Signal Processing for Astronomy: From Wavelets to Deep Learning

J.L. Starck, Florent Sureau, Francois Lanusse

Institut de Recherche sur les lois Fondamentales de l'Univers (IRFU)
CEA Saclay, France

www.cosmotat.org

COSMOSTAT

Outline

1. Sparsity and Multi-Scale Representations
2. Sparsity and Inverse Problems
3. Deep Generative models for Inverse problems

Representation of data

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

$$\mathbf{s} = \sum_{k=1}^{K} \underbrace{\alpha_k}_{\text{coefficients}} \underbrace{\phi_k}_{\text{basis, frame, dictionary}}$$

- Allows to **analyze the signals through the statistical properties of coefficients** (e.g. Fourier basis to analyze the frequency content)

- **Sparsity** of the coefficients in appropriate dictionary key to many results in **approximation theory**

- Requirement in practice: **Fast computation** of coefficients α_k (structured dictionary)

Multi-scale Representation

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

Dictionary
(basis, frame)

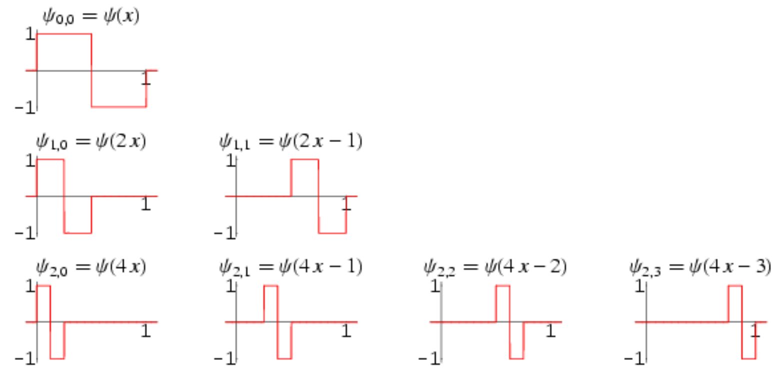
$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

Atoms

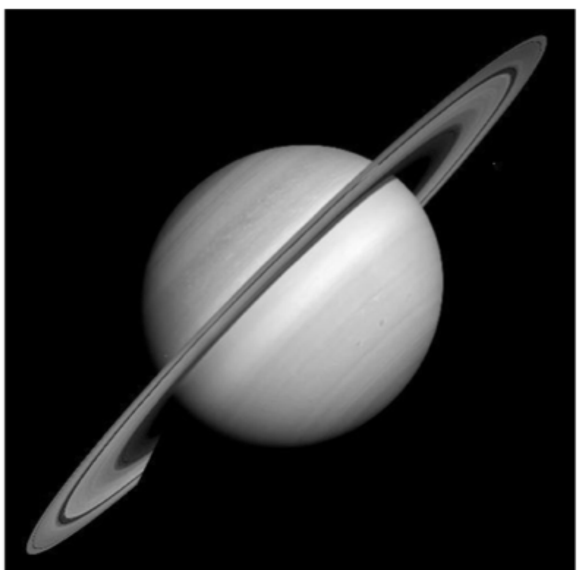
coefficients

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

Ex: Haar wavelet



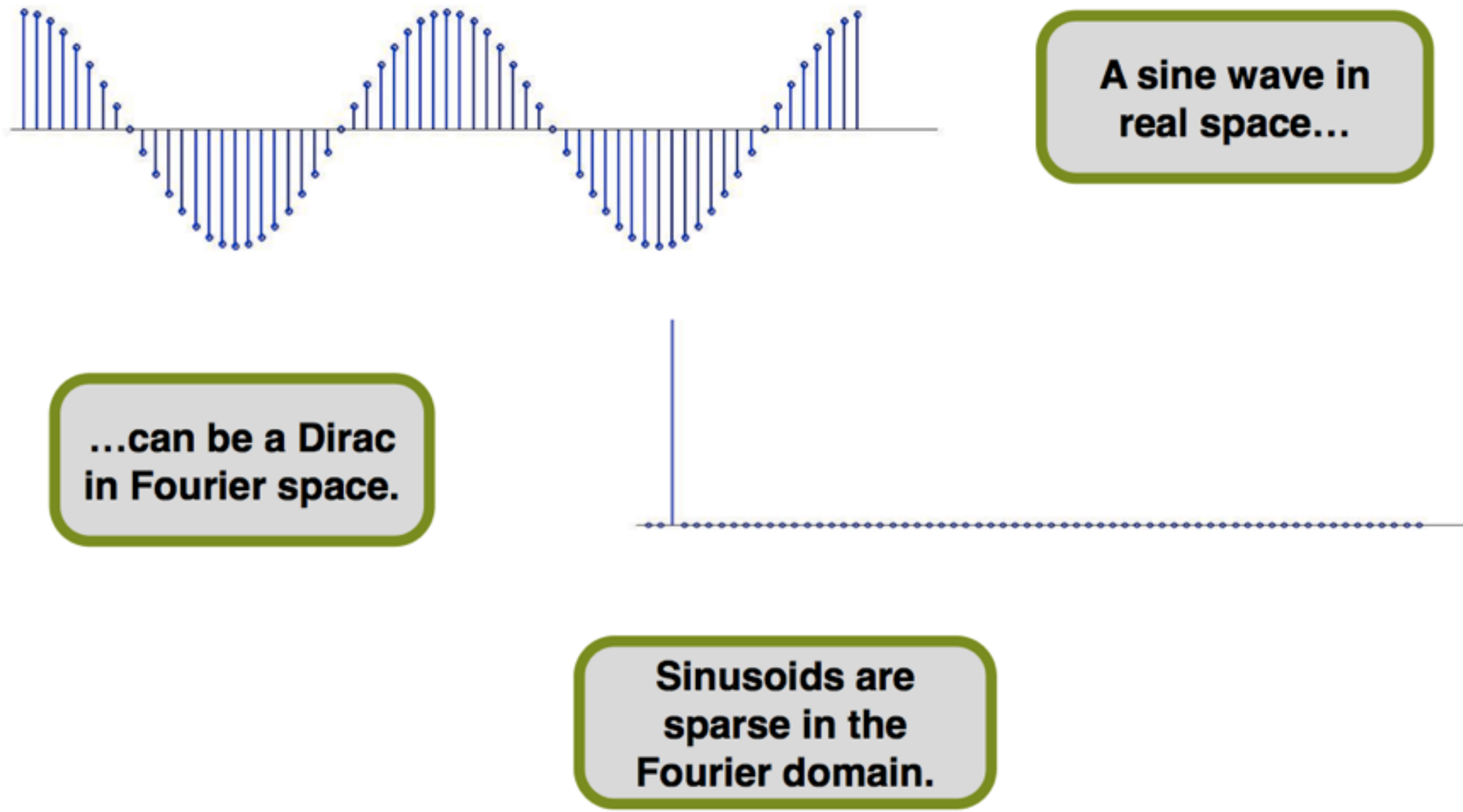
$$f(x) = c_0 + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} c_{jk}\psi_{jk}(x).$$



Multiscale analysis of Saturn via the wavelet transform

Strict Sparsity: K-sparse signals

K-sparse : K coefficient different from zero



How to measure Sparsity? Sparse decomposition?

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

$$\text{(P0)} \qquad \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \qquad s.t. \qquad \mathbf{s} = \boldsymbol{\phi}\boldsymbol{\alpha}$$

with the convention $0^0 = 0$, $\|\boldsymbol{\alpha}\|_0 = \sum_k \alpha_k^0 = |\{\alpha_k \neq 0\}|$

It has been proposed (to relax and) to replace the ℓ_0 norm by the ℓ_1 norm (Chen95):

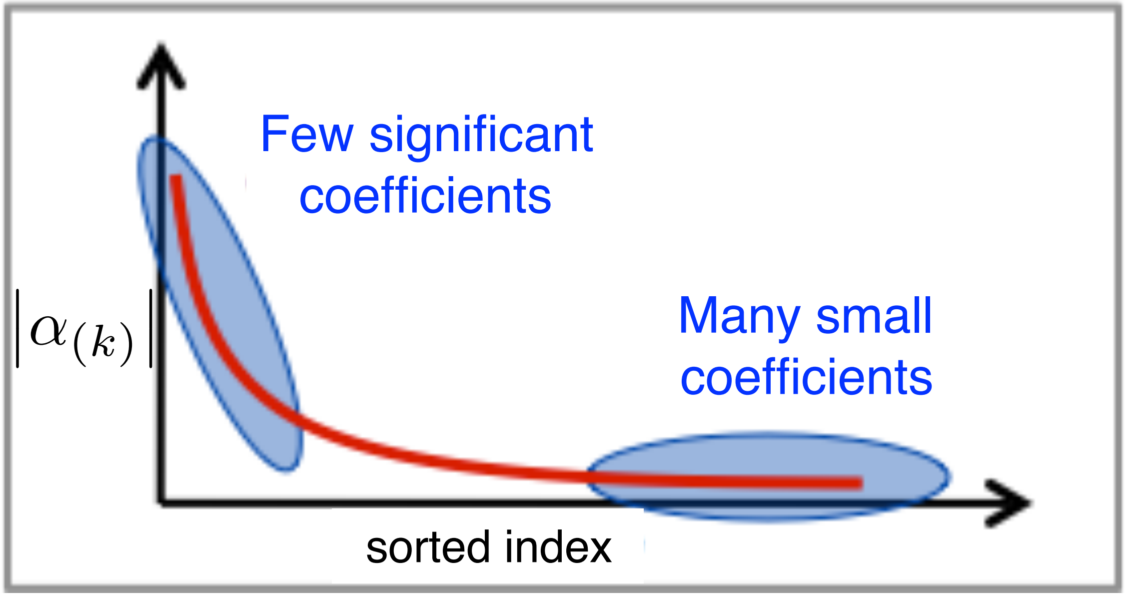
$$\text{(P1)} \qquad \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \qquad s.t. \qquad \mathbf{s} = \boldsymbol{\phi}\boldsymbol{\alpha}$$

with $\|\boldsymbol{\alpha}\|_1 = \sum_k |\alpha_k|$

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

It has been shown (Donoho99) that for certain dictionary, if there exists a highly sparse solution to (P1), then it is identical to the solution of (P0).
The key point is the **mutual coherence** of the dictionary ("two atoms should not look too much alike").

From Sparse to Compressible Signals



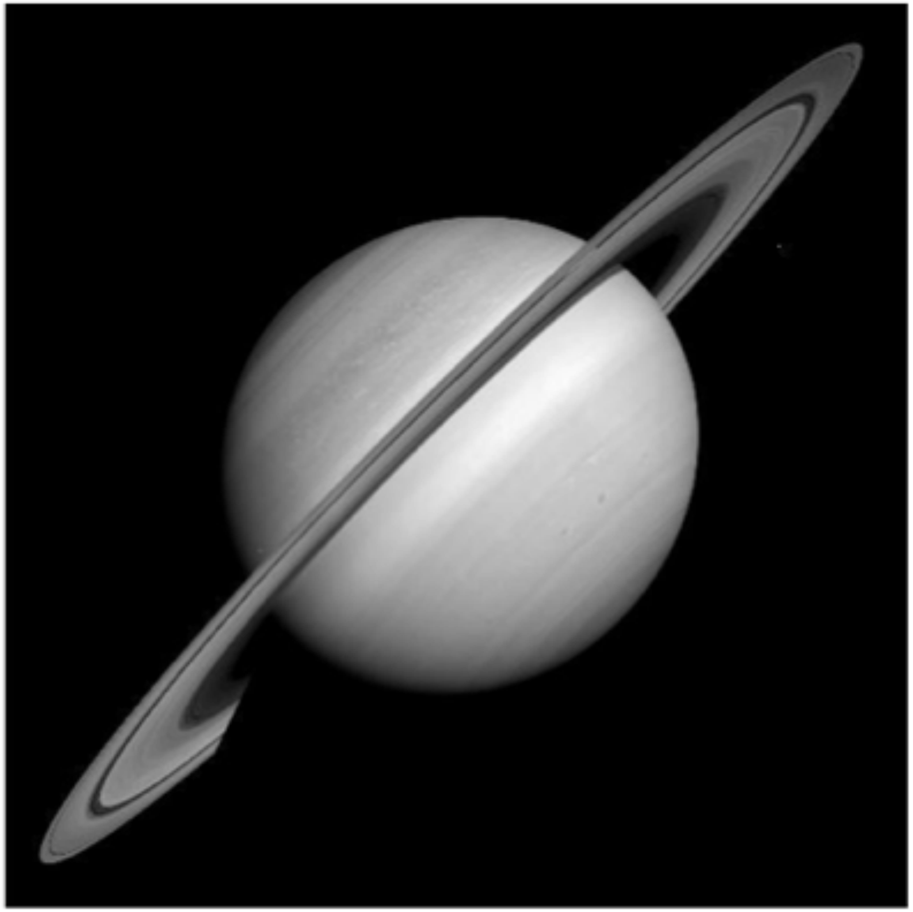
$$|\alpha_{(k)}| \leq C k^{-1/q} \quad \Leftrightarrow \quad ||\alpha - \alpha_K||_2 \leq C_q K^{1/2-1/q} \qquad q < 2$$

Best K-term approximation

Non-Linear Approximation curve: $f(K) = ||\alpha - \alpha_K||_2$

A compressible signal can be closely approximated by a sparse signal

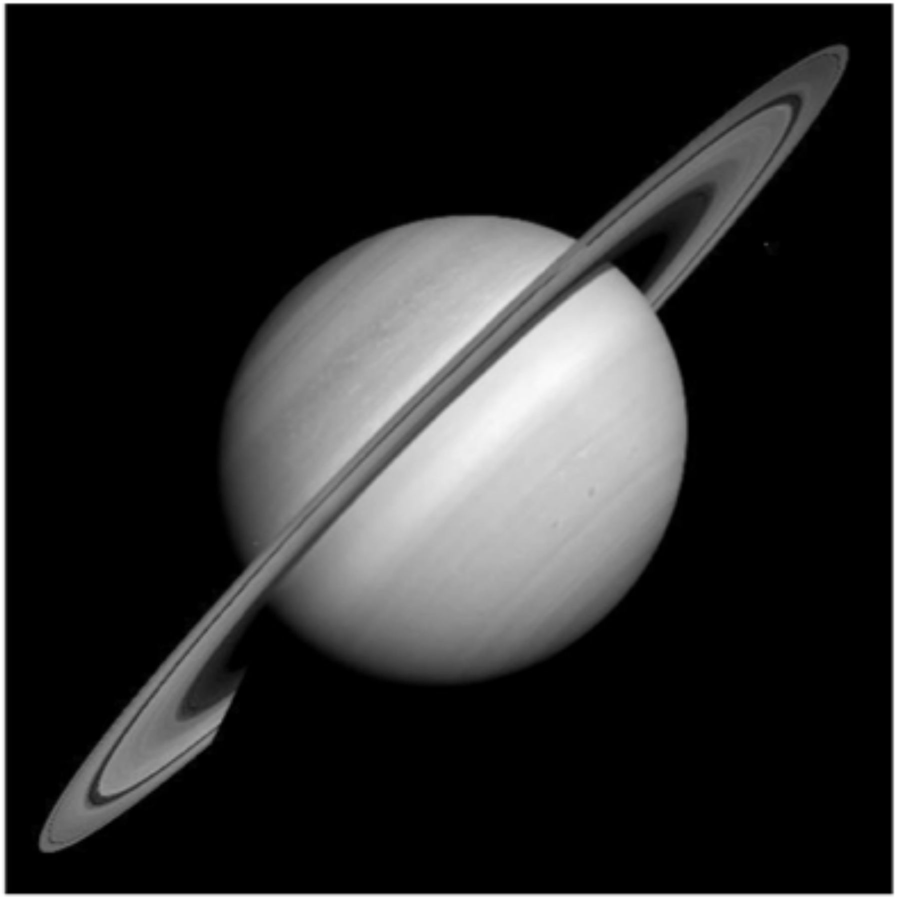
An hint of how helpful sparsity is....



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

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

An hint of how helpful sparsity is....

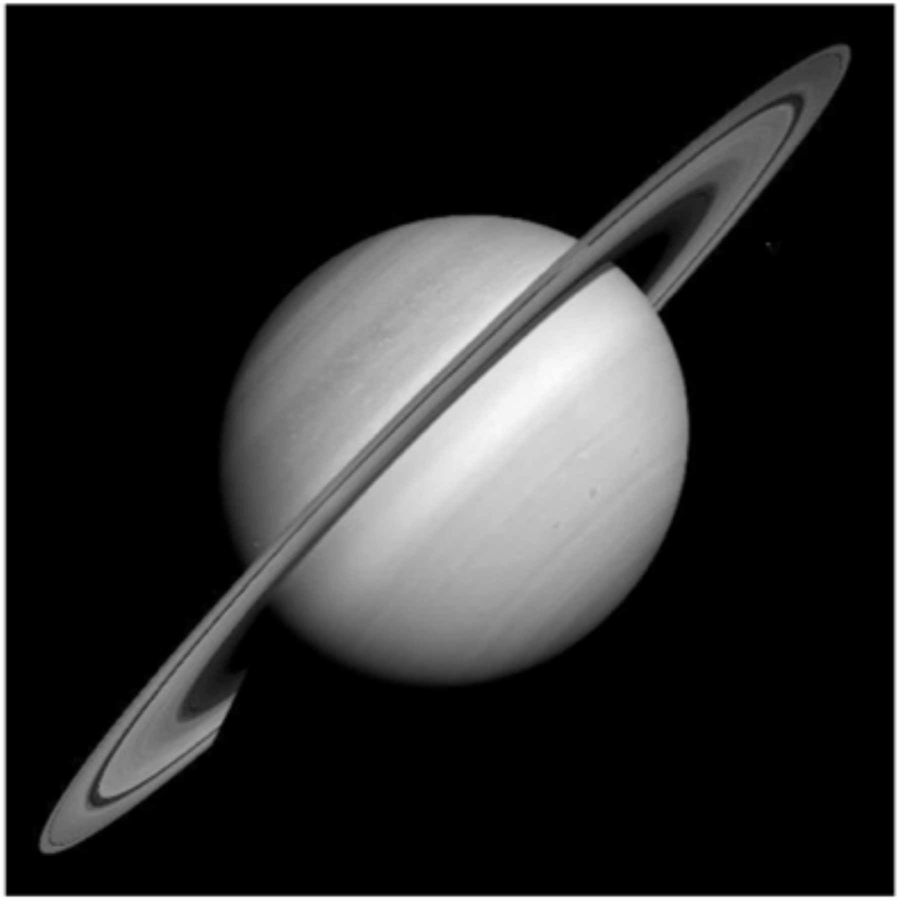


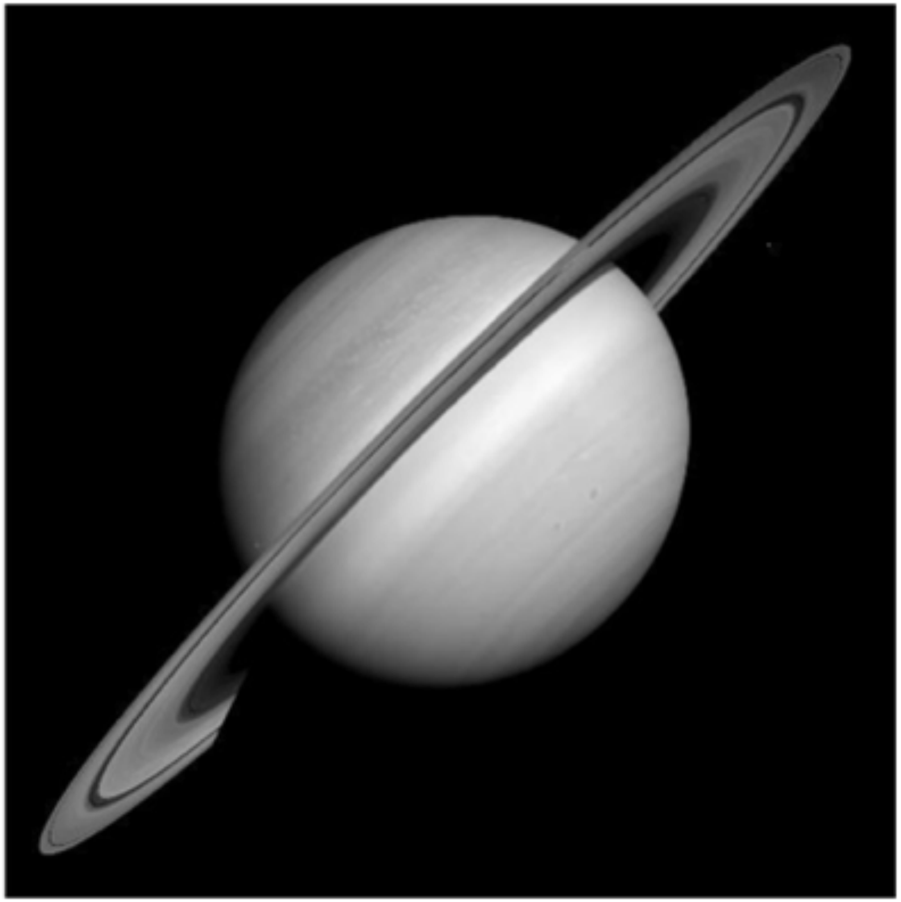
1% largest coefficients in wavelet space
(the others are set to 0)

The wavelet coefficients encode edges and large scale information.

Wavelet transform

An hint of how helpful sparsity is....





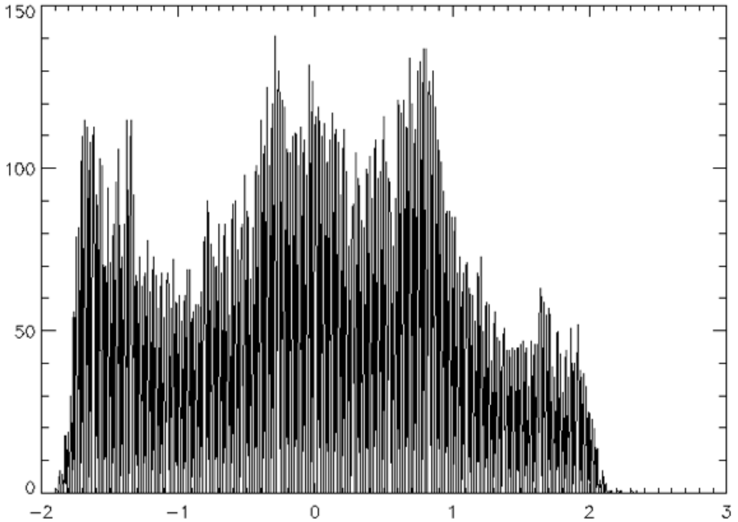
1% of the wavelet coefficients concentrate 99.96% of the energy: This can be used as a *prior.*

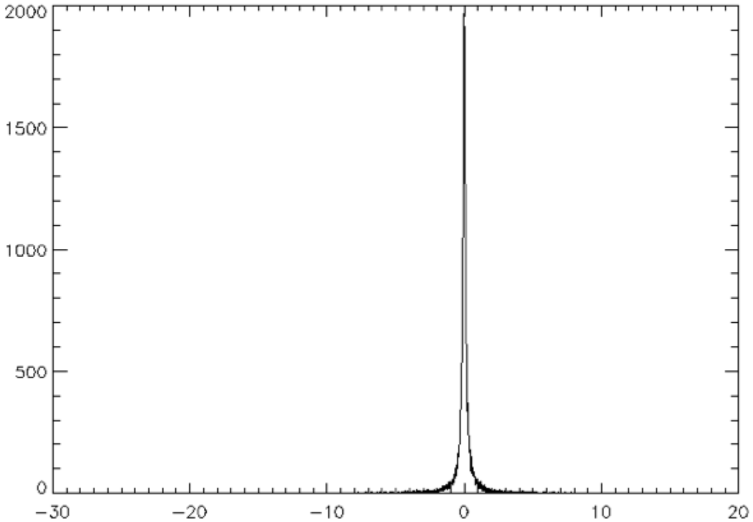
Reconstruction, after throwing away 99% of the wavelet coefficients

Representing Barbara

Direct Space

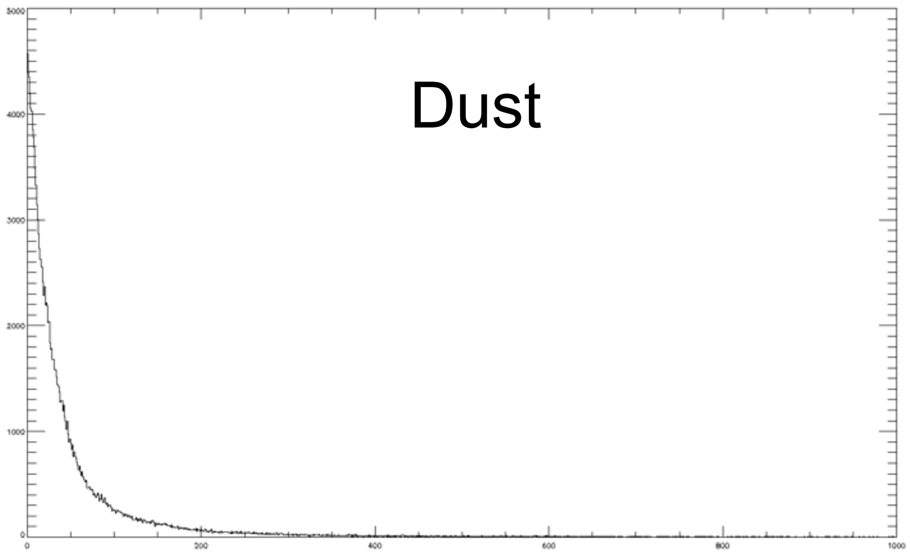
Curvelet Space

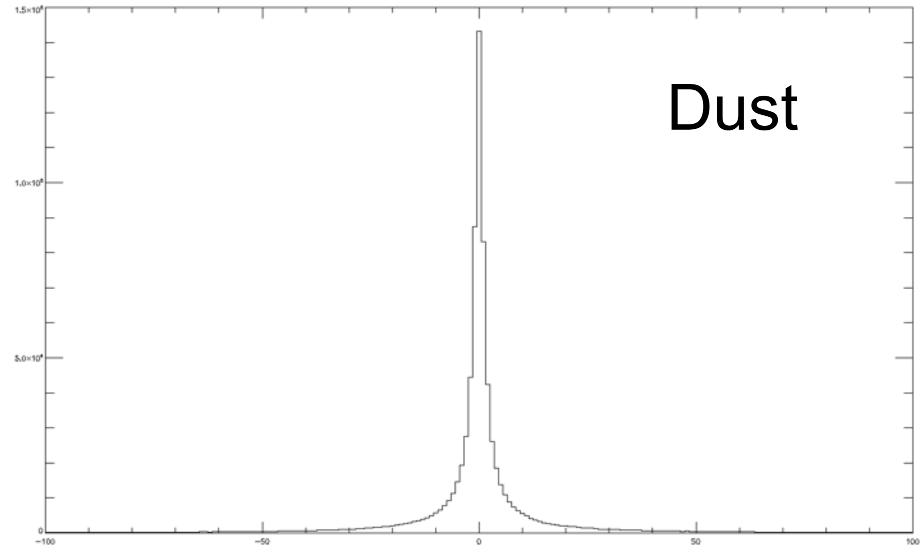


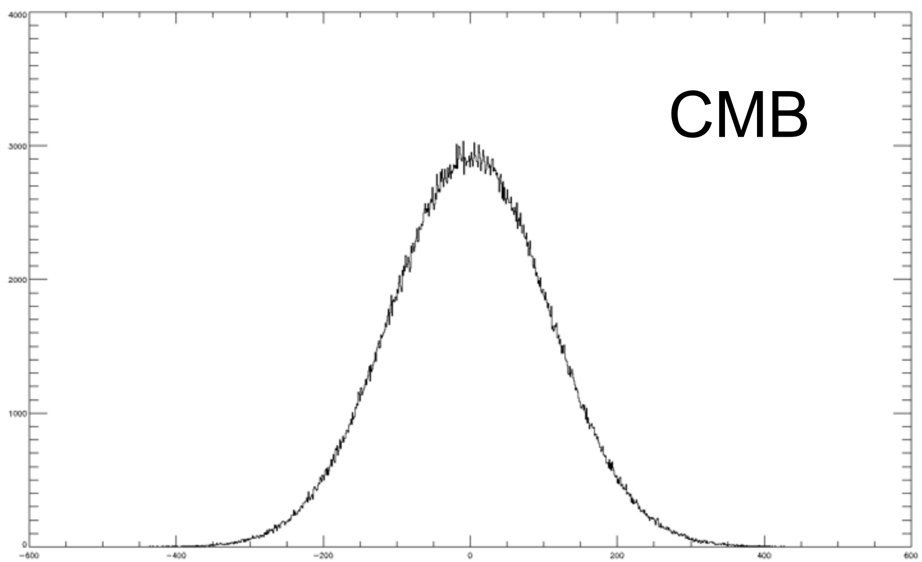


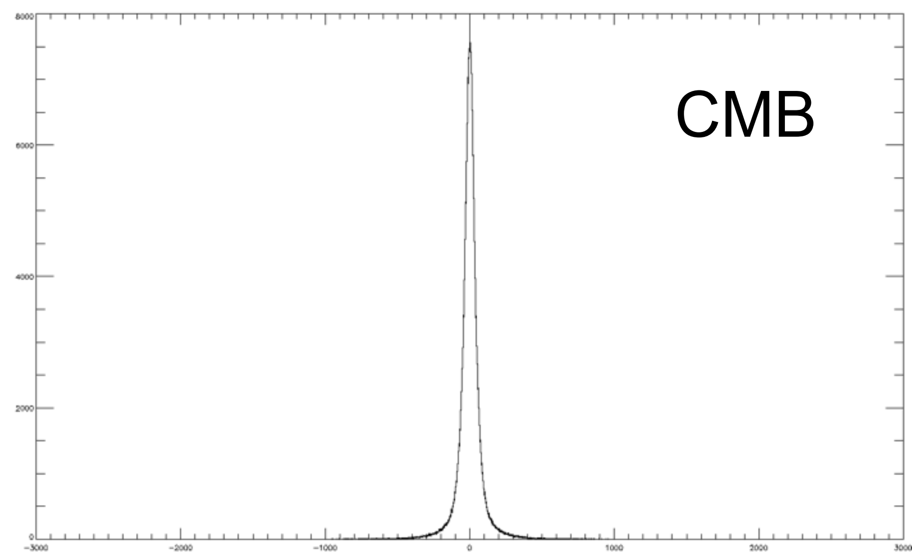
Planck Galactic and Cosmologic Components

Spatial Domain | Wavelet on the Sphere









Representation of data

$$\mathbf{s} = \sum_{k=1}^{K} \alpha_k \boldsymbol{\phi}_k$$

coefficients

basis, frame, dictionary

We want a sparse α. How to choose ϕ ?

A little bit of history on harmonic analysis

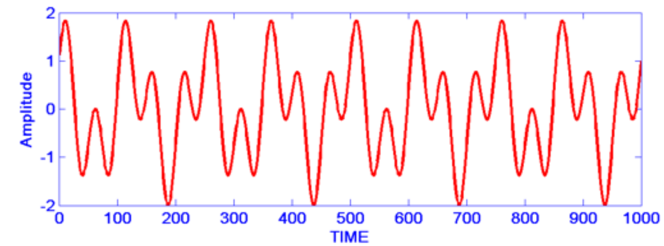
Jean-Baptiste-Joseph Fourier
(1768-1830)

Any integrable function can be expressed as a linear combination of basic trigonometric functions (sine and cosine)

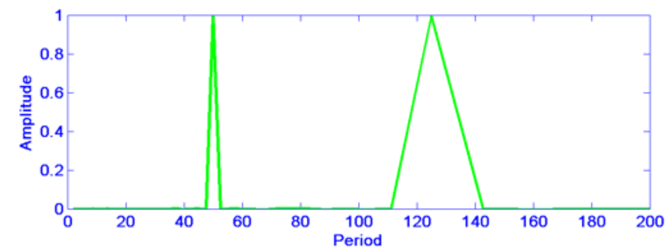
$$X(f) = \int_{-\infty}^{+\infty} x(t)\mathrm{e}^{-2i\pi f t}\mathrm{d}t$$

$$x(t) = \int_{-\infty}^{+\infty} X(f)\mathrm{e}^{2i\pi f t}\mathrm{d}f$$

Time Domain



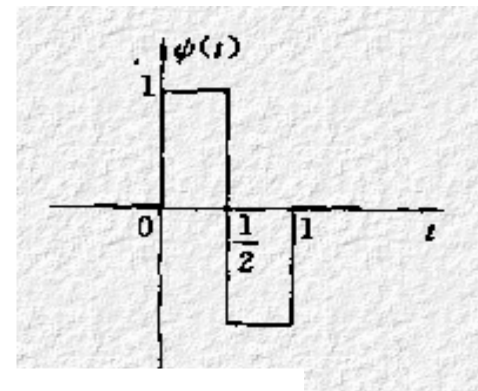
Frequency Domain



A little bit of history: first Wavelets

First mention of wavelets: appendix to the thesis of Haar (1909)

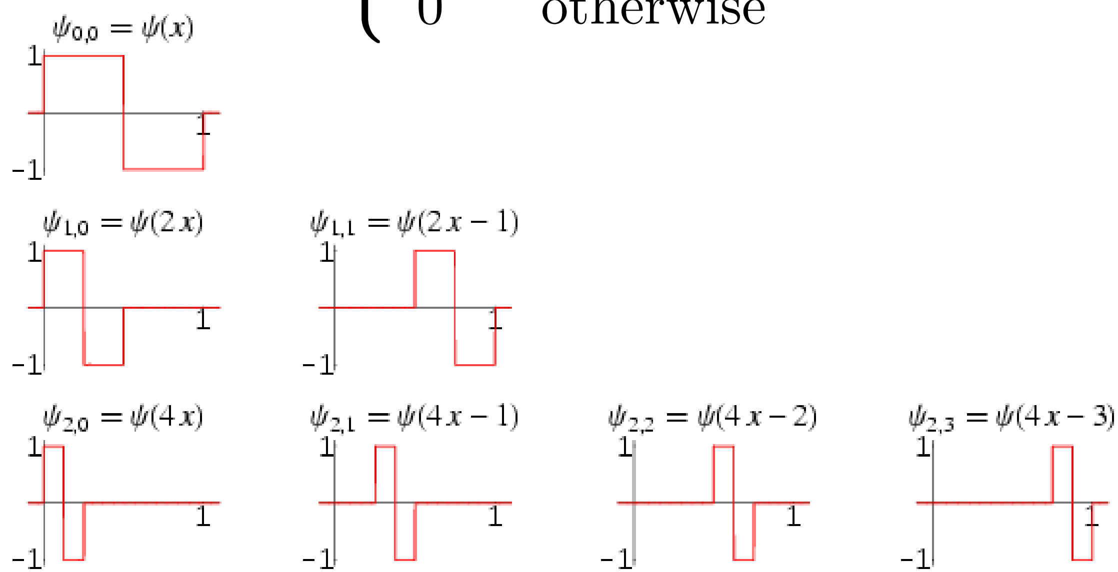
- With compact support, vanishes outside of a finite interval
- Not continuously differentiable
- Wavelets are functions defined over a finite interval
  and having an average value of zero.



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

$$\psi_{j,k}(x) = 2^j \psi\left(2^j x - k\right) \qquad \psi(x) = \begin{cases} 1 & 0 \le x < \frac{1}{2} \\ -1 & \frac{1}{2} \le x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Haar Wavelet



A little bit of history: Uncertainty Principle

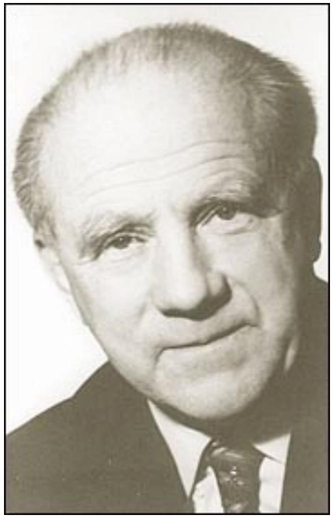
We want transforms that have good time and frequency resolution

Choice of decomposition:

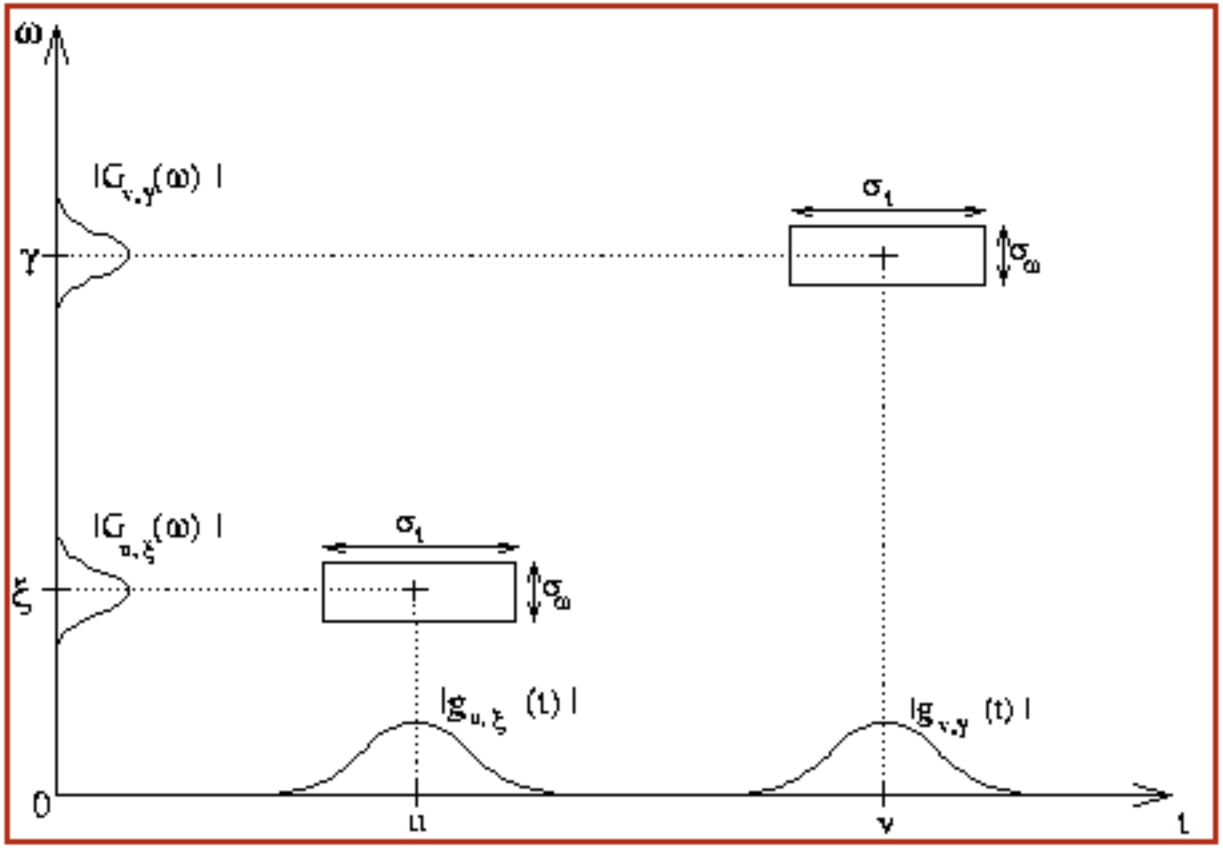
- dirac basis: best time resolution
- Sinusoids (Fourier): best frequency resolution

HOWEVER: Heisenberg, 1930
Uncertainty Principle:
There is a lower bound for $\sigma_t \cdot \sigma_\omega$

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



Example of Heisenberg boxes



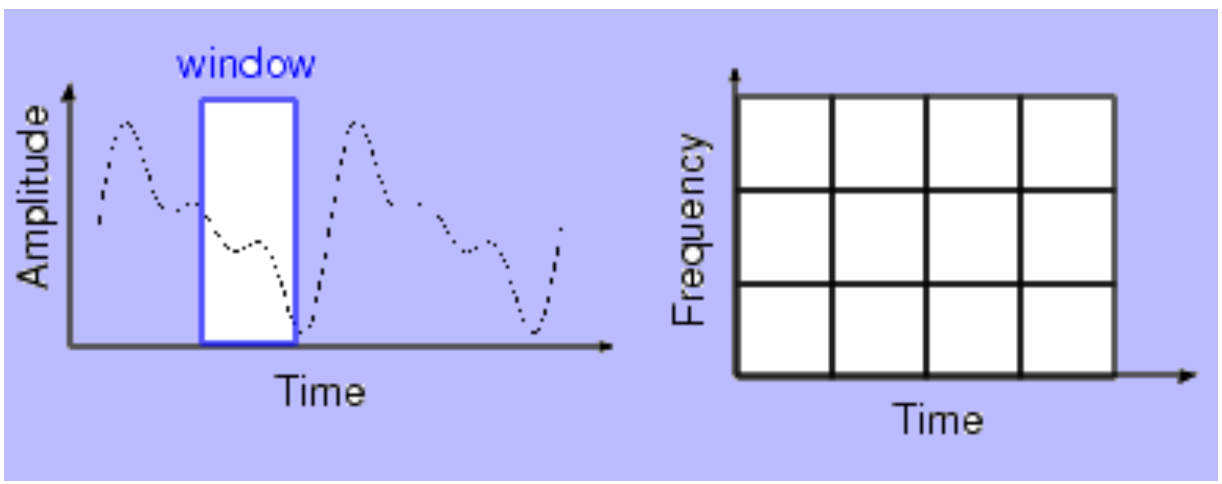
$$\sigma_t^2 \sigma_\omega^2 \geq \frac{1}{4}$$

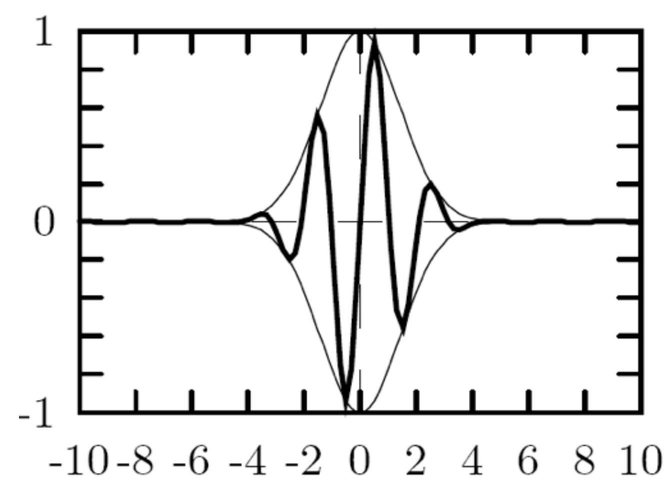
A little bit of history: Short Time Fourier Transform

- Windowing the Signal:
  Dennis Gabor (1946) used STFT to analyze only a small section of the signal at a time.

