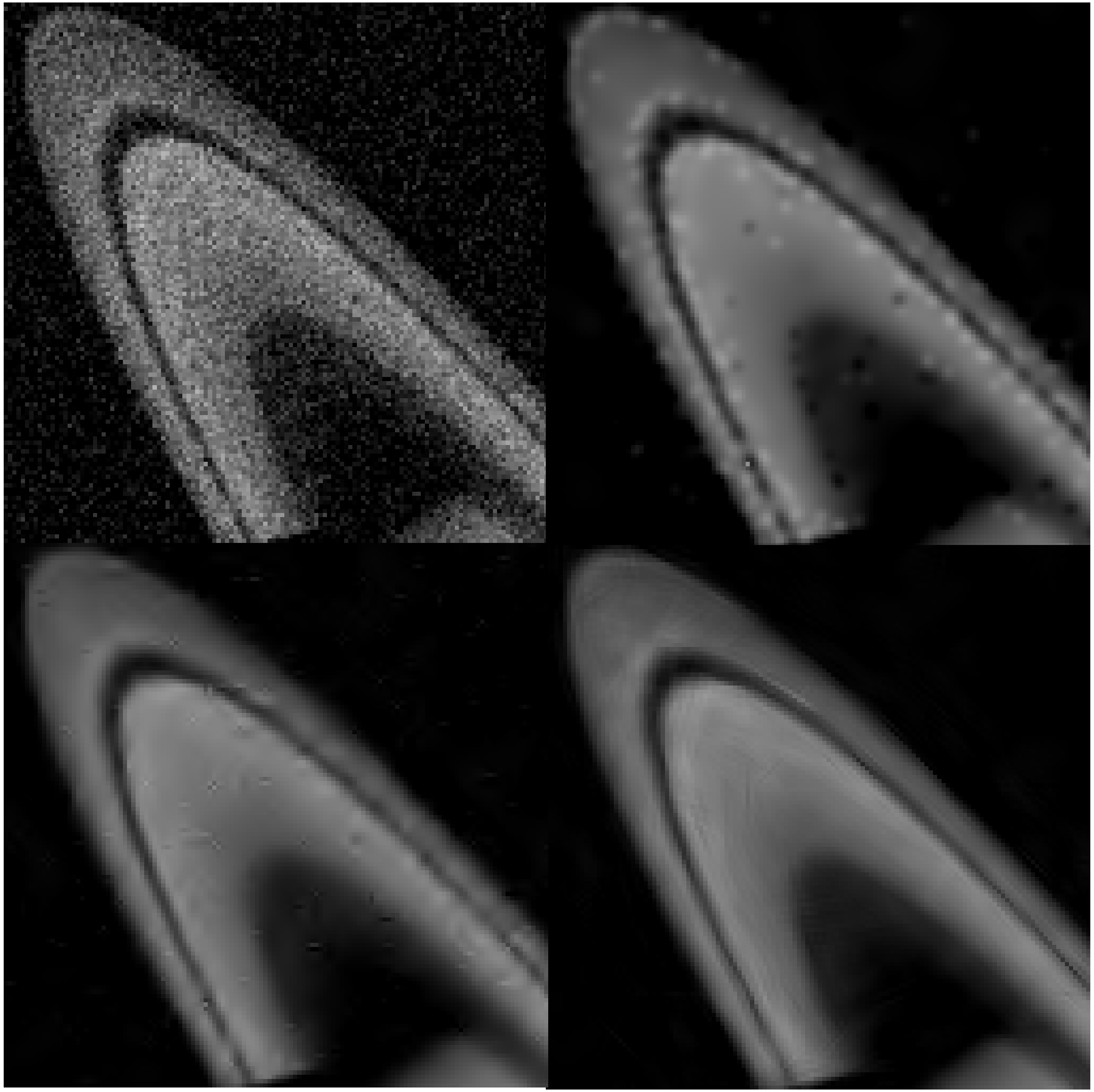


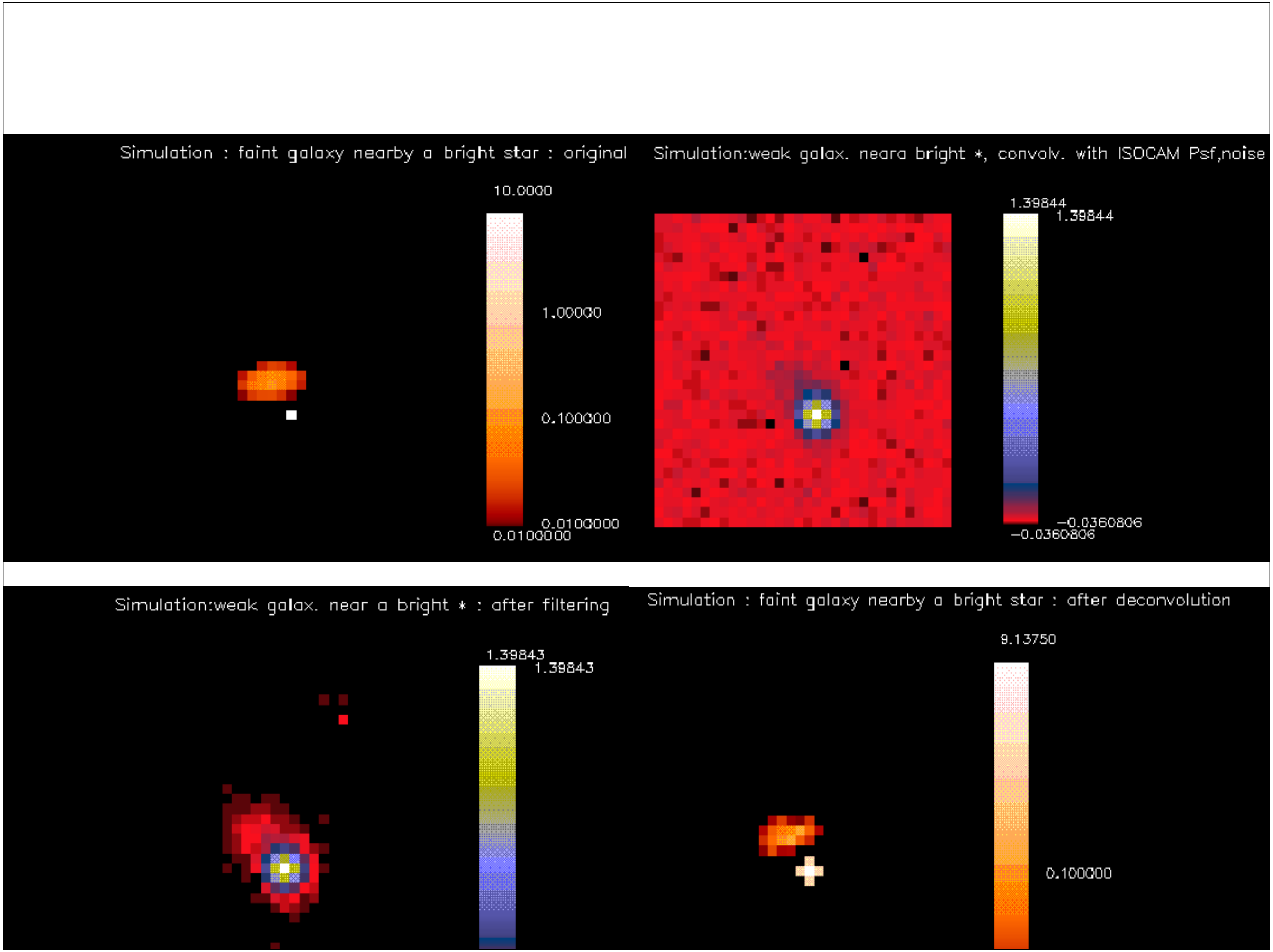
Given a threshold t:
if P > t, the coefficient could be due to the noise.
if P < t, the coefficient cannot be due to the noise,
and a **significant coefficient** is detected.

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

Hard Thresholding:

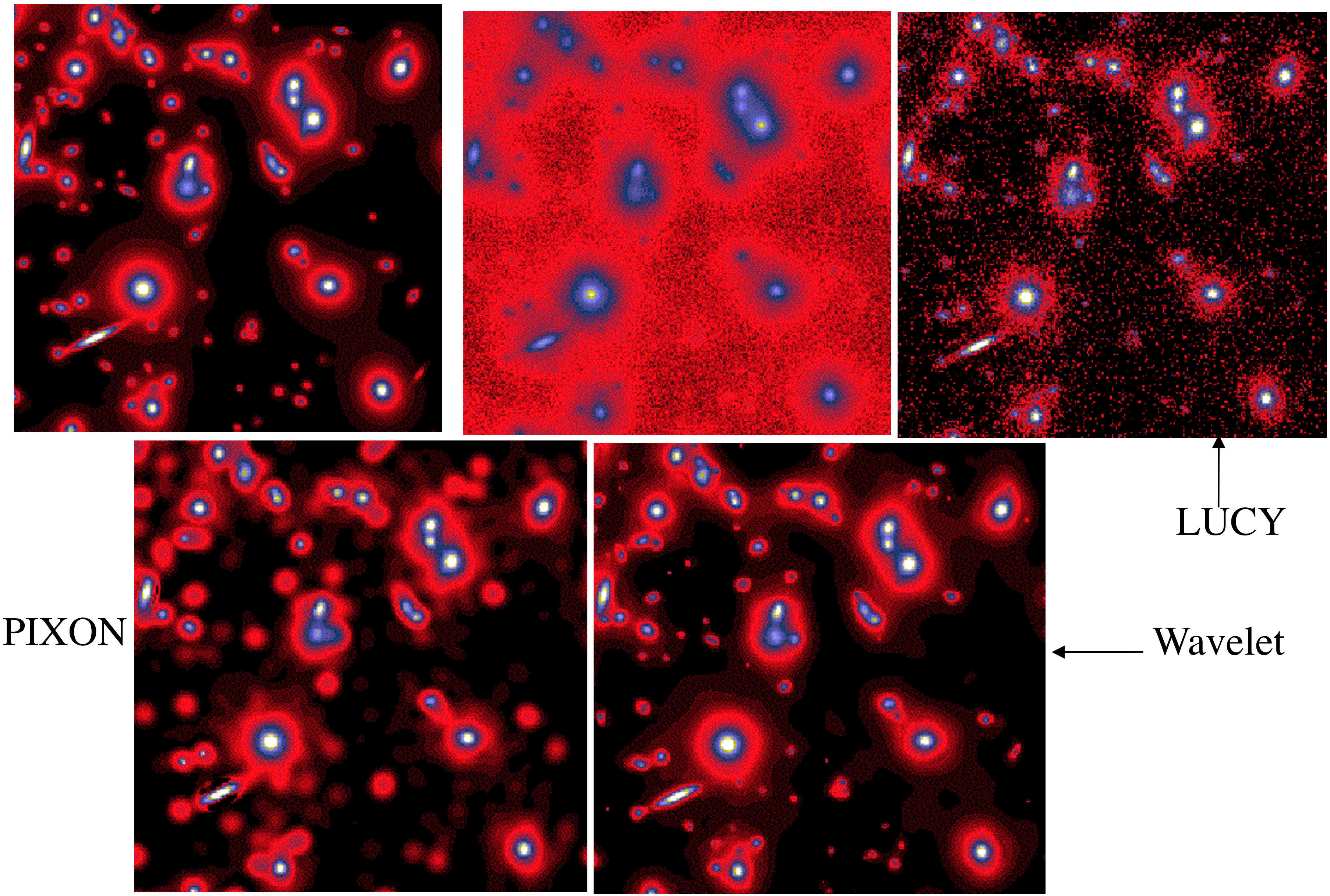
$$\begin{aligned} \delta(c) &= c \qquad if\ |c| \geq t \\ &= 0 \qquad if\ |c| < t \end{aligned}$$





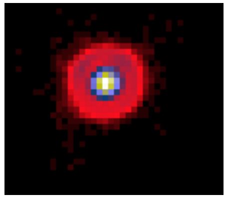
Simulation : faint galaxy nearby a bright star : original
10.0000
1.00000
0.100000
0.0100000
0.0100000
Simulation:weak galax. neara bright *, convolv. with ISOCAM Psf,noise
1.39844
1.39844
-0.0360806
-0.0360806
Simulation:weak galax. near a bright * : after filtering
1.39843
1.39843
Simulation : faint galaxy nearby a bright star : after deconvolution
9.13750
0.100000

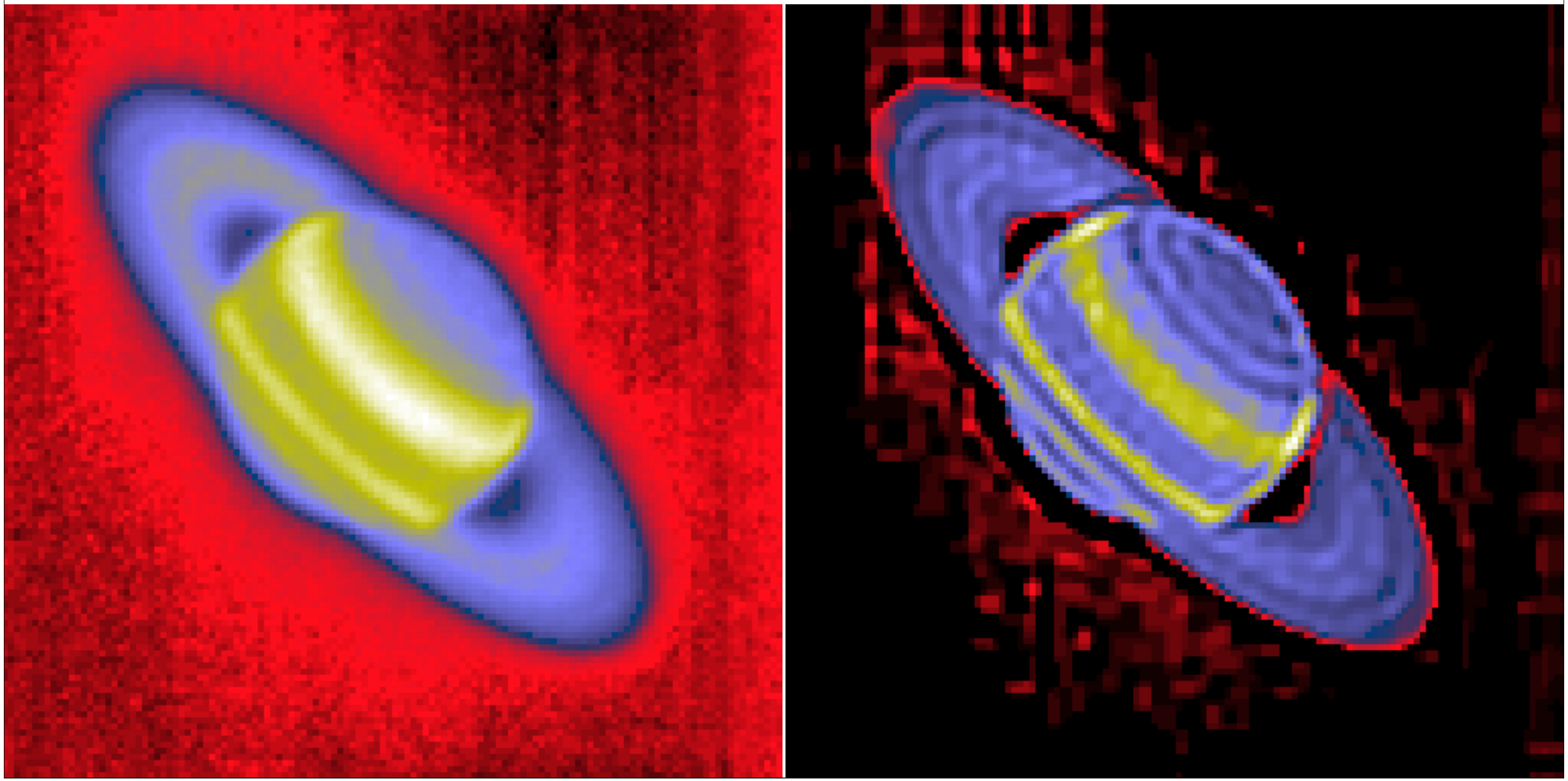
DECONVOLUTION SIMULATION



DECONVOLUTION

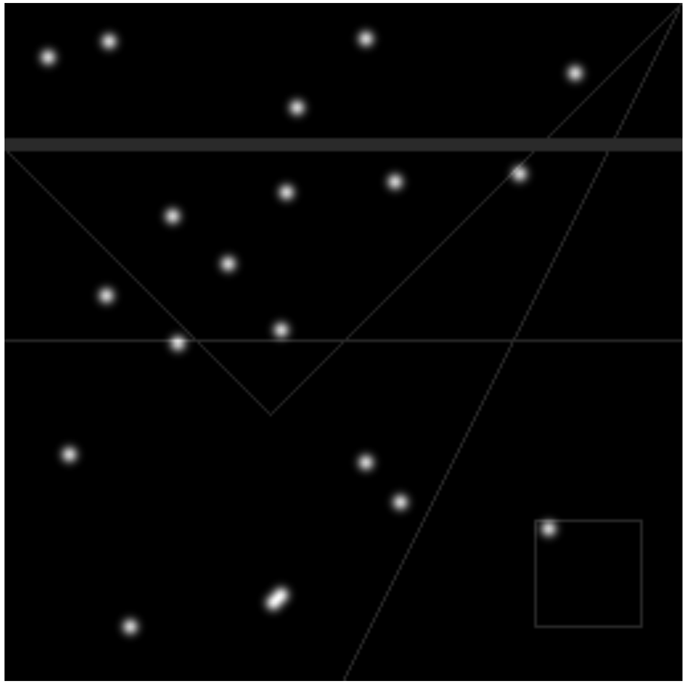
- E. Pantin, J.-L. Starck, and F. Murtagh, "Deconvolution and Blind Deconvolution in Astronomy", in Blind image deconvolution: theory and applications, pp 277--317, 2007.
- J.-L. Starck, F. Murtagh, and M. Bertero, "The Starlet Transform in Astronomical Data Processing: Application to Source Detection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