- The segment of signal is assumed stationary

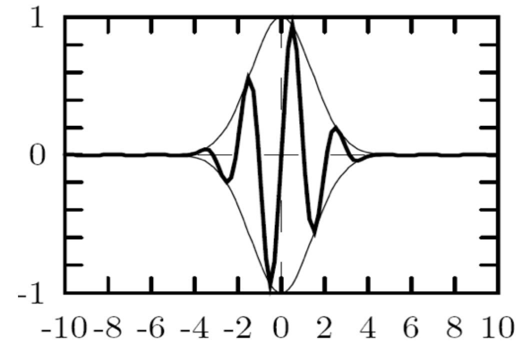
- Windowed Fourier Transform or Gaborlets:





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

Analyzing functions for Piecewise smooth signals



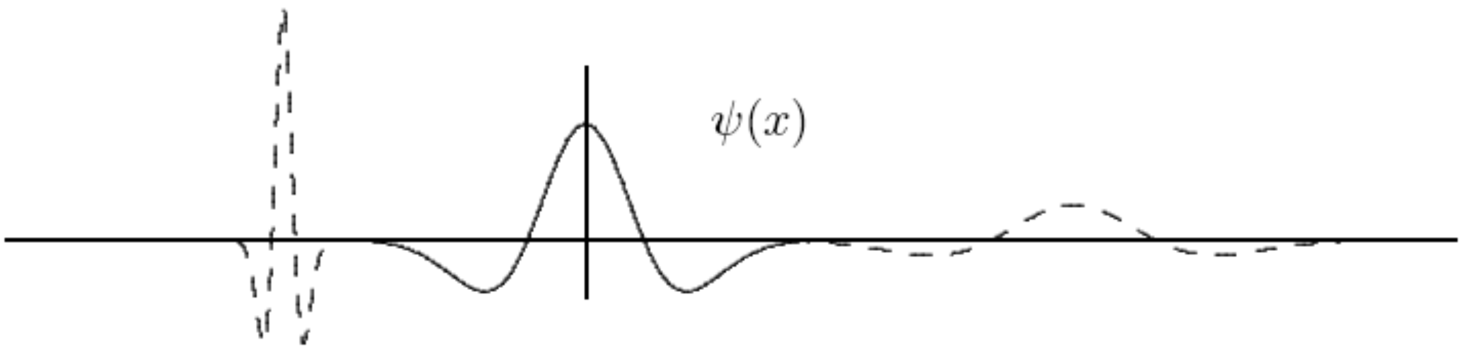
- Windowed Fourier Transform or Gaborlets:

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

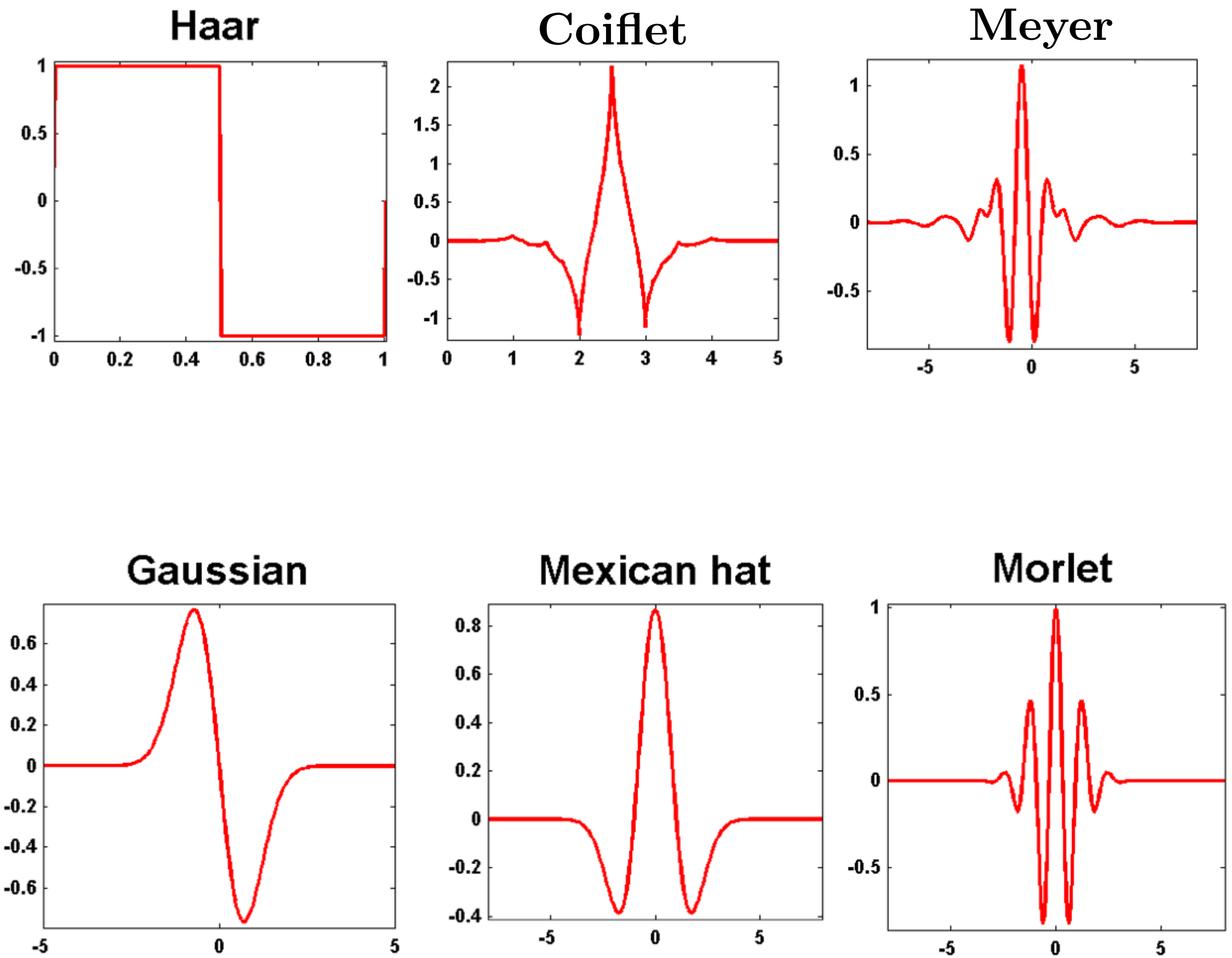
- Wavelets:

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

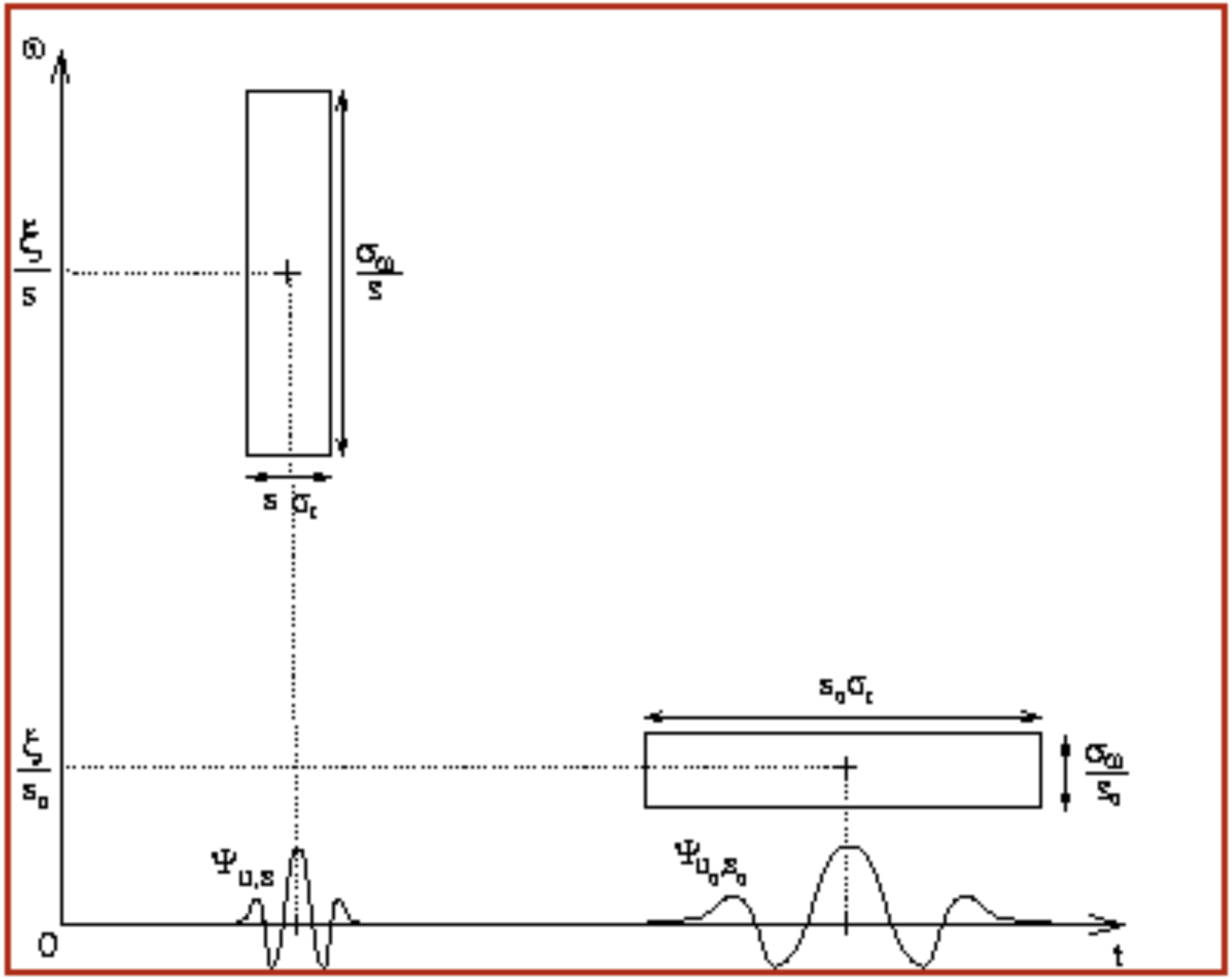
ψ: mother wavelet



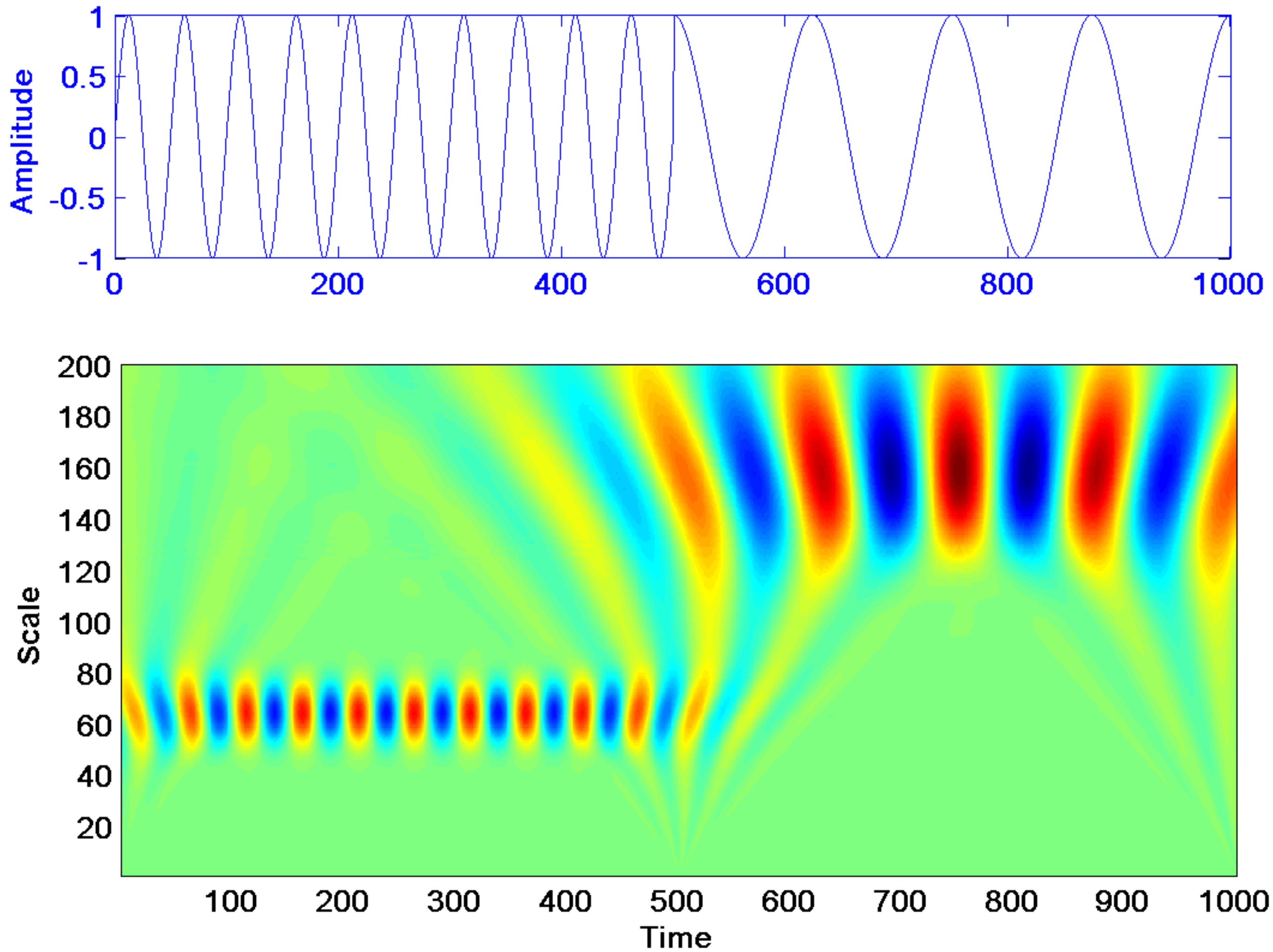
Typical Mother Wavelets



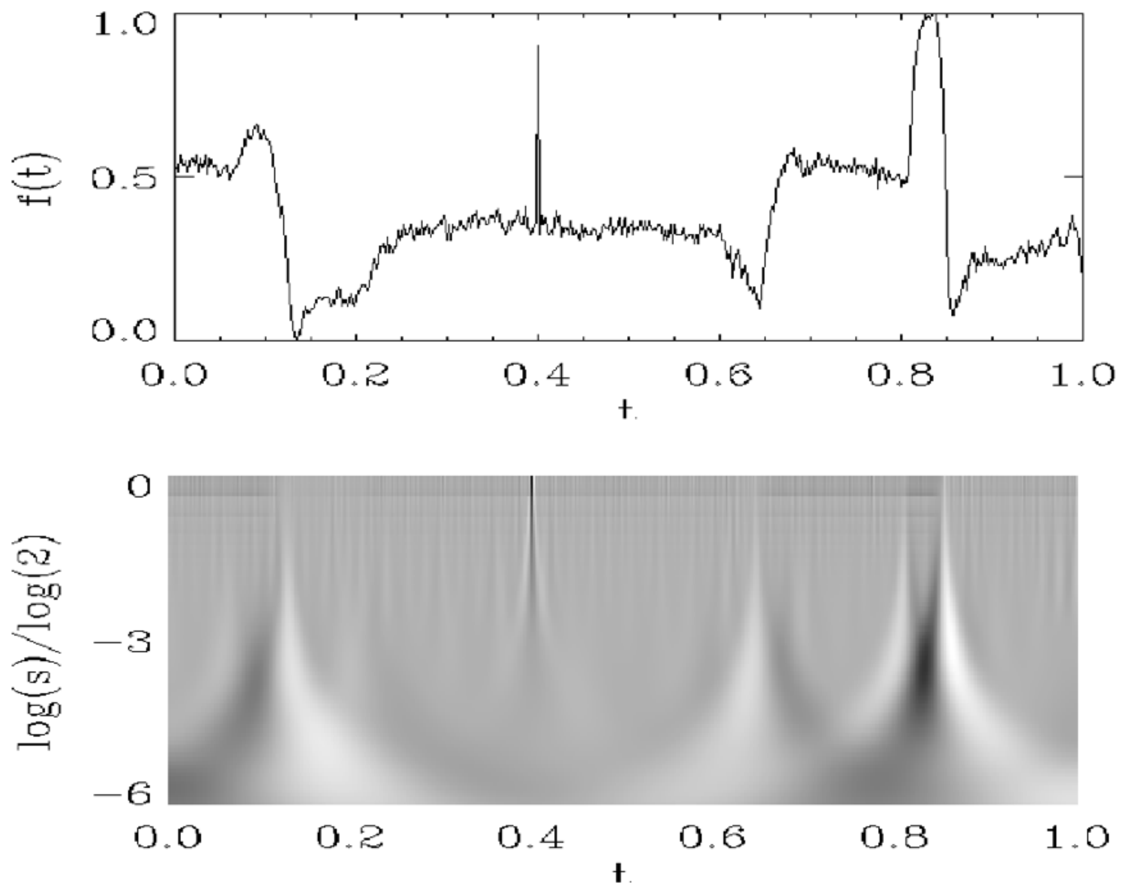
Heisenberg boxes for Wavelets



Typical spectrogram (1)

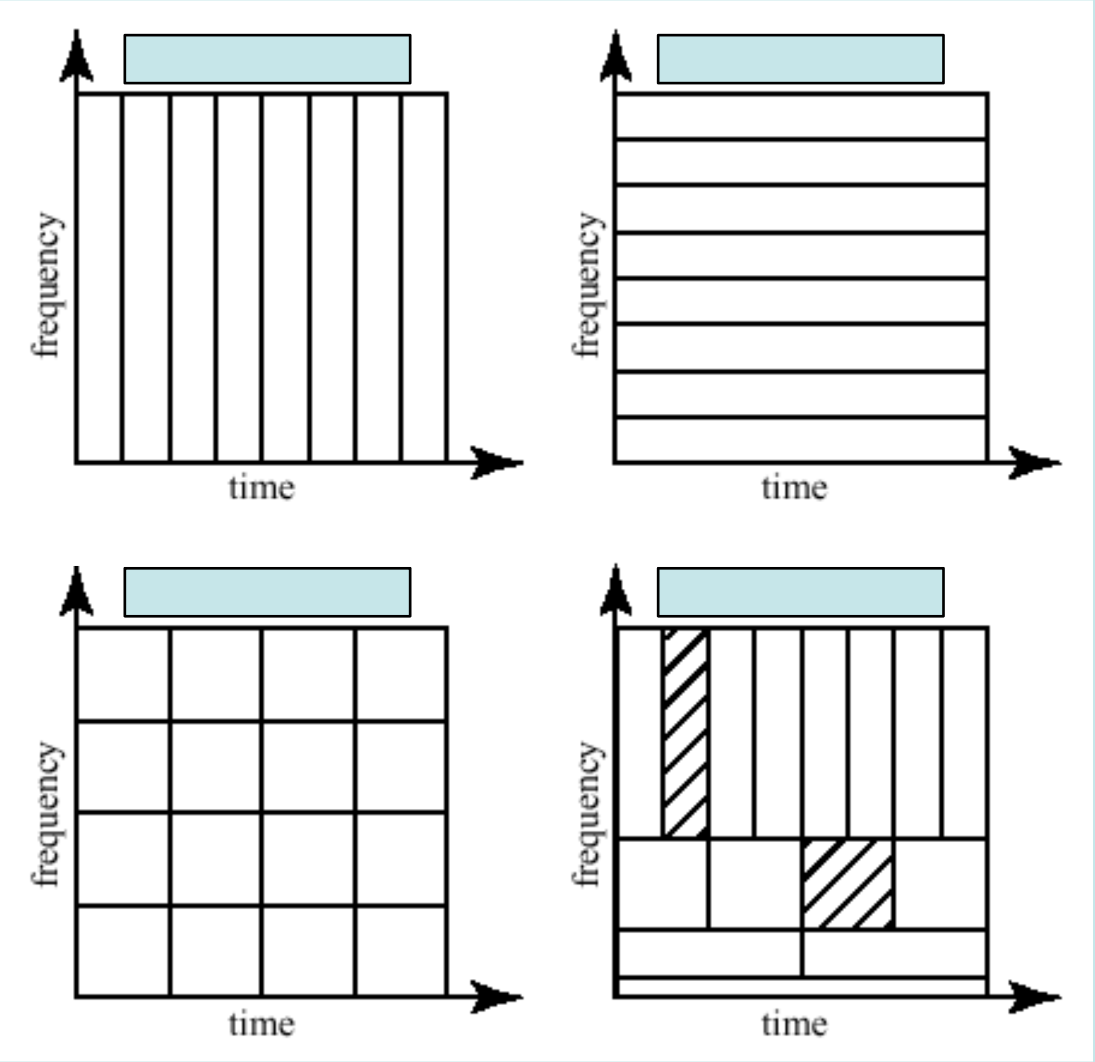


Typical spectrogram (2)



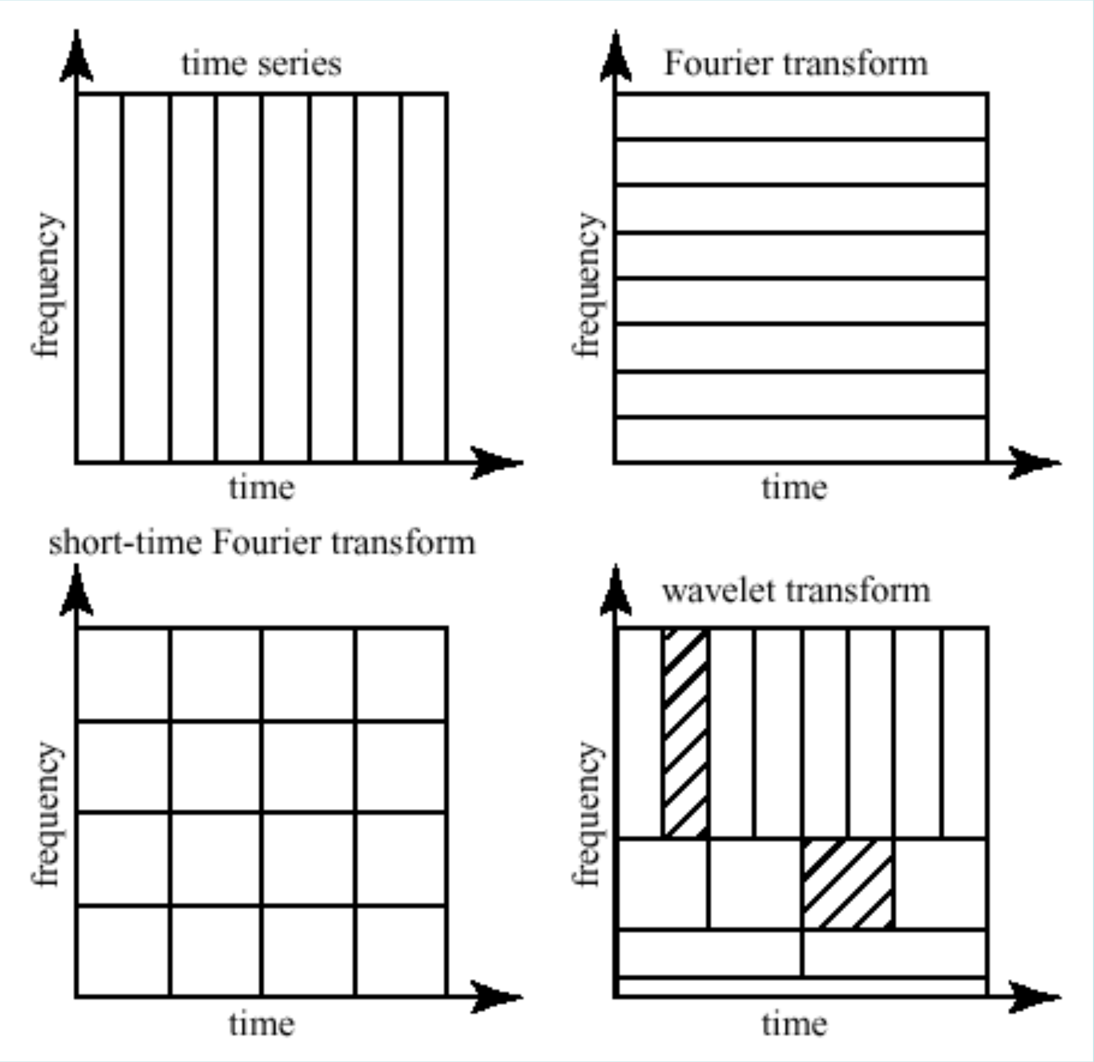
Black, grey, and white points correspond respectively to positive, zero and negative wavelet coefficients.

QUIZ: what is what?



From http://www.cerm.unifi.it/EUcourse2001/Gunther_lecturenotes.pdf, p.10

QUIZ: what is what?



From http://www.cerm.unifi.it/EUcourse2001/Gunther_lecturenotes.pdf, p.10

Bestiary of Multi-scale X-let 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

Major Breakthrough

Yves Meyer

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

(Bi-)Orthogonal Wavelet Transform (OWT)

$$f(x) = \sum_k c_J[k]\phi_{J,k}(x) + \sum_k \sum_{j=1}^{J} w_j[k]\psi_{j,k}(x)$$

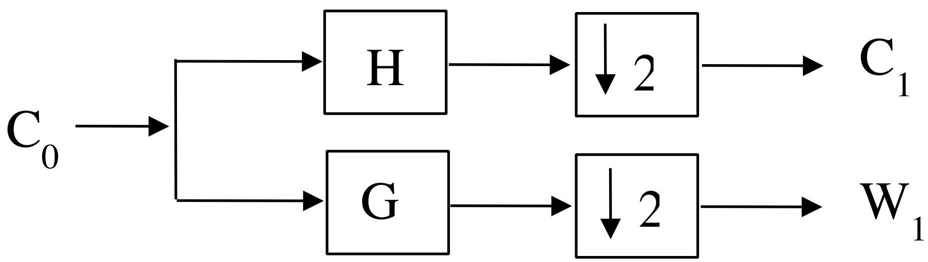
$$c_j[k] = \langle f(x), \phi_{j,k}(x)\rangle = \langle f(x), 2^{-j}\phi(2^{-j}x - k)\rangle$$

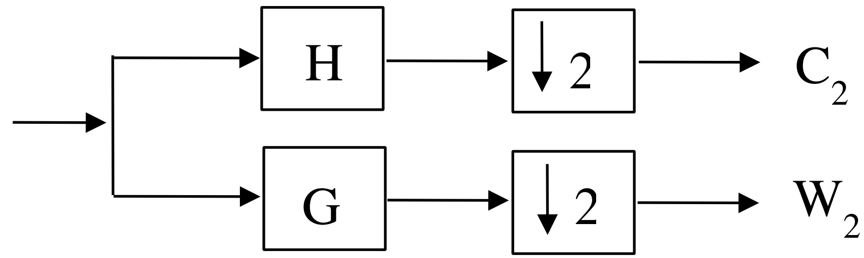
$$w_j[k] = \langle f(x), \psi_{j,k}(x)\rangle = \langle f(x), 2^{-j}\psi(2^{-j}x - k)\rangle$$

$$\tfrac{1}{2}\phi\left(\tfrac{t}{2}\right) = \sum_k h[k]\phi(t-k)$$

$$\tfrac{1}{2}\psi\left(\tfrac{t}{2}\right) = \sum_k g[k]\phi(t-k)$$

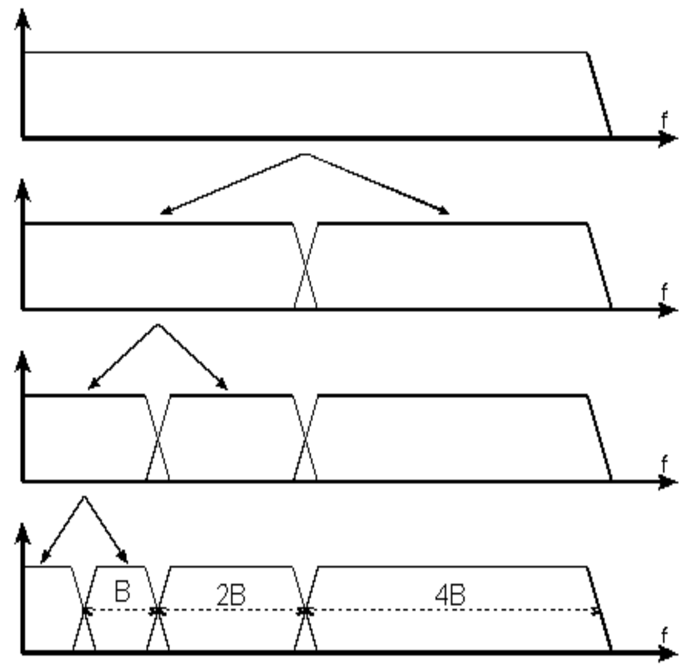
Transformation





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

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



Reconstruction:

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

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

Dealiasing + exact reconstruction conditions

Separable filters for 2D Wavelet Transform

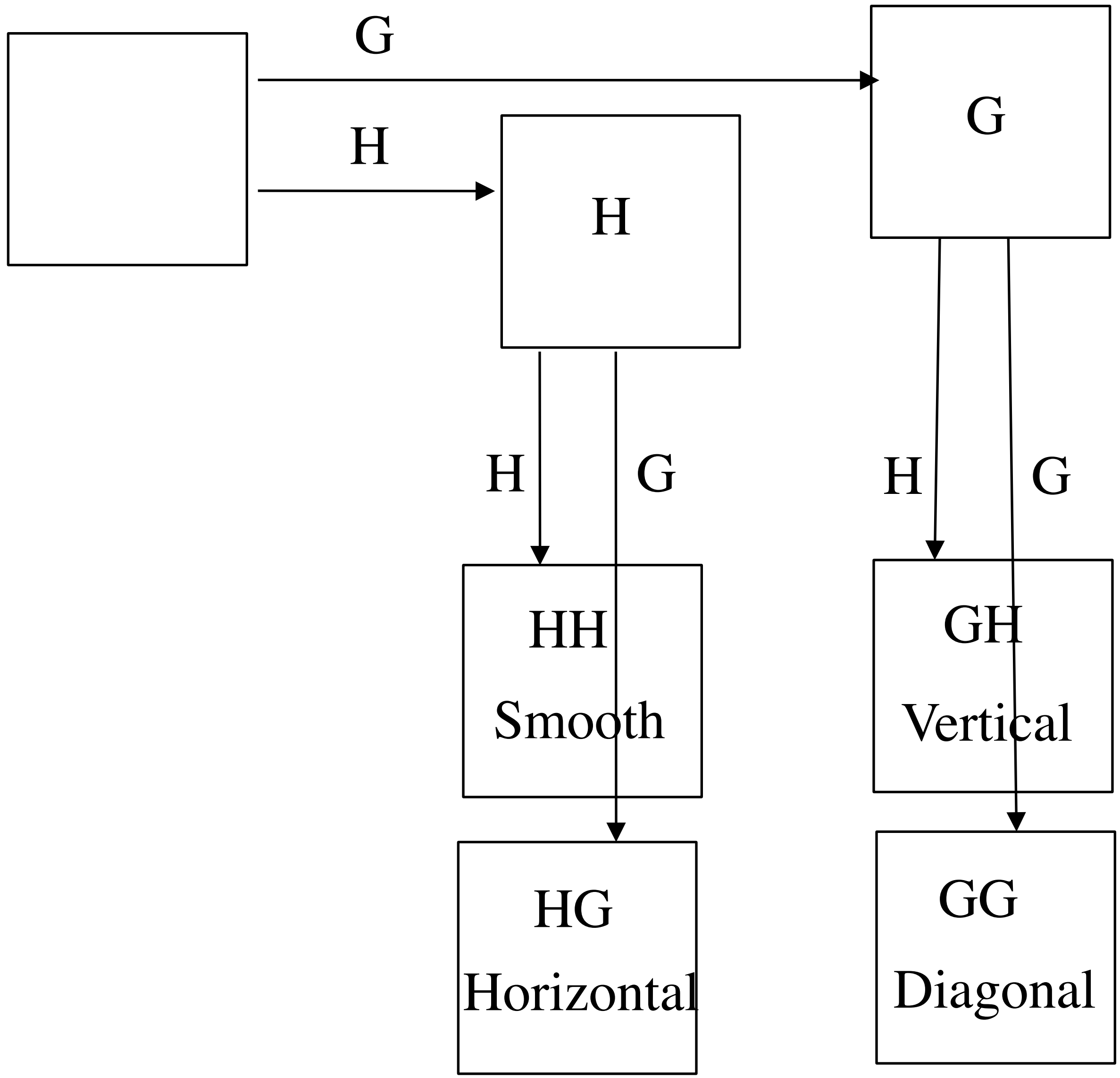
At two dimensions, we separate the variables x,y:

- vertical wavelet: $\psi^1(x,y) = \phi(x)\psi(y)$
- horizontal wavelet: $\psi^2(x,y) = \psi(x)\phi(y)$
- diagonal wavelet: $\psi^3(x,y) = \psi(x)\psi(y)$

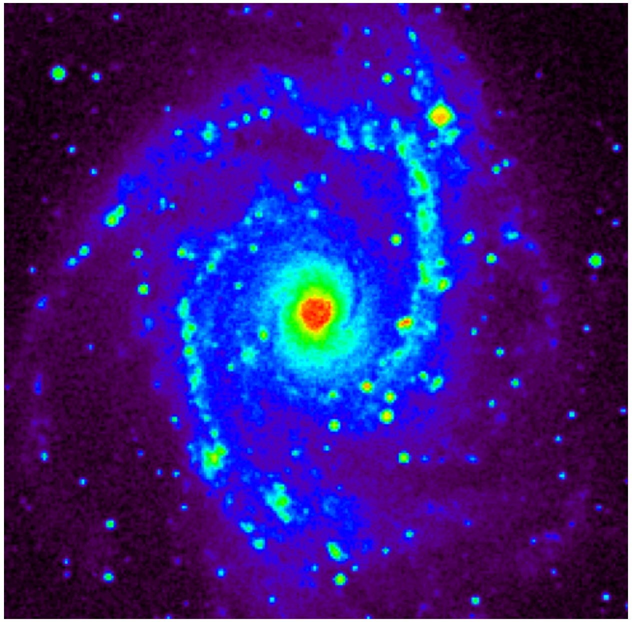
The detail signal is contained in three sub-images

$$
\begin{aligned}
w_j^1(k_x,k_y) &= \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} g(l_x - 2k_x)h(l_y - 2k_y)c_{j+1}(l_x,l_y) \\
w_j^2(k_x,k_y) &= \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} h(l_x - 2k_x)g(l_y - 2k_y)c_{j+1}(l_x,l_y) \\
w_j^3(k_x,k_y) &= \sum_{l_x=-\infty}^{+\infty} \sum_{l_y=-\infty}^{+\infty} g(l_x - 2k_x)g(l_y - 2k_y)c_{j+1}(l_x,l_y)
\end{aligned}
$$

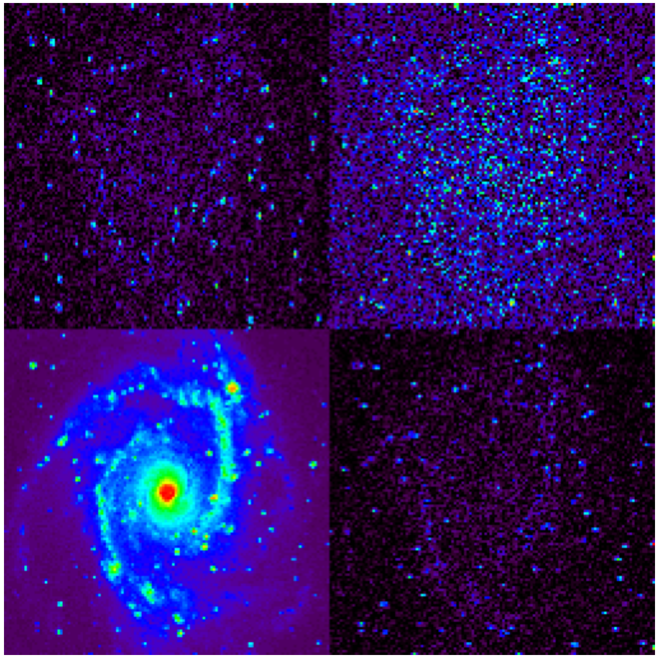
Schema for separable filter banks



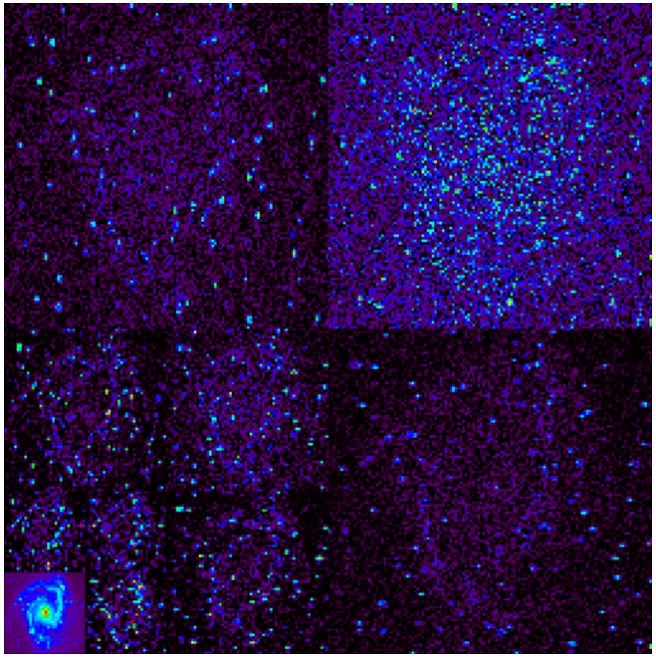
NGC2997



NGC2997 WT



NGC2997 WT



JPEG VS JPEG 2000

Original BMP

300x300x24

270056 bytes

JPEG 1:68

3983 bytes

JPEG2000 1:70

3876 bytes

Bestiary of MultiScale Transforms

Critical Sampling

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

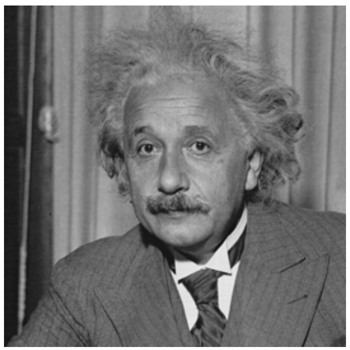
Redundant Transforms

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)

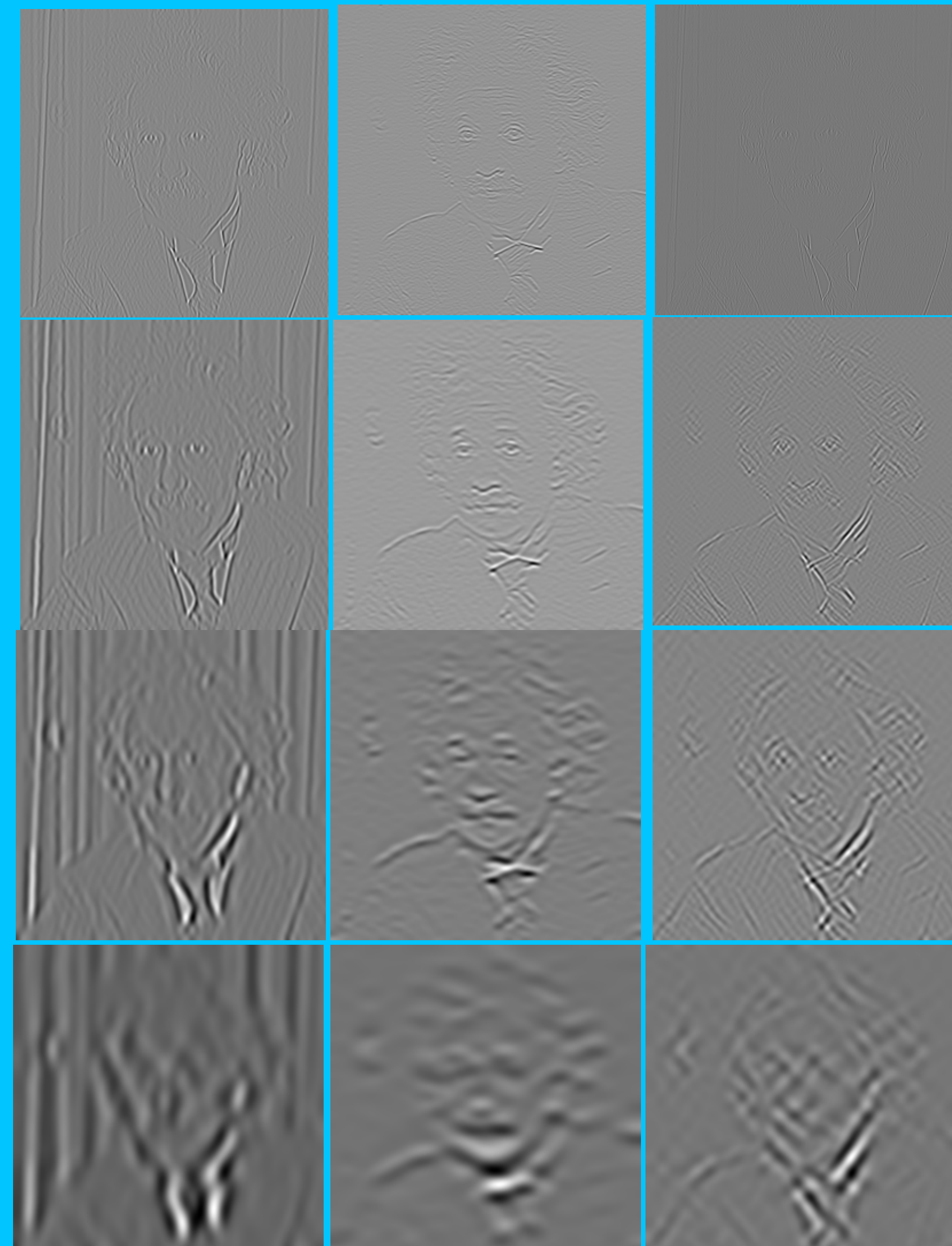
New Multiscale Construction

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

Ridgelet
Curvelet (Several implementations)
Wave Atom



Undecimated Wavelet Transform



Bestiary of 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 Multiscale Construction

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

Ridgelet
Curvelet (Several implementations)
Wave Atom

Isotropic Undecimated Wavelet Transform

- Separable wavelets do not represent well isotropic singularities
- Astronomical images: typically isotropic features (sources)
- Isotropic transform required in many practical applications in Astronomy
  - Diadic Scales.
  - Invariance per translation.