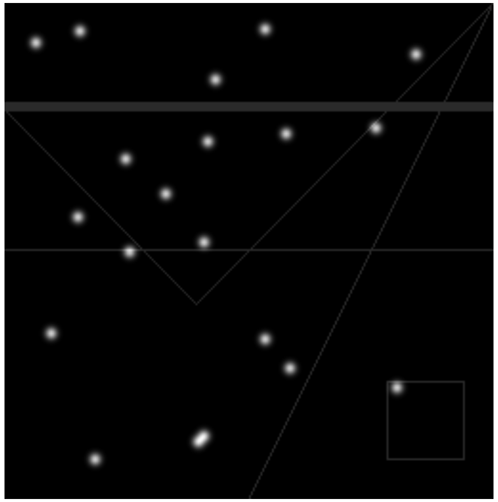


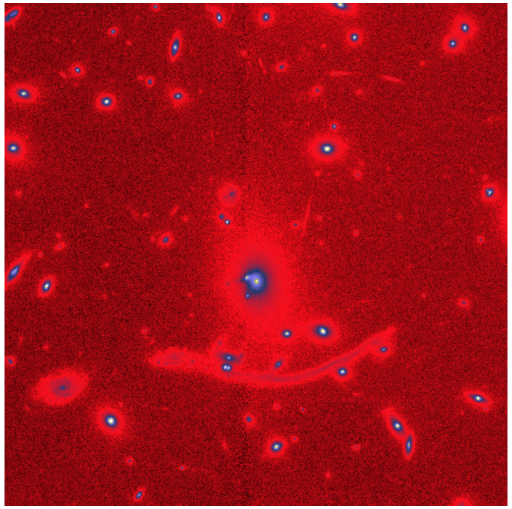
A difficult issue

Is there any representation that well represents the following image ?

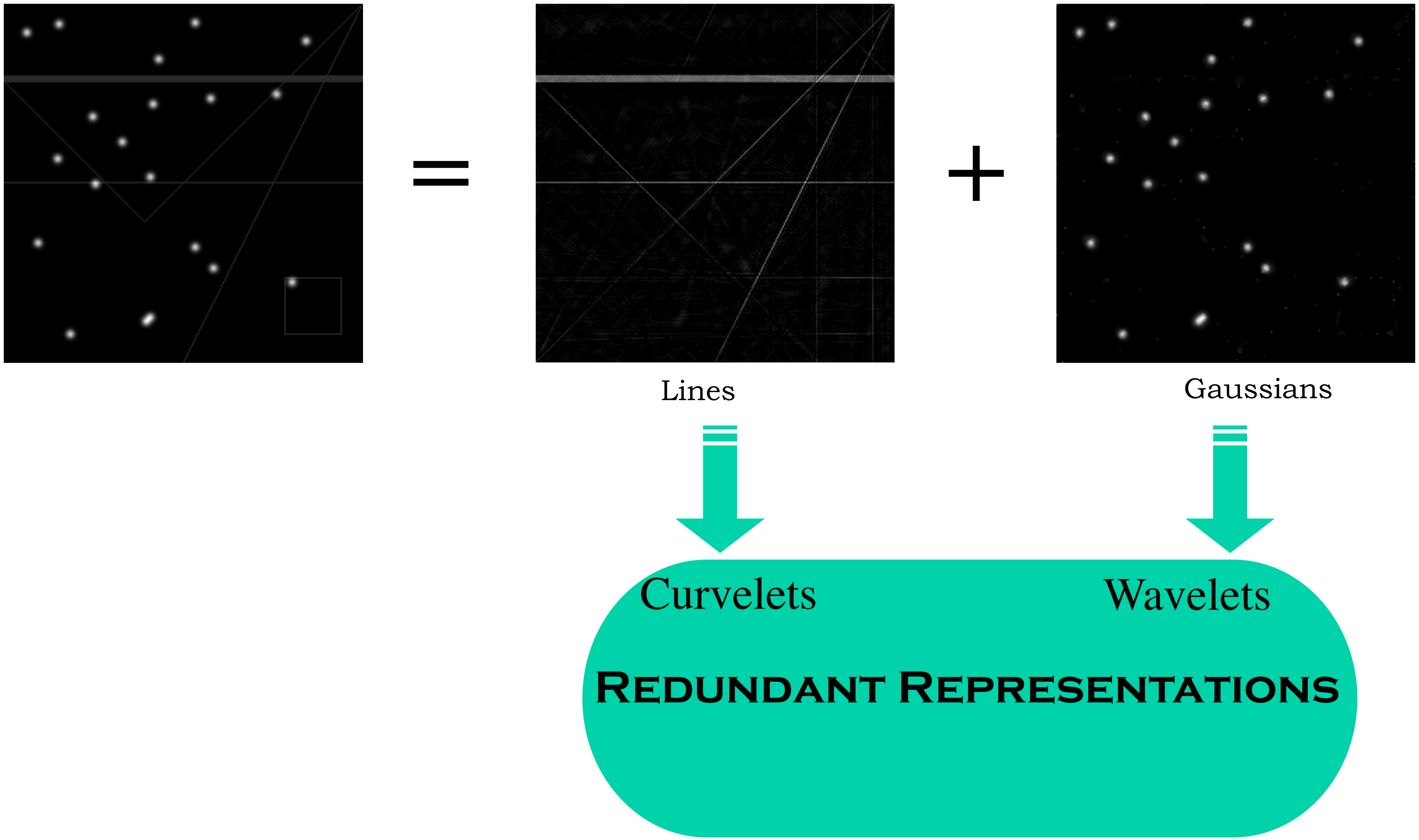


PB: a given transform does not necessary provide a good dictionary for all features contained in the data.



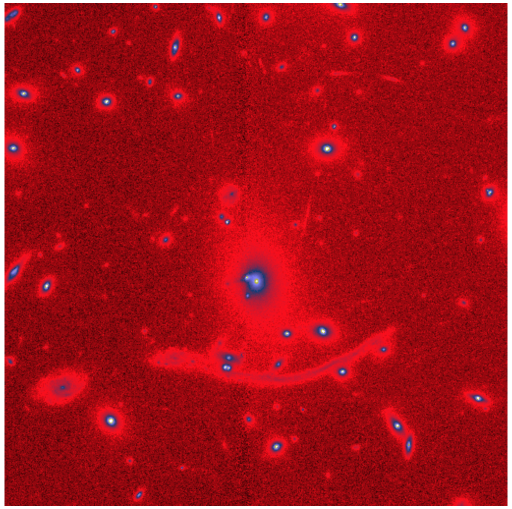


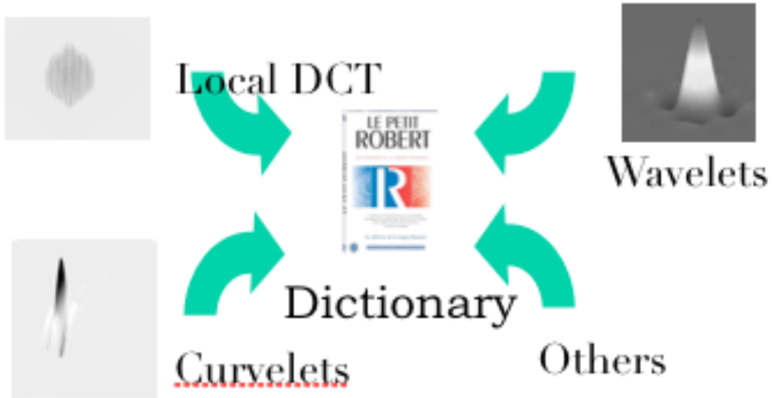
Going further



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 :

$$s_k = \Phi_k \alpha_k$$

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

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.

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

Morphological Component Analysis (MCA)

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

- Initialize all s_k to zero
- Iterate j=1,...,Niter
  - Iterate k=1,..,L

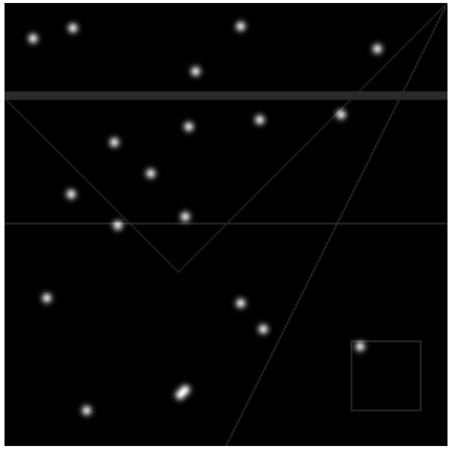
    Update the kth part of the current solution by fixing all other parts and minimizing:

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

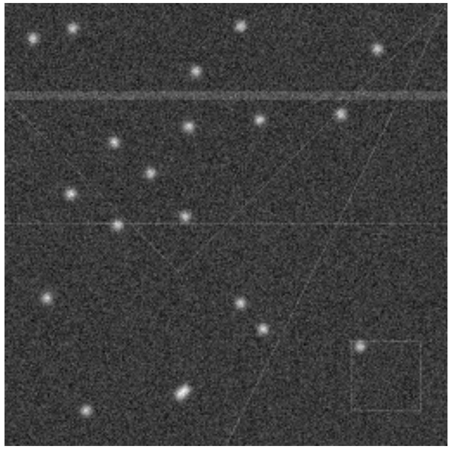
Which is obtained by a simple **hard**/soft thresholding of : $s_r = s - \sum_{i=1, i\neq k}^{L} s_i$

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

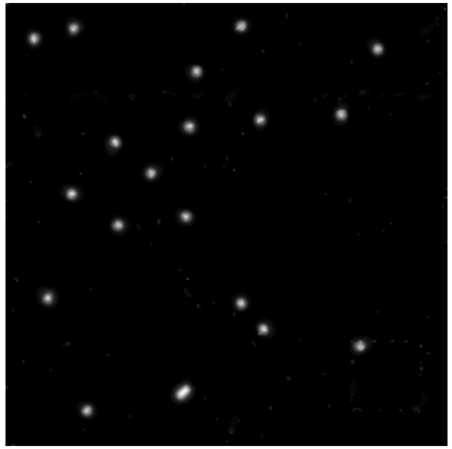
$$MIN_{s_1,s_2}\left(\|Ws_1\|_p + \|Cs_2\|_p\right) \text{ subject to } \|s-(s_1+s_2)\|_2^2 < \varepsilon$$



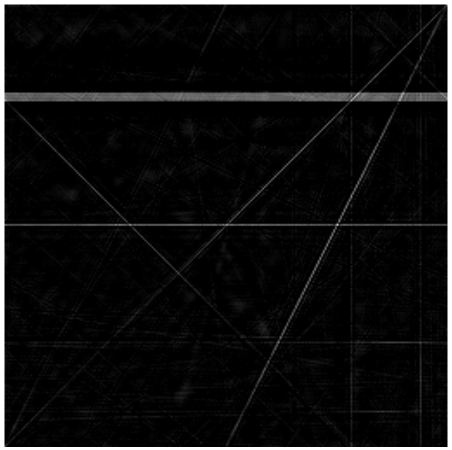
a) Simulated image (gaussians+lines)



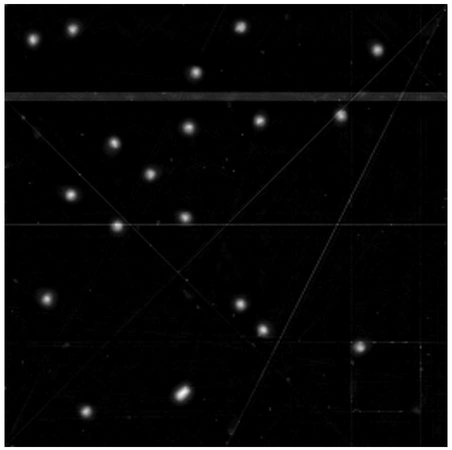
b) Simulated image + noise



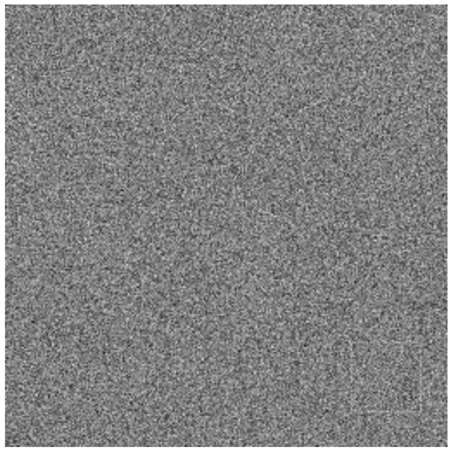
c) A trous algorithm



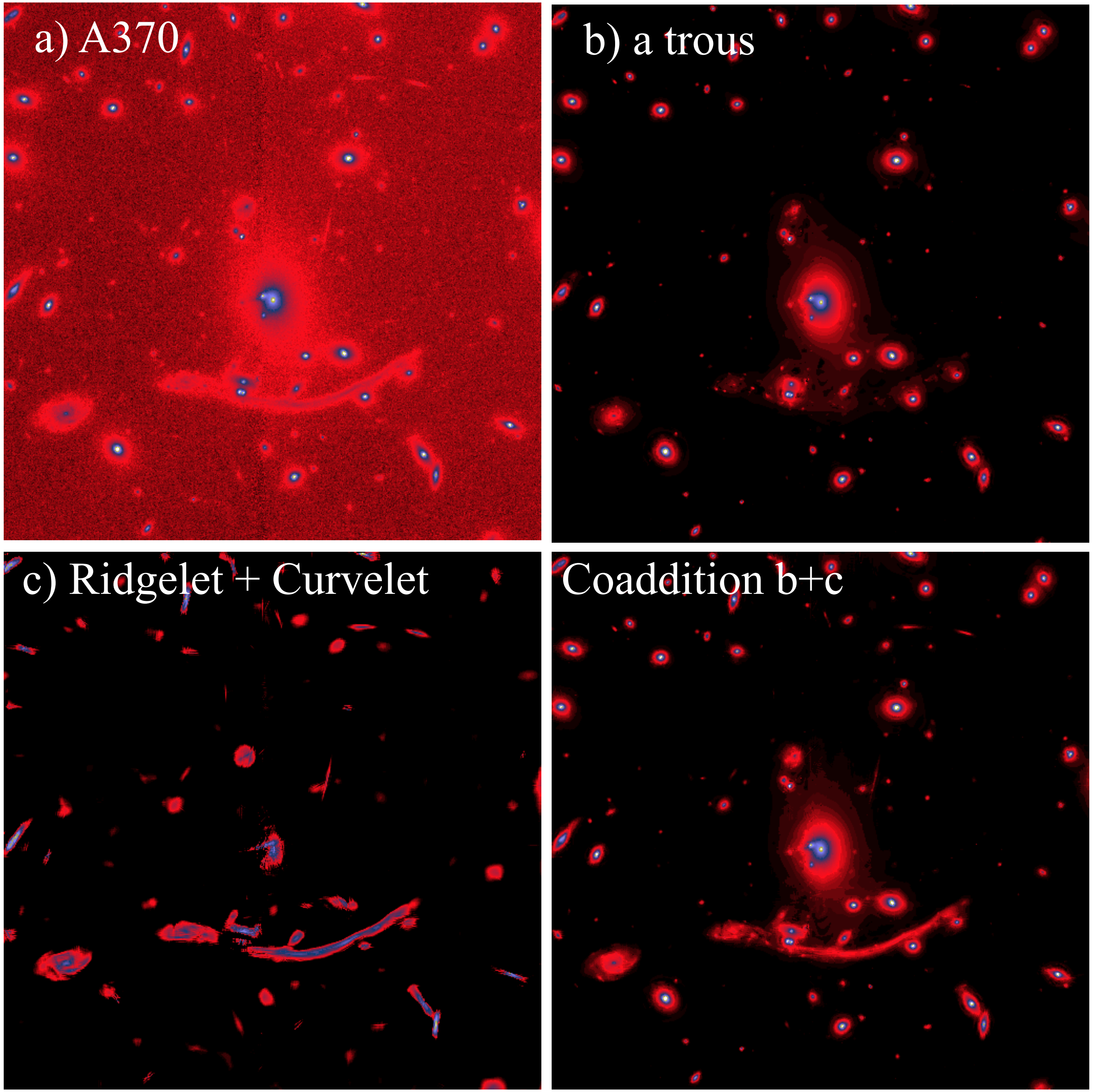
d) Curvelet transform

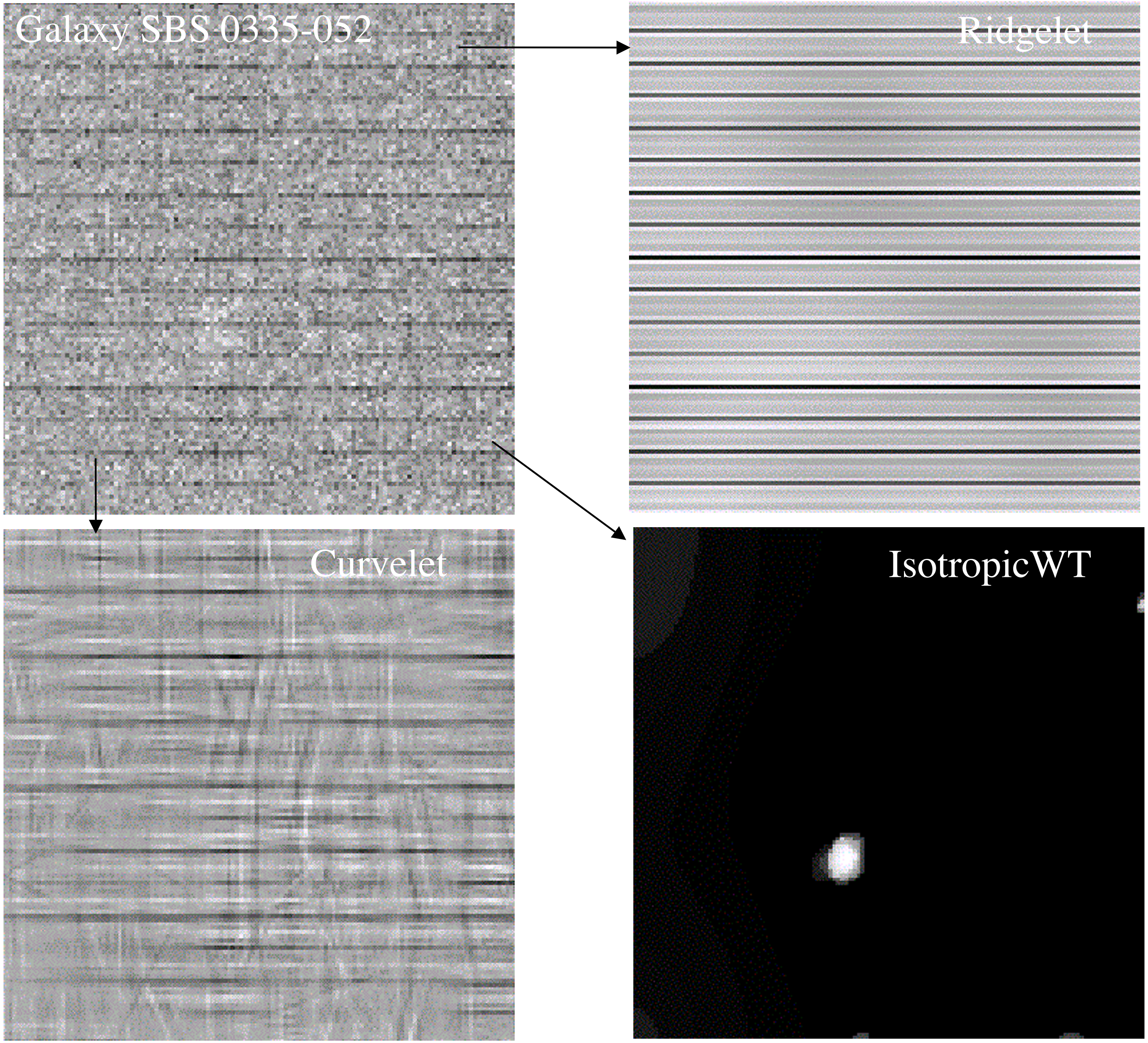


e) coaddition c+d

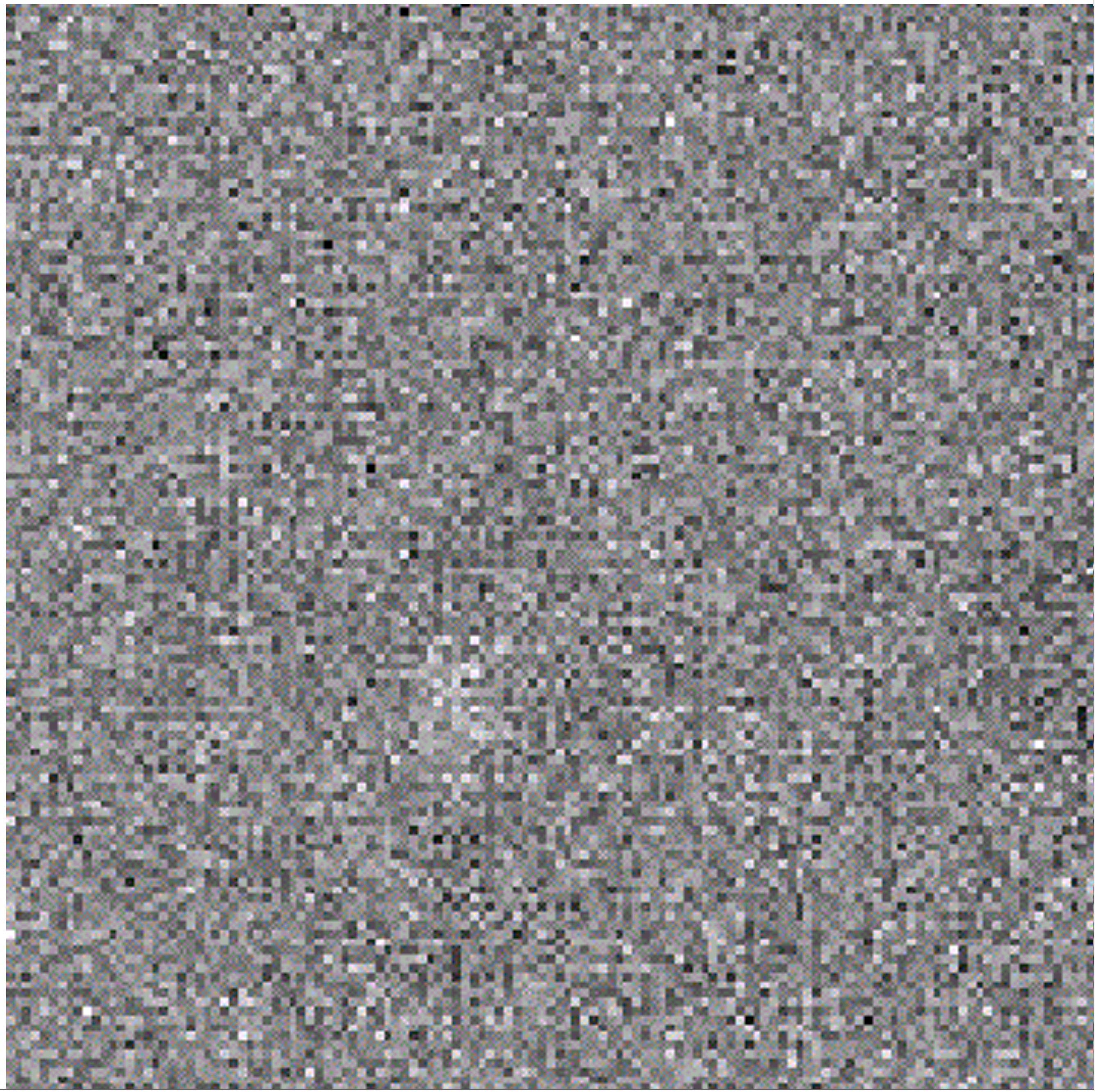


f) residual = e-b

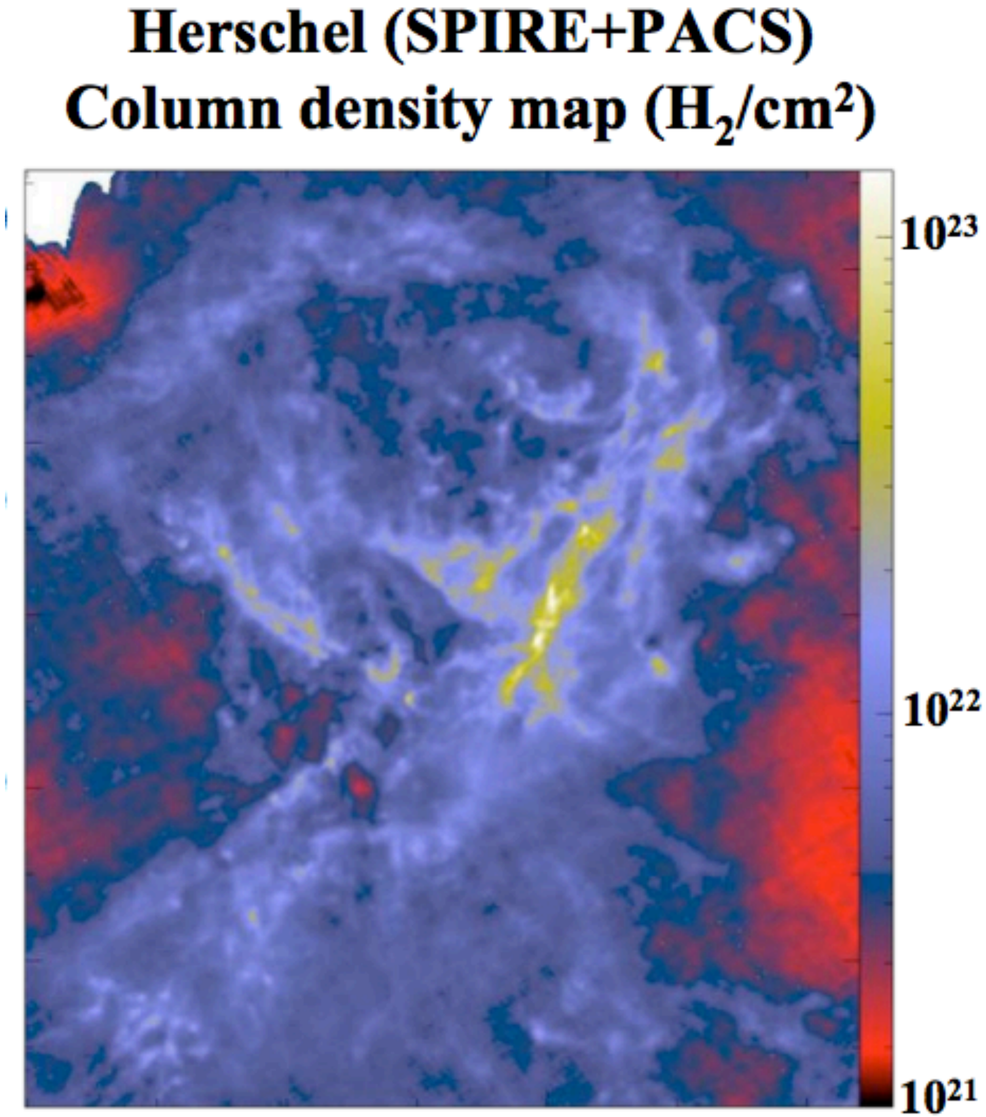

a) A370
b) a trous
c) Ridgelet + Curvelet
Coaddition b+c


Galaxy SBS 0335-052
Ridgelet
Curvelet
IsotropicWT

Galaxy SBS 0335-052
10 micron
GEMINI-OSCIR



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



Dense cores form primarily in filaments

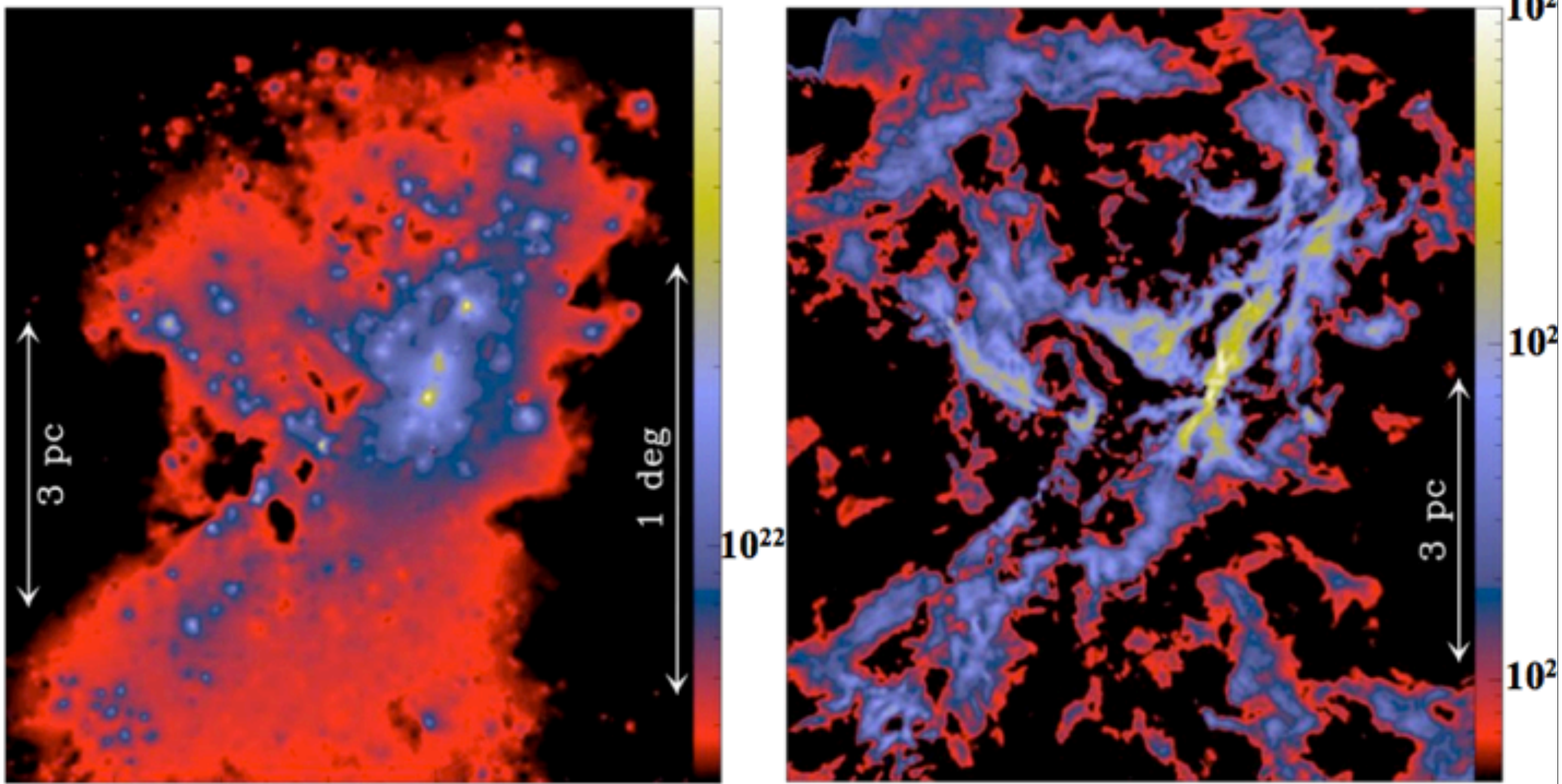
Morphological Component Analysis:

***Herschel* Column density map**

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

=

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

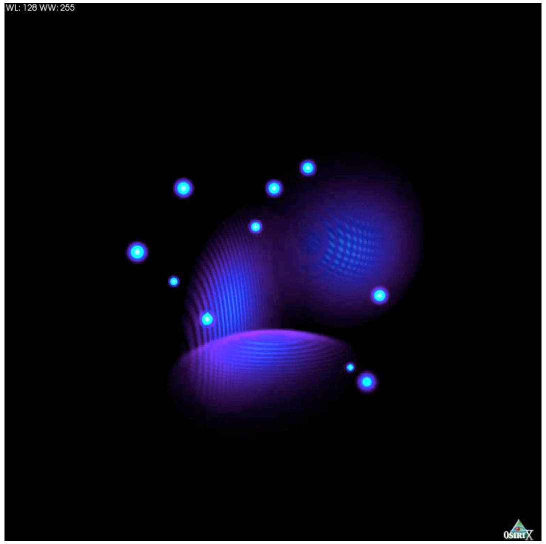


A. Menshchikov, Ph.André. P. Didelon, et al, "Filamentary structures and compact objects in the Aquila and Polaris clouds observed by Herschel", A&A, 518, id.L103, 2010.

3D Morphological Component Analysis

A. Woiselle

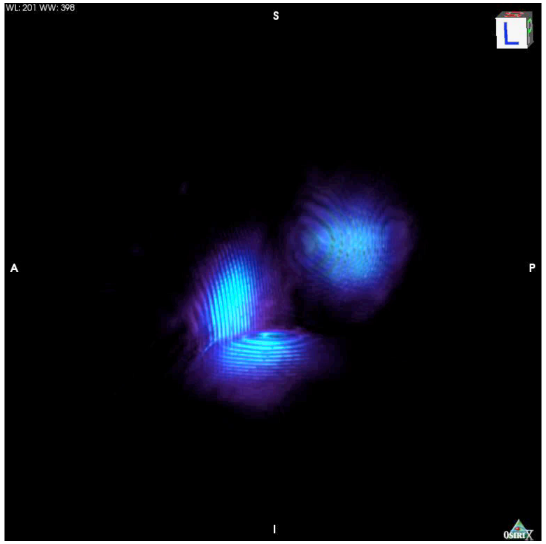
Original (3D shells + Gaussians)



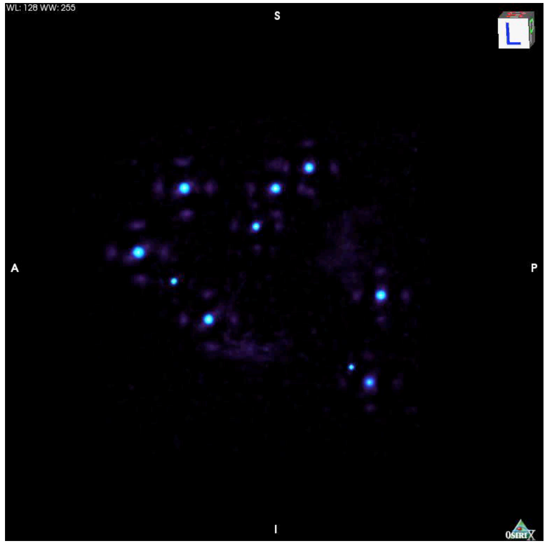
Dictionary

RidCurvelets + 3D UDWT.

Shells



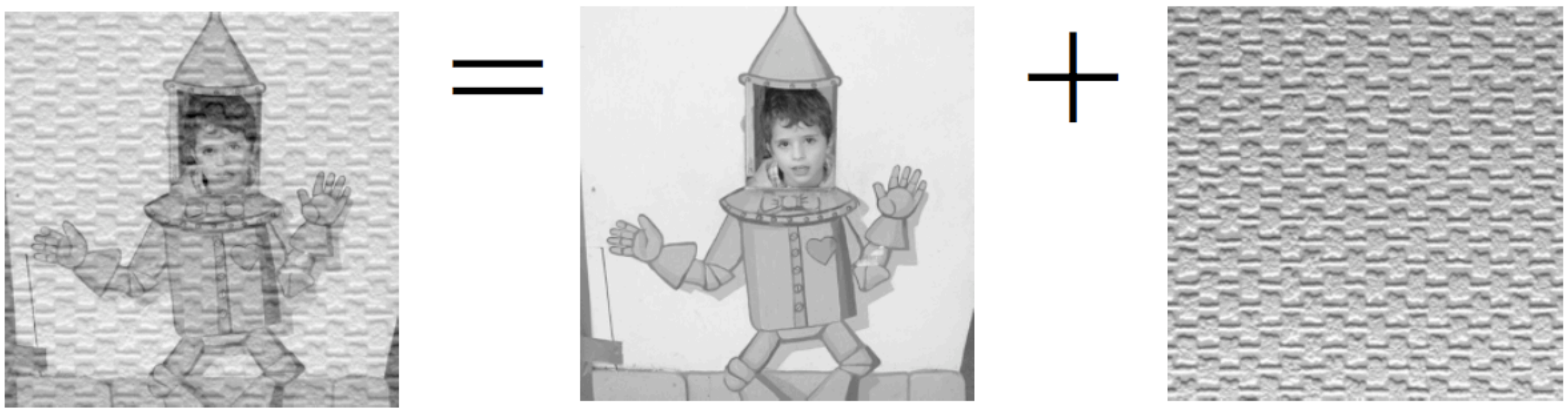
Gaussians



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

Separation of Texture from Piecewise Smooth Content

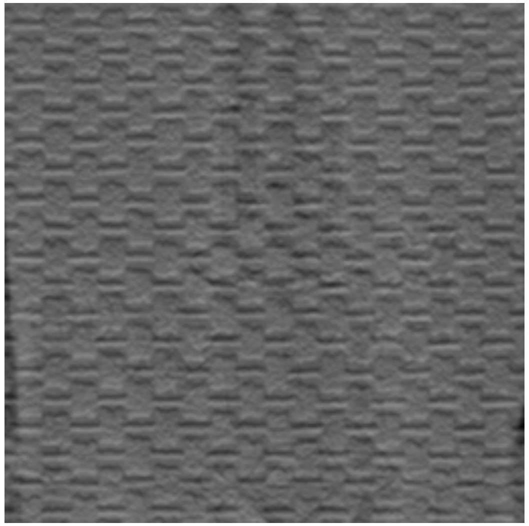
<u>The separation task</u>: decomposition of an image into a texture and a natural (piecewise smooth) scene part.



•***Image Decomposition Via the Combination of Sparse Representation and a Variational Approach***, *IEEE Trans. on Image Proces., 14, 10, pp 1570--1582, 2005*

Texture Separation using MCA: Curvelet + DCT

X_n



X_t

Edge Detection

Inp  inting

- *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", The Computer Journal, 52, 1, pp 64-79, 2009.*

$$\min_{\alpha} \|\alpha\|_{\ell_0} \text{ s.t. } y = Mx$$

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

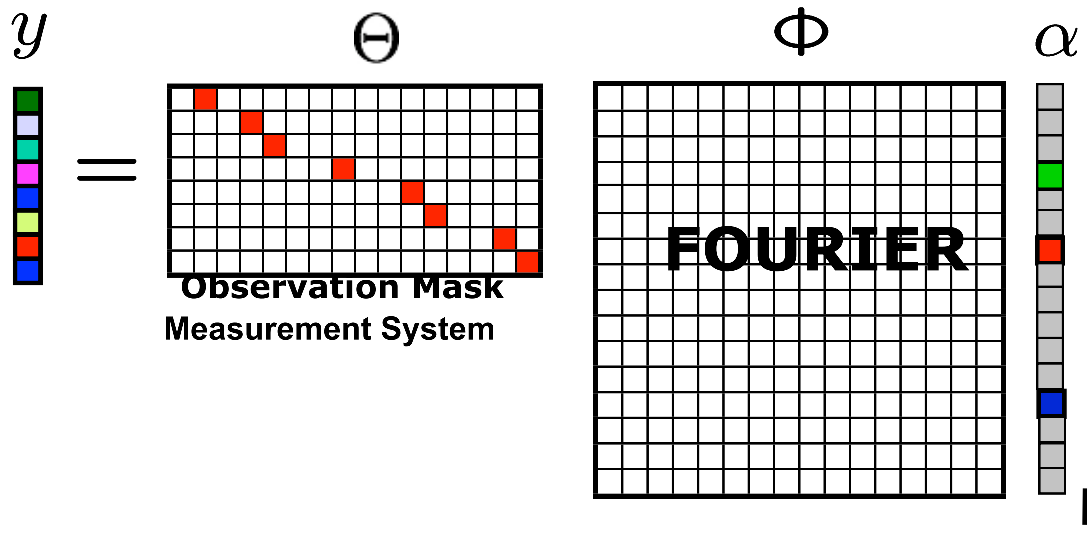
$$x^{(n+1)} = S_{\Phi,\lambda^{(n)}} \left\{ x^{(n)} + M\left(y - x^{(n)}\right) \right\}$$

Iterative Hard Thresholding with a decreasing threshold.

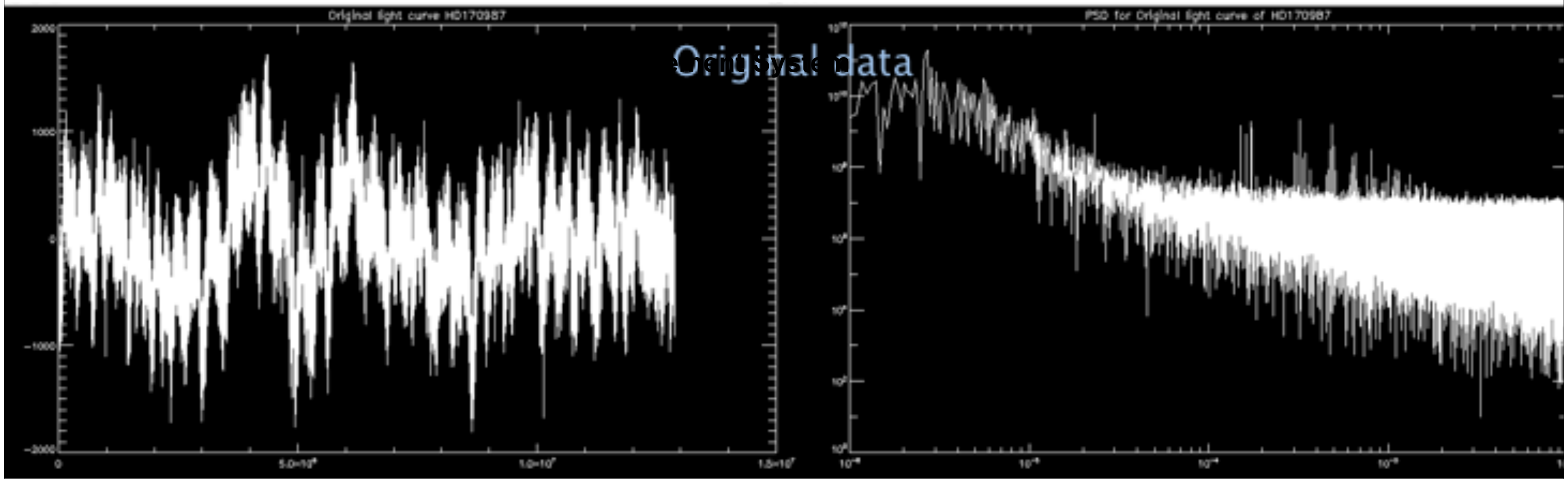
MCAlab available at: http://www.greyc.ensicaen.fr/~jfadili

Missing Data

Period detection in temporal series



COROT: HD170987



. Initialize all s_k to zero

. Iterate j=1,...,Niter

- Iterate k=1,..,L

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

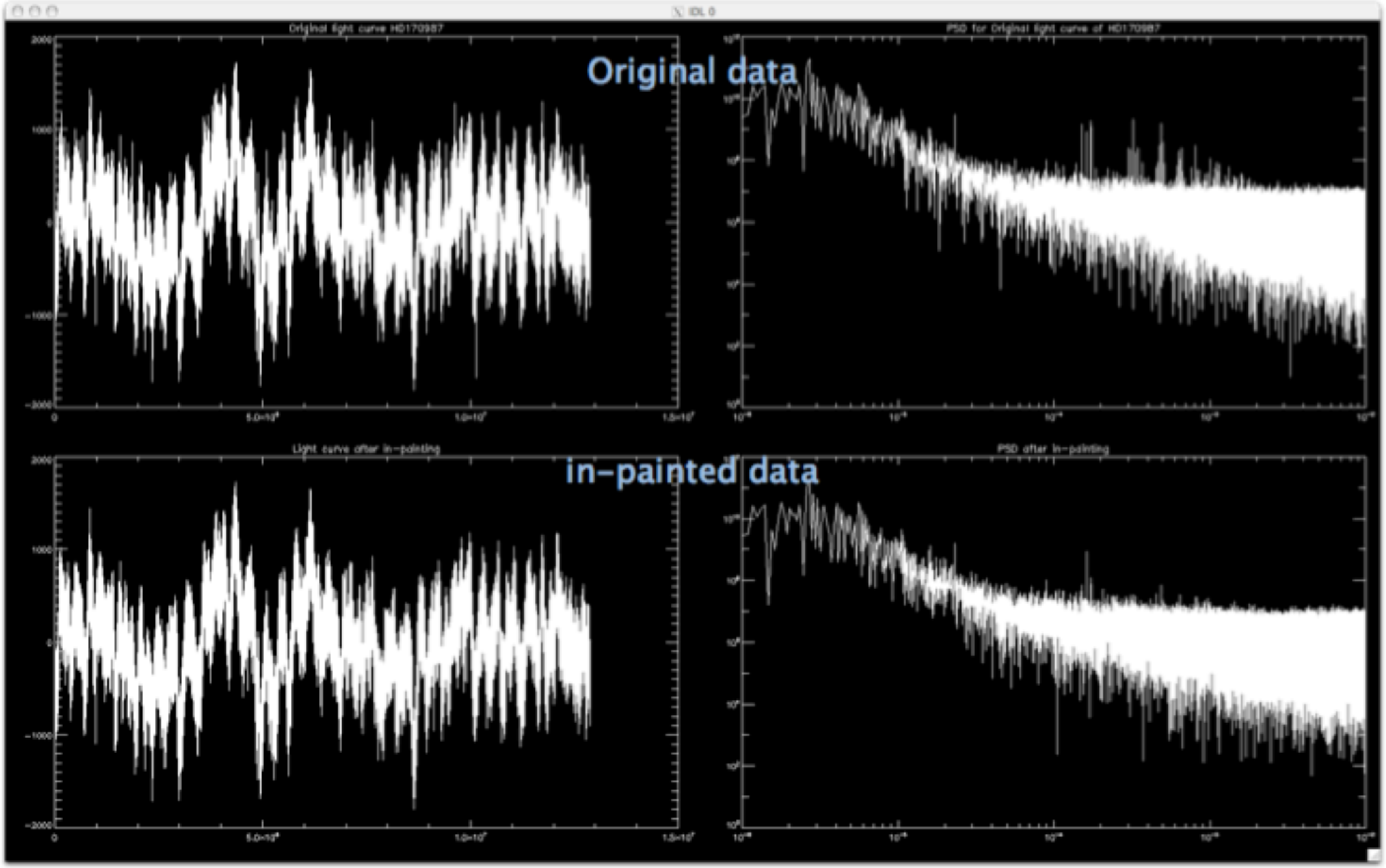
$$J(s_k) = \left\| M(s - \sum_{i=1, i \neq k}^{L} s_i - s_k) \right\|_2^2 + \lambda \left\| T_k s_k \right\|_1$$