Scaling function and dilation equation:

$$\frac{1}{4}\phi\left(\frac{x}{2},\frac{y}{2}\right) = \sum_{l,k} h(l,k)\phi(x-l,y-k)$$

Wavelet function decomposition:

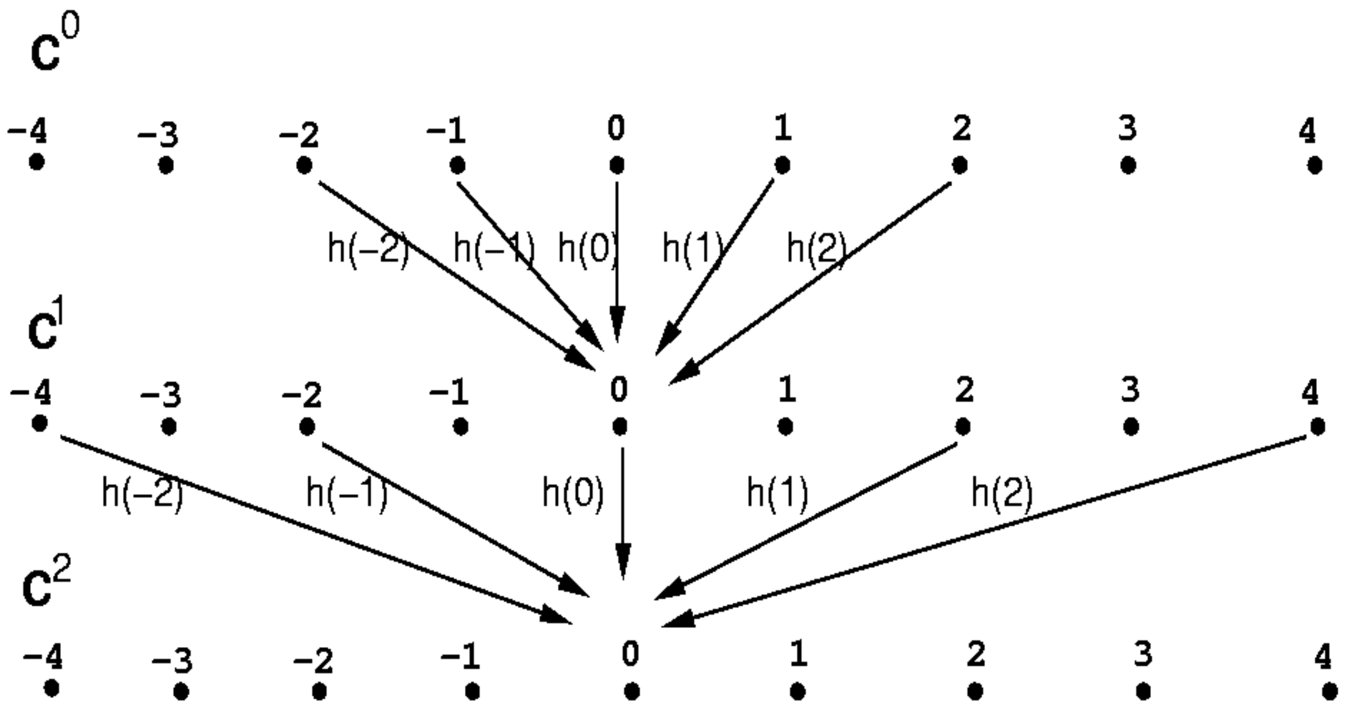
$$\frac{1}{4}\psi\left(\frac{x}{2},\frac{y}{2}\right) = \sum_{l,k} g(l,k)\phi(x-l,y-k)$$

A trous wavelet transform:

$$c_j(x,y) = \langle f(x,y), \frac{1}{4^j}\phi\left(\frac{x-l}{2^j},\frac{y-k}{2^j}\right)\rangle$$

“A trous” algorithm: scale coefficients

Passage from c_0 to c_1, and from c_1 to c_2



"A trous" algorithm: Wavelet Coefficients

Generally, the wavelet resulting from the difference between two successive approximations is applied:

$$w_{j+1,k} = c_{j,k} - c_{j+1,k}$$

The associated wavelet is $\psi(x)$.

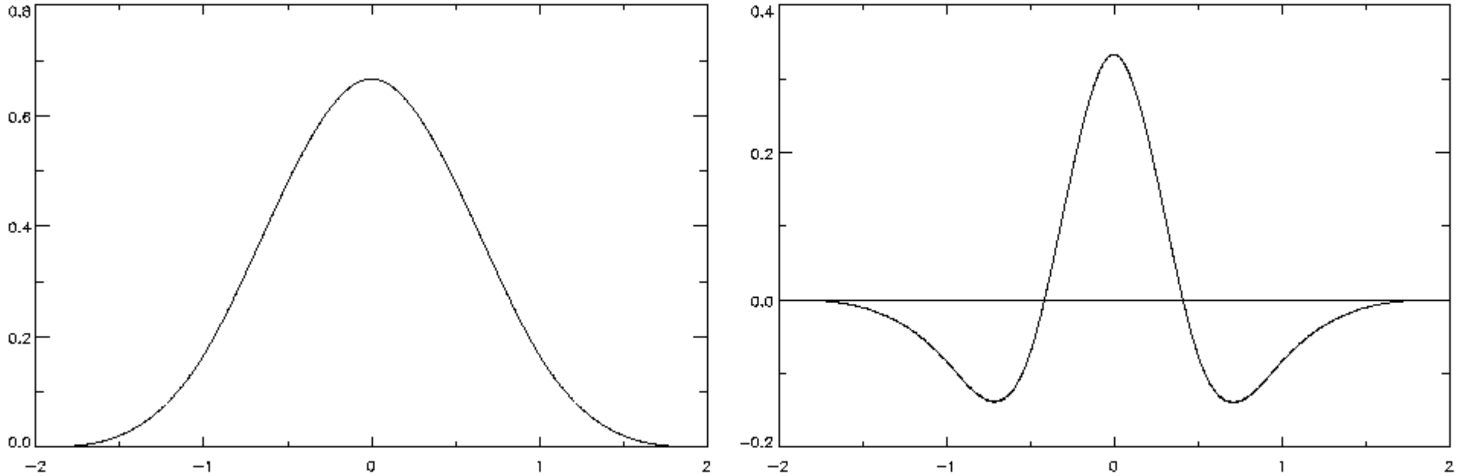
$$\frac{1}{2}\psi(\frac{x}{2}) = \phi(x) - \frac{1}{2}\phi(\frac{x}{2})$$

The reconstruction algorithm is immediate:

$$c_{0,k} = c_{J,k} + \sum_{j=1}^{J} w_{j,k}$$

Nearly Isotropic Separable Filter Bank (1)

$$
\begin{aligned}
B_3(x) &= \frac{1}{12}(\mid x-2\mid^3 - 4\mid x-1\mid^3 + 6\mid x\mid^3 - 4\mid x+1\mid^3 + \mid x+2\mid^3) \\
\phi(x,y) &= B_3(x)B_3(y) \\
\frac{1}{4}\psi(\frac{x}{2},\frac{y}{2}) &= \phi(x,y) - \frac{1}{4}\phi(\frac{x}{2},\frac{y}{2})
\end{aligned}
$$

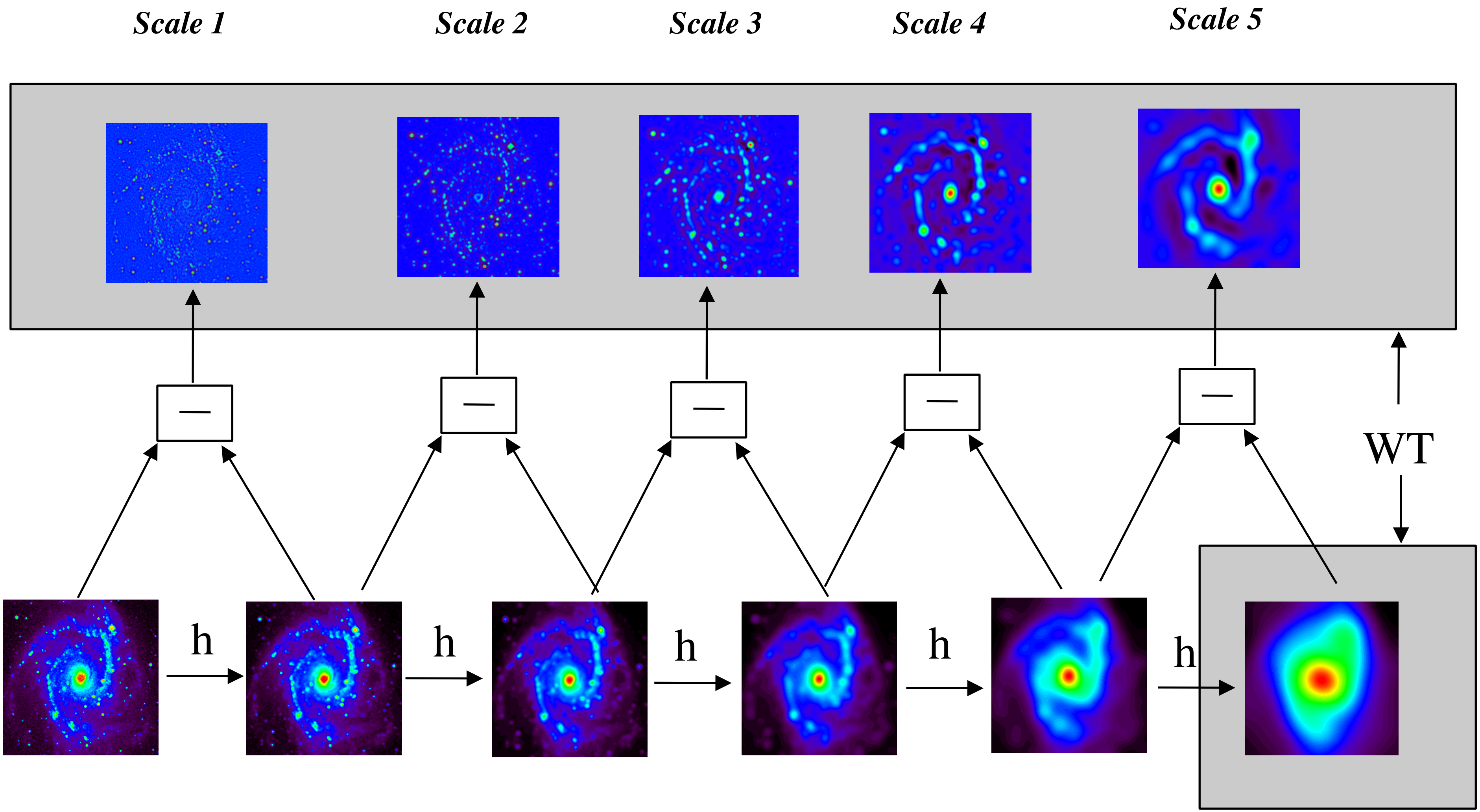


Nearly Isotropic Separable Filter Bank (2)

$$\begin{pmatrix} \frac{1}{16} & \frac{1}{4} & \frac{3}{8} & \frac{1}{4} & \frac{1}{16} \end{pmatrix} \otimes \begin{pmatrix} 1/16 \\ 1/4 \\ 3/8 \\ 1/4 \\ 1/16 \end{pmatrix} = \begin{pmatrix} \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{3}{128} & \frac{3}{32} & \frac{9}{64} & \frac{3}{32} & \frac{3}{128} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \end{pmatrix}$$

In the 2-dimensional case, we assume the separability, which leads to a row-by-row convolution with $(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$; followed by column-by-column convolution.

Isotropic Undecimated Wavelet Transform

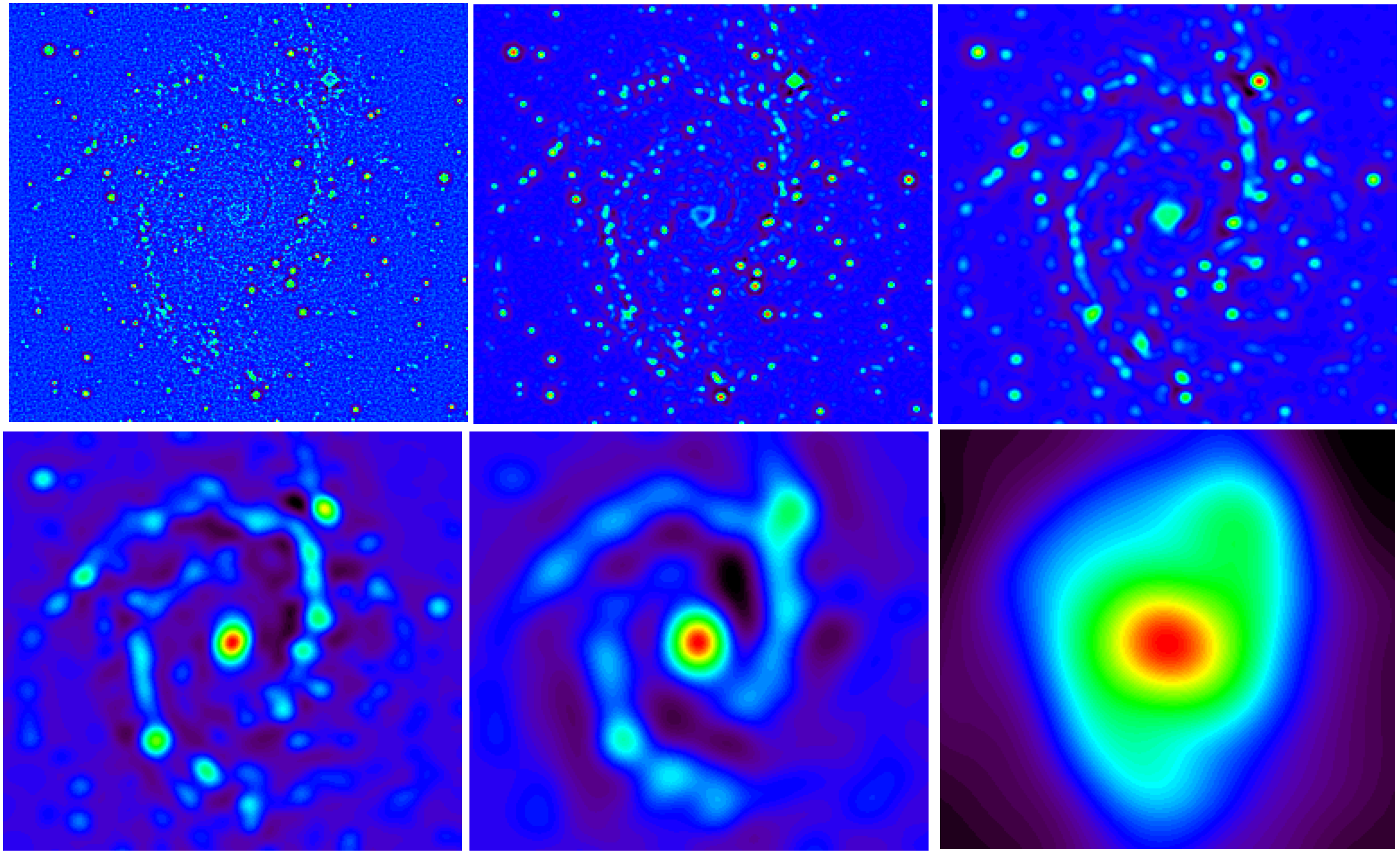


Starlet Transform (a trous algorithm)

ϕ: B_3-spline, $\frac{1}{2}\psi(\frac{x}{2}) = \frac{1}{2}\phi(\frac{x}{2}) - \phi(x)$

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

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



Bestiary of 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

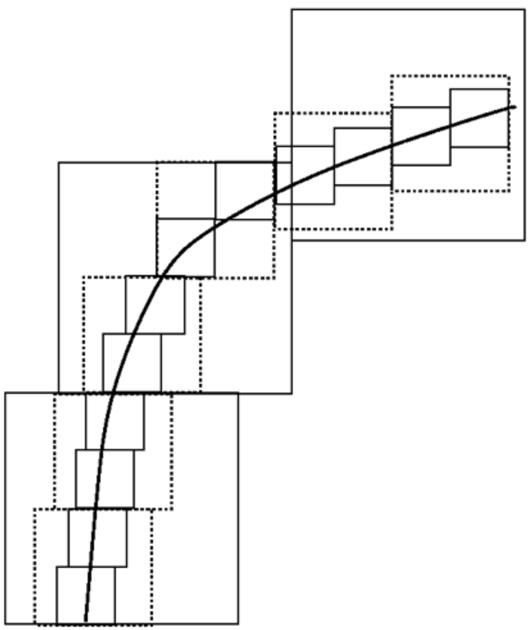
Contourlet
Bandelet
Finite Ridgelet Transform
Platelet
(W-)Edgelet
Adaptive Wavelet

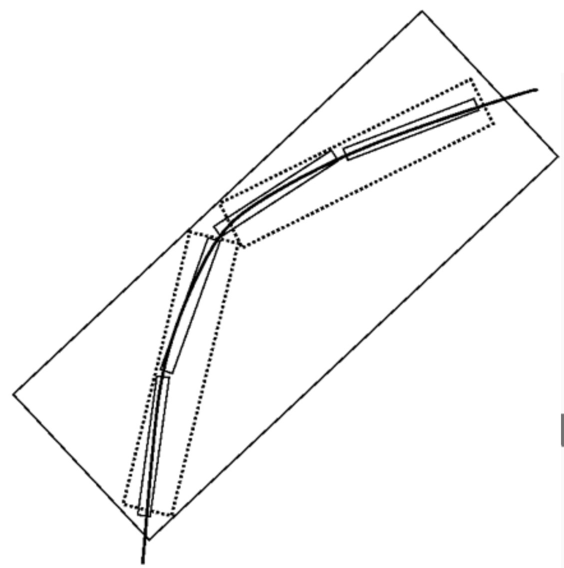
Ridgelet
Curvelet (Several implementations)
Wave Atom

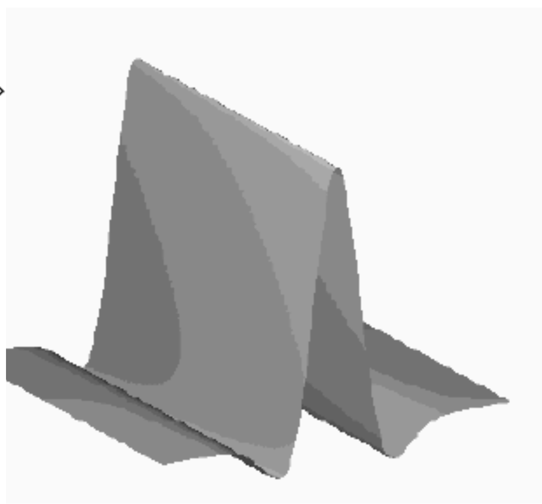
Wavelet and Edges

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

- need dictionaries of strongly anisotropic atoms :







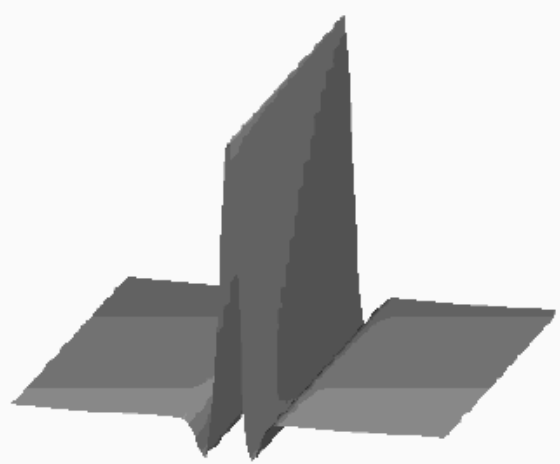
ridgelets, curvelets, contourlets, bandelettes, etc.

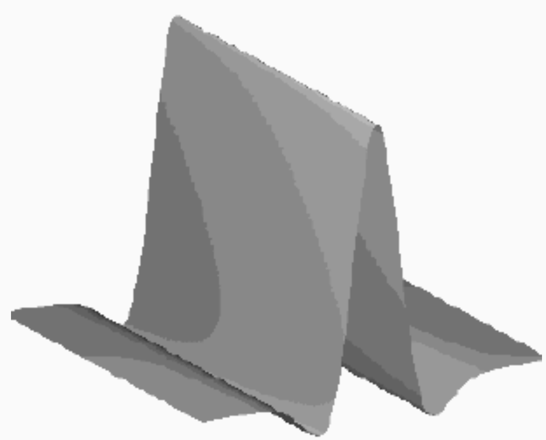
Continuous Ridgelet Transform

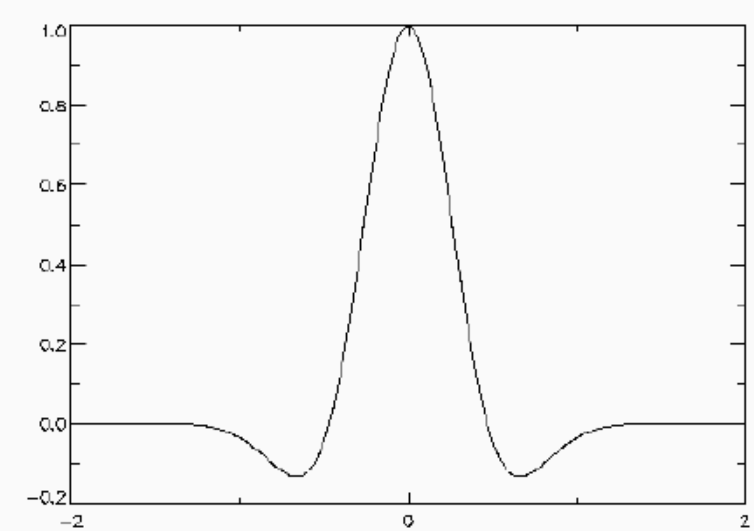
Ridgelet Transform (Candes, 1998): $R_f(a,b,\theta) = \int \psi_{a,b,\theta}(x) f(x) \mathrm{d}x$

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

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





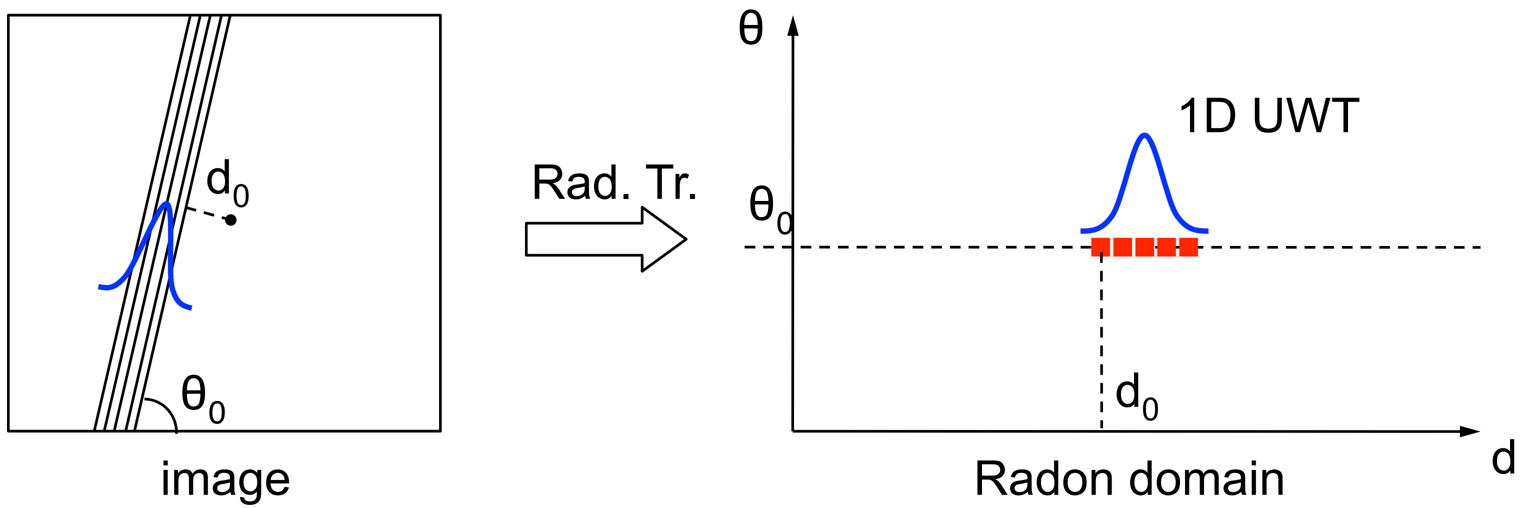


Ridgelets via Radon Transform

The ridgelet coefficients of an object f can be obtained using the Radon transform as:

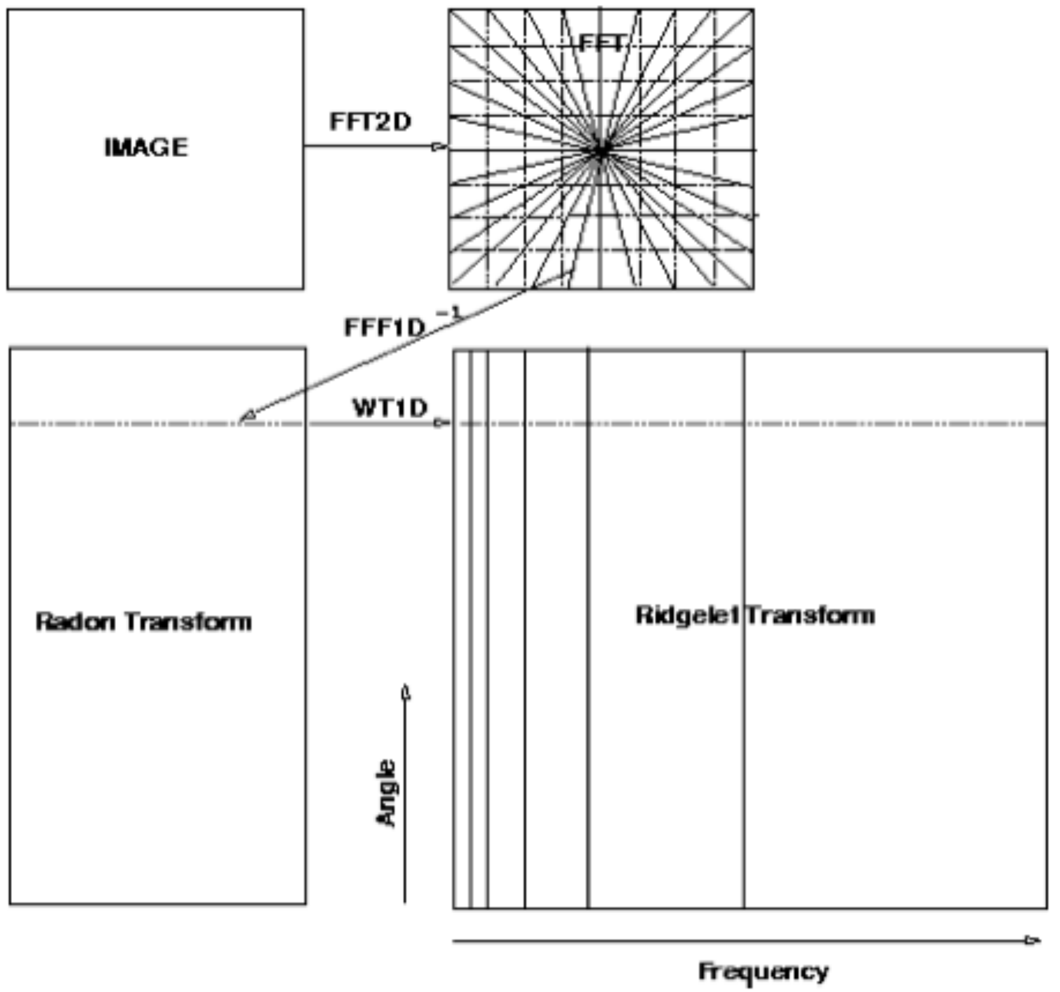
$$R_f(a, b, \theta) = \int \psi \left(\frac{t - b}{a} \right) Rf(\theta, t) \mathrm{d}t$$

Ridgelet transform: Radon + 1D Wavelet



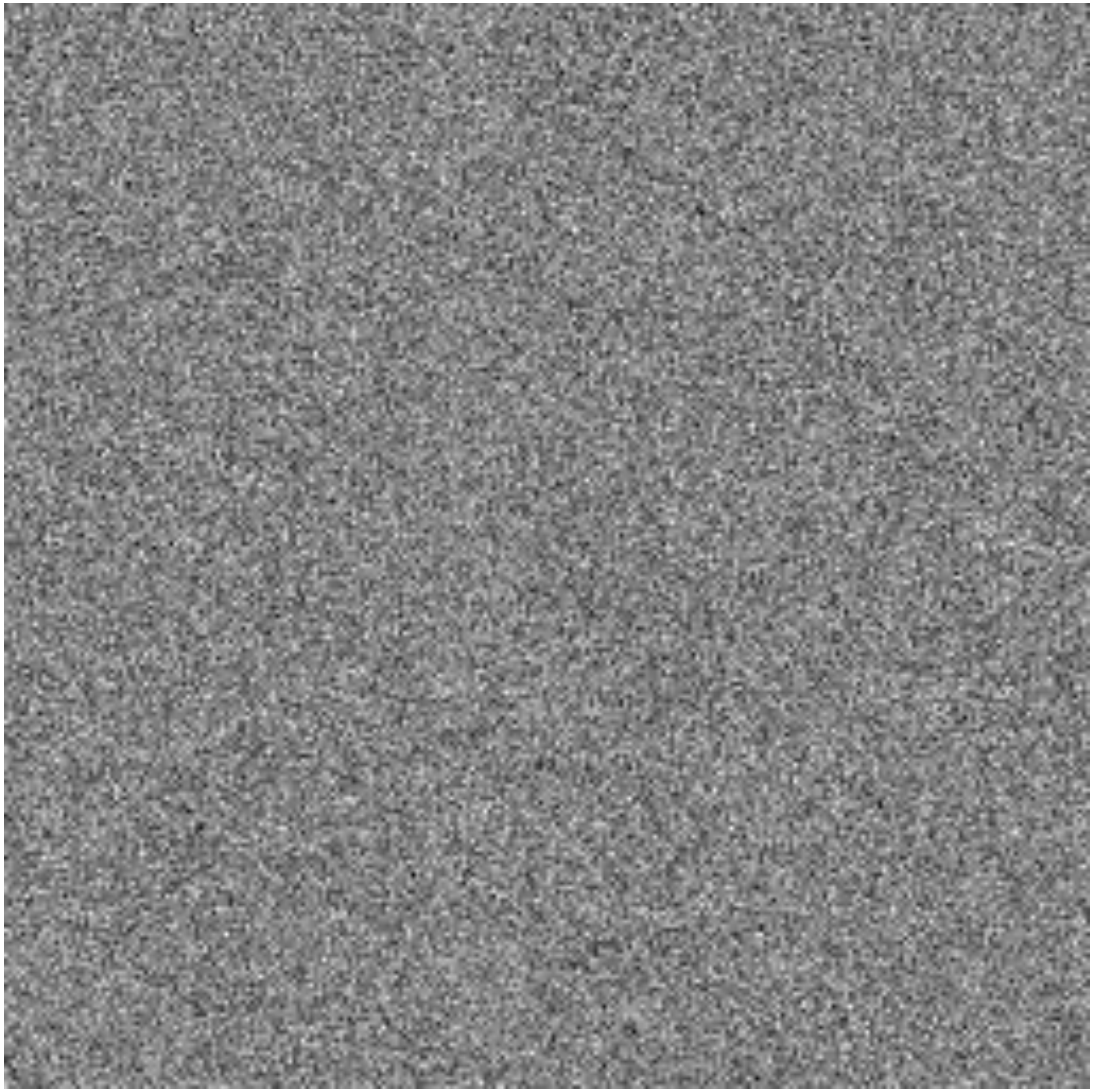
Implementation of Ridgelets

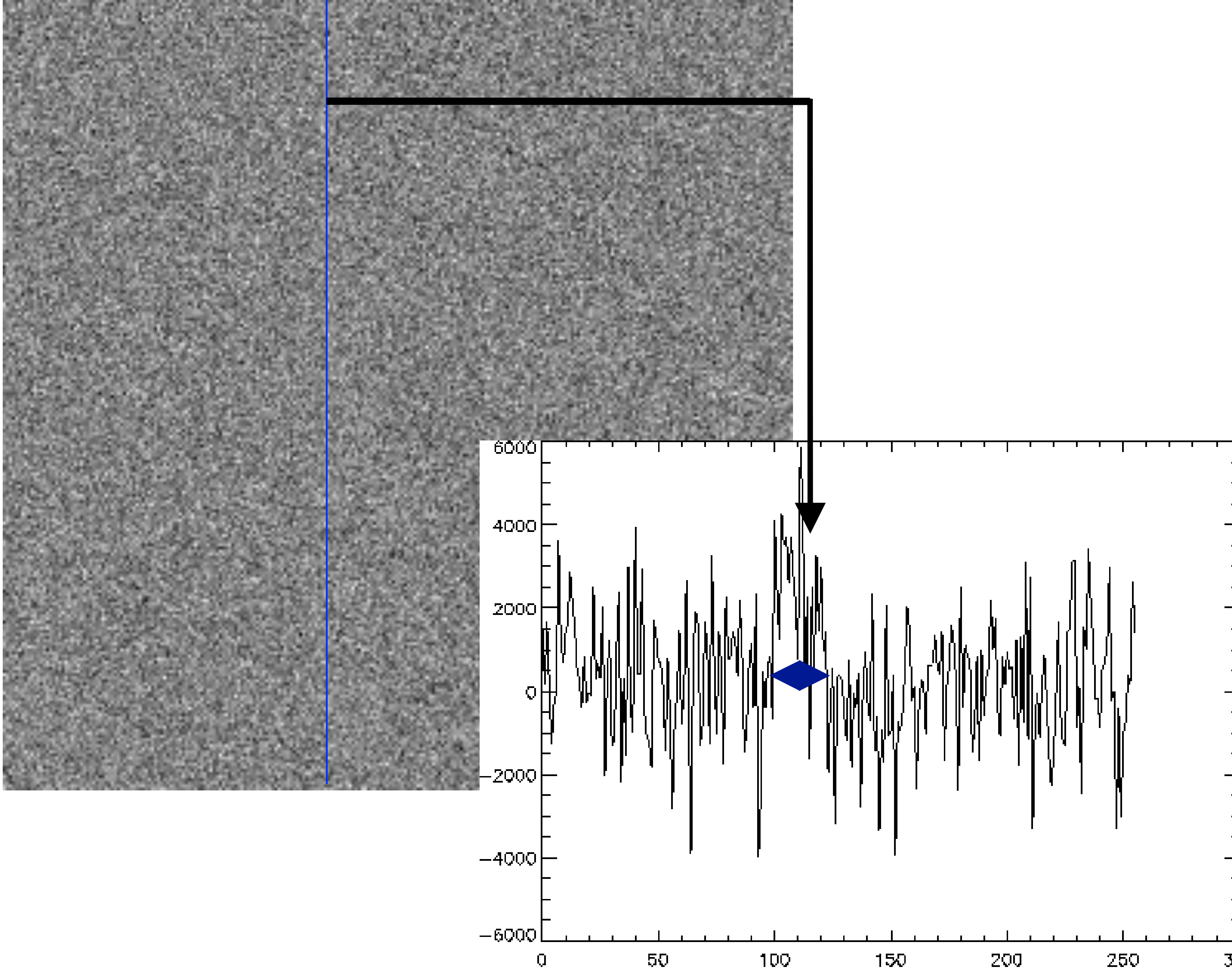
Fourier slice theorem:
2D Fourier transform = 1D Fourier transform of a slice of Radon transform



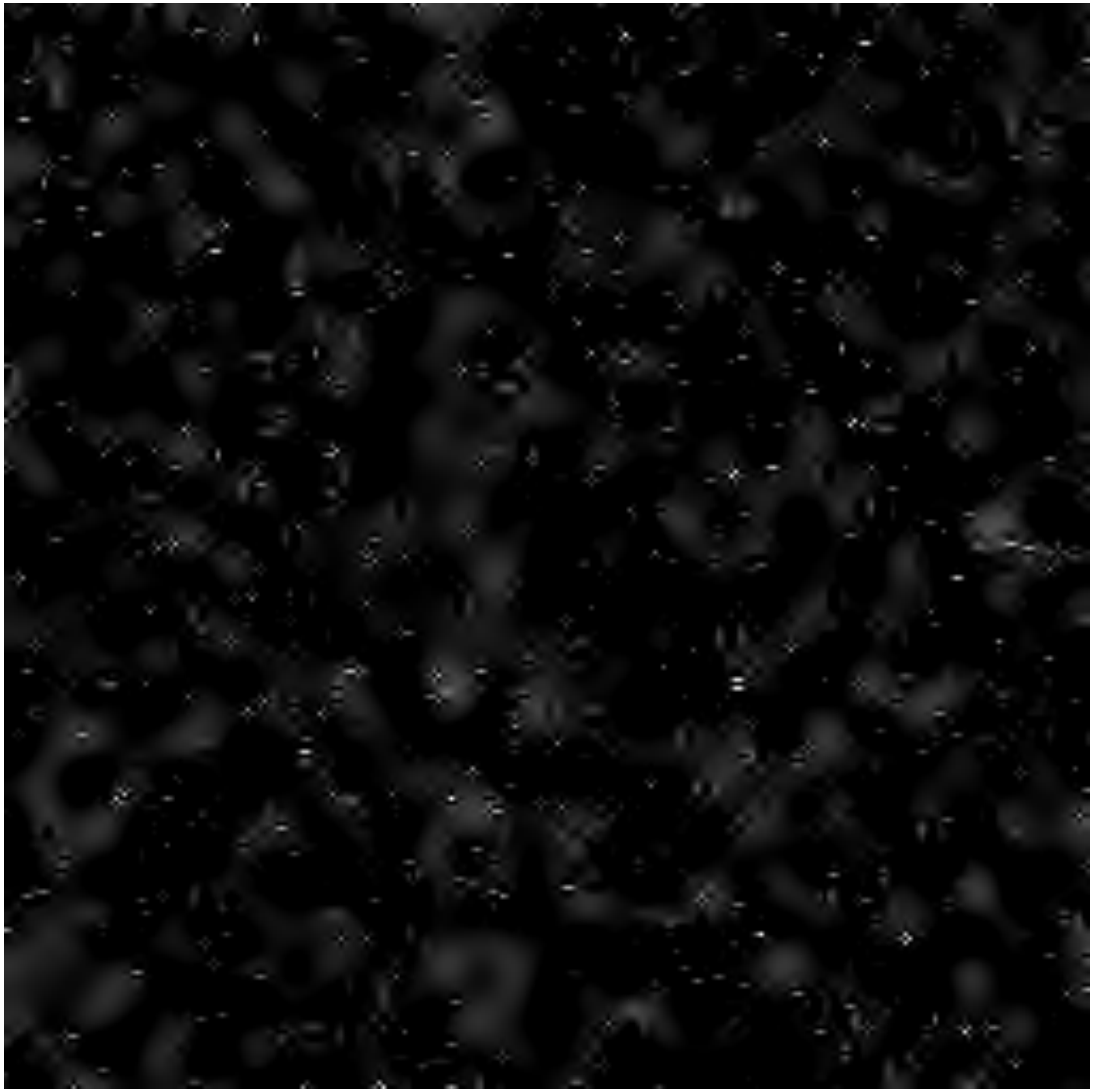
Example of denoising with Ridgelets

SNR = 0.1

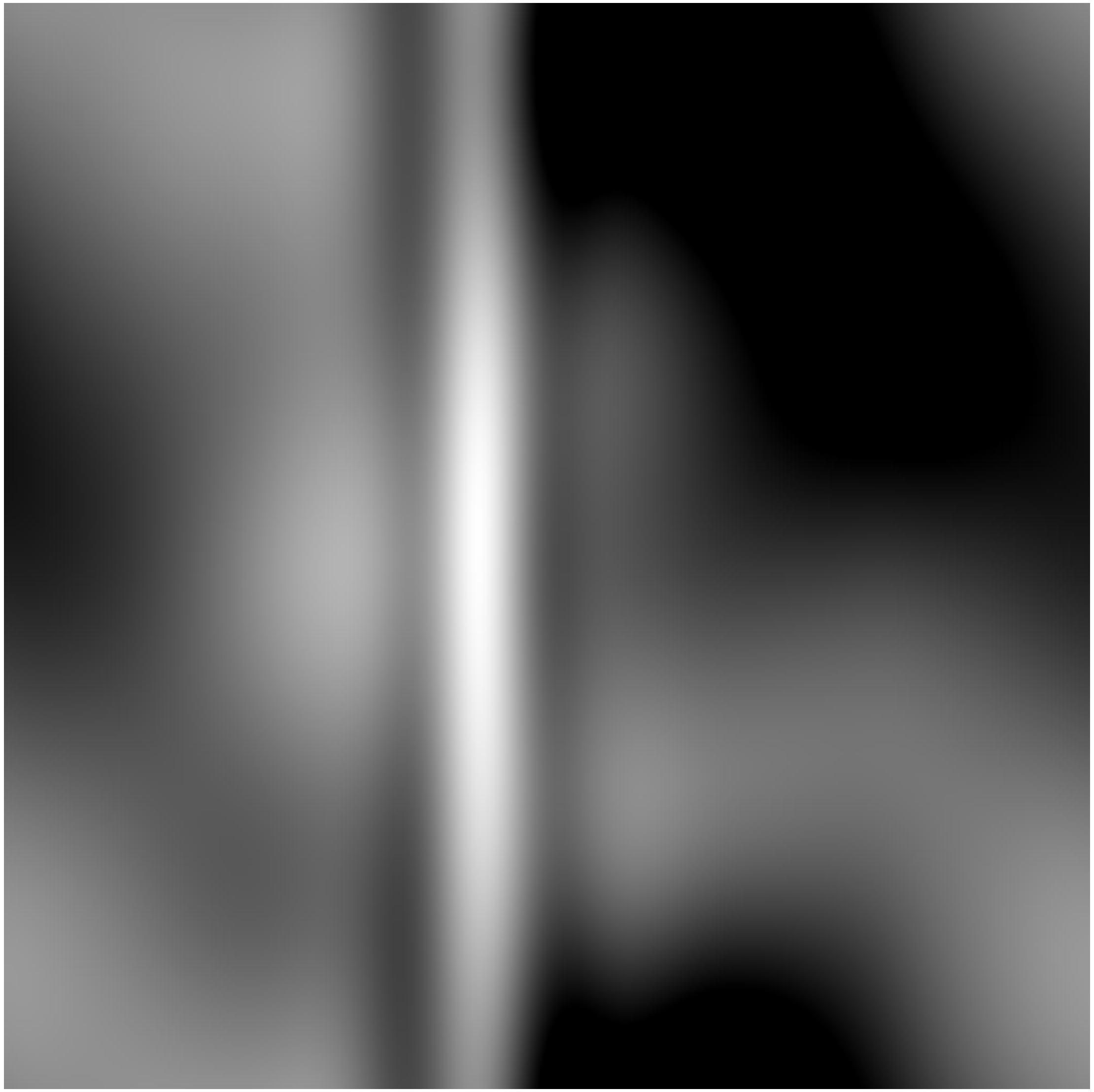



6000
4000
2000
0
−2000
−4000
−6000
0
50
100
150
200
250
300

Undecimated Wavelet Filtering (3 sigma)



Ridgelet Filtering (5 sigma)



Bestiary of 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)

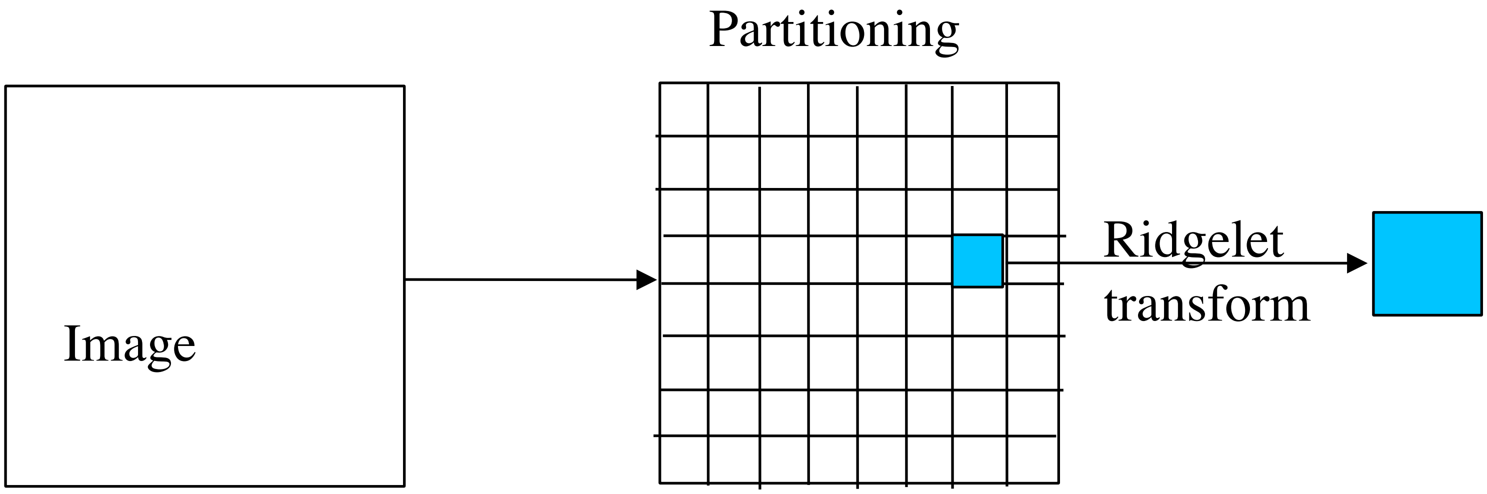
New Multiscale Construction

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

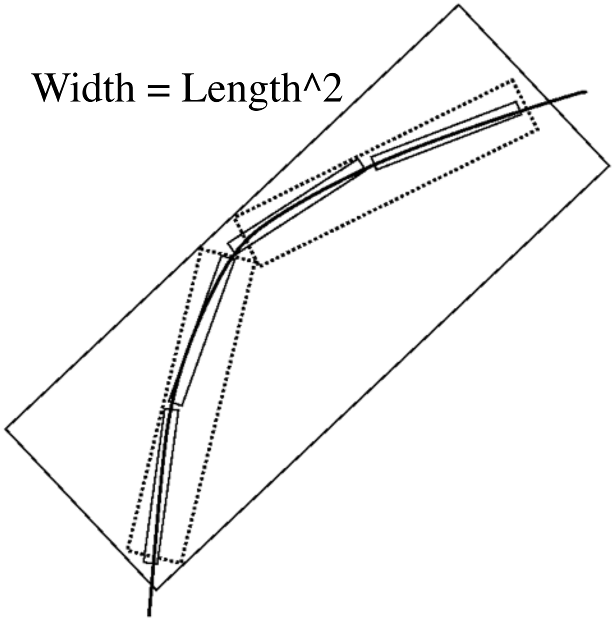
Ridgelet
Curvelet (Several implementations)
Wave Atom

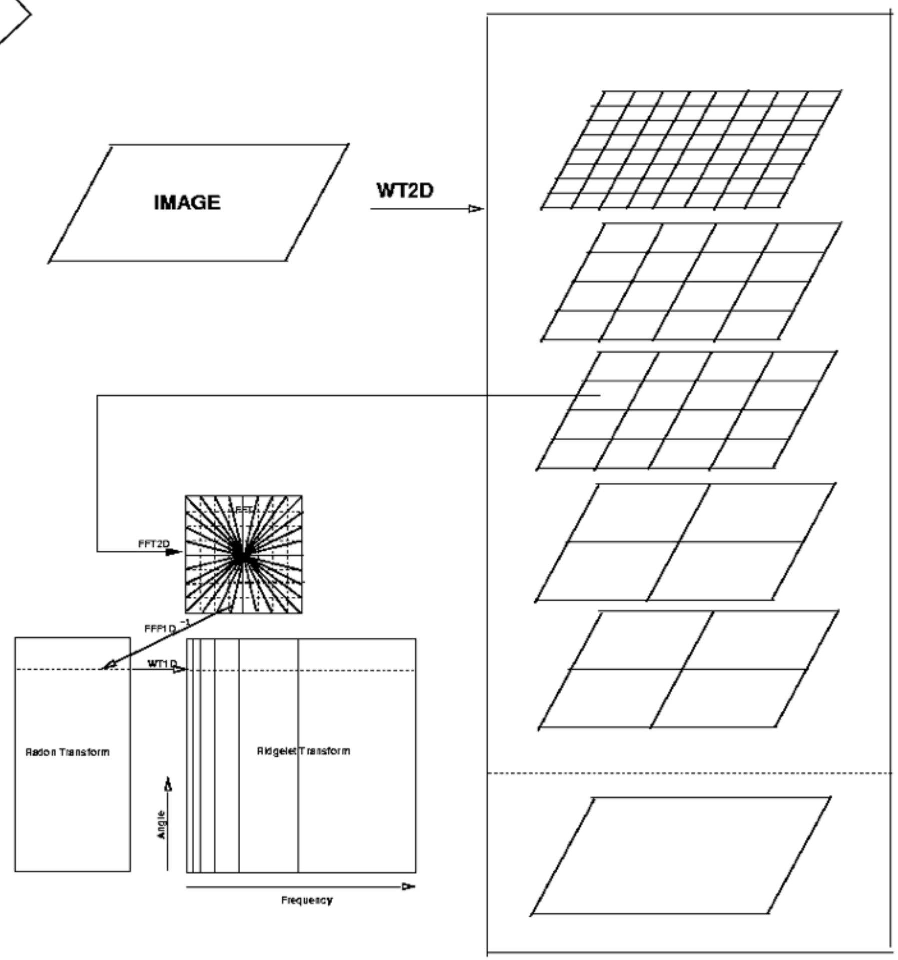
The Local Ridgelet Transform

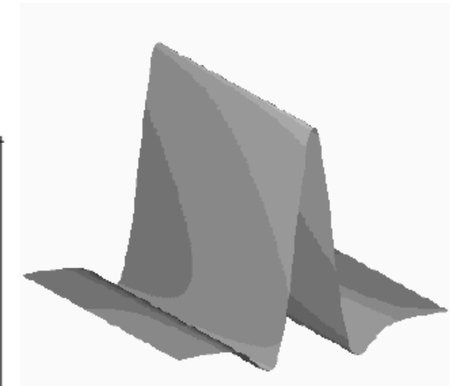
Ridgelet transform suited to represent sparsely lines of the size of the image. For line segments, need a partitioning: image decomposed in blocks, and a ridgelet transform performed per block



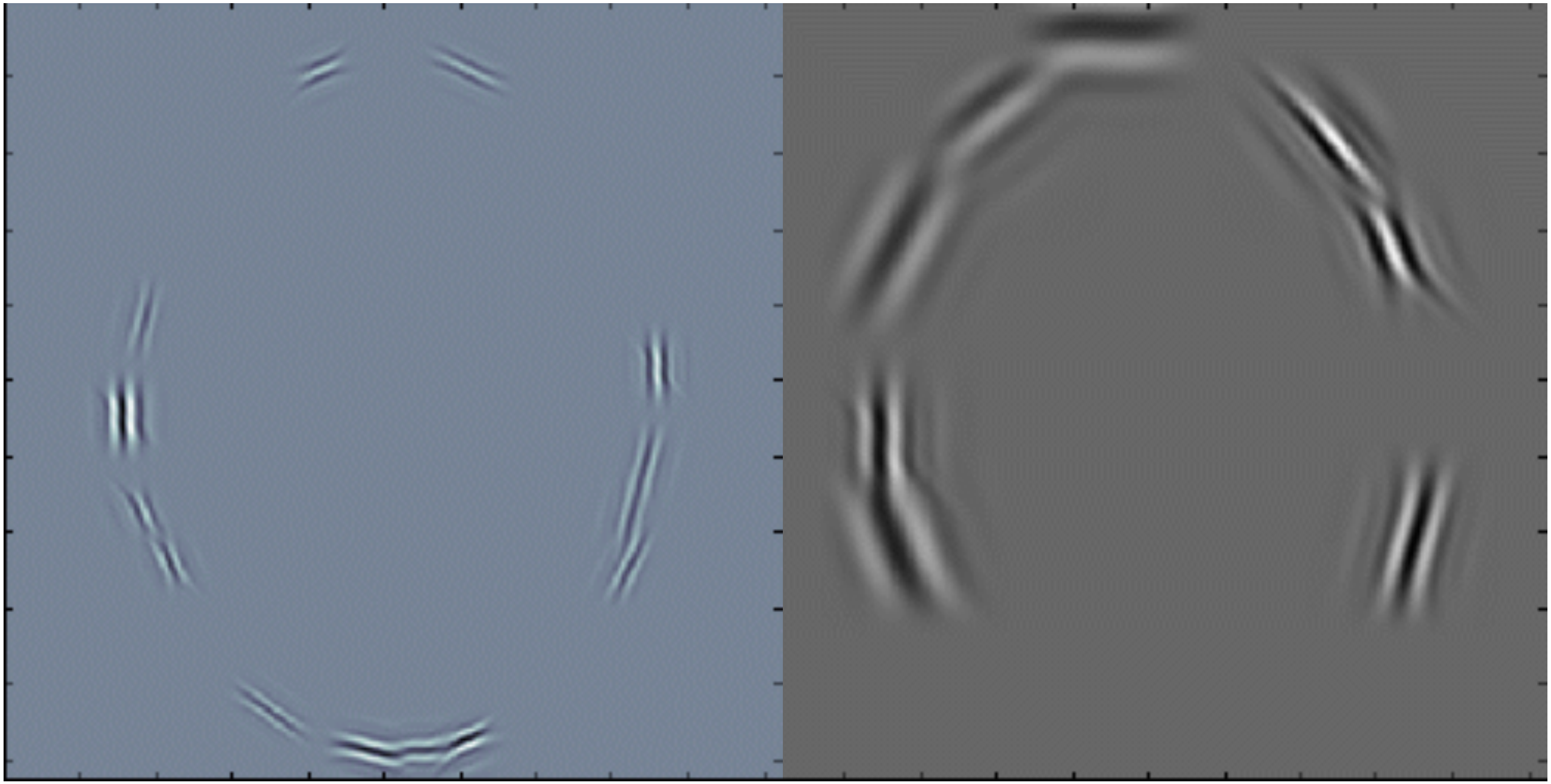
The Curvelet Transform (CUR01)







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



•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 via the Curvelet Transform

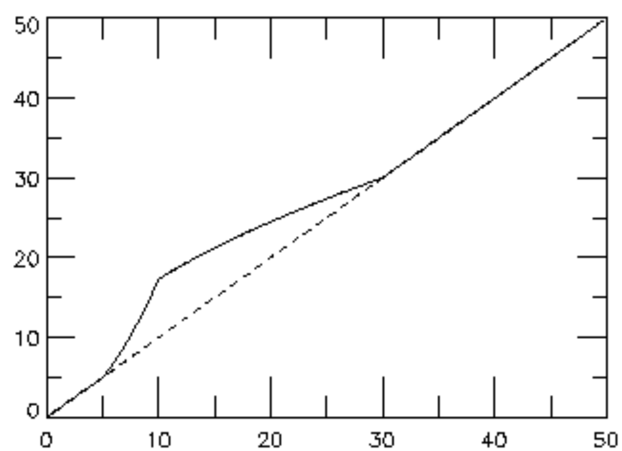
J.-L Starck, F. Murtagh, E. Candes and D.L. Donoho, ***"Gray and Color Image Contrast Enhancement by the Curvelet Transform"***,

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

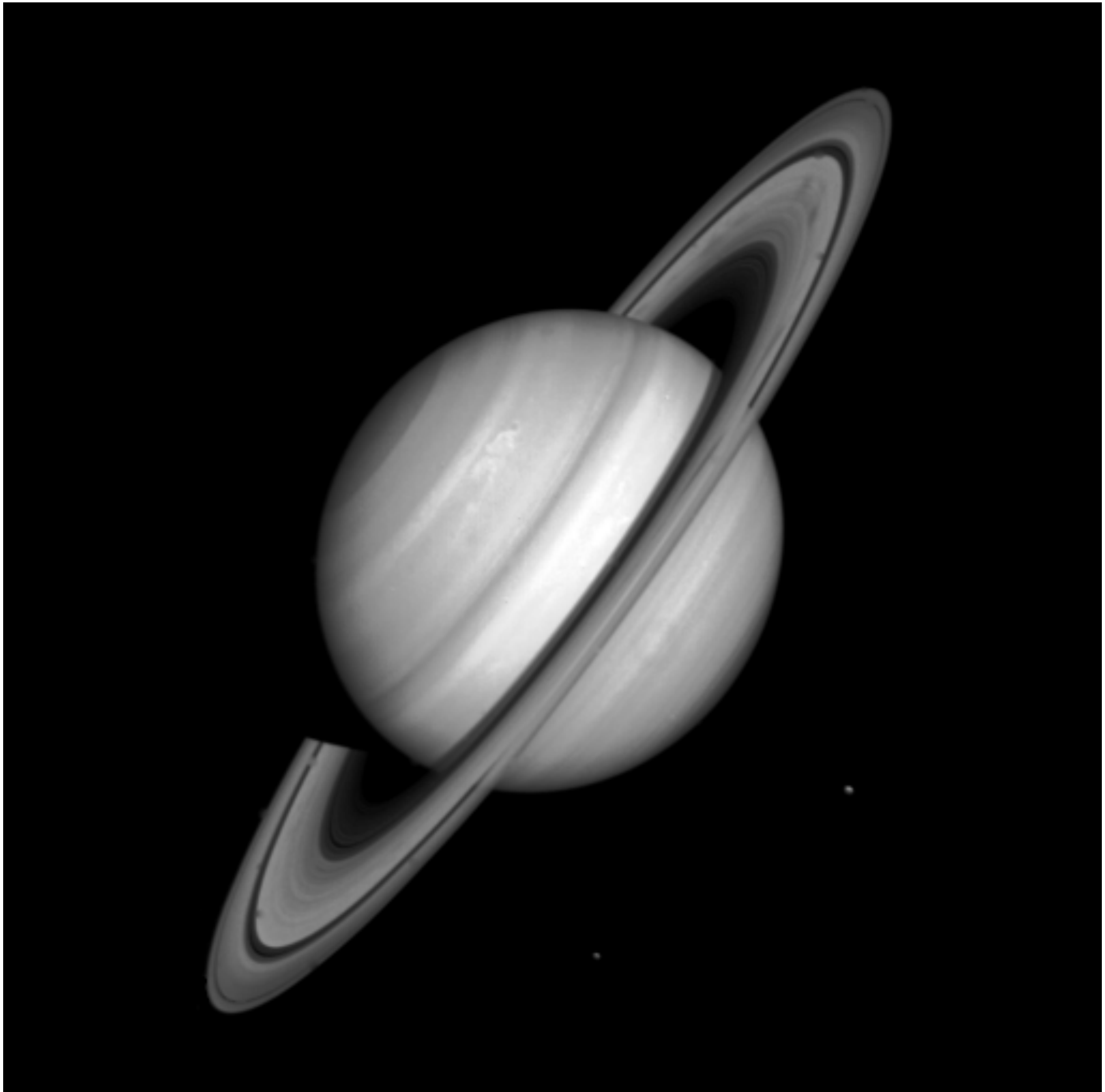
$$\tilde{I} = C_R\left(y_c\left(C_T I\right)\right)$$

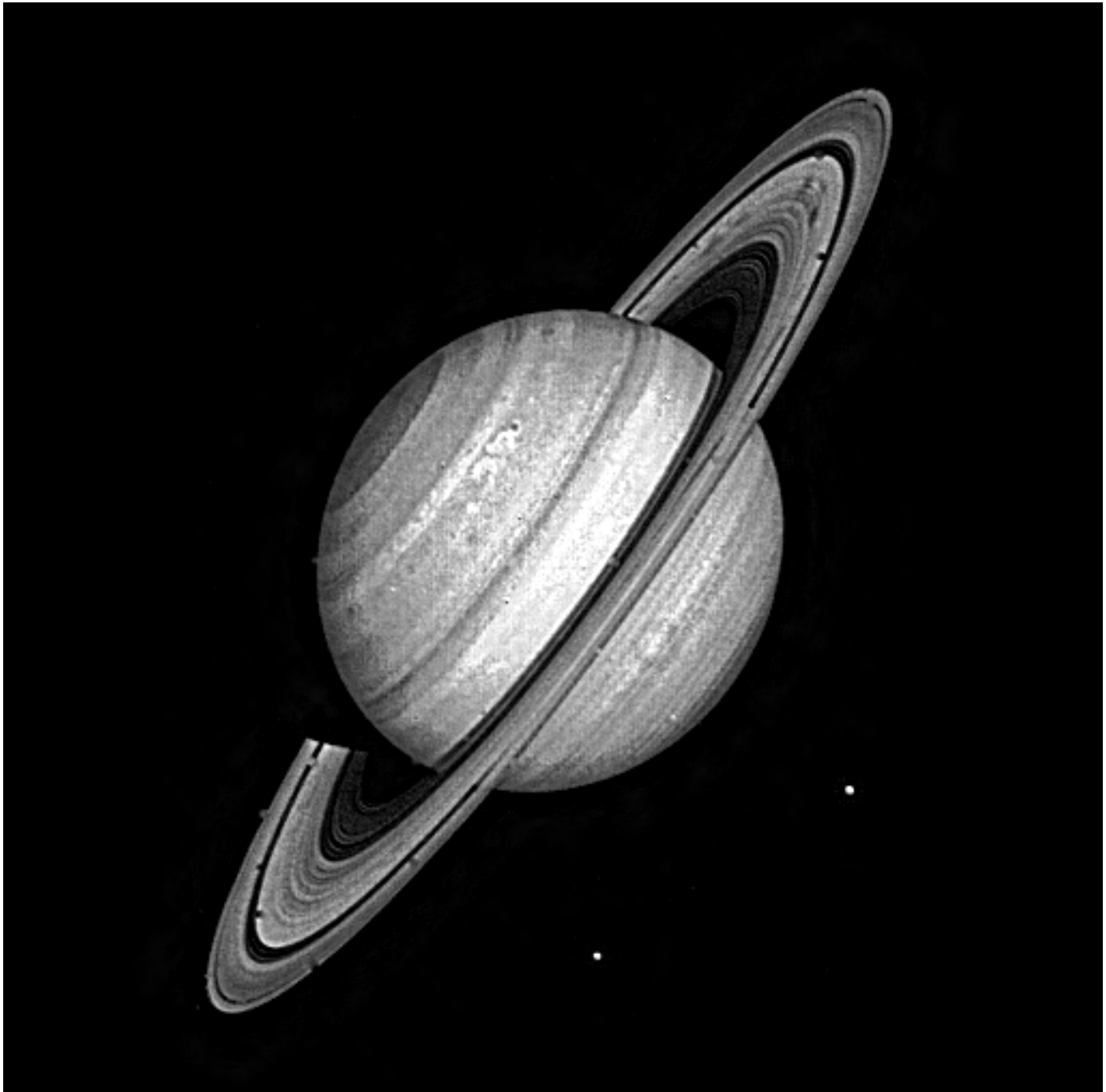
$$\begin{cases} y_c(x,\sigma) = 1 & \text{if} \quad x < c\sigma \\ y_c(x,\sigma) = \dfrac{x - c\sigma}{c\sigma}\left(\dfrac{m}{c\sigma}\right)^p + \dfrac{2c\sigma - x}{c\sigma} & \text{if} \quad x < 2c\sigma \\ y_c(x,\sigma) = \left(\dfrac{m}{x}\right)^p & \text{if} \quad 2c\sigma \le x < m \\ y_c(x,\sigma)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

Modified curvelet coefficient



Curvelet coefficient





Some elements to answer to choice of dictionary

Sparsity Model 1: we consider a dictionary which also have a fast transform and reconstruction operator:

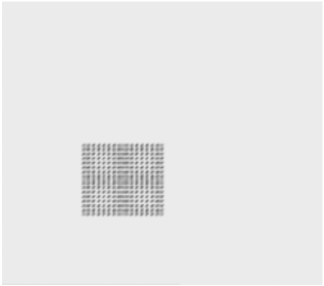
$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

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

<u>Local DCT</u>

Stationary textures

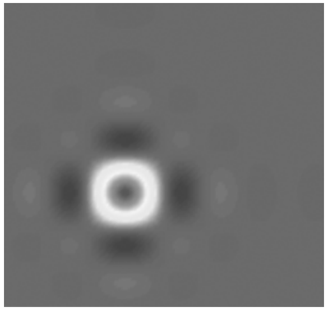
Locally oscillatory



<u>Wavelet transform</u>

Piecewise smooth

Isotropic structures



<u>Curvelet transform</u>

Piecewise smooth, edge

Outline

1. Sparsity and Multi-Scale Representations

2. Sparsity and Inverse Problems

3. Deep Generative models for Inverse problems

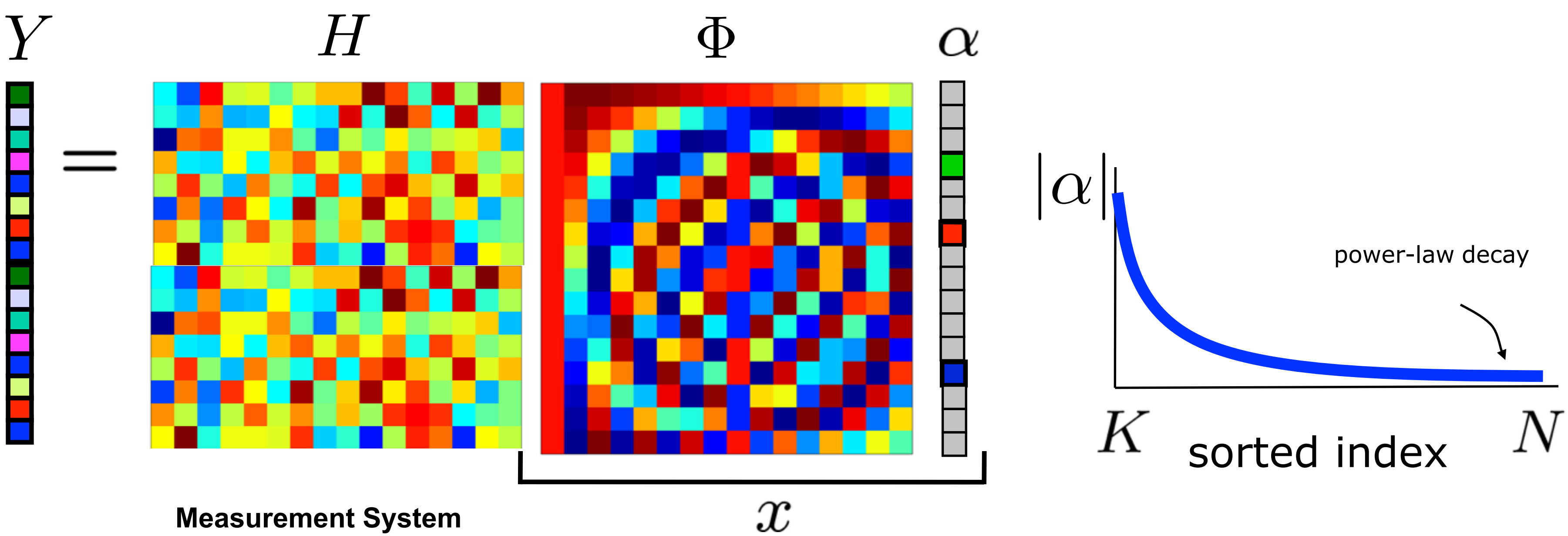
Inverse Problem tour and Sparse Recovery

$$Y = HX + N$$

$X = \Phi\alpha$, and α is sparse
(or compressible)

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

$$\min_{\alpha} \|\alpha\|_p \quad \text{subject to} \quad \|Y - H\Phi\alpha\|_2^2 \leq \epsilon$$



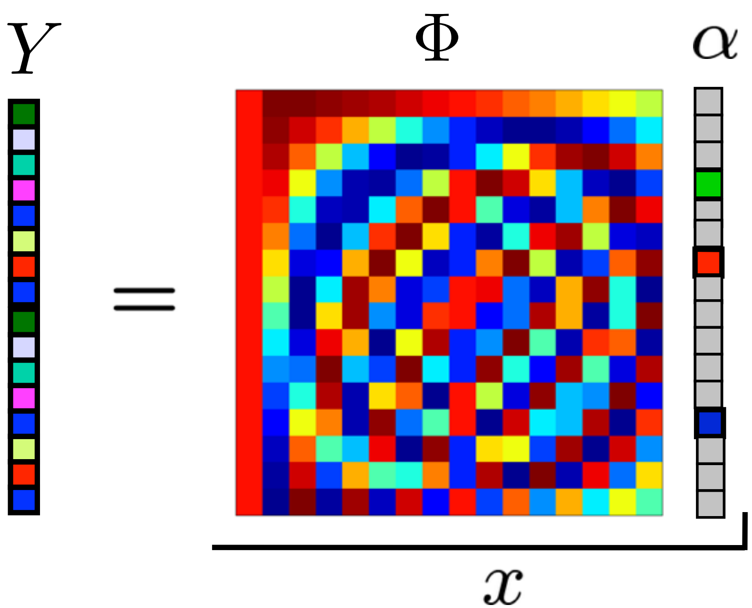
Inverse Problem tour and Sparse Recovery

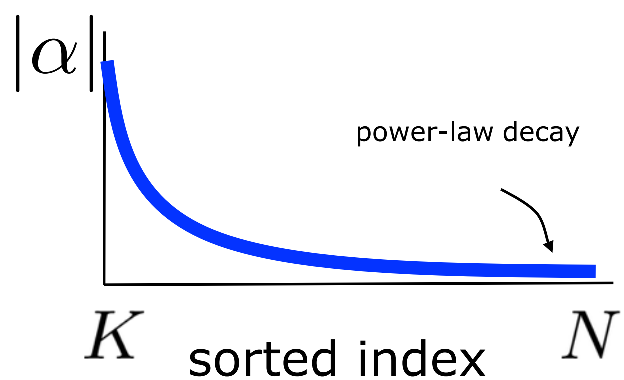
$$Y = X + N$$

$X = \Phi\alpha$, and α is sparse (or compressible)

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

$$\tilde{\alpha} \in \arg\min_{\alpha} \|Y - \Phi\alpha\|_2^2 + t\|\alpha\|_p, \qquad 0 \leq \mathrm{p} \leq 1$$





Algorithm for the Inverse Problem with p=0

$$\tilde{\alpha} \in \arg\min_{\alpha} \|Y - \Phi\alpha\|_2^2 + t^2\|\alpha\|_0$$

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.

Algorithm solving the Inverse Problem for p=1

$$\tilde{\alpha} \in \arg\min_{\alpha} \|Y - \Phi\alpha\|_2^2 + t\|\alpha\|_1$$

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.

What value for the threshold to choose?

Heuristic: Look at the orthonormal case, or at the fixed point criterion (ie at convergence)

For IHT, at convergence:

$$\alpha^* = \mathrm{HardTresh}_{\mu t}\left(\alpha^* + \mu\phi^T\left(Y - \phi\alpha^*\right)\right)$$

$$\alpha^* = \mathrm{HardTresh}_{\mu t}\left(\alpha^* + \mu\phi^T N^*\right)$$

$$\text{Only Noise: } \alpha^* = \mathrm{HardTresh}_{t}\left(\phi^T N^*\right) \quad \leftarrow 0$$

And similar heuristic for IST

Threshold depends on noise

Detection of significant wavelet coefficients

Noise Statistics in Data

The noise in the data follows a distribution law which can be:

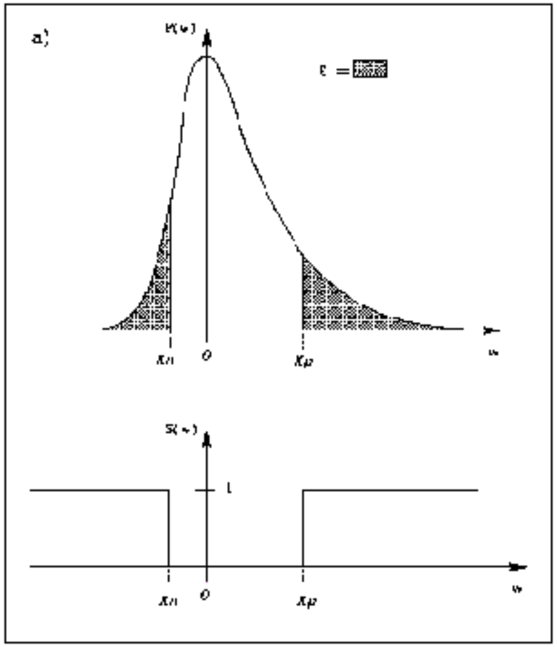
- a White Gaussian Noise
- Correlated Noise
- a Poisson Noise
- a Poisson + Gaussian distribution (noise in the CCD)
- Poisson noise with few events (Galaxies counting, X ray images, ...)
- Speckle noise
- Root Mean Square map: we have a noise standard deviation of each data value.

Detection in the Wavelet Domain

NOISE MODELING

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

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



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.