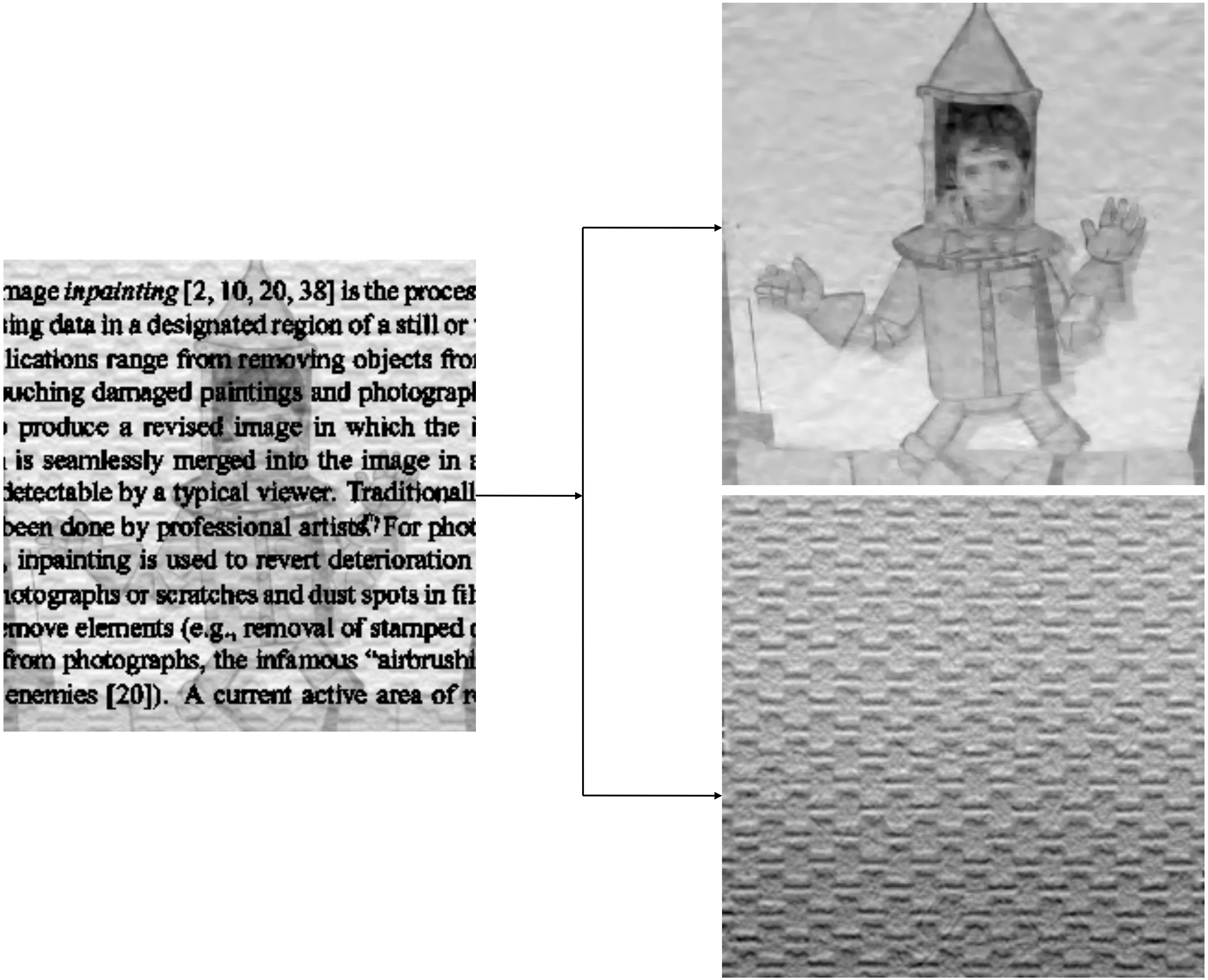
Which is obtained by a simple soft thresholding of :

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

COROT: HD170987 with in-painting

arXiv:1003.5178

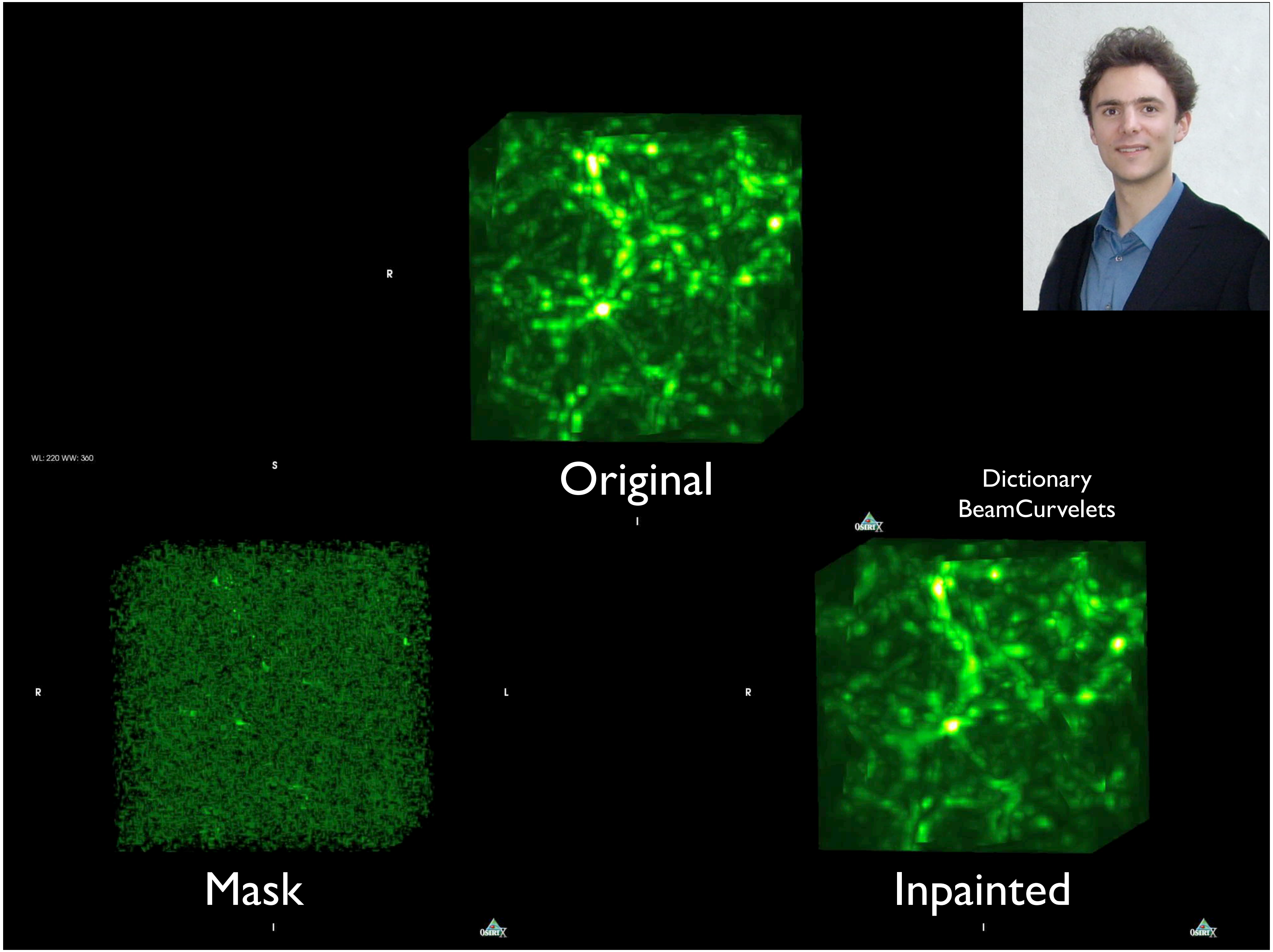


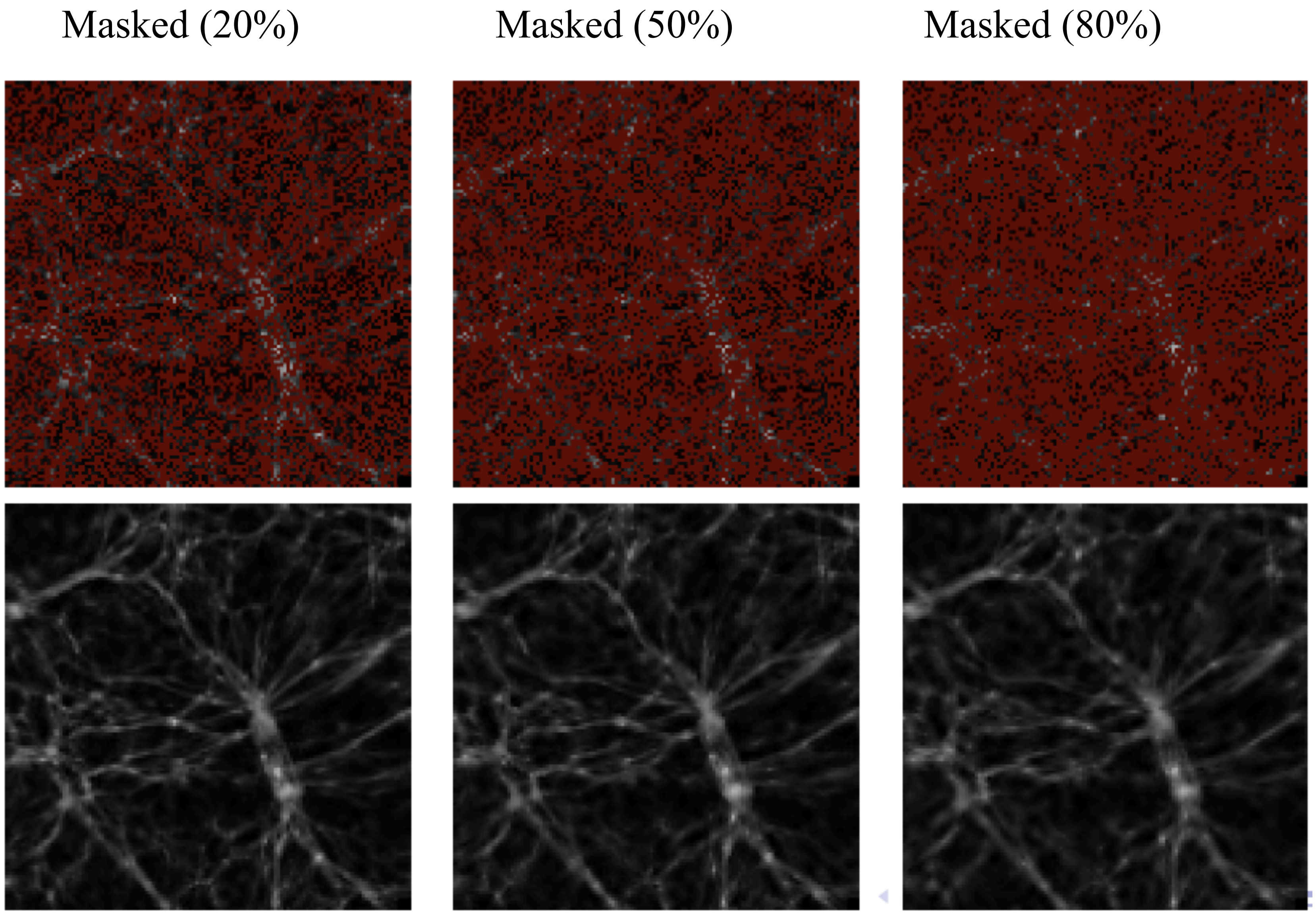

mage inpainting [2, 10, 20, 38] is the proces
sing data in a designated region of a still or
lications range from removing objects fro
uching damaged paintings and photograp
produce a revised image in which the i
is seamlessly merged into the image in
detectable by a typical viewer. Traditionall
been done by professional artists. For phot
, inpainting is used to revert deterioration
otographs or scratches and dust spots in fil
emove elements (e.g., removal of stamped
from photographs, the infamous "airbrushi
enemies [20]). A current active area of r