Ex. : Gaussian Centered Additive Noise

$$p(w_{j,l}) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-w_{j,l}^2/2\sigma_j^2}$$

Rejection of hypothesis $\mathcal{H}$, depends (for a positive coefficient value) on:

$$P = Prob(w_{j,l} > W) = \frac{1}{\sqrt{2\pi}\sigma_j} \int_{w_{j,l}}^{+\infty} e^{-W^2/2\sigma_j^2} dW$$

and if the coefficient value is negative, it depends on

$$P = Prob(w_{j,l} < W) = \frac{1}{\sqrt{2\pi}\sigma_j} \int_{-\infty}^{w_{j,l}} e^{-W^2/2\sigma_j^2} dW$$

Given stationary Gaussian noise, it suffices to compare $w_{j,l}$ to $k\sigma_j$.

if $\mid w_j \mid \geq k\sigma_j$ then w_j is significant

if $\mid w_j \mid < k\sigma_j$ then w_j is not significant

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 $\omega_{j,l}$ at scale j and position l using the noise level σ_j . A user parameter α determines the number of false detections as a percentage of the number of true detections. The FDR fixes the threshold.

Other Strategies for wavelet denoising

Wavelet denoising can be improved by:

1. Taking into account the neighboors of each wavelet coefficient. **Local Wiener Filter**

$$\tilde{\omega}_{j,l} = \frac{s^2}{s^2 + \sigma^2_{j,\mathcal{N}_l}} \omega_{j,l}$$

   OWT, PSNR = 28.90 dB ==> 30.63 dB

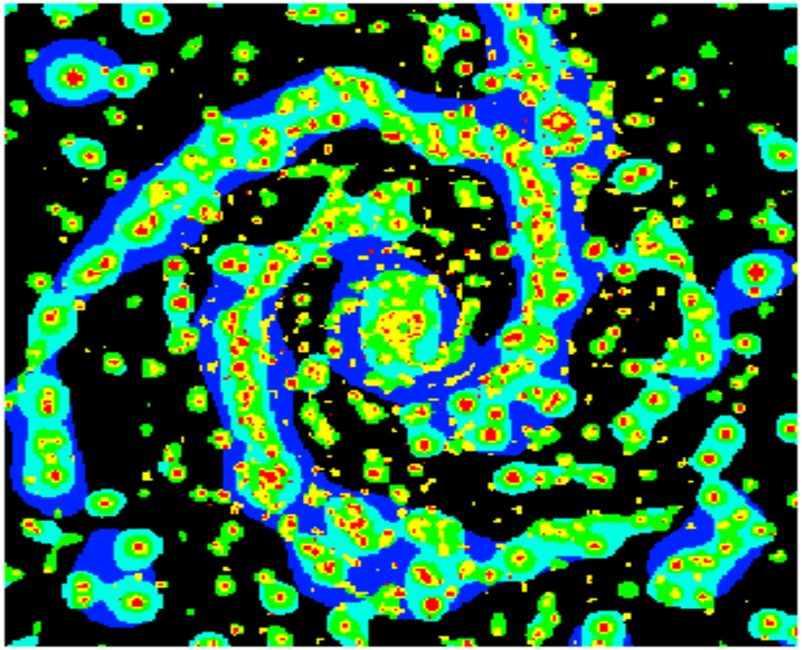
2. Taking into account the interscale correlation. **bivariate shrinkage:**

   OWT, PSNR = 30.95
   UWT, PSNR = 32.28

$$\tilde{\omega}_j = \frac{\left(\sqrt{\omega_j^2 + \omega_{j+1}^2} - \frac{\sqrt{3}\sigma^2}{s}\right)_+}{\sqrt{\omega_j^2 + \omega_{j+1}^2}} \omega_j$$

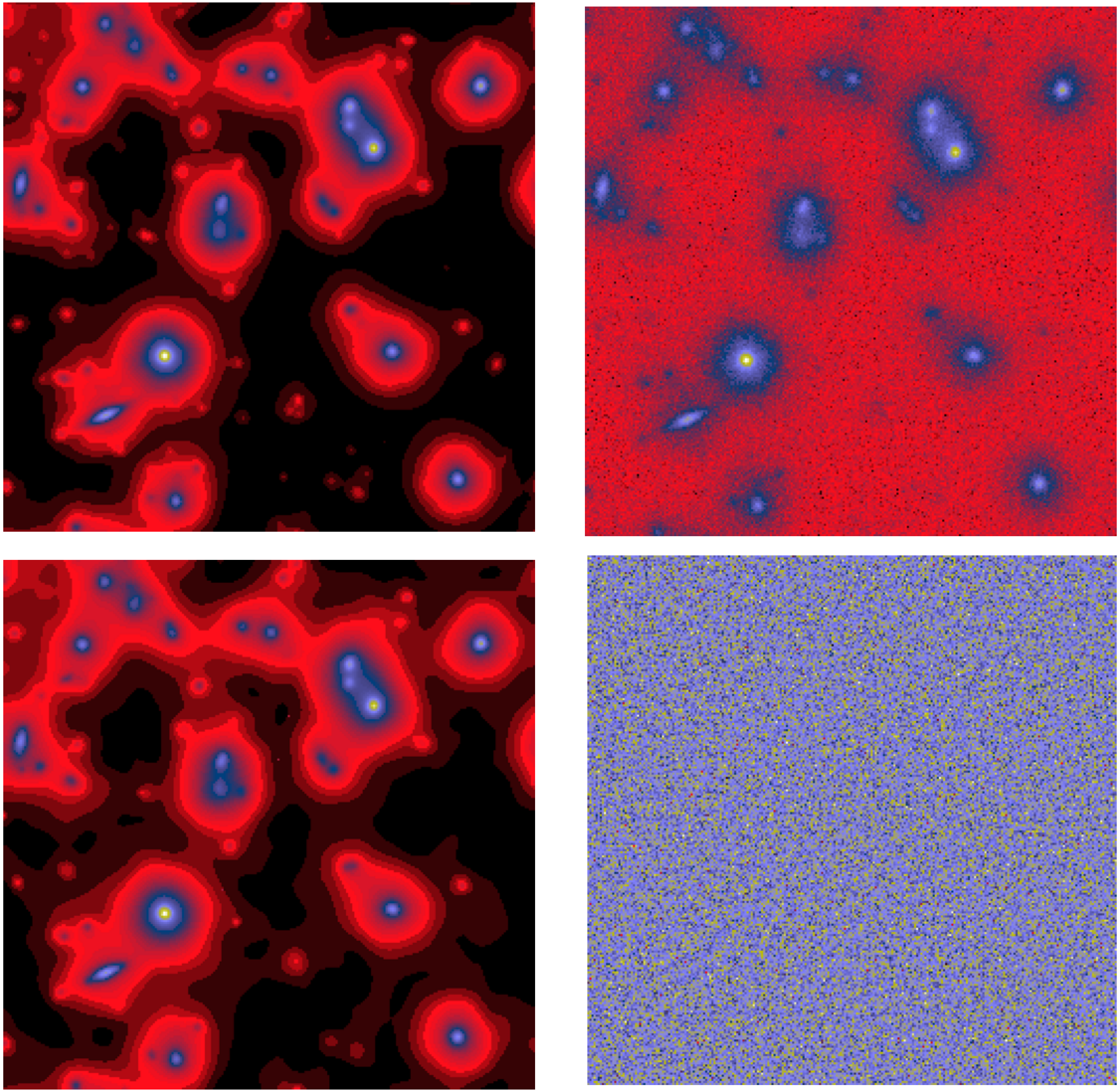
Denoising with Multiresolution Support

NGC2997
Multiresolution
Support



For IHT: $\alpha^{(n+1)} = \mathrm{HT}_{\mu t}\left(\alpha^{(n)} + \mu \mathcal{M} \Phi^T \left(Y - \Phi \mathcal{M} \alpha^{(n)}\right)\right)$

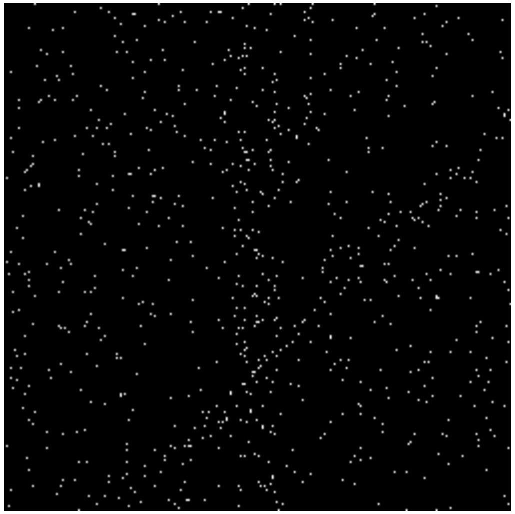
For IST: $\alpha^{(n+1)} = \mathrm{ST}_{\mu t}\left(\alpha^{(n)} + \mu \mathcal{M} \Phi^T \left(Y - \Phi \mathcal{M} \alpha^{(n)}\right)\right)$



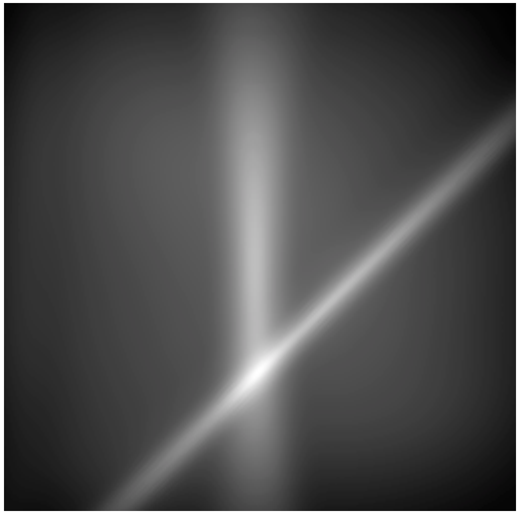
Poisson noise and Line-Like Sources

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

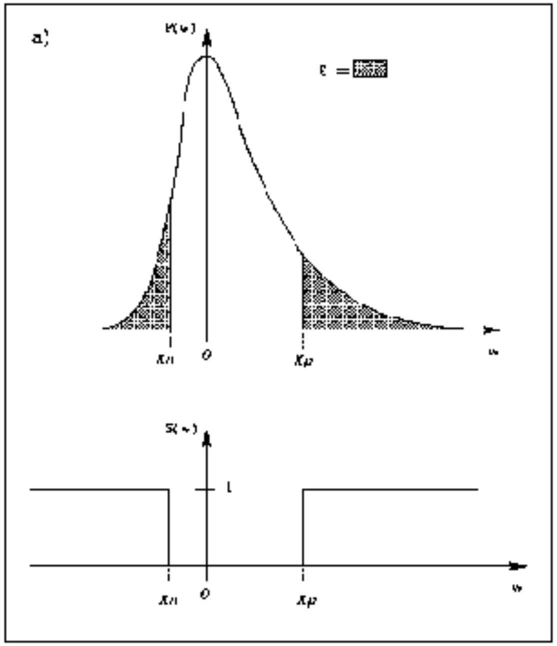
Max Intensity
background = 0.01
vertical bar = 0.03
inclined bar = 0.04

Curvelet Filtering

NOISE MODELING

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

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

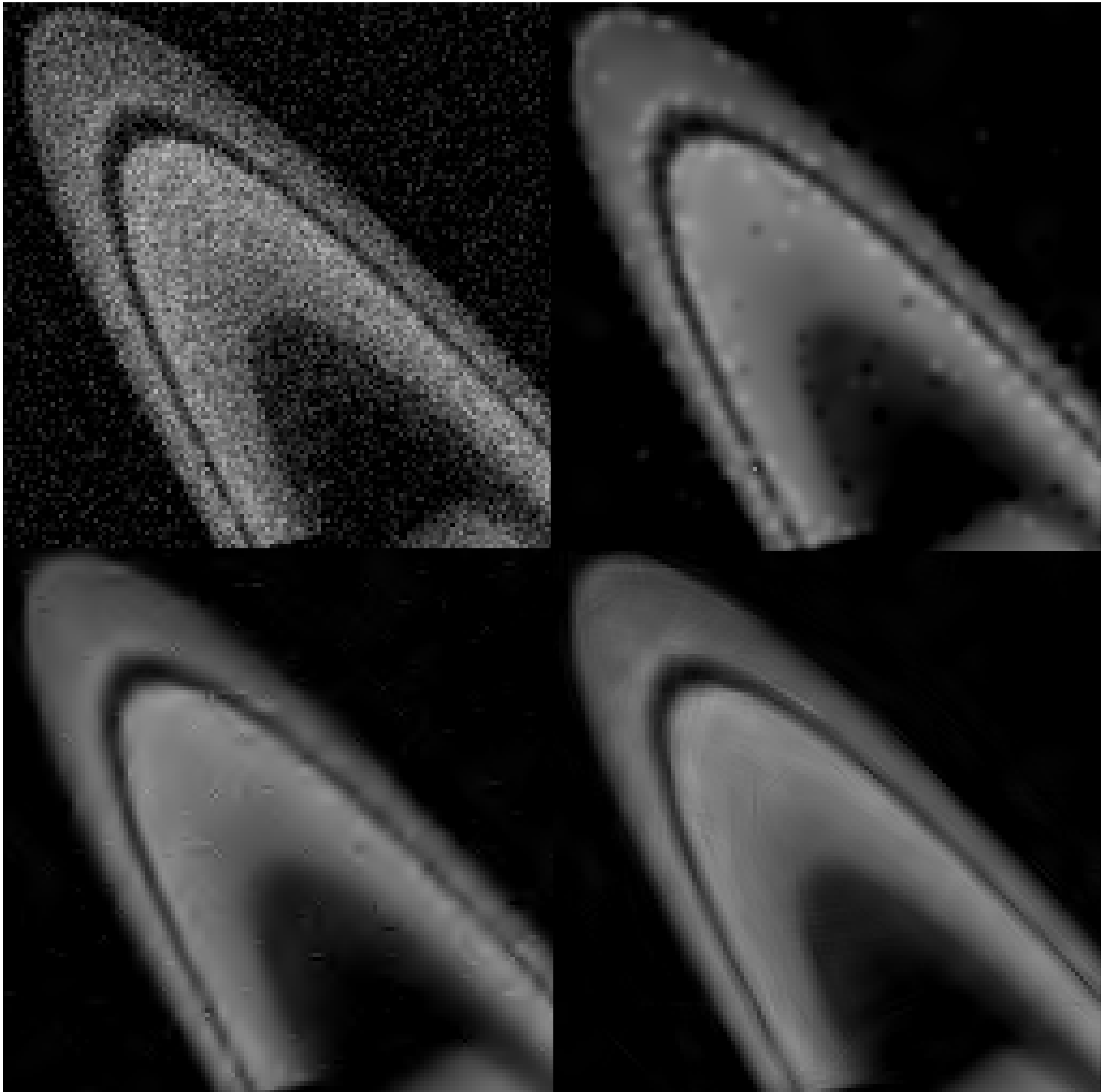


Given a threshold t:

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.

$$\tilde{y} = C_R\left[\delta(C_T y)\right],\ \delta(c)\text{: Hard Thresholding}$$



Undecimated
Bi-orthogonal

“A-trous”
wavelet
Transform

Curvelet
Transform

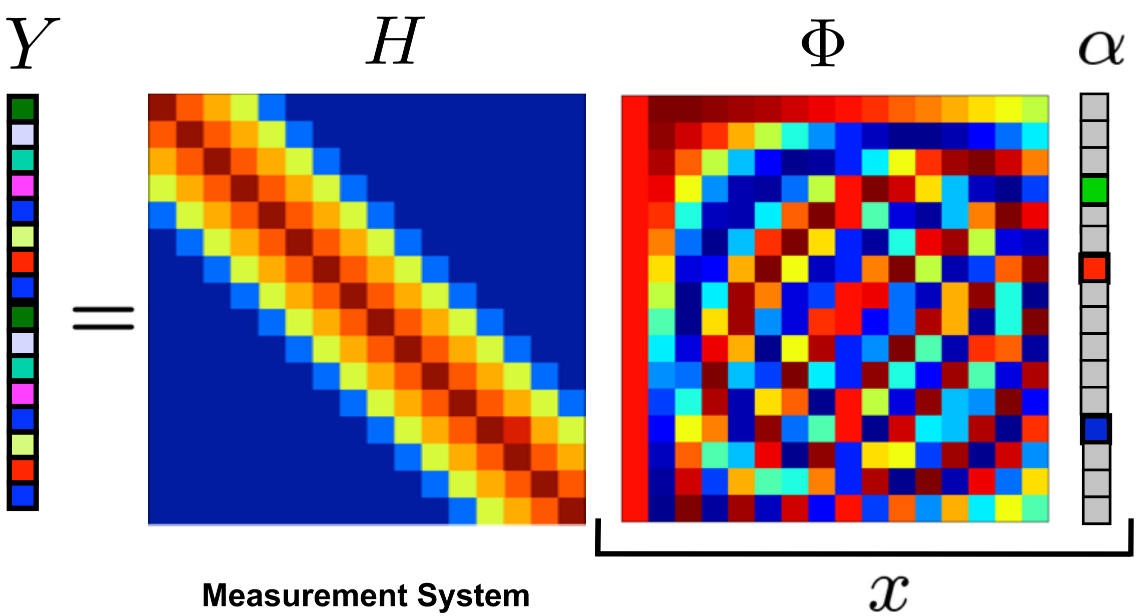
Inverse Problem tour and Sparse Recovery

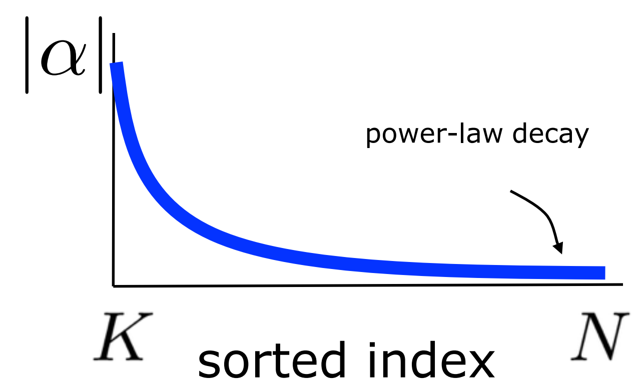
$$Y = HX + N$$

$X = \Phi\alpha$, and α is sparse (or compressible)

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

$$\min_{\alpha} \|\alpha\|_p \quad \text{subject to} \quad \|Y - H\Phi\alpha\|_2^2 \leq \epsilon$$





Deconvolution Problems

$$Y = HX + N$$

PB 1 : Find X knowing Y,H and the statistical properties of the noise N:
Astronomical image deconvolution

PB 2 : Find X and H knowing Y and the statistical properties of the noise N
Blind deconvolution
Multichannel Data (PCA, ICA, etc)

Ill posed problem, i.e. not an unique and stable solution: sparse Regularization

$$\widetilde{\alpha} \in \arg\min_{\alpha} \|Y - H\Phi\alpha\|_2^2 + t\|\alpha\|_p, \quad 0 \leq \mathrm{p} \leq 1$$

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

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

Decreasing Threshold per Iteration

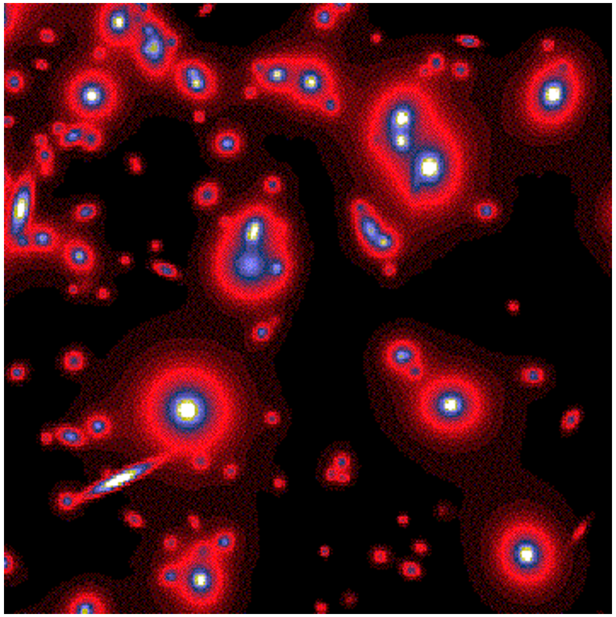
Iterative thresholding with a varying threshold was proposed in (Starck et al, 2004; Elad et al, 2005) for sparse signal decomposition in order to accelerate the convergence. The idea consists in using a different threshold λ^n at each iteration.

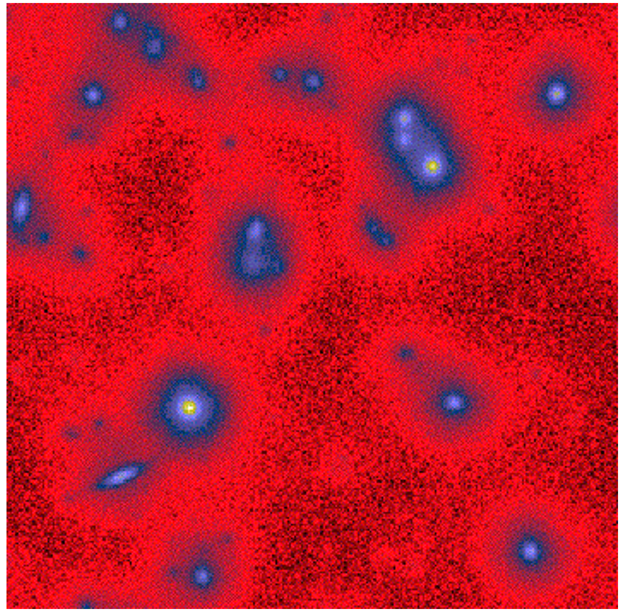
For IHT: $$\alpha^{(n+1)} = \mathrm{HT}_{\mu\lambda^{(n)}}\left(\alpha^{(n)} + \mu\Phi^T H^T\left(Y - H\Phi\alpha^{(n)}\right)\right)$$

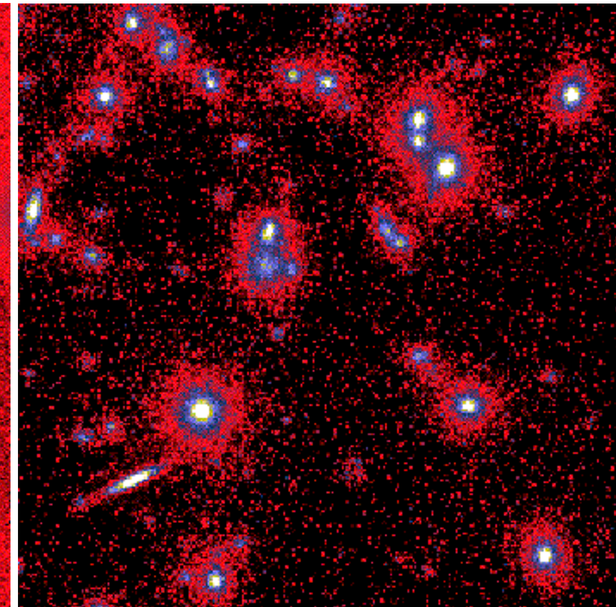
For IST: $$\alpha^{(n+1)} = \mathrm{ST}_{\mu\lambda^{(n)}}\left(\alpha^{(n)} + \mu\Phi^T H^T\left(Y - H\Phi\alpha^{(n)}\right)\right)$$

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

DECONVOLUTION SIMULATION

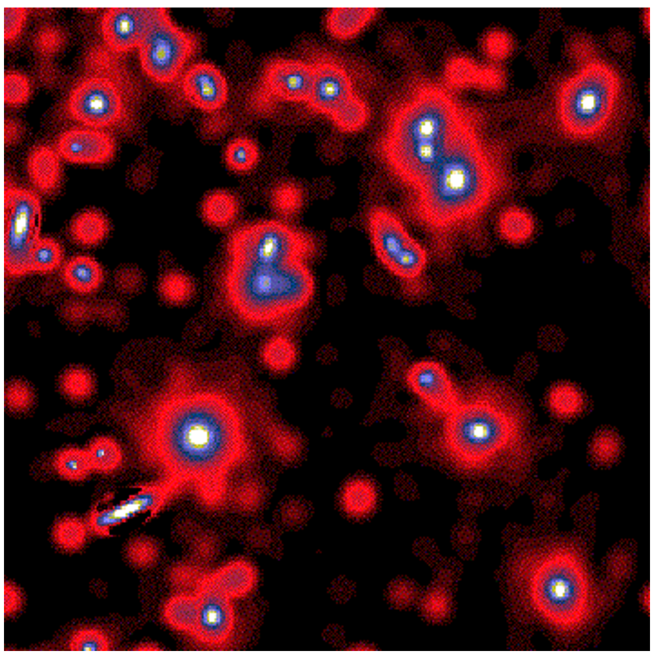


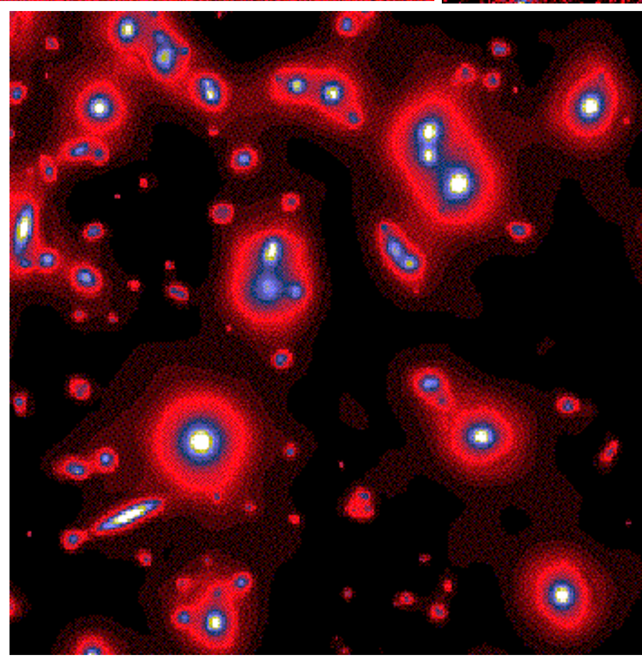




LUCY

PIXON





Wavelet

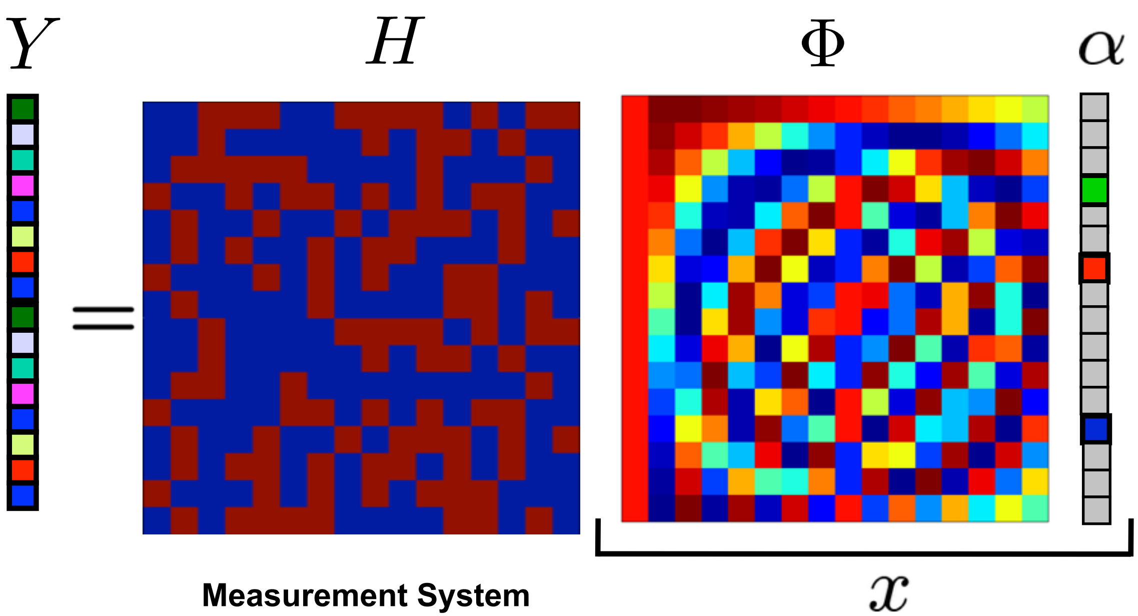
Inverse Problem tour and Sparse Recovery

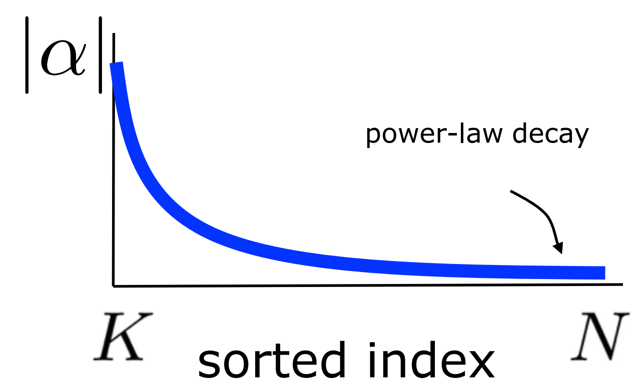
$$Y = HX + N$$

$X = \Phi\alpha$, and α is sparse (or compressible)

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

$$\tilde{\alpha} \in \arg\min_{\alpha} \|Y - H\Phi\alpha\|_2^2 + t\|\alpha\|_p, \quad 0 \leq \text{p} \leq 1$$

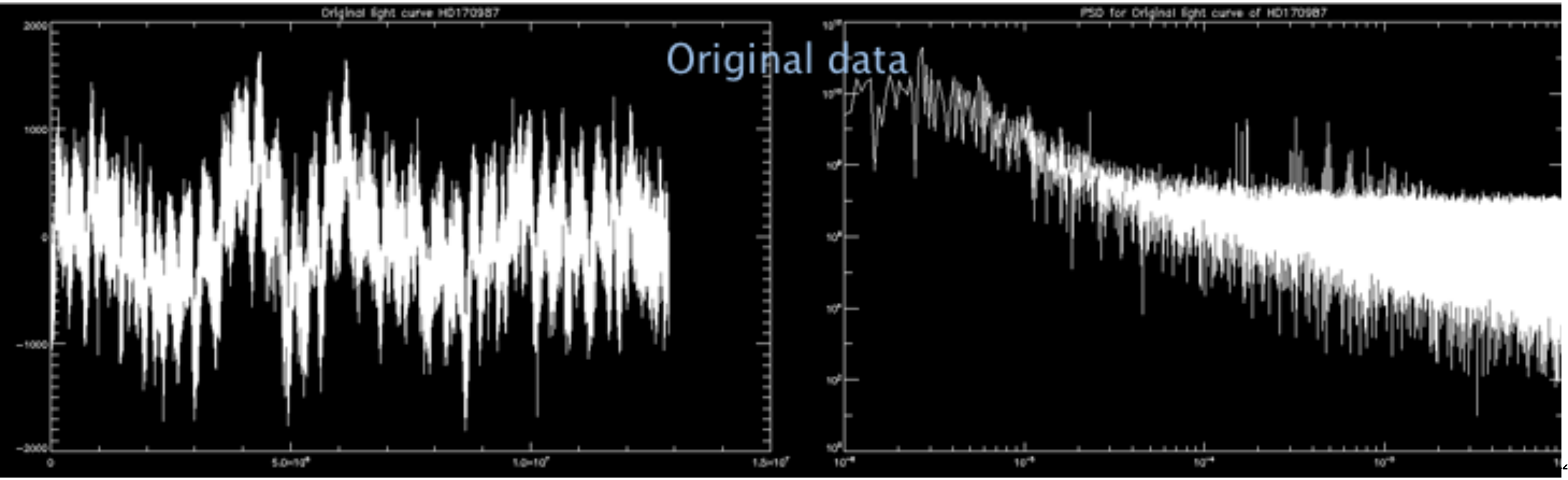




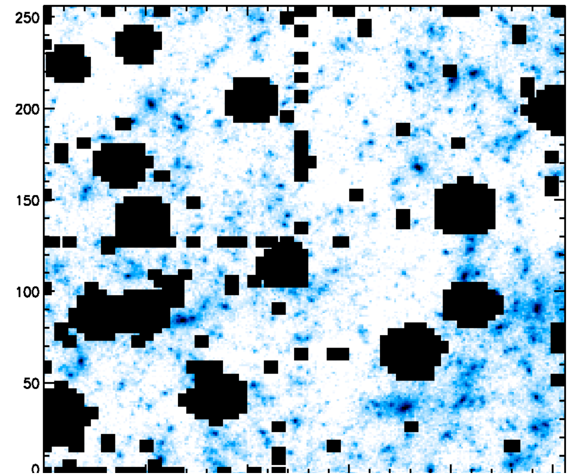
Problem of Missing Data

- Period detection in temporal series

COROT: HD170987



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



Inpainting: Missing Data

- *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.*

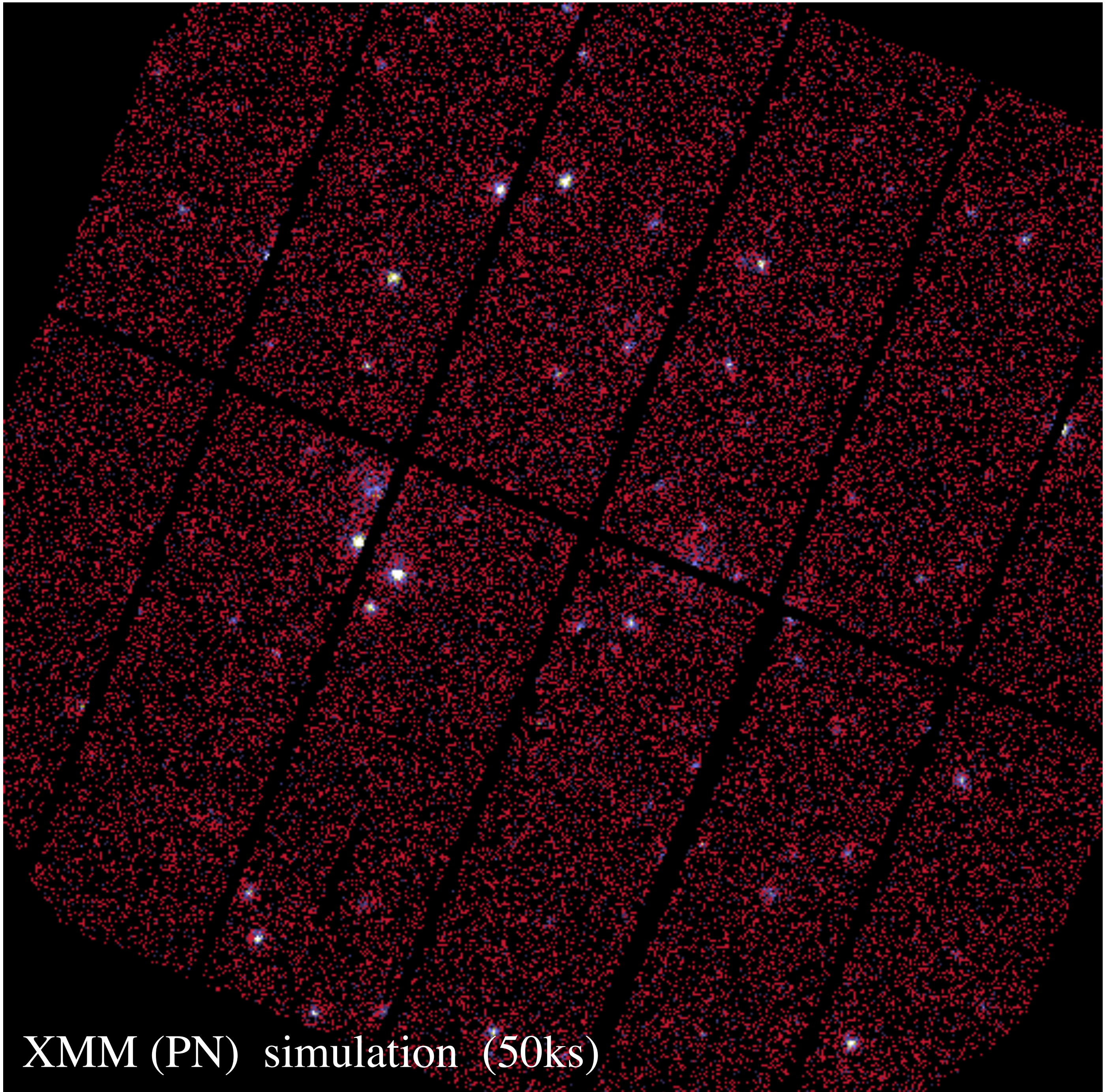
$$\Phi_\Lambda = \mathrm{Id}_\Lambda \qquad \underset{\alpha}{\text{minimize}} \; ||\alpha||_p \text{ s.t. } y = Mx = M\Phi\alpha$$

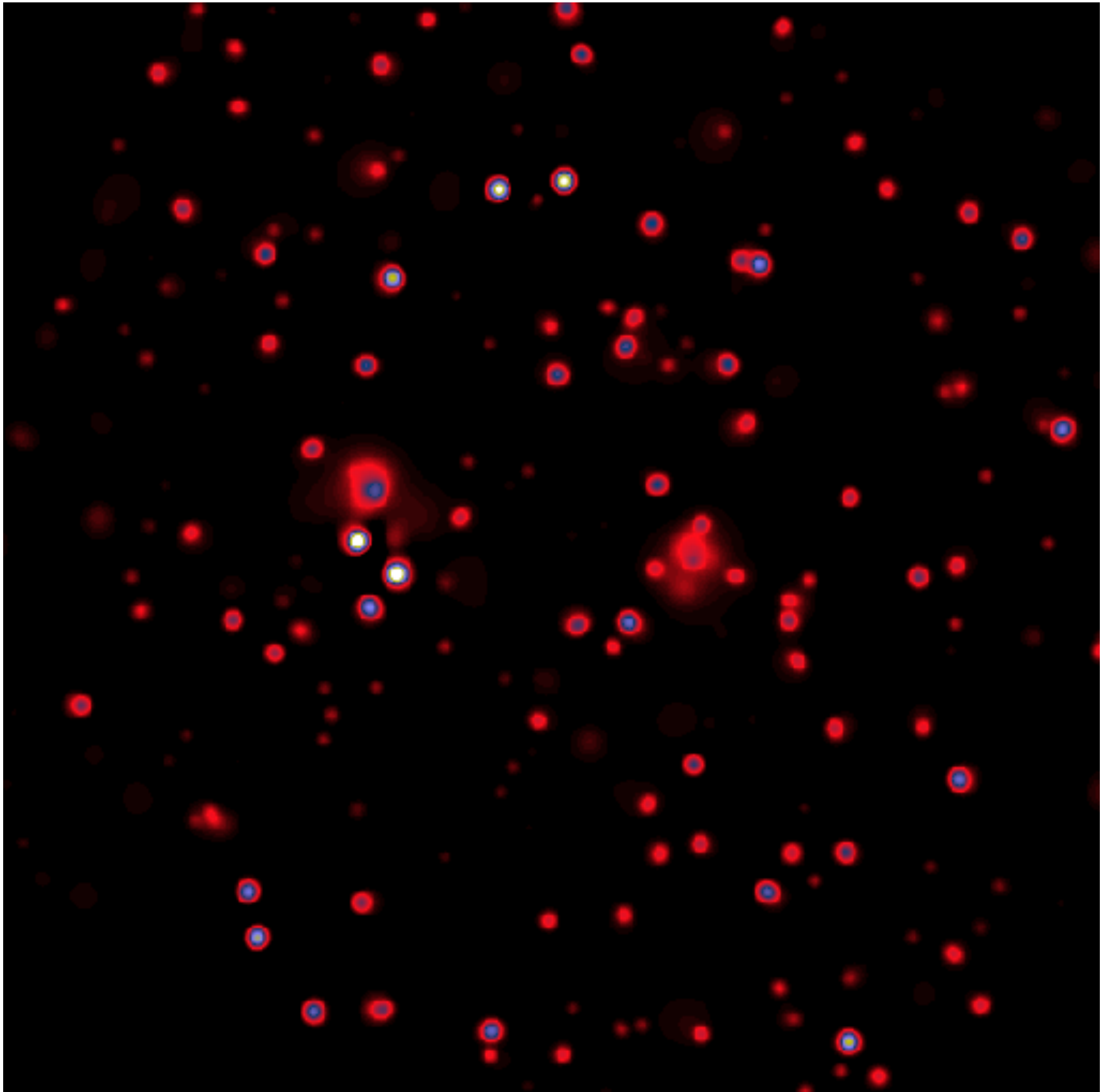
where M is the mask: $M(i,j) = 0$ for missing data, $M(i,j) = 1$ for good data

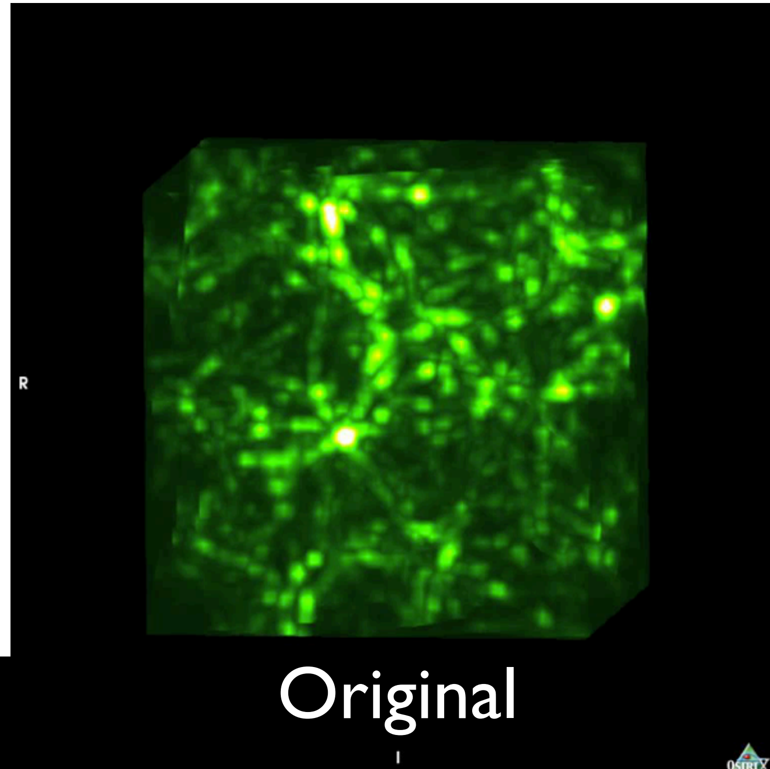
$$x^{(n+1)} = HT_{\phi,\lambda^{(n)}} \left\{ x^{(n)} + M\left(y - x^{(n)}\right) \right\}$$

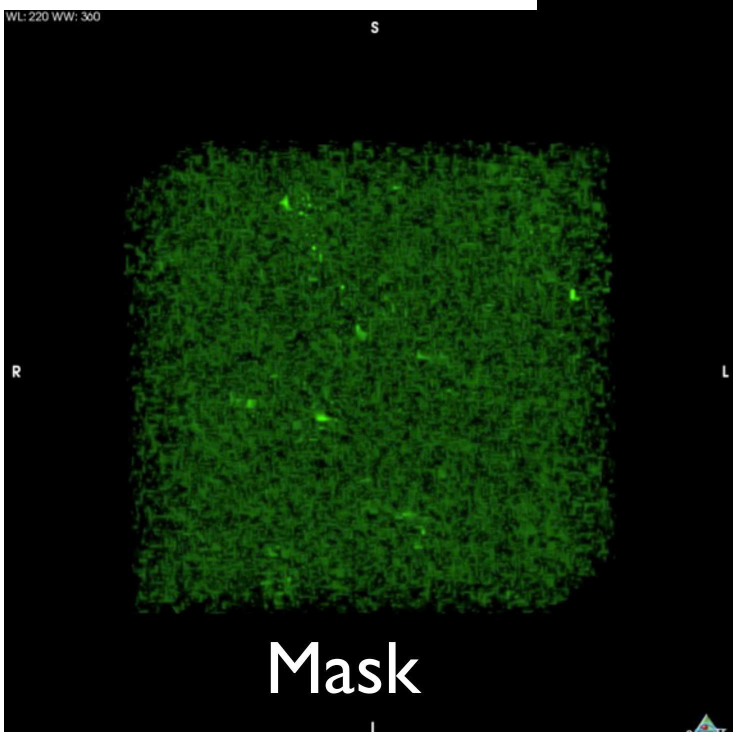
Iterative Hard Thresholding with a decreasing threshold.