20%

50%

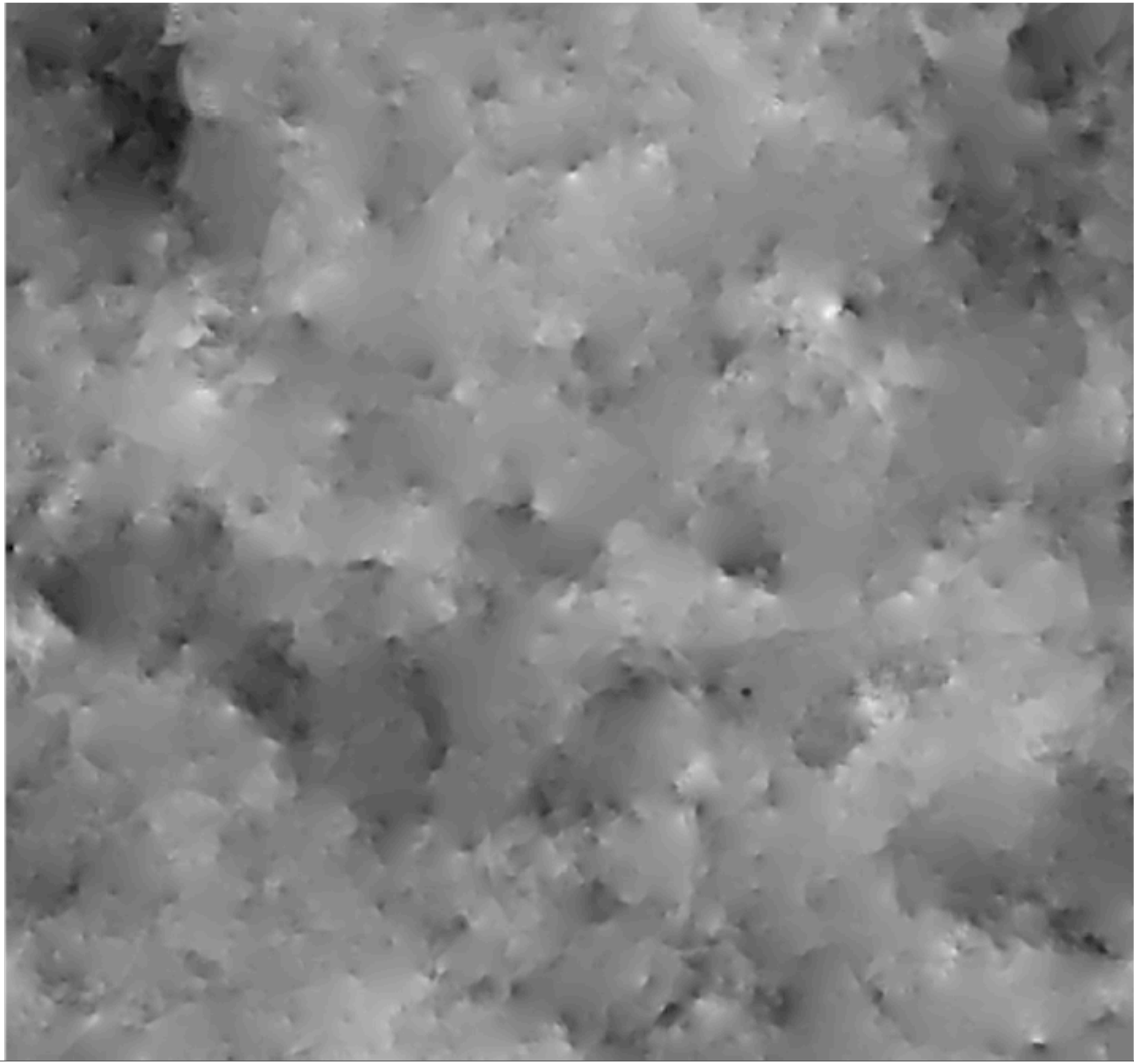
80%


Original
Dictionary
BeamCurvelets
Mask
Inpainted
WL: 220 WW: 360
R
S
I
R
L
R
I
I

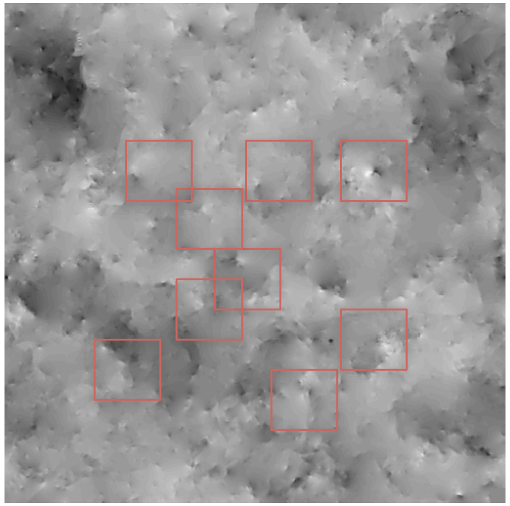


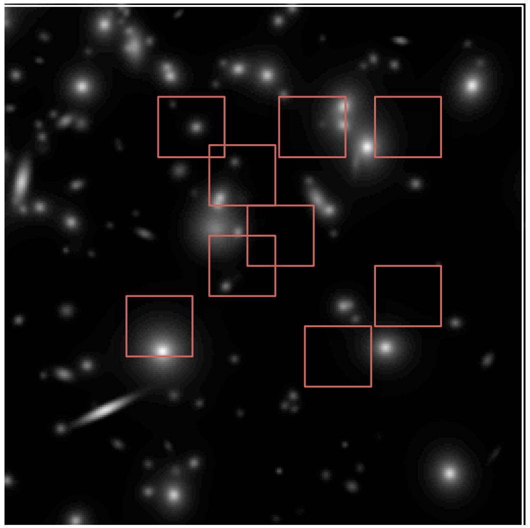
Central slice of the masked CDM data with 20, 50, and 80% missing voxels, and the inpainted maps. The missing voxels are dark red.

Simulated Cosmic String Map



Dictionary Learning





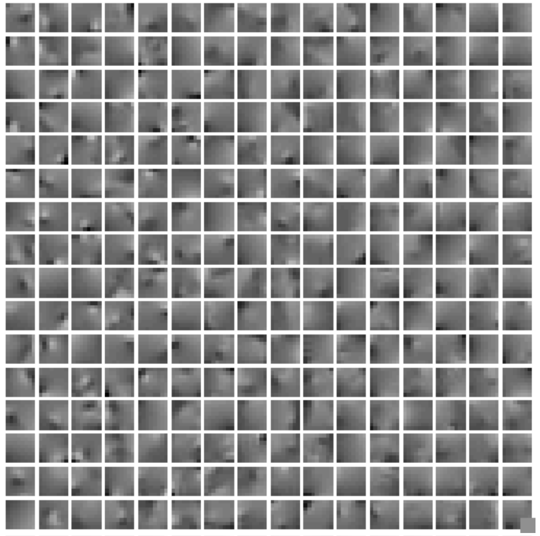
Training basis.

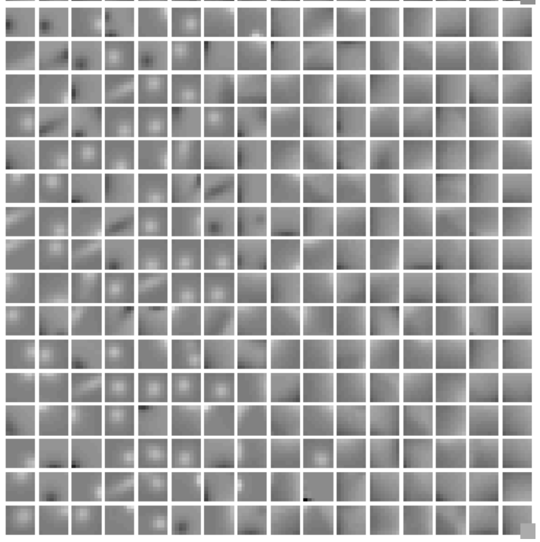
$$\left(\hat{D},\hat{A}\right) = \underset{\substack{D\in C_1 \\ A\in C_2}}{\arg\min}\left(Y = DA\right)$$

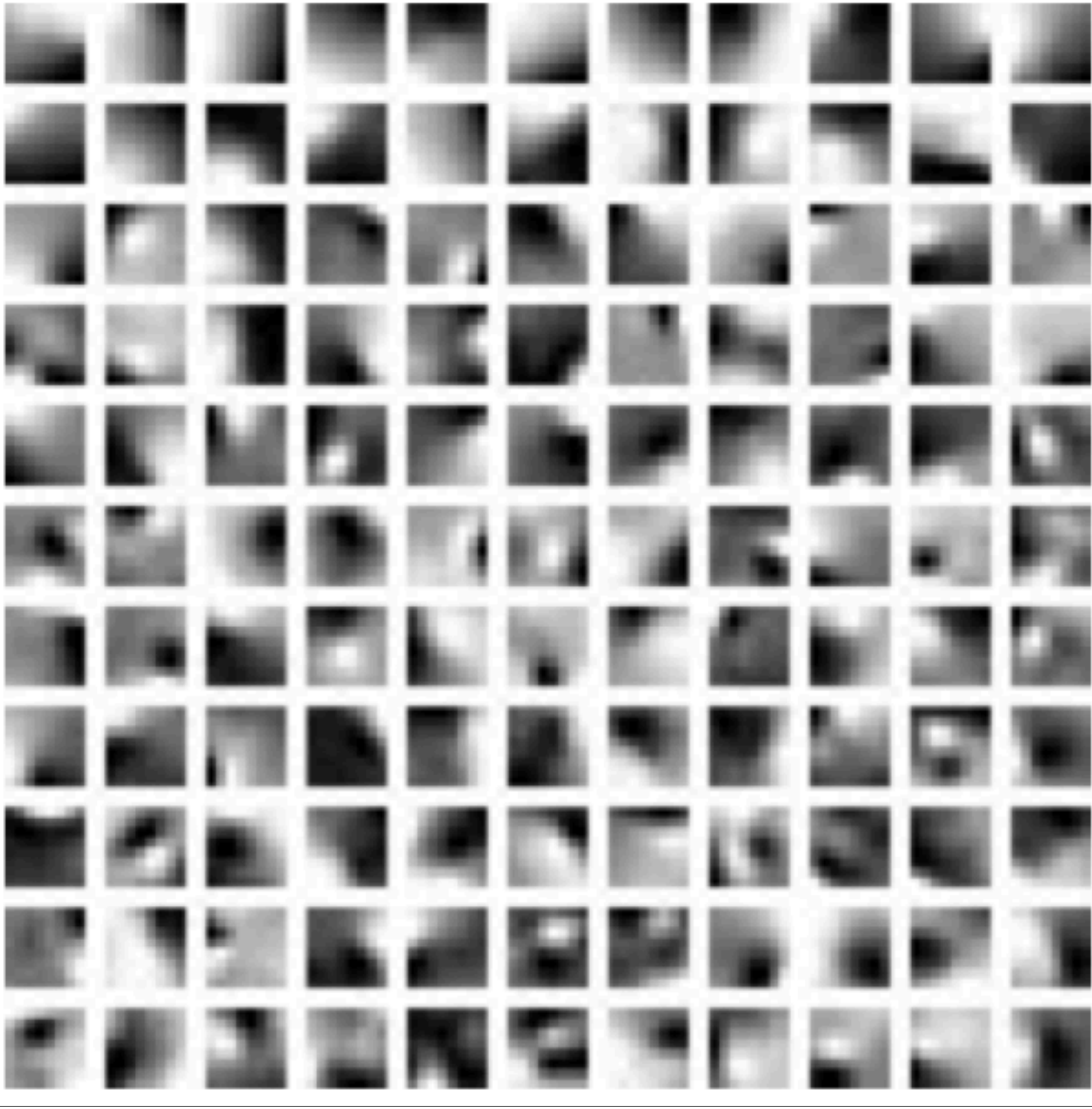
DL: Matrix Factorization problem

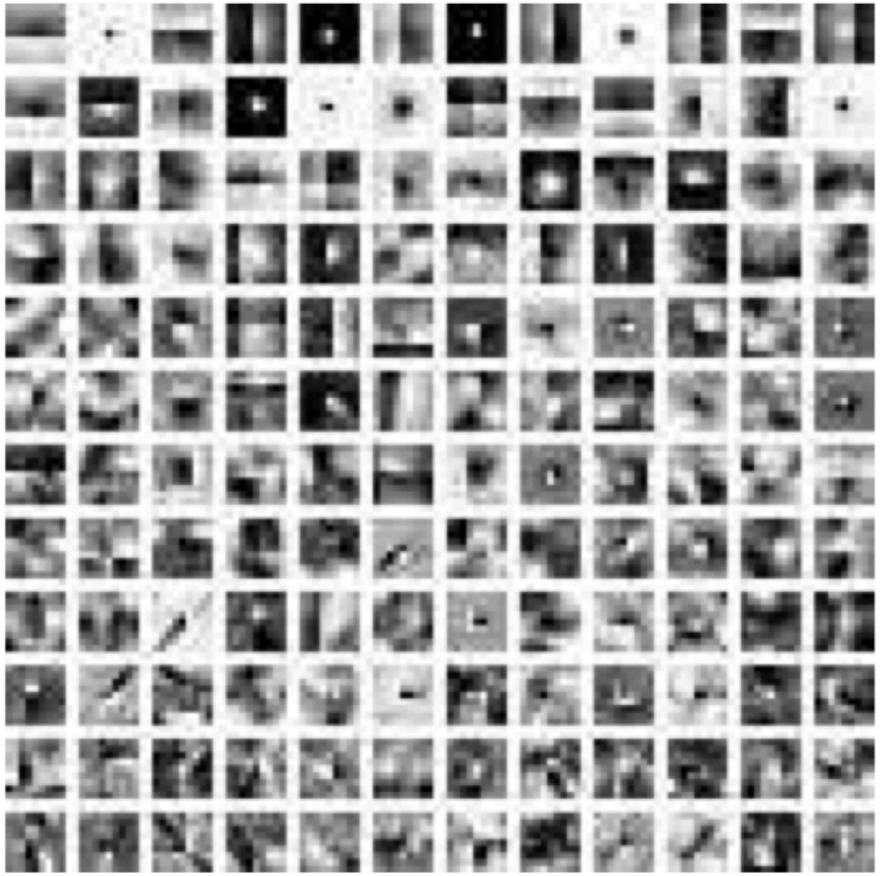
C_1: Constraints on the Sparsifying dictionary D

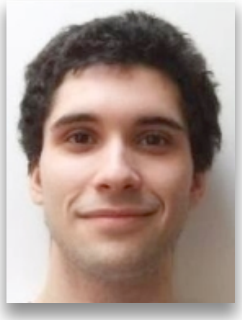
C_2: Constraints on the Sparse codes





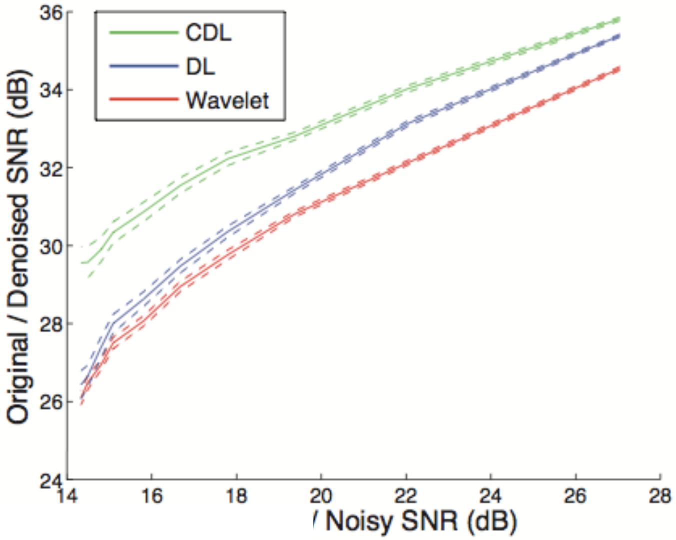


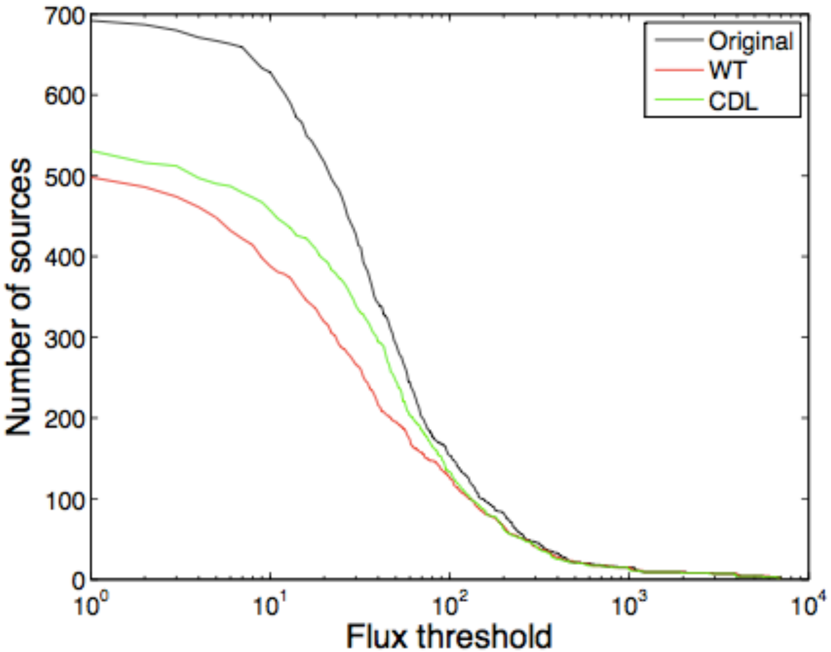



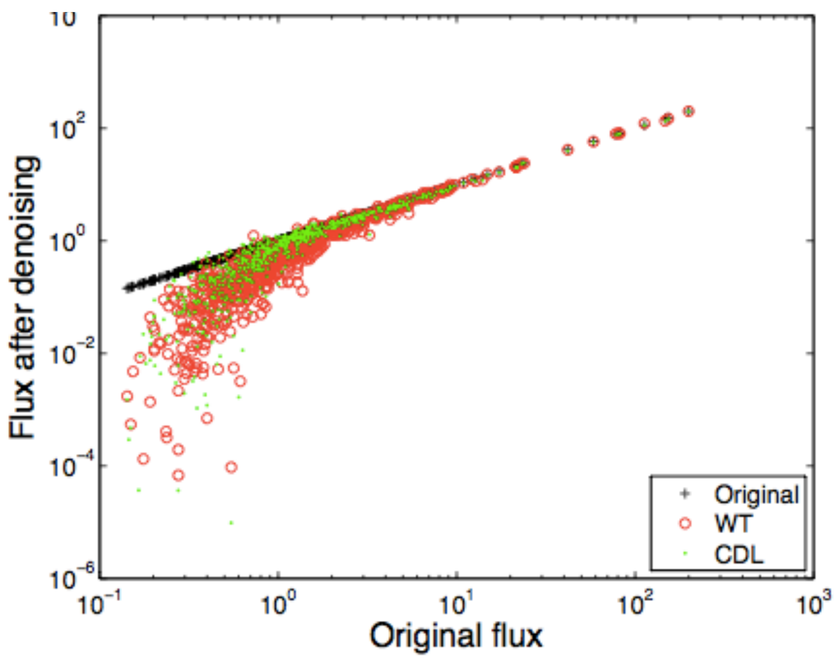


S. Beckouche

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







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

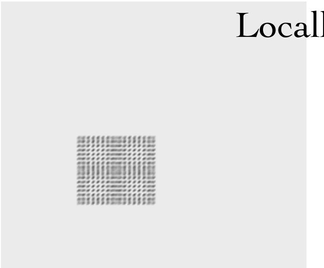
$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