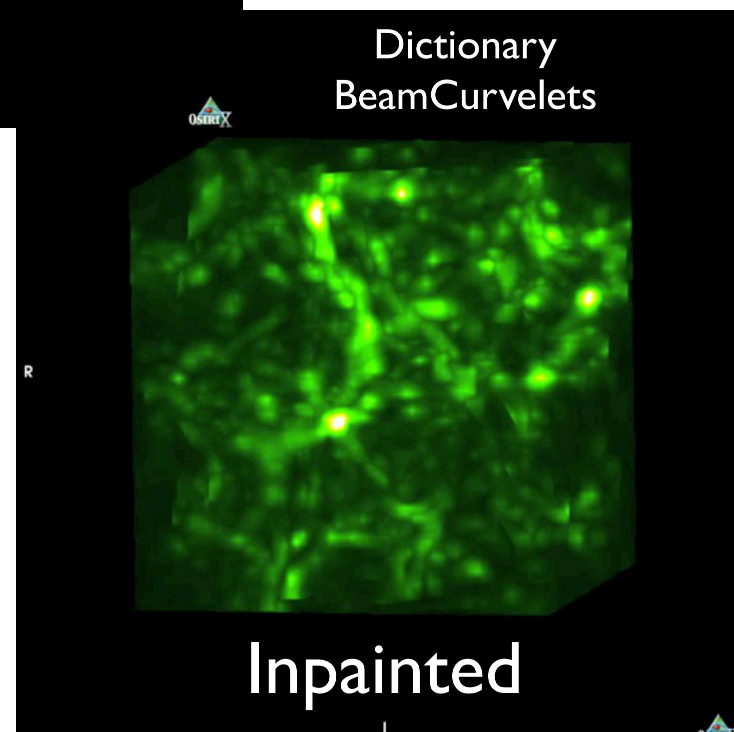
MCAlab available at: http://www.greyc.ensicaen.fr/~jfadili


XMM (PN) simulation (50ks)




R
Original
I


WL: 220 WW: 360
S
R
L
Mask
I


Dictionary
BeamCurvelets
R
Inpainted
I

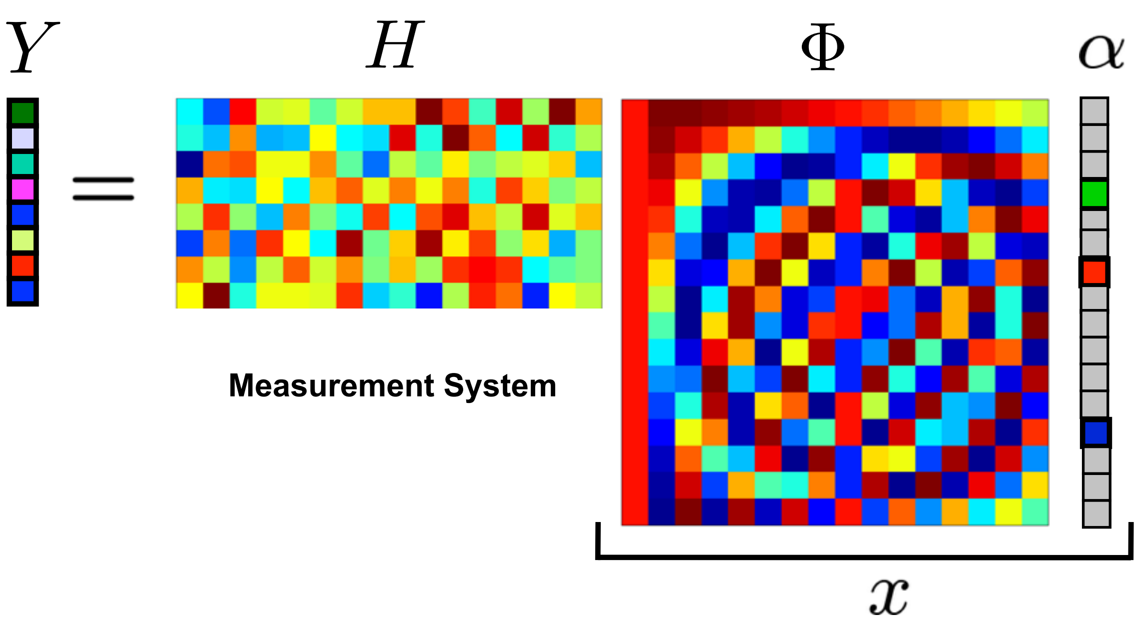
Inverse Problem tour and Sparse Recovery

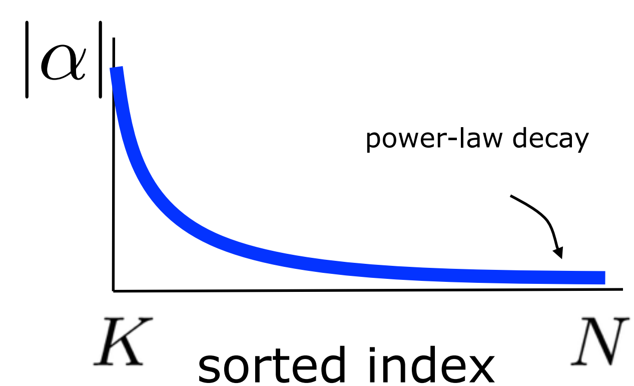
$$Y = HX + N$$

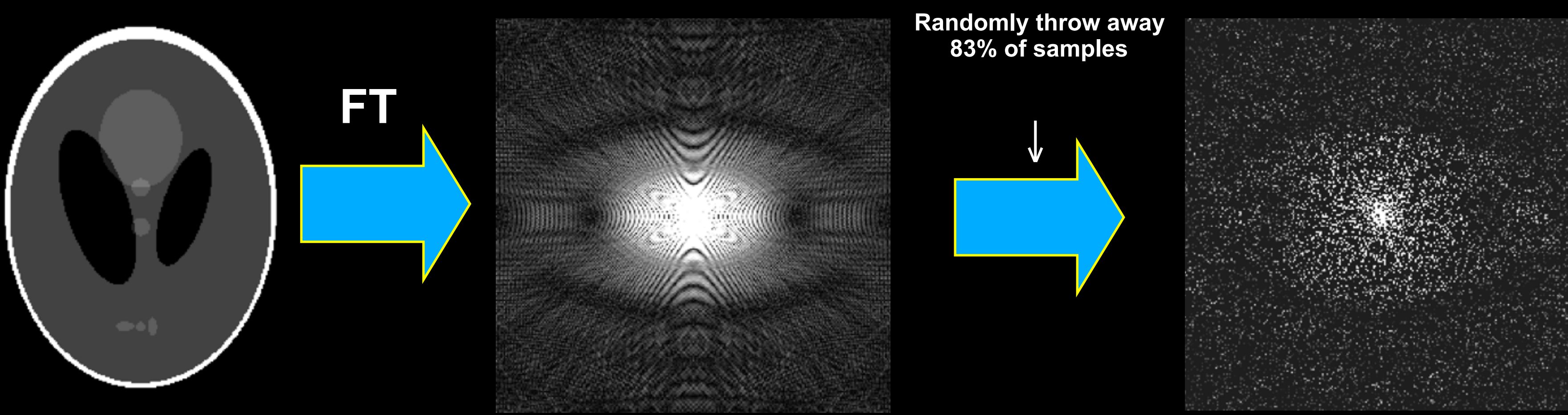
$X = \Phi\alpha$, and α is sparse (or compressible)

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

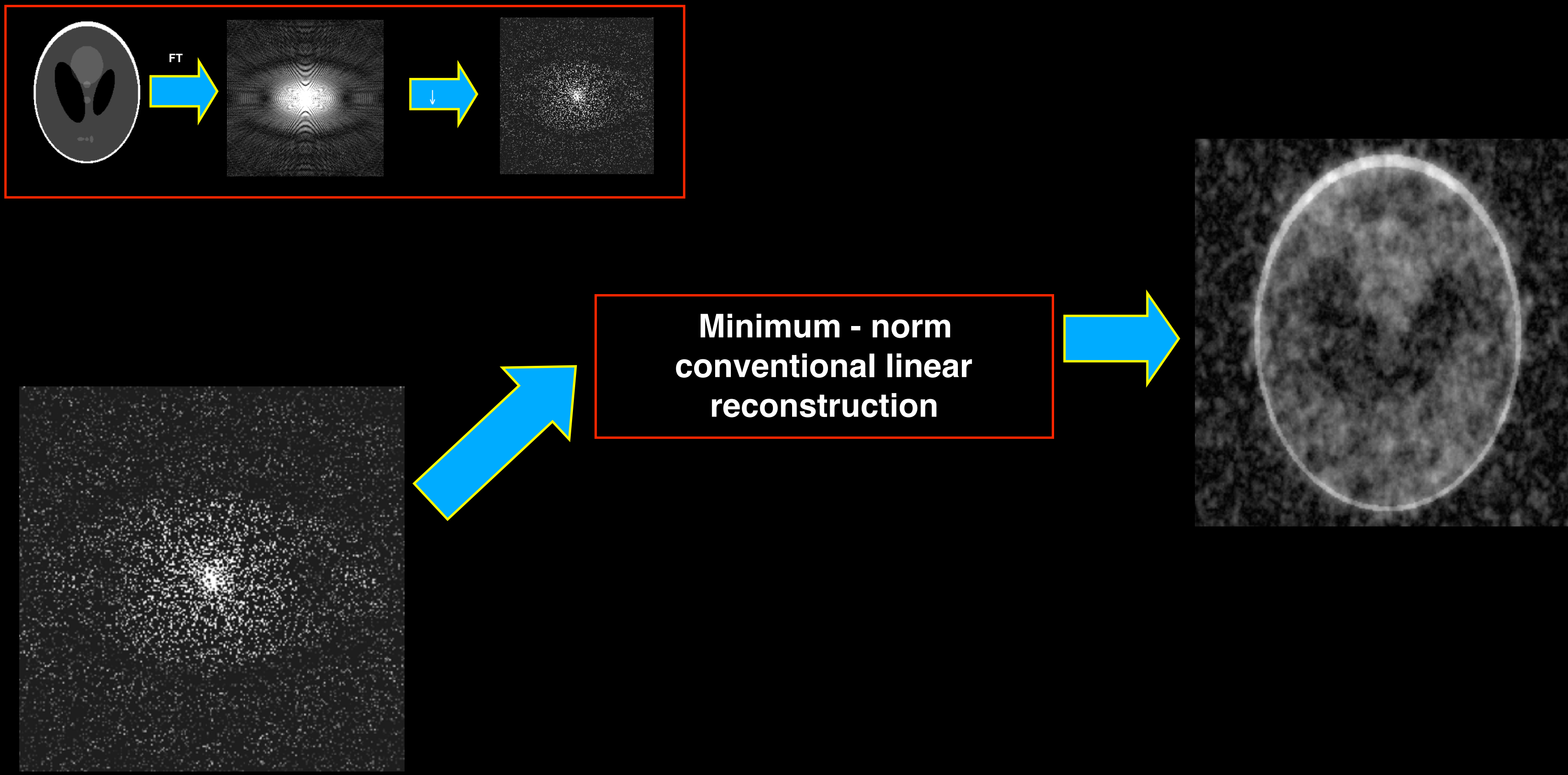
$$\tilde{\alpha} \in \arg\min_{\alpha} \|Y - H\Phi\alpha\|_2^2 + t\|\alpha\|_p, \quad 0 \leq \mathrm{p} \leq 1$$



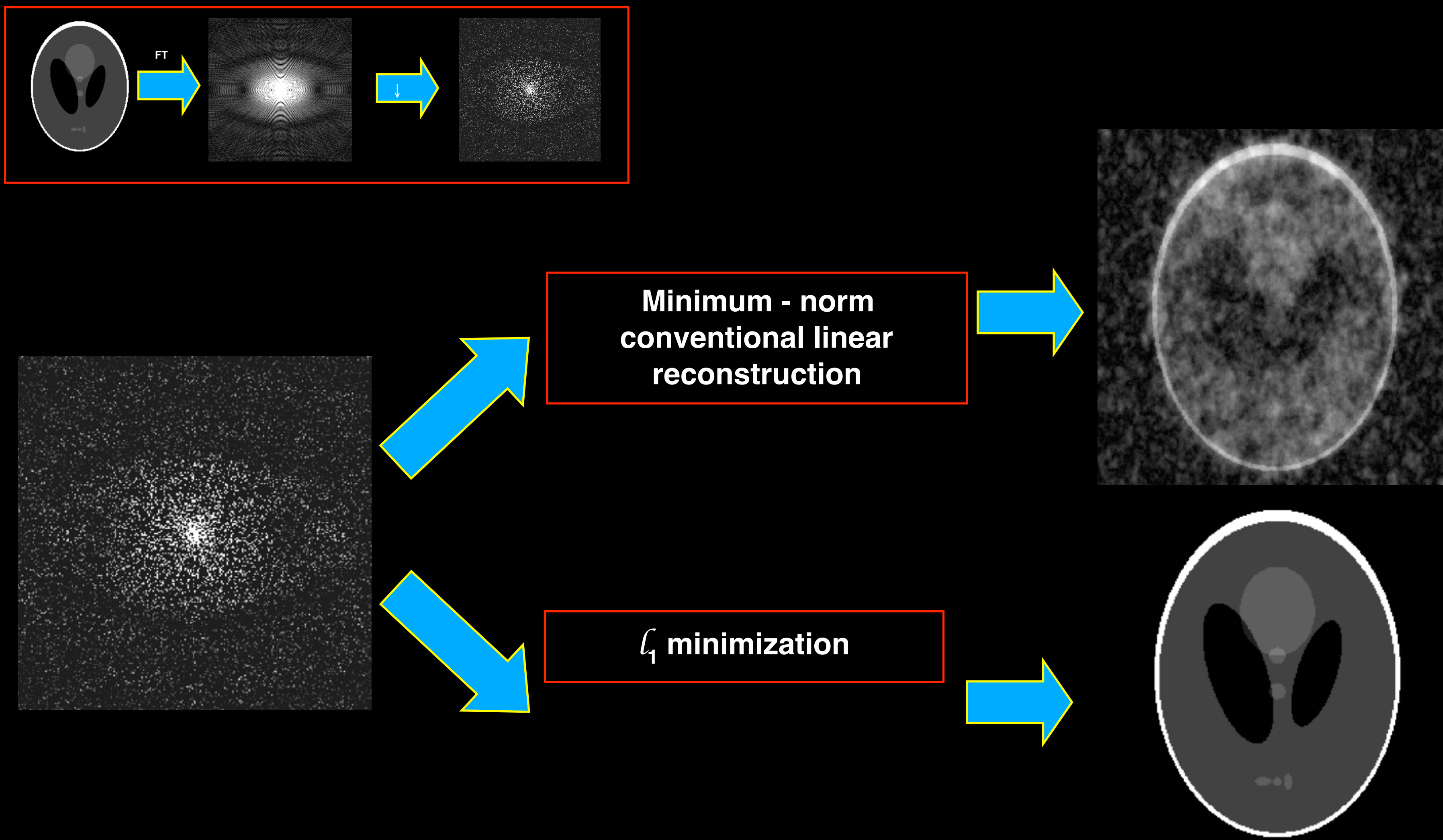




* E.J. Candes, J. Romberg and T. Tao.



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


FT
Minimum - norm conventional linear reconstruction
ℓ_1 minimization

E.J. Candes

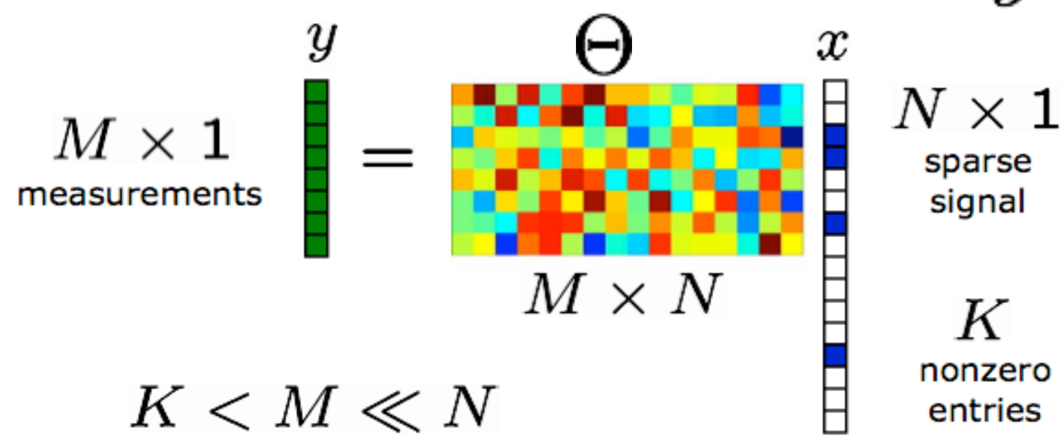
Compressed sensing

* 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"

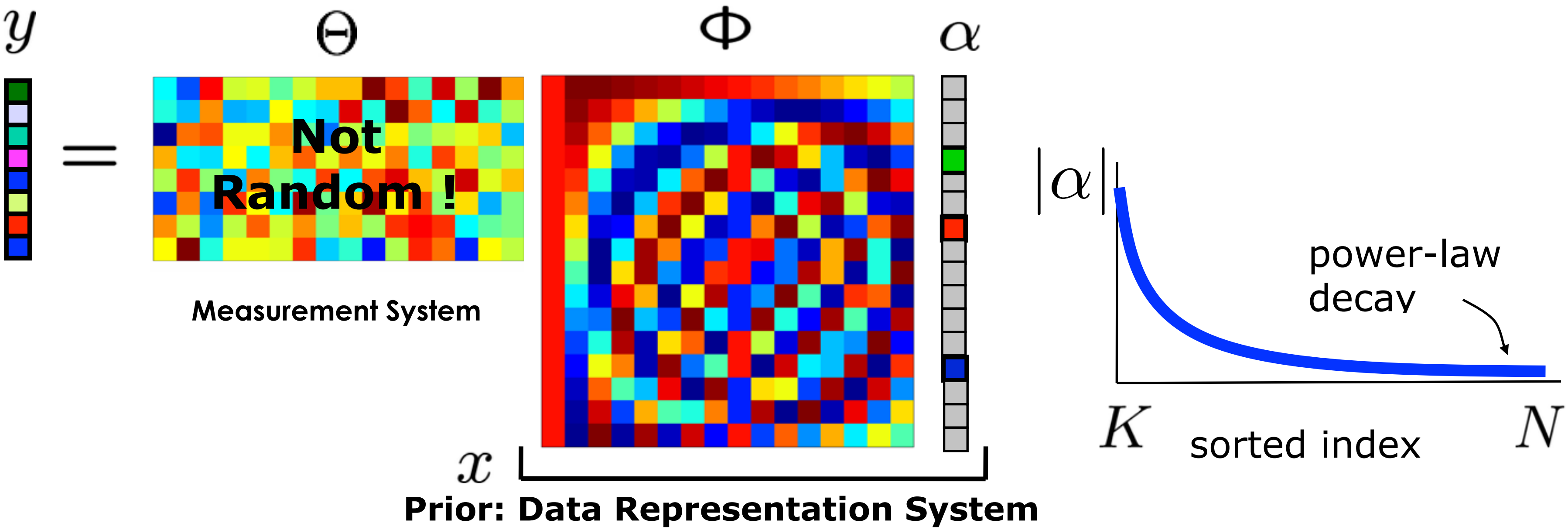
Replace samples with *few linear projections* $y = \Theta x$



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

Soft compressed sensing reconstruction

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

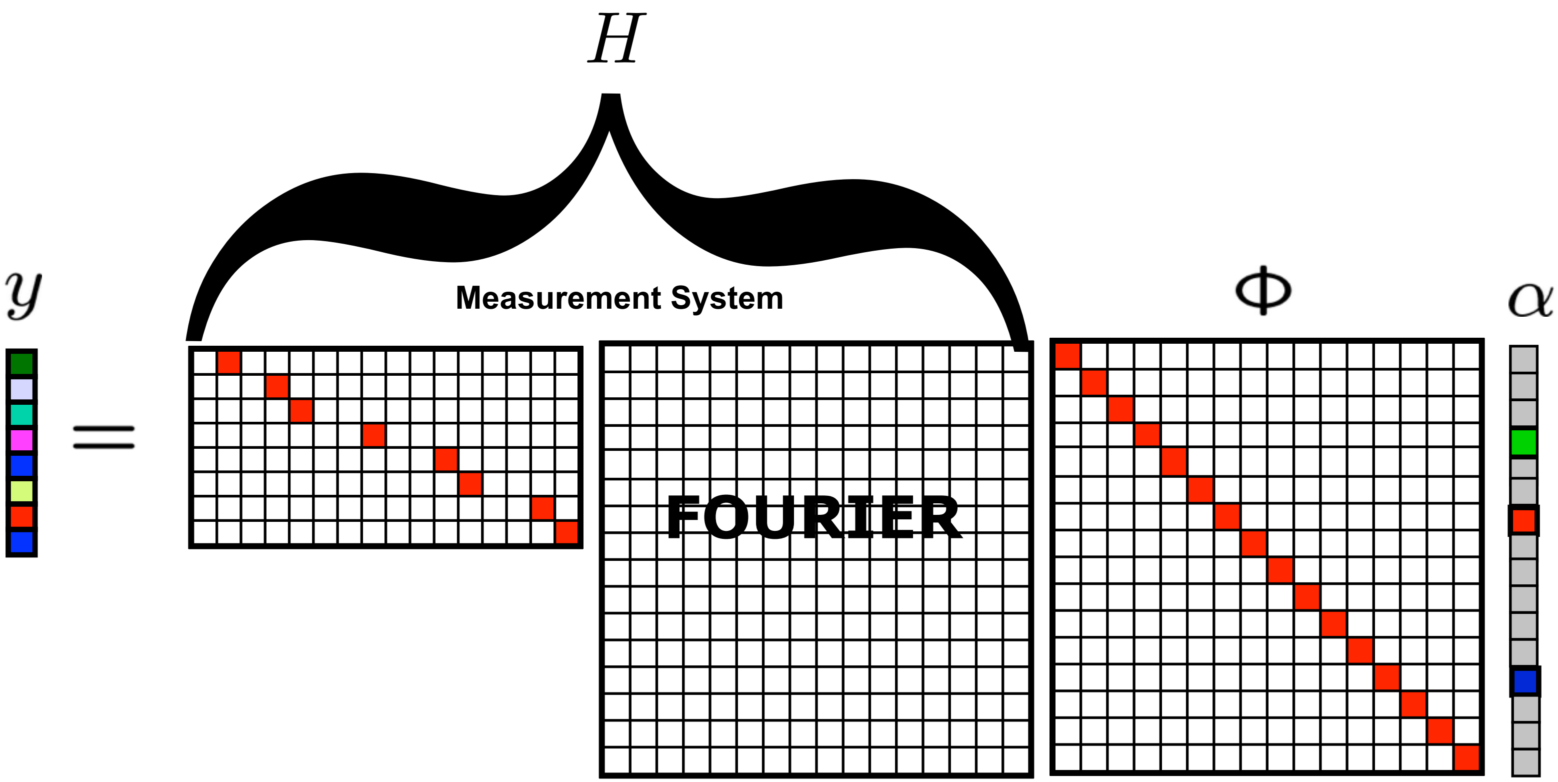


Mutual coherence: $\mu_{\Theta,\Phi} = \max_{i,k} \left| \left\langle \Theta_i , \Phi_k , \right\rangle \right|$ $m \geq C \mu^2_{\Theta,\Phi} K \log n$

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

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

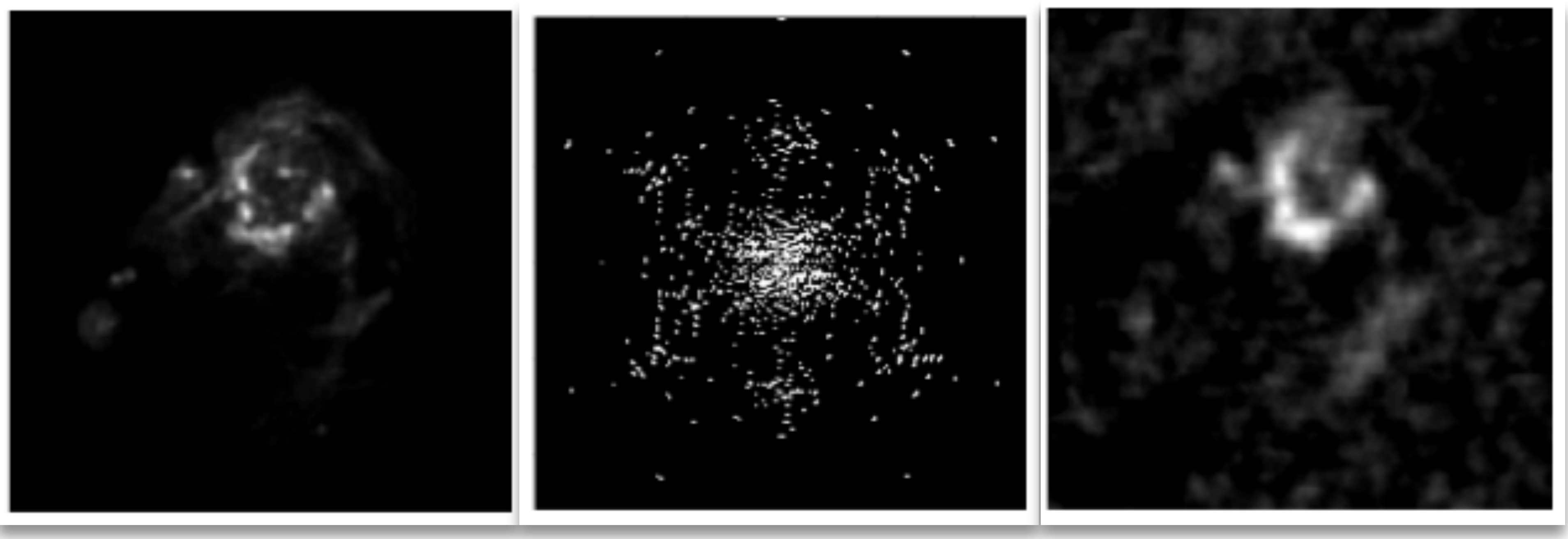
Radio-Interferometry and Compressed Sensing



==> See (McEwen et al, 2011; Wenger et al, 2010; Wiaux et al, 2009; Cornwell et al, 2009; Suskimo, 2009; Feng et al, 2011, Garsden et al 2014).

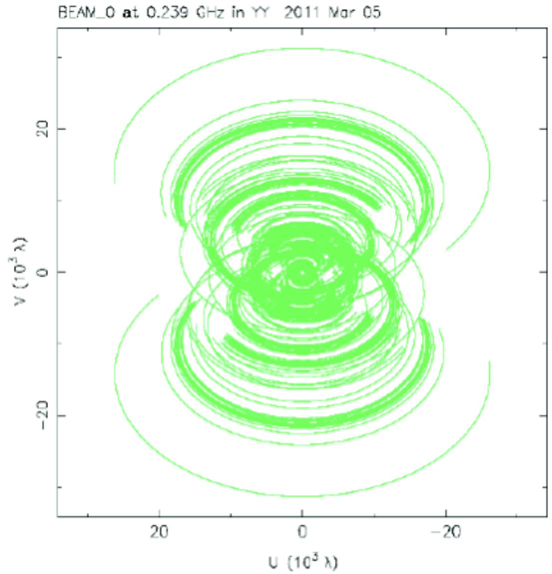
CS- Radio Astronomy

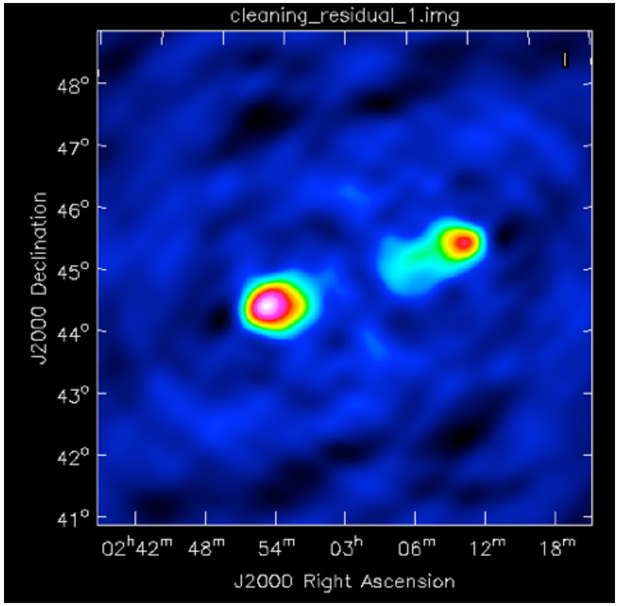
The Applications of Compressive Sensing to Radio Astronomy: I Deconvolution
Feng Li, Tim J. Cornwell and Frank De hoog, ArXiv:1106.1711, Volume 528, A31,2011.

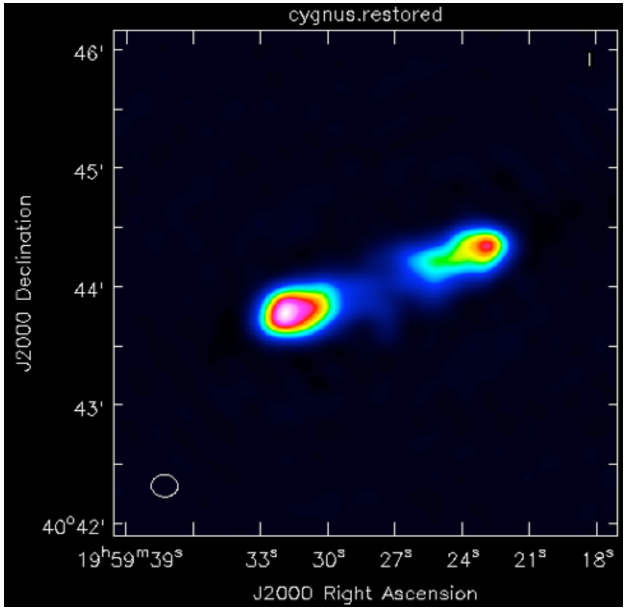


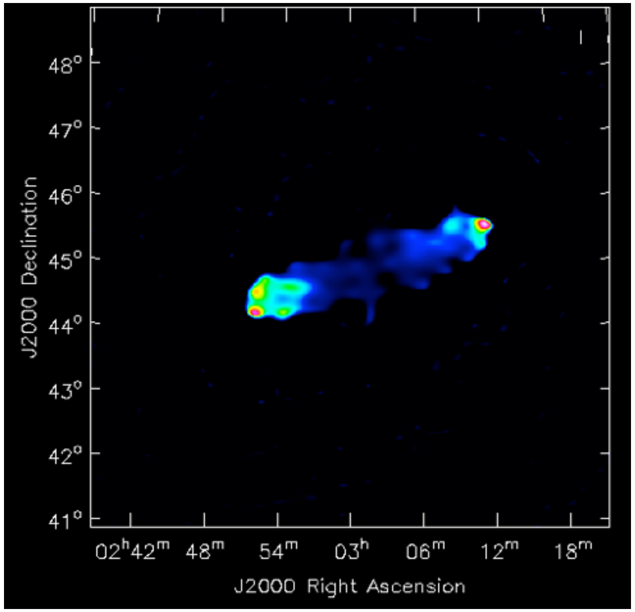
Australian Square Kilometer Array Pathfinder (ASKAP) radio telescope.