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

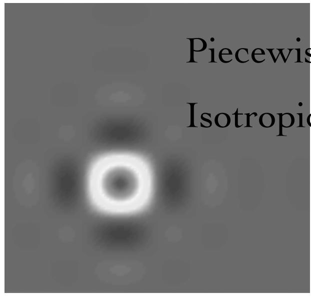
Local DCT

Stationary textures

Locally oscillatory



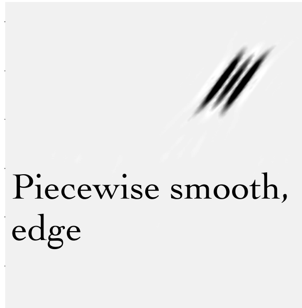
Wavelet transform



Piecewise smooth

Isotropic structures

Curvelet transform



Piecewise smooth, edge

Sparsity Model 2: Morphological Diversity:

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

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

Model 3 can be also combined with model 2:

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

Advantages of model 1 (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

Morpho-Spectral Diversity

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

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

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

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

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

$\mathbf{\Phi_A} = [\mathbf{\Phi_{A,1}}, \mathbf{\Phi_{A,2}}]$ Spatial Dictionary with

$\mathbf{\Phi_S}$ Spectral Dictionary

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

Generalized MCA (GMCA)

•J. Bobin, J.-L. Starck, M.J. Fadili, and Y. Moudden, "Sparsity, Morphological Diversity and Blind Source Separation", IEEE Trans. on Image Processing, Vol 16, No 11, pp 2662 - 2674, 2007.
•J. Bobin, J.-L. Starck, M.J. Fadili, and Y. Moudden, "Blind Source Separation: The Sparsity Revolution", Advances in Imaging and Electron Physics , Vol 152, pp 221 -- 306, 2008.

Source: $S = [s_1, \ldots, s_n]$ Data: $X = [x_1, \ldots, x_m] = AS$

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

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

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

==> GMCA searches a sparse solution S in the dictionary ϕ subject to the constraint that the norm $\| X - AS \|^2$ is minimal.

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

GMCA aims at solving the following minimization:

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

Sparse Component Separation: the GMCA Method

A and S are estimated alternately and iteratively in two steps :

1) Estimate S assuming A is fixed (iterative thresholding) :

$$\{S\} = \mathrm{Argmin}_S \sum_j \lambda_j \| s_j \mathbf{W} \|_1 + \| \mathbf{X} - \mathbf{AS} \|^2_{F,\mathbf{\Sigma}}$$

2) Estimate A assuming S is fixed (a simple least square problem) :

$$\{A\} = \mathrm{Argmin}_A \| \mathbf{X} - \mathbf{AS} \|^2_{F,\mathbf{\Sigma}}$$

BSS experiment : Noiseless case

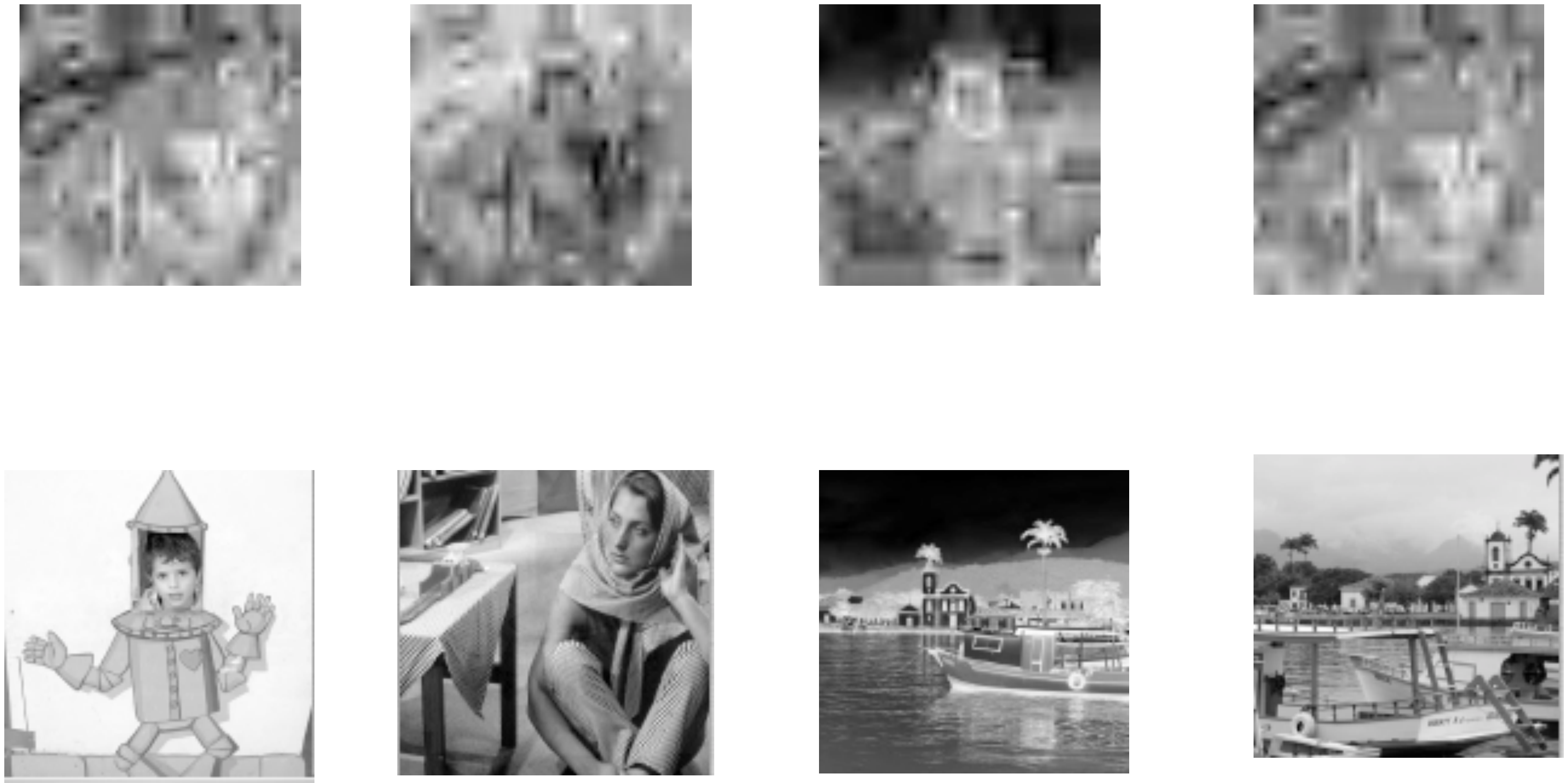
Original Sources

2 of 4 Mixtures

Noiseless experiment, 4 random mixtures, 4 sources

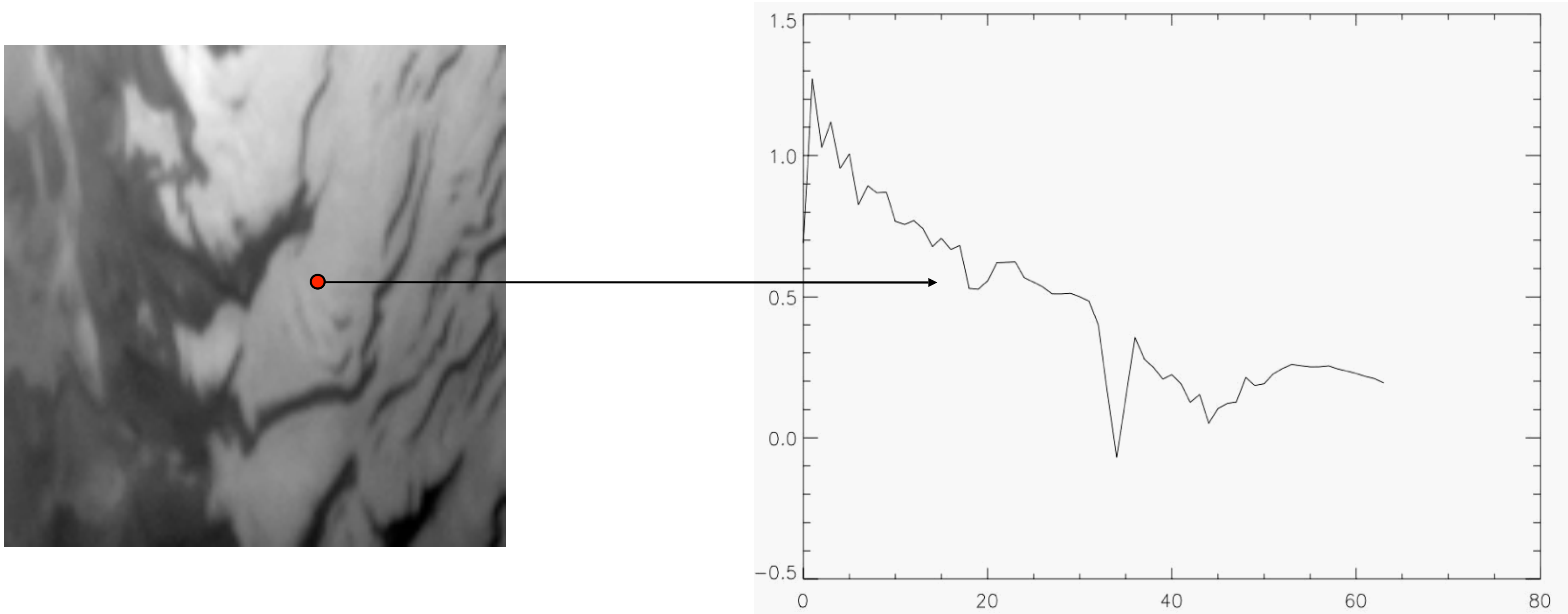
GMCA Experiment

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



Hyperspectral Data

Morphological Component Analysis for Sparse Multichannel Data: Application to Inpainting,
Journal of Mathematical Imaging and Vision, submitted.

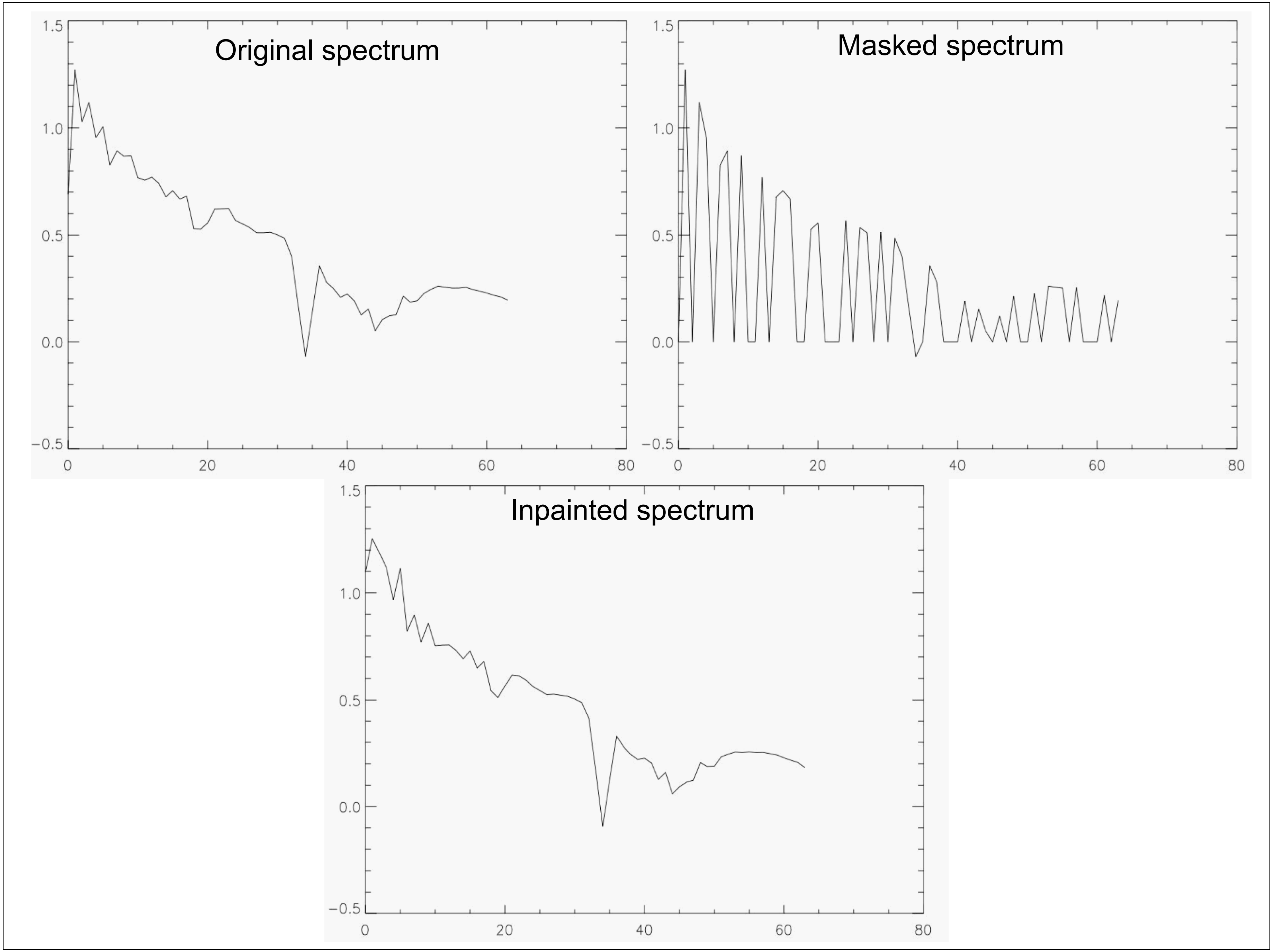


$$\min_{A,c_{1,1},\ldots,c_{1,K},\ldots,c_{n,1},\ldots,c_{n,K}} = \sum_{l=1}^{m} \left\| M_l \left(X_l - \sum_{i=1}^{n} A_{i,l} \sum_{k=1}^{K} c_{i,k} \right) \right\|_2^2 + \lambda \sum_{i=1}^{n} \sum_{k=1}^{K} \left\| T_{i,k} c_{i,k} \right\|_p + \lambda \sum_{i=1}^{n} \left\| W^{(1D)} A_i \right\|_p$$