Compressed Sensing & LOFAR Cygnus A



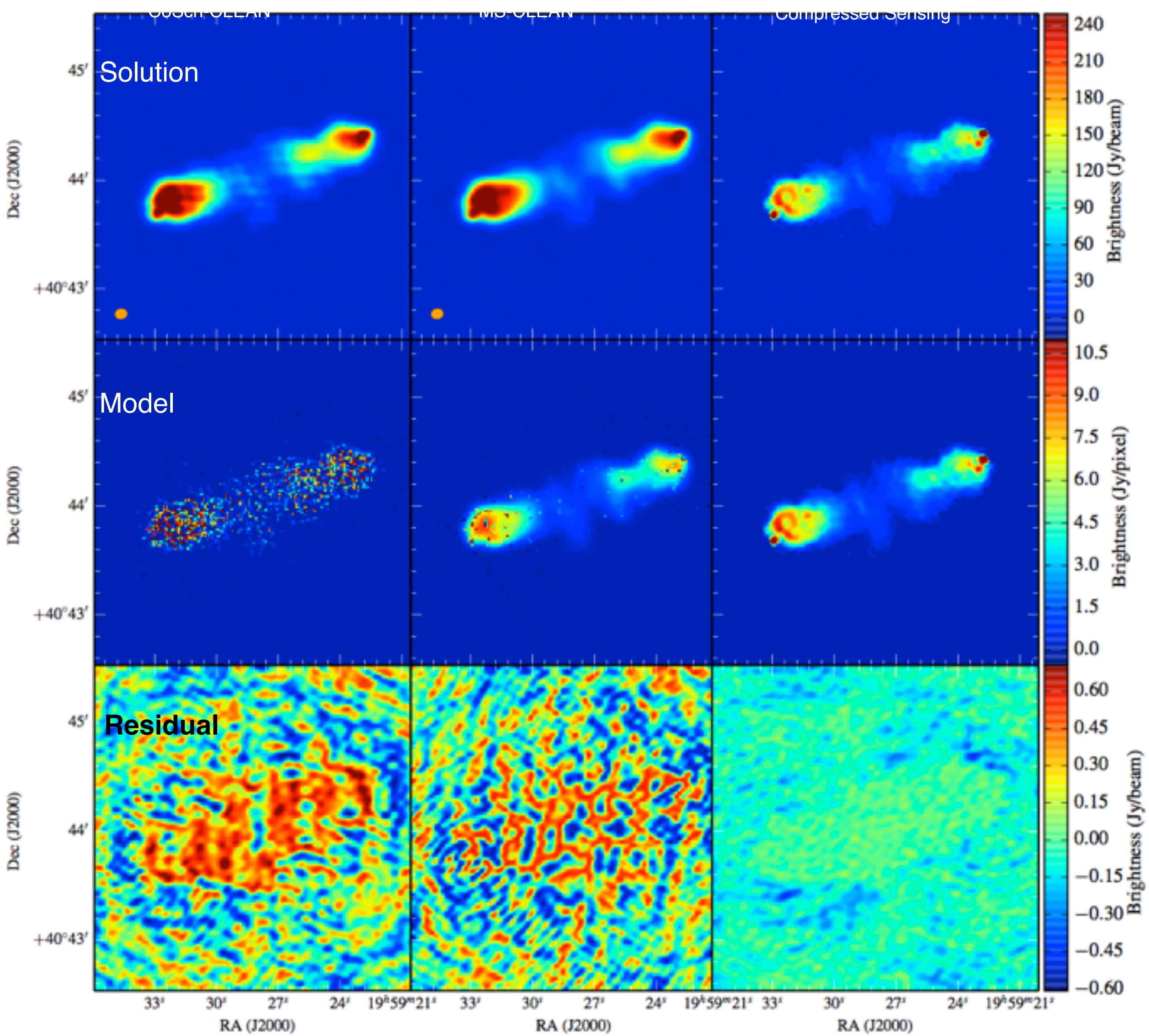




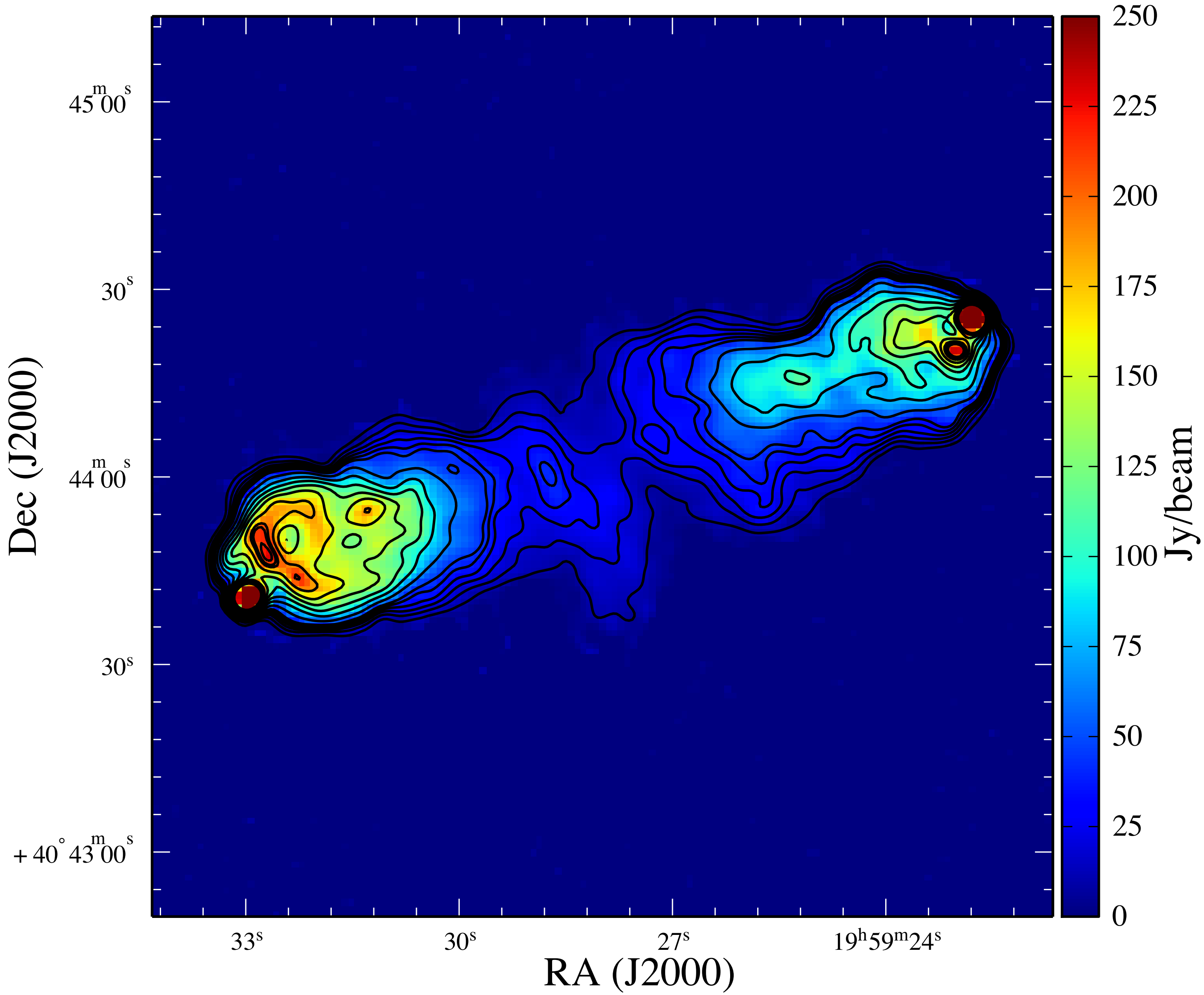


H. Garsden et al “LOFAR Sparse Image Reconstruction”, A&A, 575, 2015.

Reconstructed Images of Cygnus A



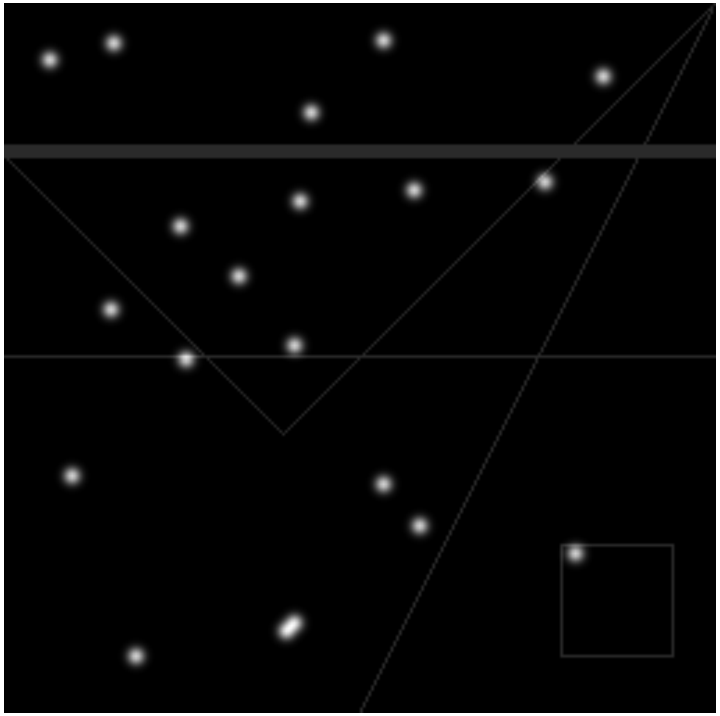
Residual std-dev = 2,65 Jy/beam, 0,26 Jy/beam, 0,05 Jy/beam



Colorscale: reconstructed 512x512 image of Cygnus A at 151 MHz (with resolution 2.8” and a pixel size of 1”). Contours levels are [1,2,3,4,5,6,9,13,17,21,25,30,35,37,40] Jy/Beam from a 327.5 MHz Cyg A VLA image (Project AK570) at 2.5” angular resolution and a pixel size of 0.5”. Most of the recovered features in the CS image correspond to real structures observed at higher frequencies.

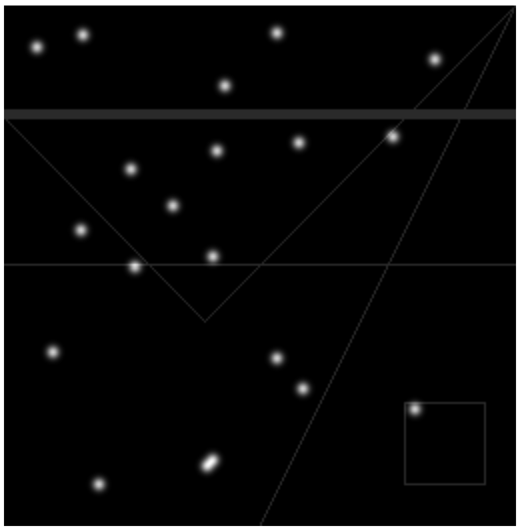
How to represent diverse data ? (1)

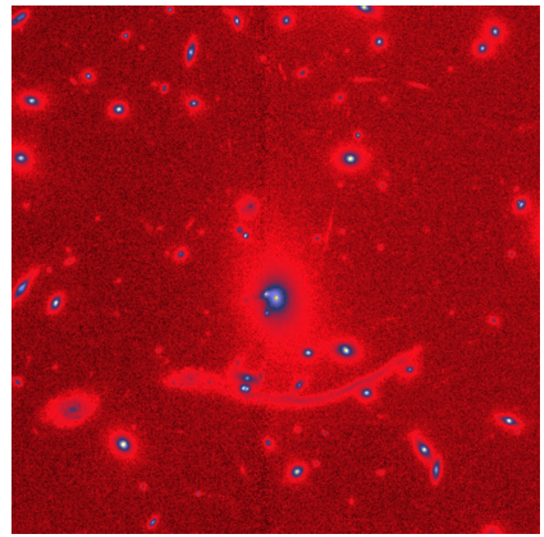
Is there any representation that well represents the following image ?



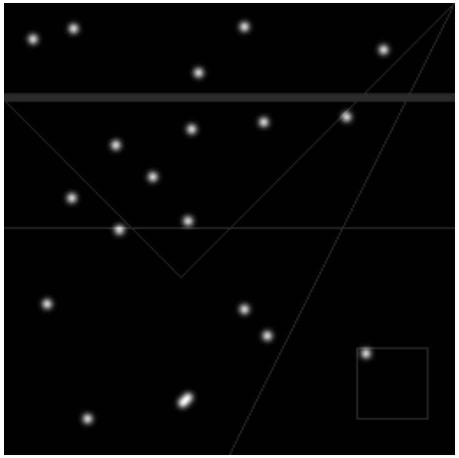
How to represent diverse data ? (2)

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

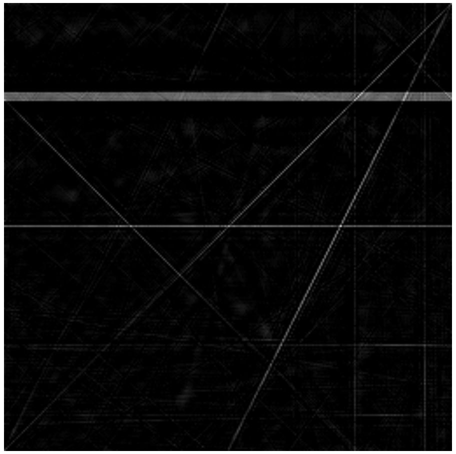




How to represent diverse data ? (3)

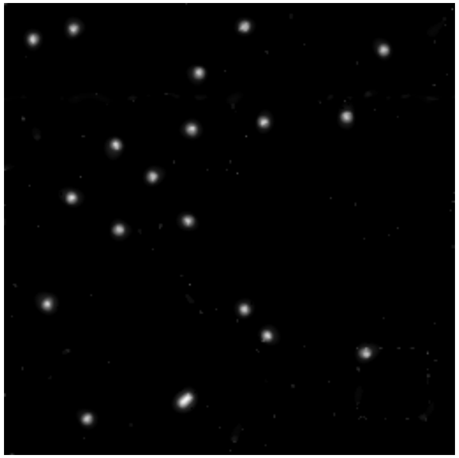


=



Lines

+



Gaussians

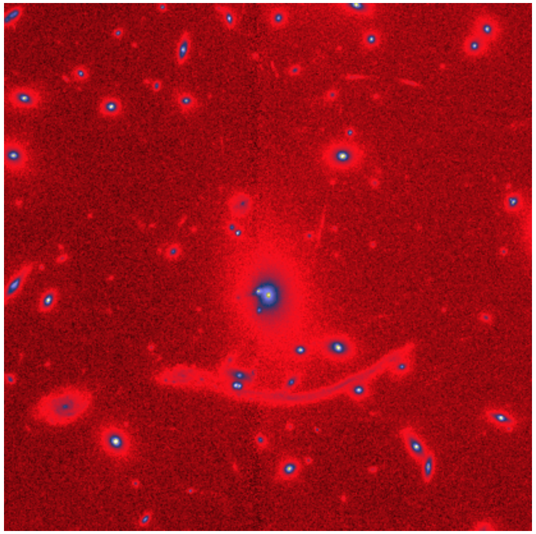
Curvelets

Wavelets

REDUNDANT REPRESENTATIONS

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.



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

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

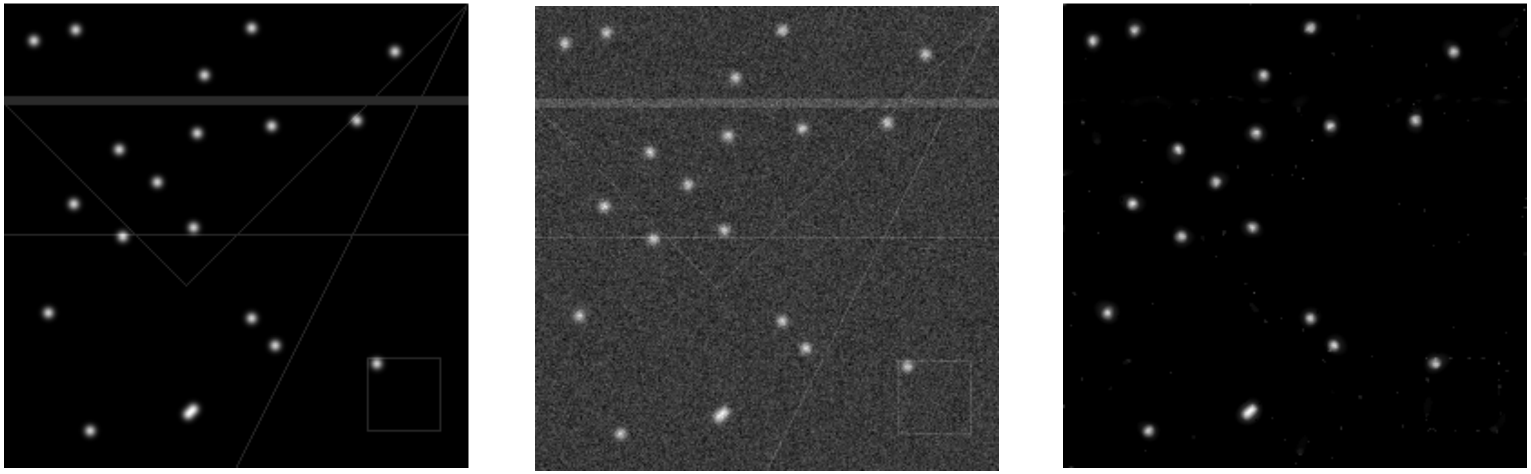
$$s_k = \phi_k \alpha_k$$
$$s = \sum_{k=1}^{L} s_k = \sum_{k=1}^{L} \phi_k \alpha_k$$

MCA Algorithm

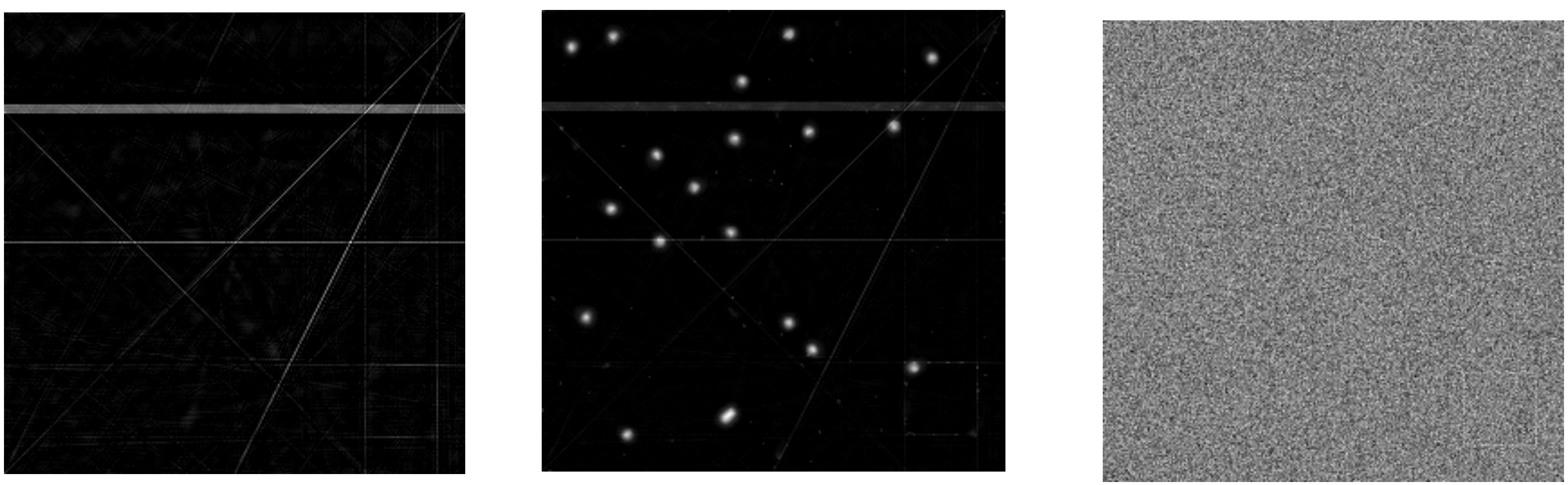
$$\underset{s_k, k=1..L}{\text{minimize}} \|s - \sum_{k=1}^{L} s_k\|_2^2 + \lambda \sum_{k=1}^{L} ||T_k s_k||_p$$

- 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:
    $\underset{s_k}{\text{minimize}} ||s - \sum_{i=1, i \neq k}^{L} s_i - s_k||_2^2 + \lambda^{(j)} ||T_k s_k||_p$
    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 $\lambda^{(j)}$

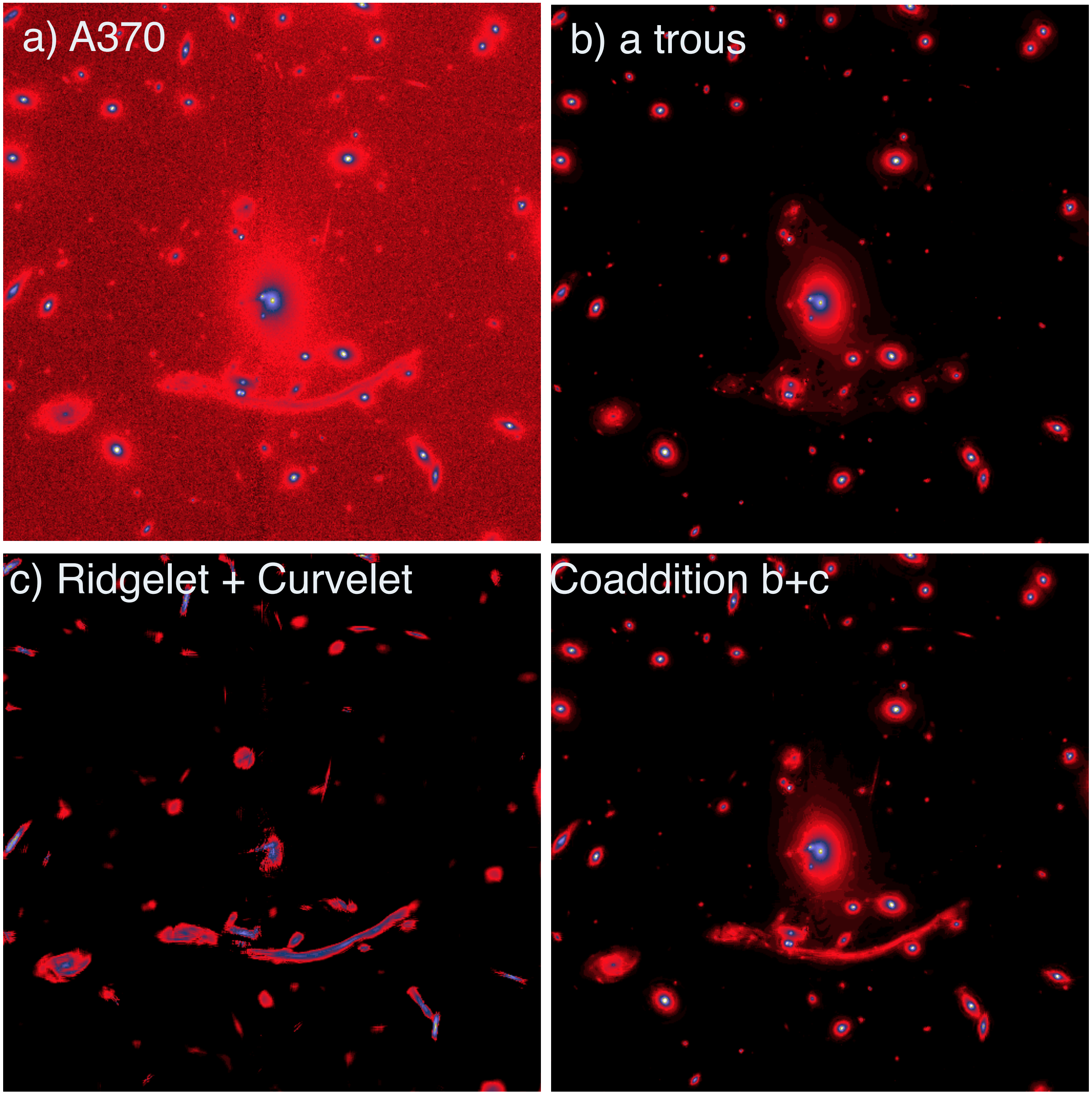
$$\underset{s_1, s_2}{\text{minimize}} \, ||W s_1||_p + ||C s_2||_p \text{ subject to } ||s - (s_1 + s_2)||_2^2 < \epsilon$$



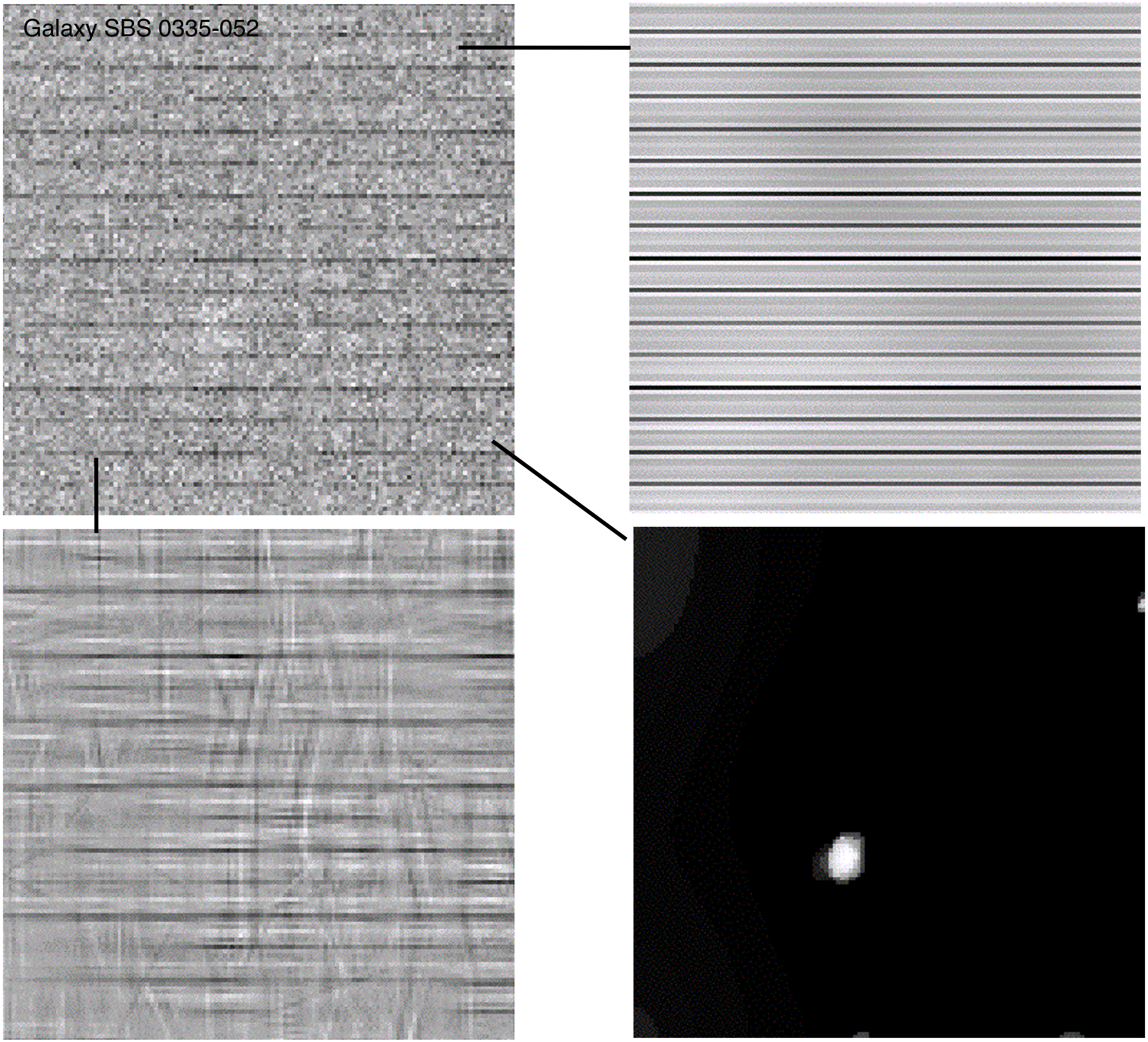
a) Simulated image (gaussians+lines) b) Simulated image + noise c) A trous algorithm



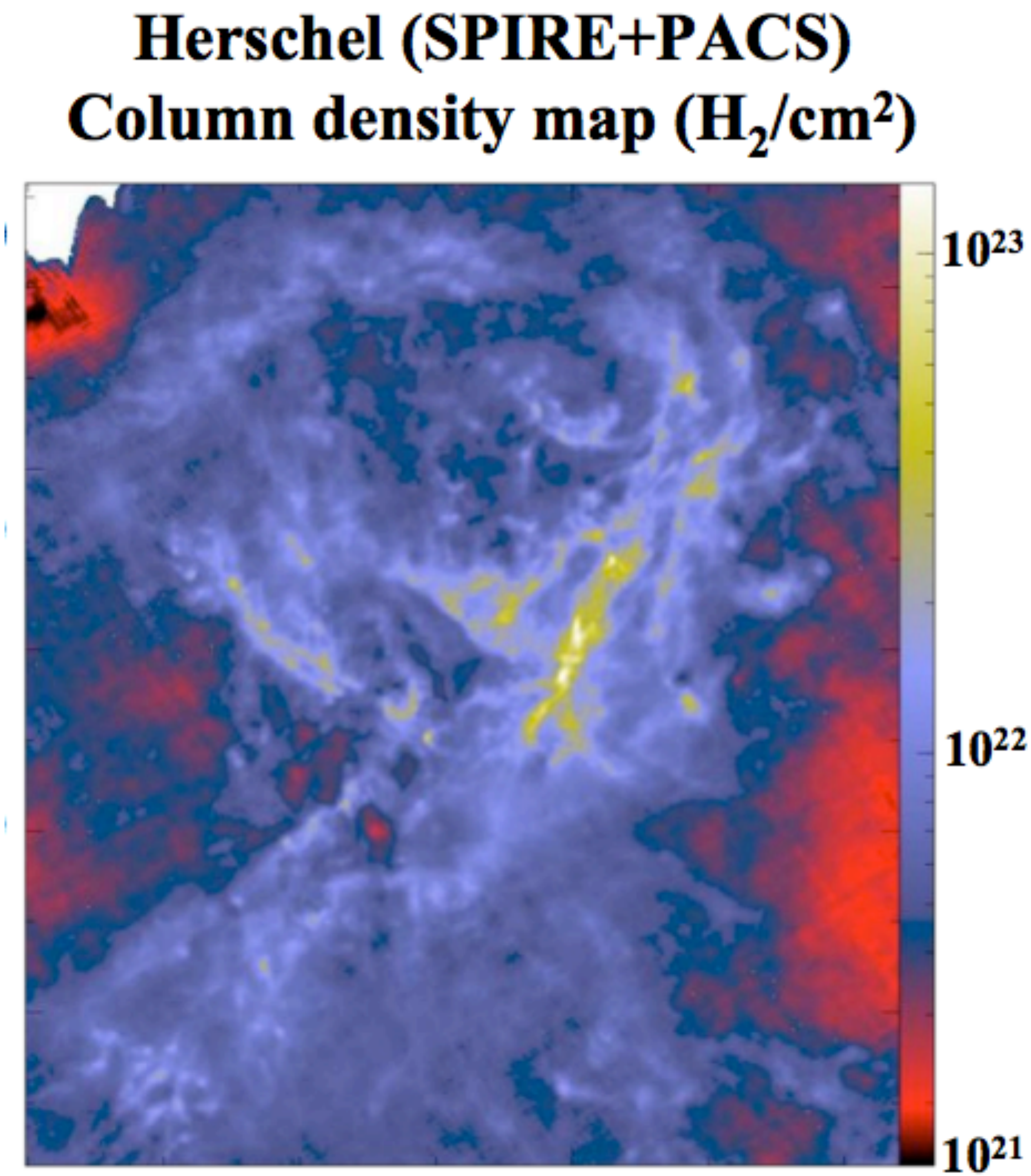
d) Curvelet transform e) coaddition c+d f) residual = e-b



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


Galaxy SBS 0335-052

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



Dense cores form primarily in filaments

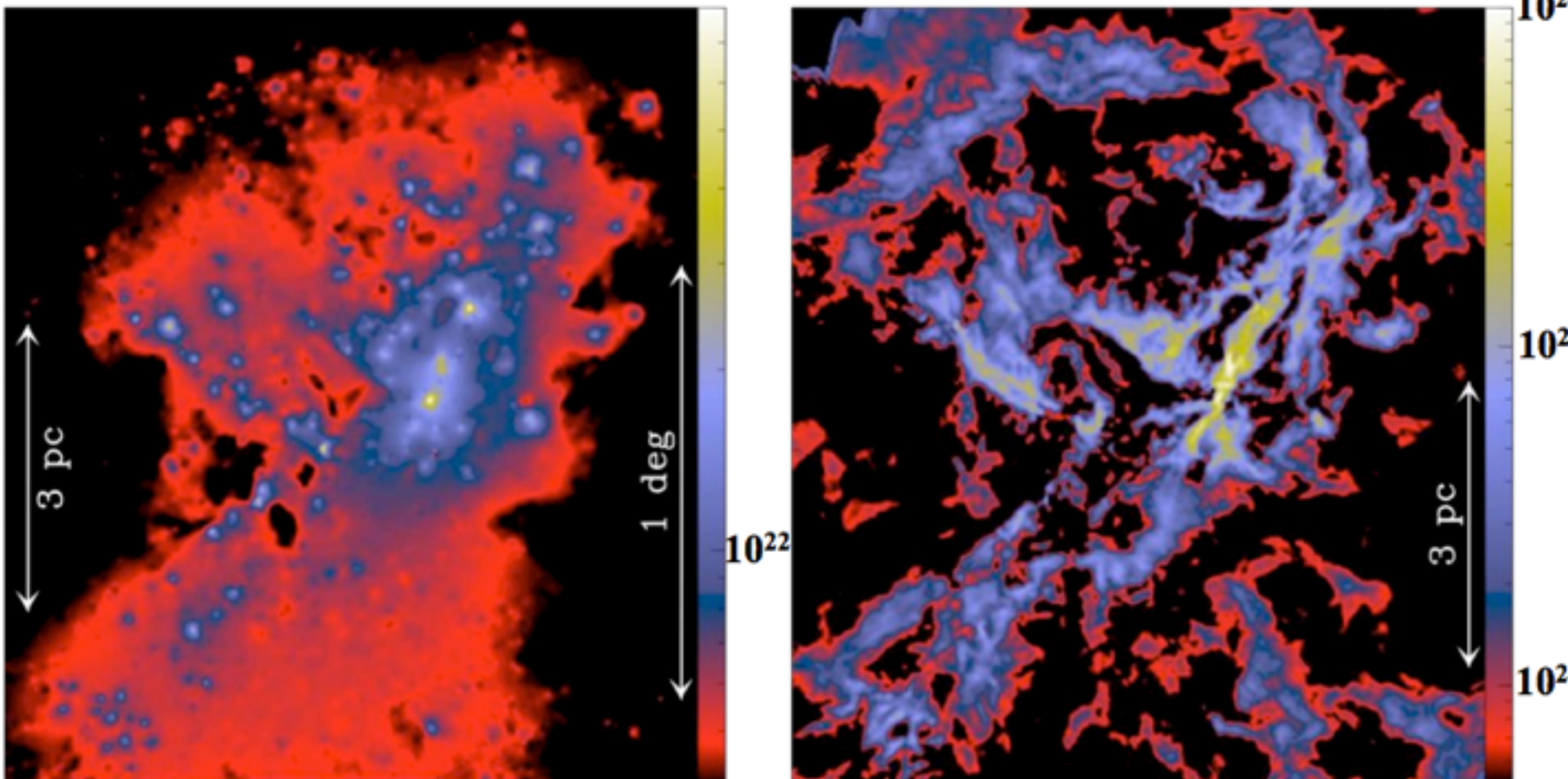
Morphological Component Analysis:

***Herschel* Column density map**

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

=

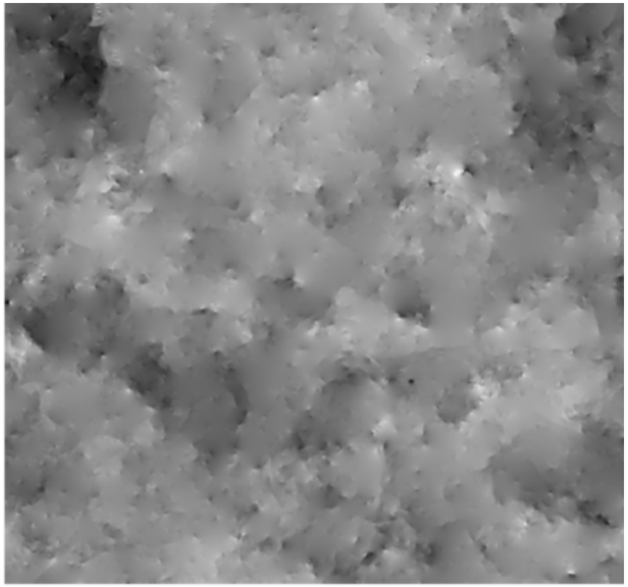
Cores **Wavelet component (H_2/cm^2)** **+** Filaments **Curvelet component (H_2/cm^2)**



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

Dictionary Learning

- Which representation best for given signals/tasks?
- When training data is available (simulations, multiple exemplars), why not learning an adapted representation?



Simulated Cosmic String Map

- Sparse dictionary learning (DL):

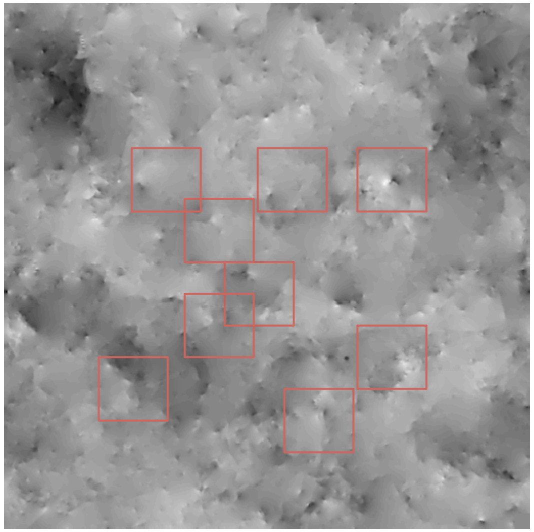
$$\underset{\mathbf{D}\in\mathcal{D},\boldsymbol{\Lambda}\in\mathcal{C}}{\arg\min} \sum_{(i,j)\in\mathcal{T}} \|\mathbf{x}_{ij} - \mathbf{D}\boldsymbol{\lambda}_{ij}\|_2^2 + \mu \cdot \|\boldsymbol{\lambda}_{ij}\|_p$$

Sparse Dictionary Learning

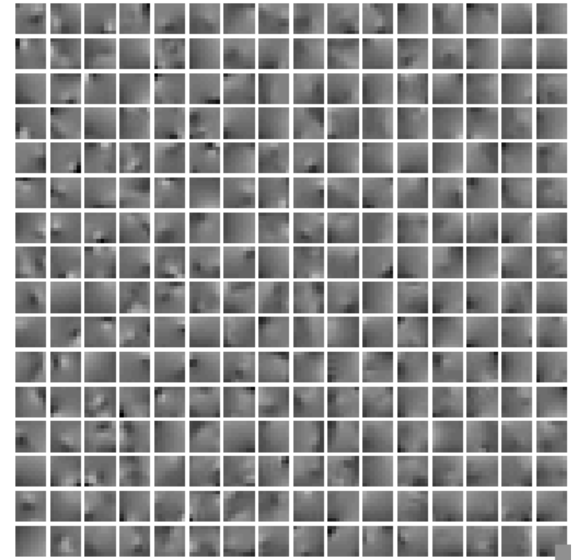
$$\underset{\mathbf{D}\in\mathcal{D},\boldsymbol{\Lambda}\in\mathcal{C}}{\arg\min} \sum_{(i,j)\in\mathcal{T}} \|\mathbf{x}_{ij} - \mathbf{D}\boldsymbol{\lambda}_{ij}\|_2^2 + \mu \cdot \|\boldsymbol{\lambda}_{ij}\|_p$$

- Non-convex matrix factorization problem
- Biconvex if $p = 1$, $\mathcal{D}$ and $\mathcal{C}$ convex sets (alternated minimization)
- Various algorithmic approaches for sparse coding/dictionary updating: MOD (Engan99), K-SVD (Aharon06, Elad06), PALM (Bolte14), online algorithms (Mairal 2010)...
- Computationally costly: small-scales problems (patch-based DL: eg Elad06, Mairal08, Yang10, Mairal14), sparse convolutional DL (Bristow13,Heide15, Wohlberg16,Papyan17)
- vast literature on subject: multivariate (eg Mairal08), multiscale (eg Mairal08, Ophir11), translation-invariant (eg Jost06,Aharon08), hierarchical (eg Jenatton11)
- various DL approaches for inverse problems: using coupled DL (eg Rubinstein14), as analysis prior (eg Rubinstein13), task-driven DL (eg Mairal10), double sparse model (eg Rubinstein10, Sulam16)...

Dictionary Learning



Training basis.