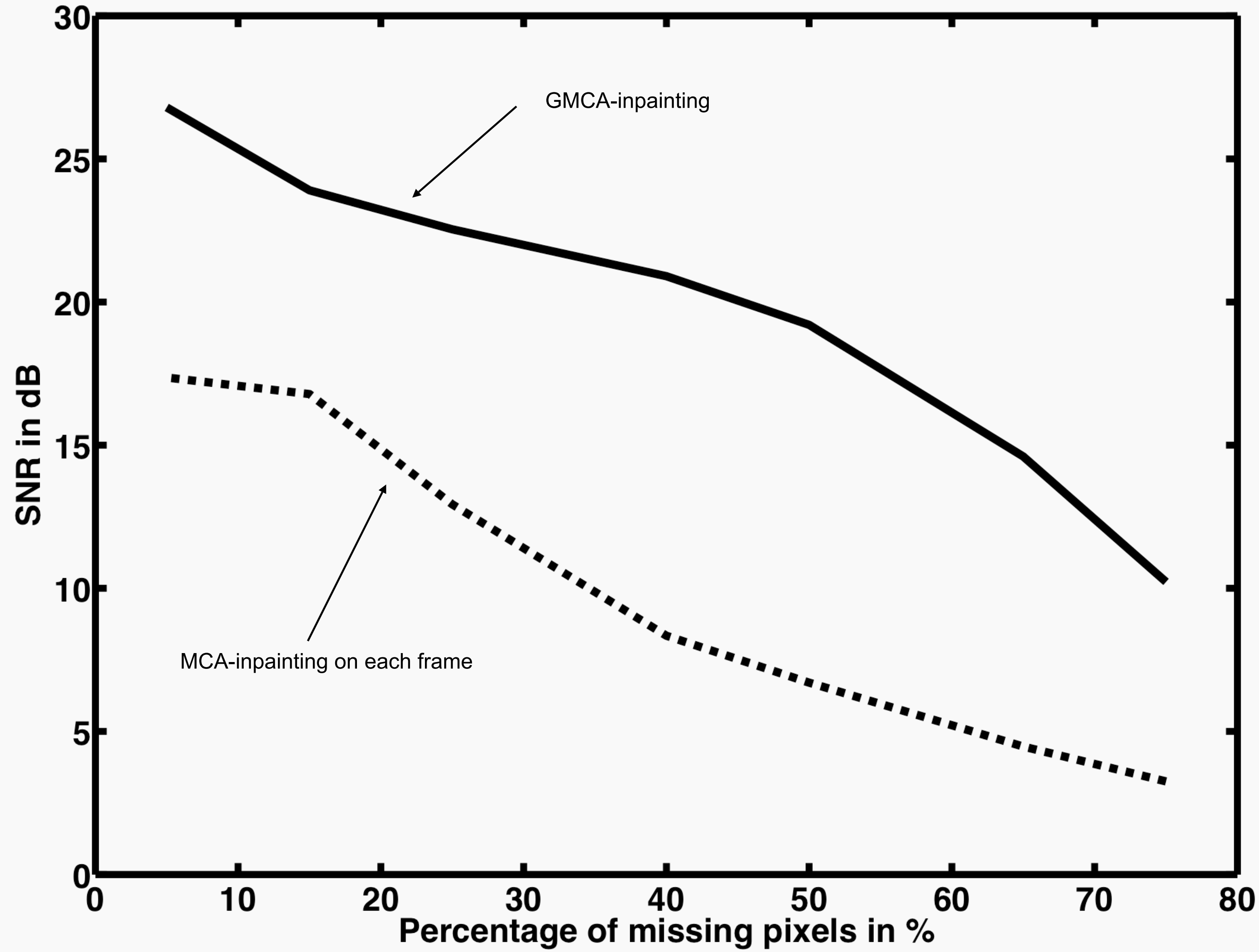
Inpainting hyperspectral data

Omega Camera on Mars Orbiter: 128 x128 x 64 channels

50% missing pixels



Original spectrum
Masked spectrum
Inpainted spectrum
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80

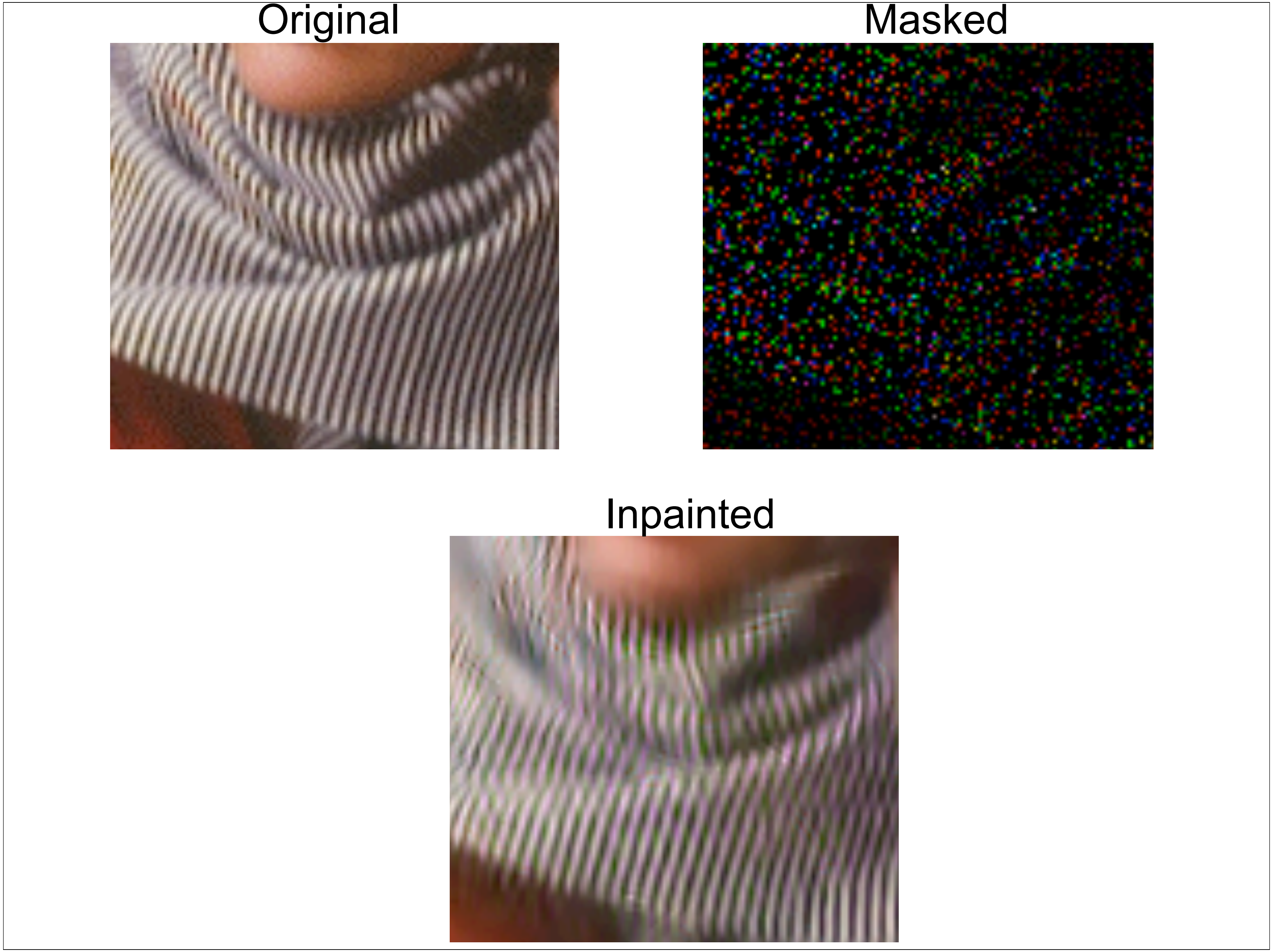


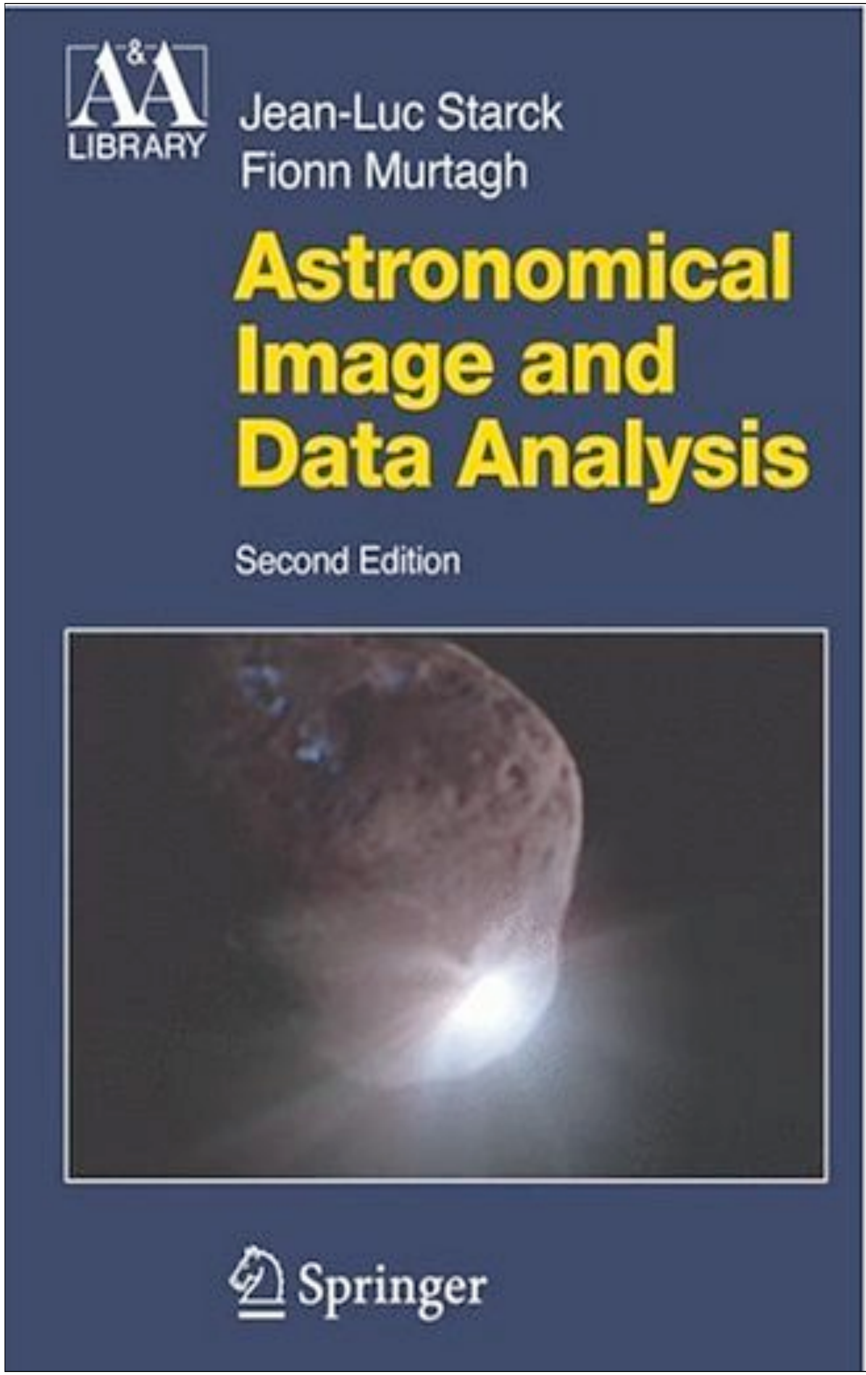
30
25
20
15
10
5
0
0
10
20
30
40
50
60
70
80
SNR in dB
Percentage of missing pixels in %
GMCA-inpainting
MCA-inpainting on each frame

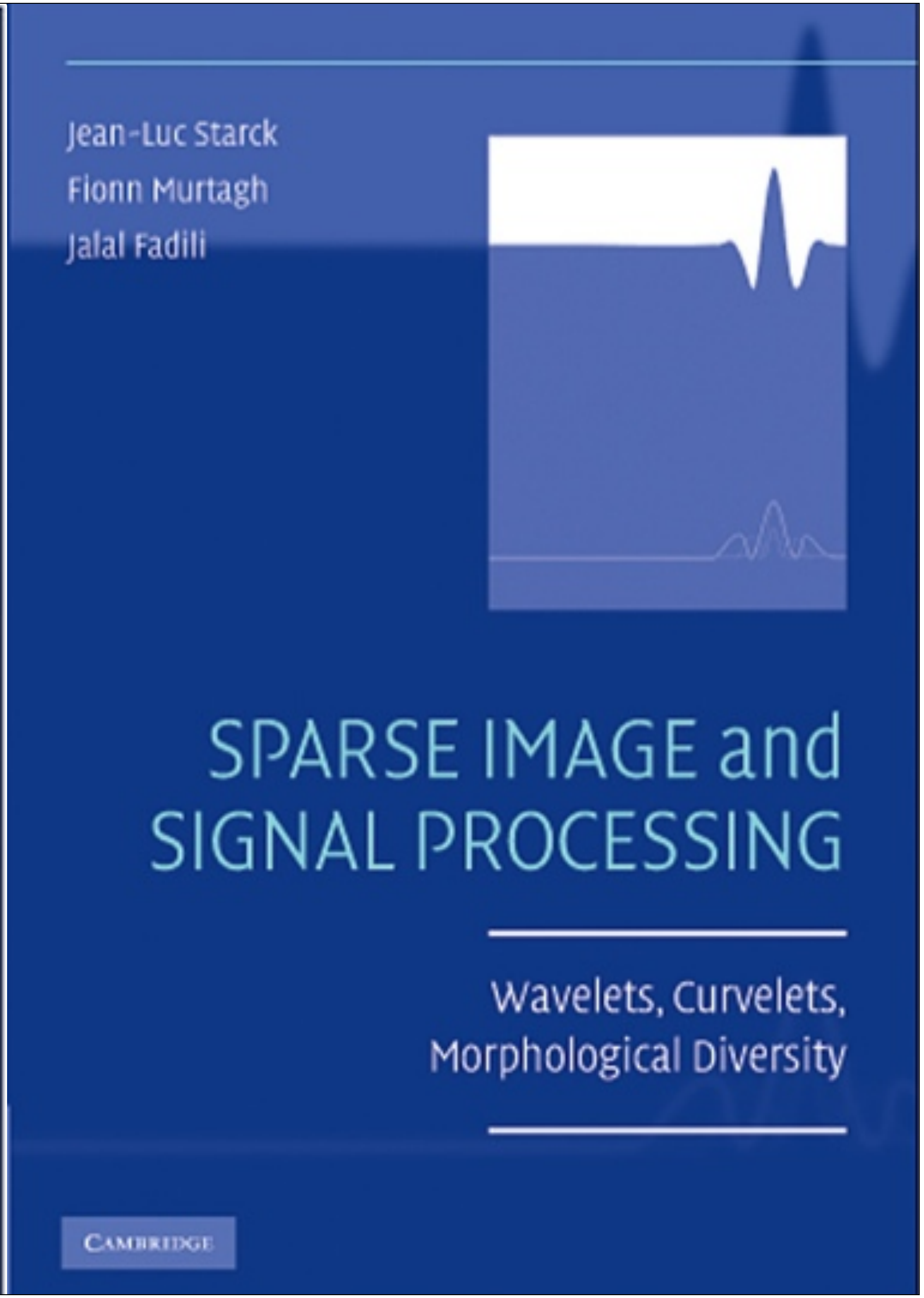
Inpainting color images

3 color channels

Dictionary
Curvelets + LDCT


Original
Masked
Inpainted


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


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