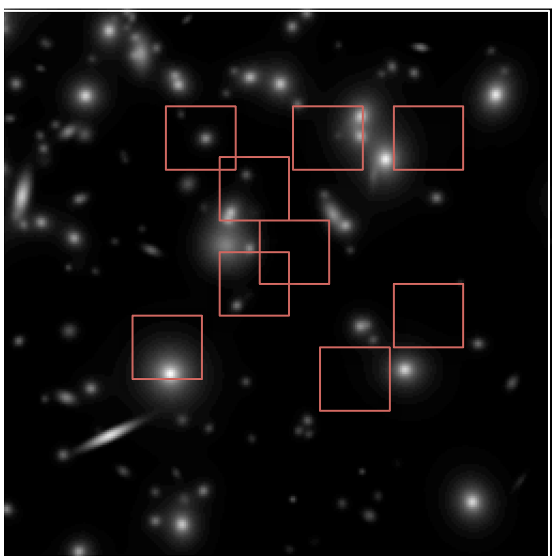


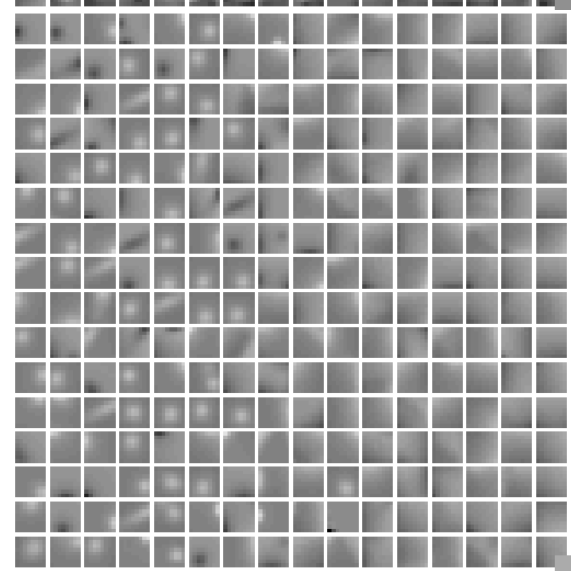
$$\left(\hat{D},\hat{A}\right) = \underset{\substack{D\in C_1\\ A\in C_2}}{\arg\min}\left(Y = DA\right)$$

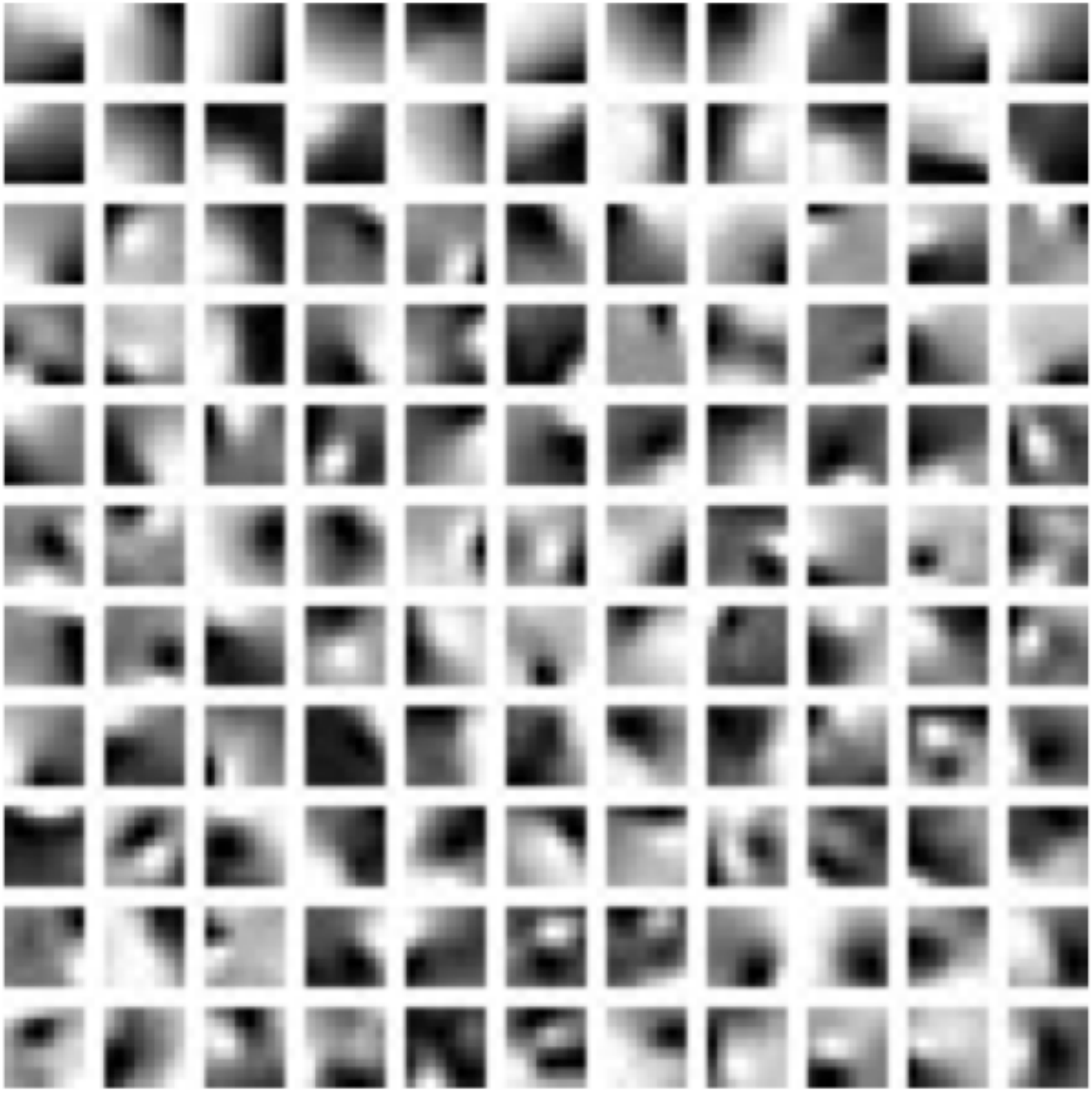
DL: Matrix Factorization problem

C_1: Constraints on the Sparsifying dictionary D

C_2: Constraints on the Sparse codes

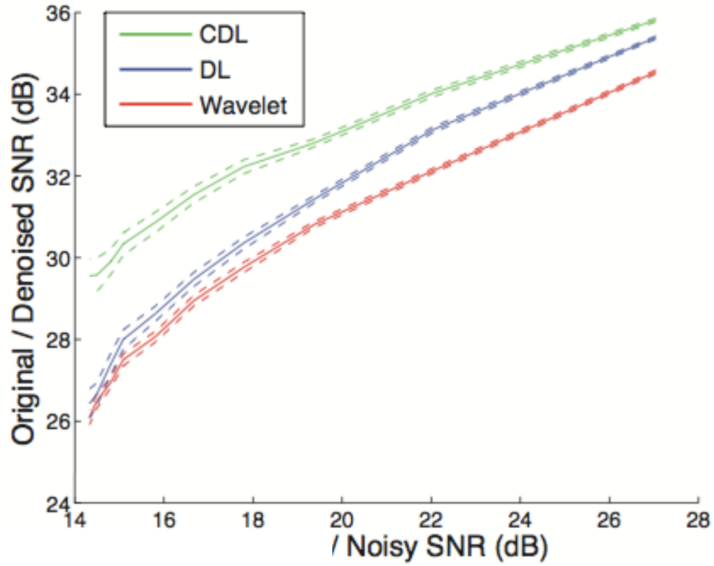


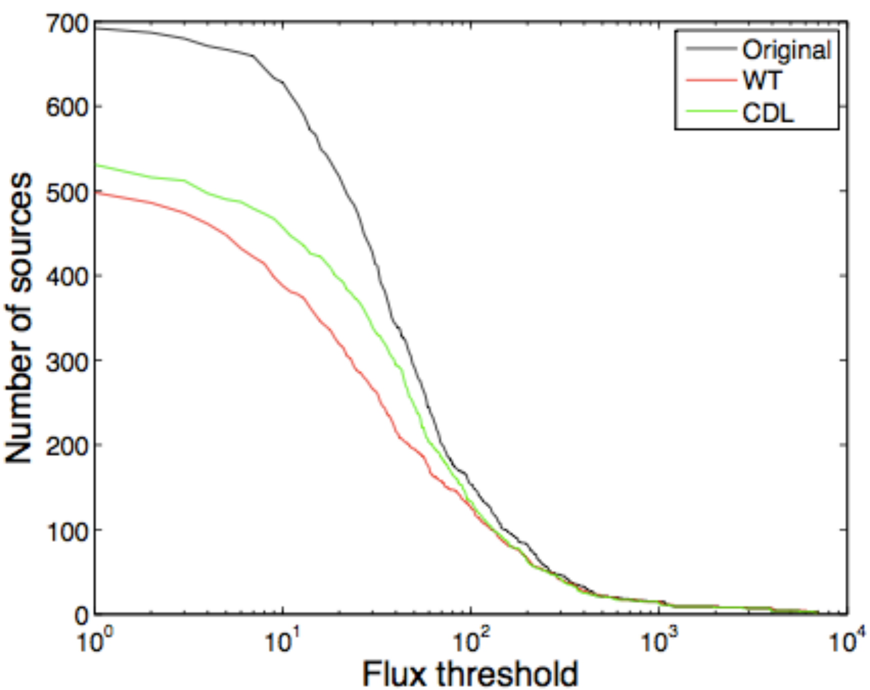


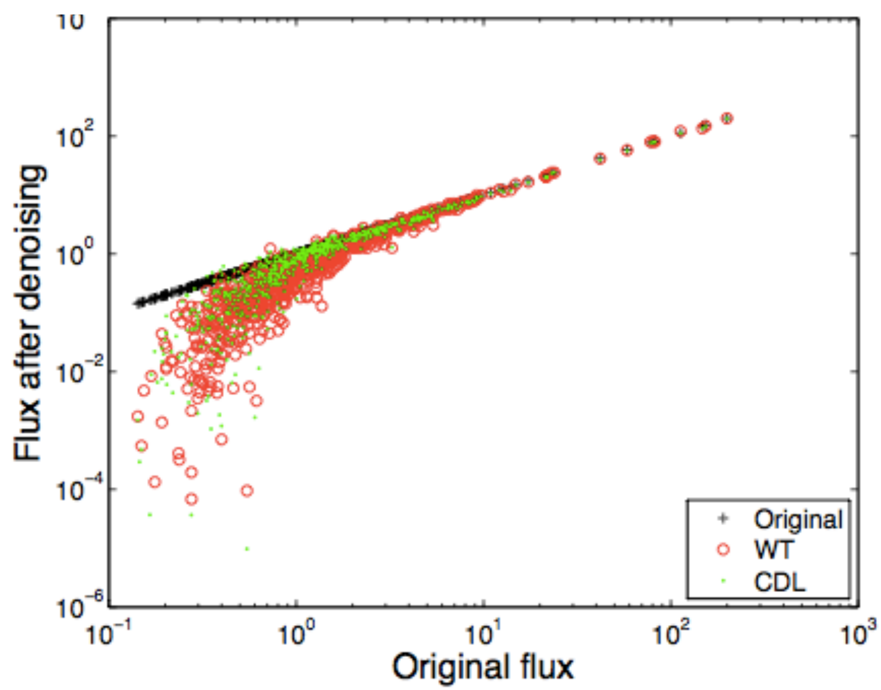


S. Beckouche

Astronomical Image Denoising Using Dictionary Learning, S. Beckouche, J.L. Starck, and J. Fadili, A&A, 2013.



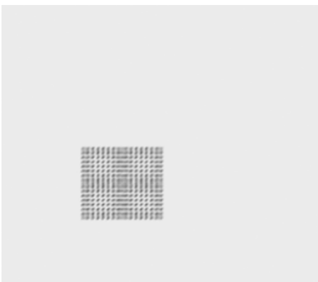




Sparsity Model 1: we consider a dictionary $\Phi = \{\phi_1, \ldots, \phi_K\}$ which has a fast transform/reconstruction operator:

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

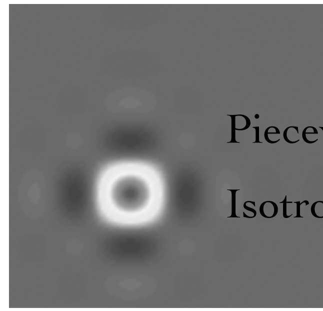
<u>Local DCT</u>



Stationary textures

Locally oscillatory

<u>Wavelet transform</u>



Piecewise smooth

Isotropic structures

<u>Curvelet transform</u>

Piecewise smooth,
edge

Sparsity Model 2: Morphological Diversity:

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

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

Model 3 can be also combined with model 2:

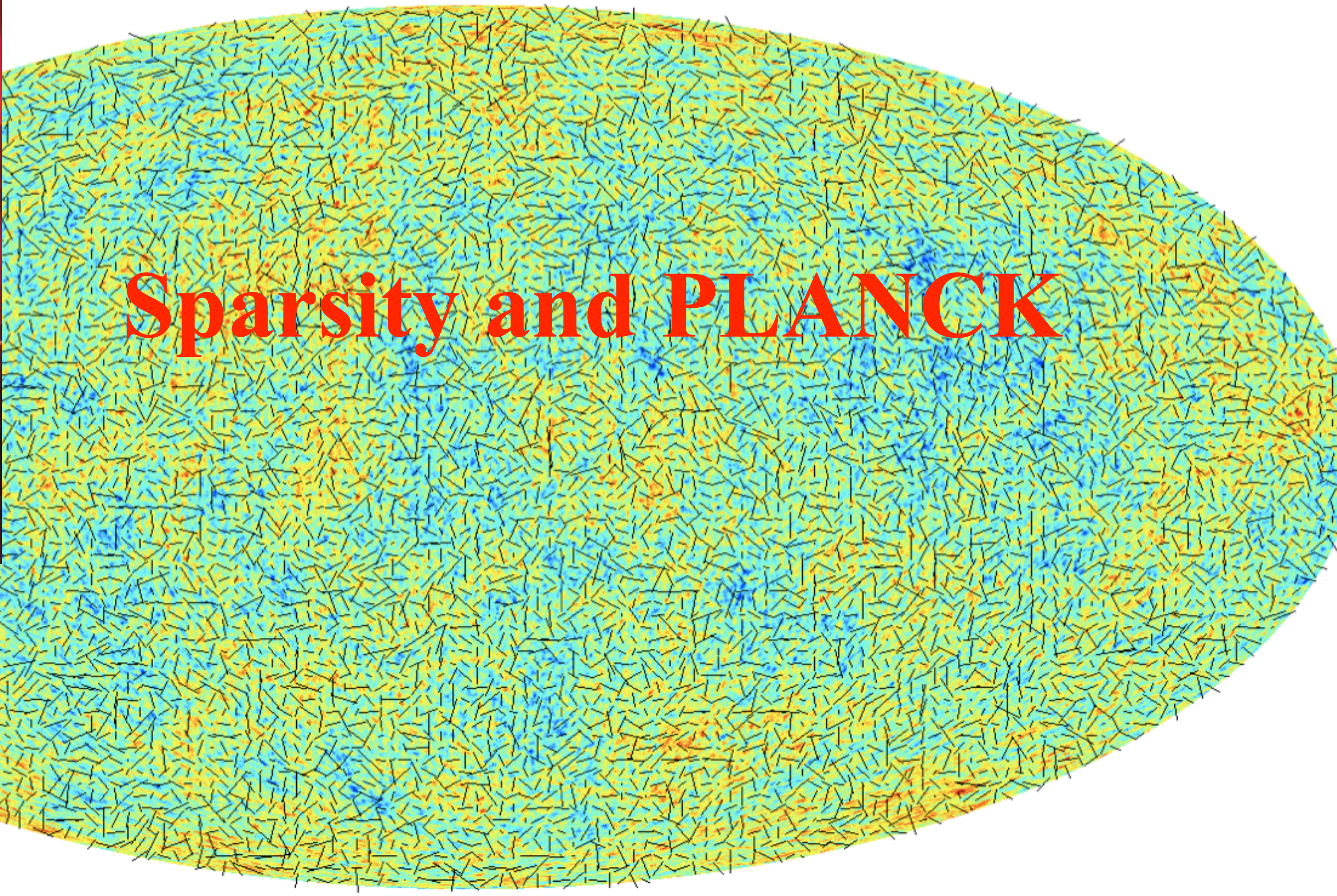
G. Peyre, M.J. Fadili and J.L. Starck, , "Learning the Morphological Diversity", **SIAM Journal of Imaging Science**, 3 (3) , pp.646-669, 2010.

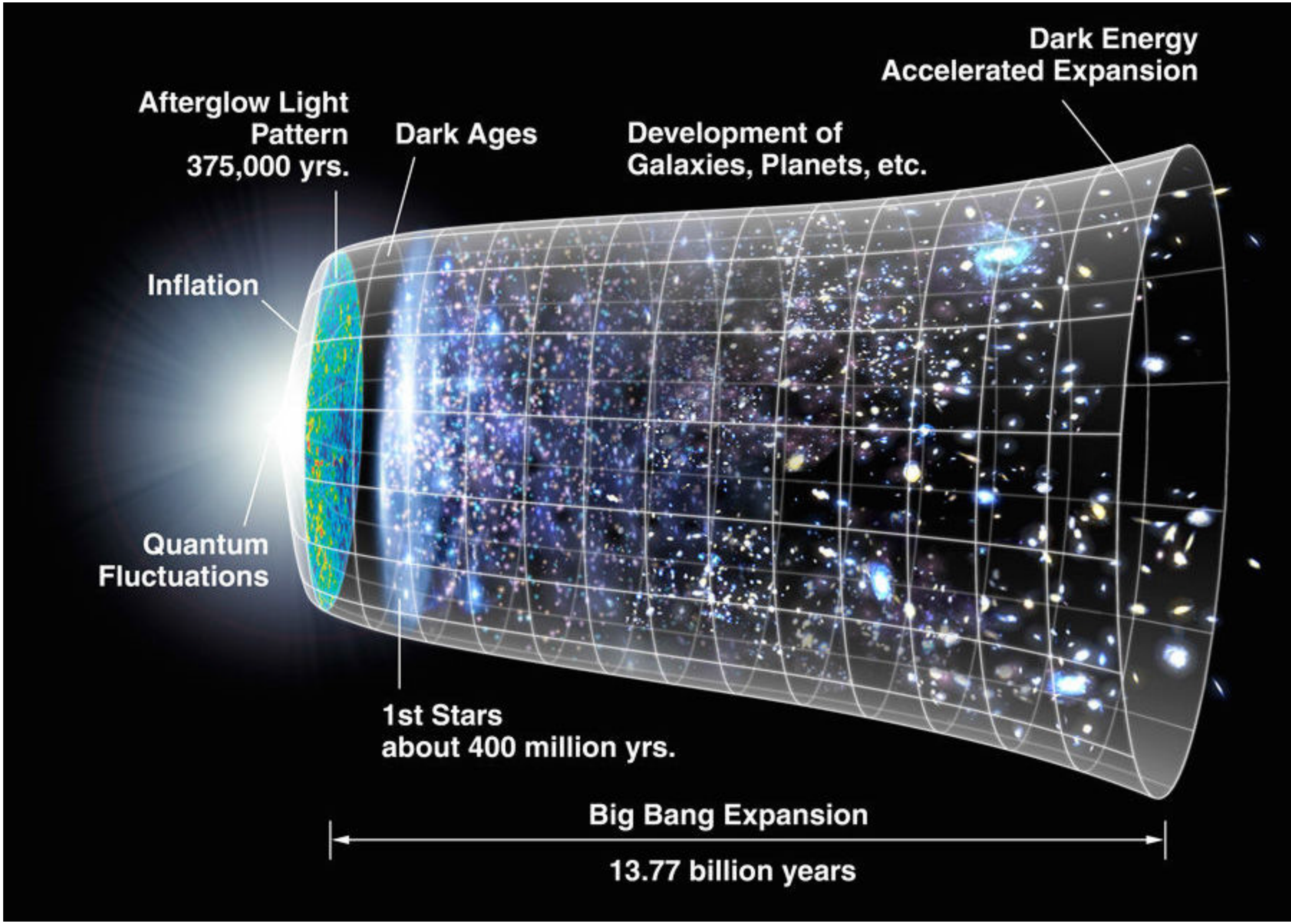
Advantages of model 1 (fixed dictionary) : extremely fast.

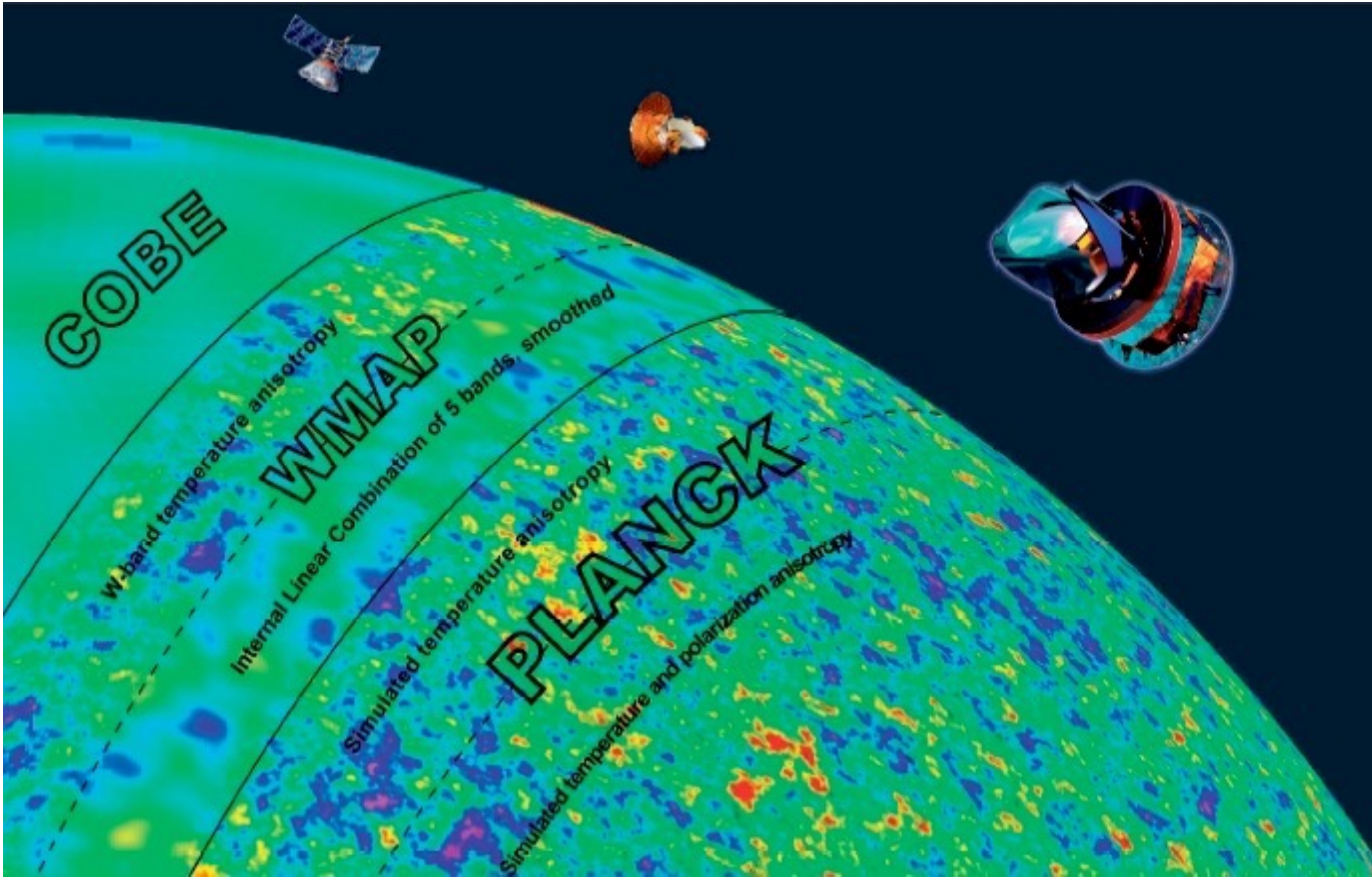
Advantages of model 2 (union of fixed dictionaries):
- more flexible than model 1.
- The coupling of local DCT+curvelet is well adapted to a relatively large class of images.

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

Sparsity and PLANCK

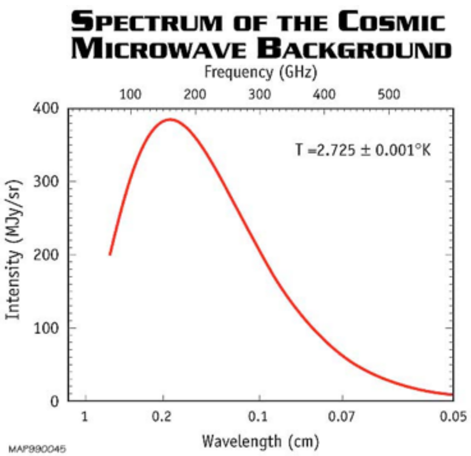



Dark Energy
Accelerated Expansion
Afterglow Light
Pattern
375,000 yrs.
Dark Ages
Development of
Galaxies, Planets, etc.
Inflation
Quantum
Fluctuations
1st Stars
about 400 million yrs.
Big Bang Expansion
13.77 billion years

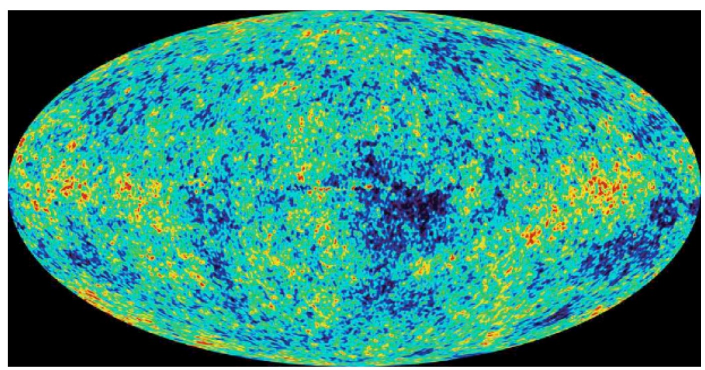

COBE
W-band temperature anisotropy
WMAP
Internal Linear Combination of 5 bands, smoothed
Simulated temperature anisotropy
PLANCK
Simulated temperature and polarization anisotropy

CMB Radiation

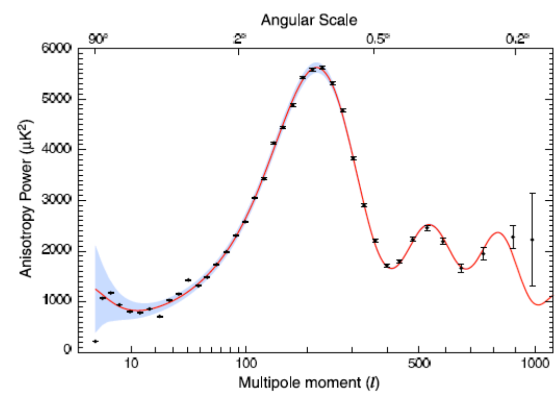
- Thermal Radiation with black body spectrum
- Highly isotropic but wealth of information in temperature anisotropies: $2.7K \pm 10^{-5}K$
- Linearly polarized signal (but one order weaker) measured by WMAP and Planck
- "Snapshot" of the universe at recombination
- Standard Model: fluctuations gaussian isotropic, i.e. fully characterized by power spectra



NASA

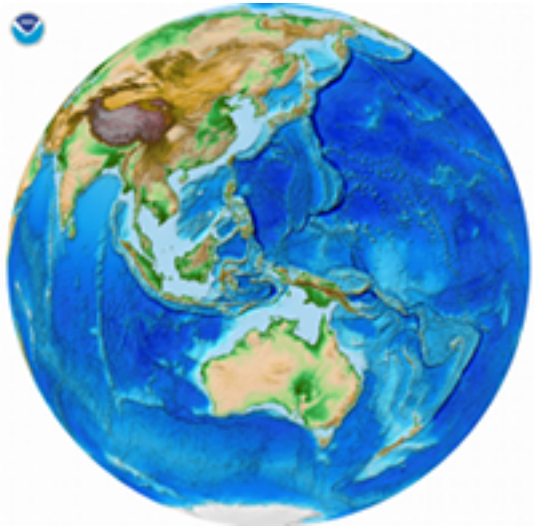


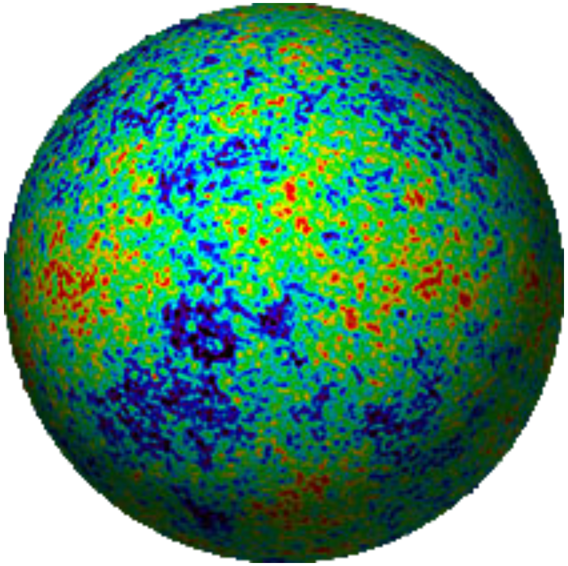
WMAP collaboration

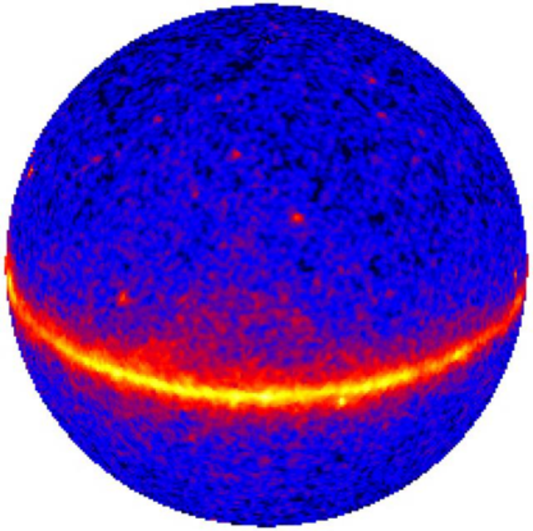


WMAP collaboration

First Challenge: Sparse Tools on the Sphere



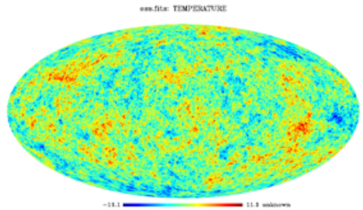


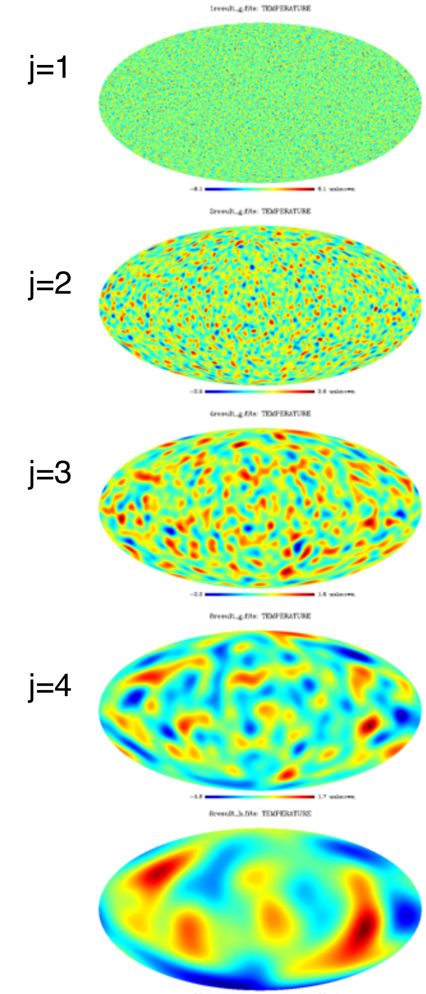


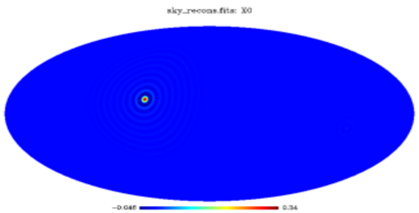
Isotropic Undecimated Wavelet on the Sphere

Wavelets, Ridgelets and Curvelets on the Sphere, Astronomy & Astrophysics, 446, 1191-1204, 2006.

Undecimated Wavelet Transform







$$\hat{\psi}_{\frac{l_c}{2^j}}(l,m) = \hat{\phi}_{\frac{l_c}{2^{j-1}}}(l,m) - \hat{\phi}_{\frac{l_c}{2^j}}(l,m)$$

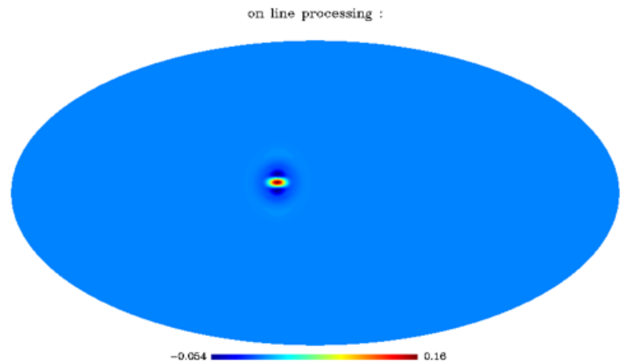
$$\hat{H}_j(l,m) = \begin{cases} \dfrac{\hat{\phi}_{\frac{l_c}{2^{j+1}}}(l,m)}{\hat{\phi}_{\frac{l_c}{2^j}}(l,m)} & \text{if } l < \frac{l_c}{2^{j+1}} \quad \text{and} \quad m = 0 \\ 0 & \text{otherwise} \end{cases}$$

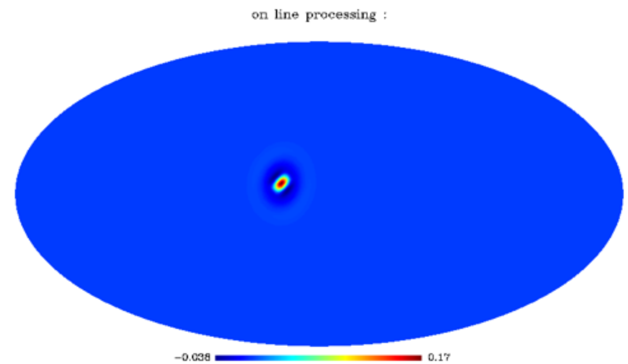
$$\hat{G}_j(l,m) = \begin{cases} \dfrac{\hat{\psi}_{\frac{l_c}{2^{j+1}}}(l,m)}{\hat{\phi}_{\frac{l_c}{2^j}}(l,m)} & \text{if } l < \frac{l_c}{2^{j+1}} \quad \text{and} \quad m = 0 \\ 1 & \text{if } l \geq \frac{l_c}{2^{j+1}} \quad \text{and} \quad m = 0 \\ 0 & \text{otherwise} \end{cases}$$

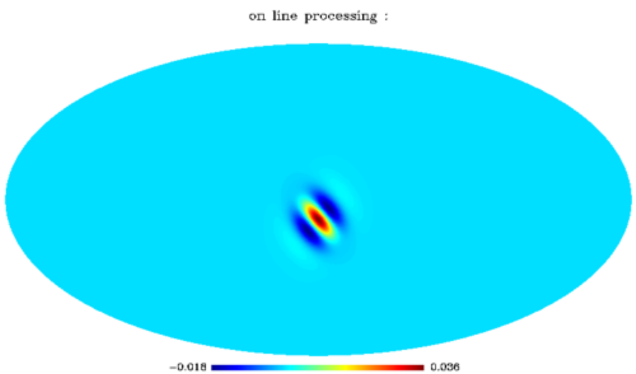
$$\hat{c}_{j+1}(l,m) = \widehat{H}_j(l,m)\hat{c}_j(l,m)$$
$$\hat{w}_{j+1}(l,m) = \widehat{G}_j(l,m)\hat{c}_j(l,m)$$

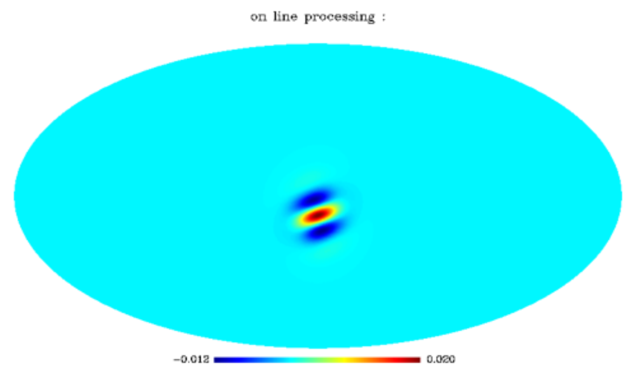
$$\boxed{c_0(\vartheta,\varphi) = c_J(\vartheta,\varphi) + \sum_{j=1}^{J} w_j(\vartheta,\varphi)}$$

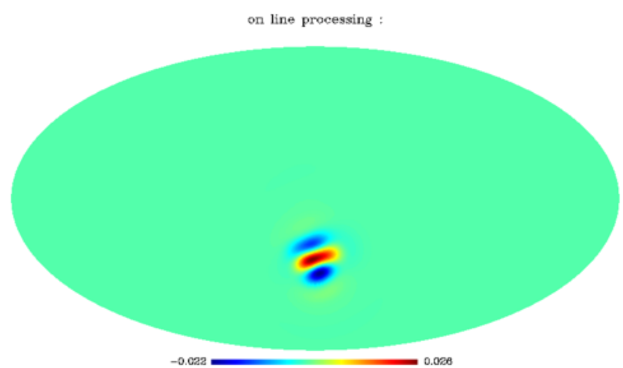
Curvelet Transform on the Sphere

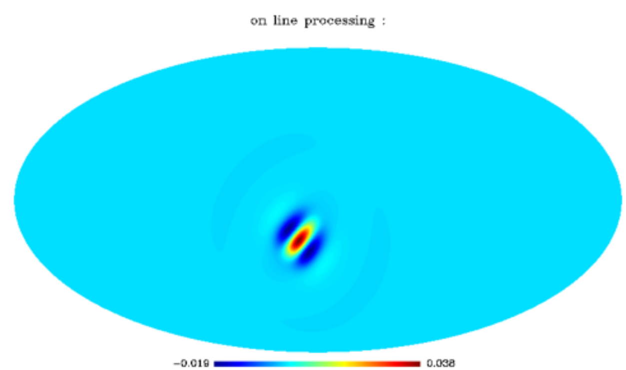




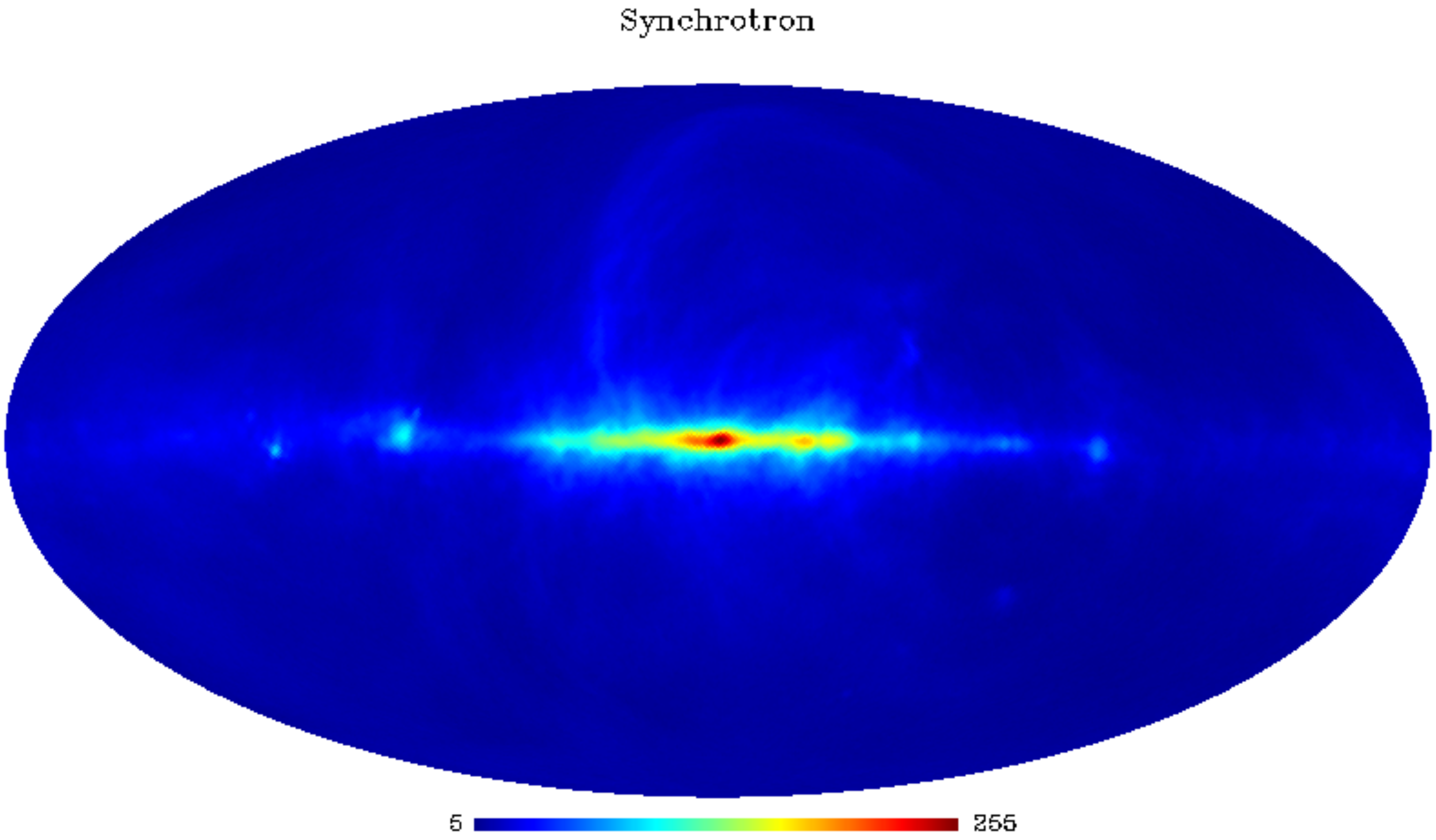




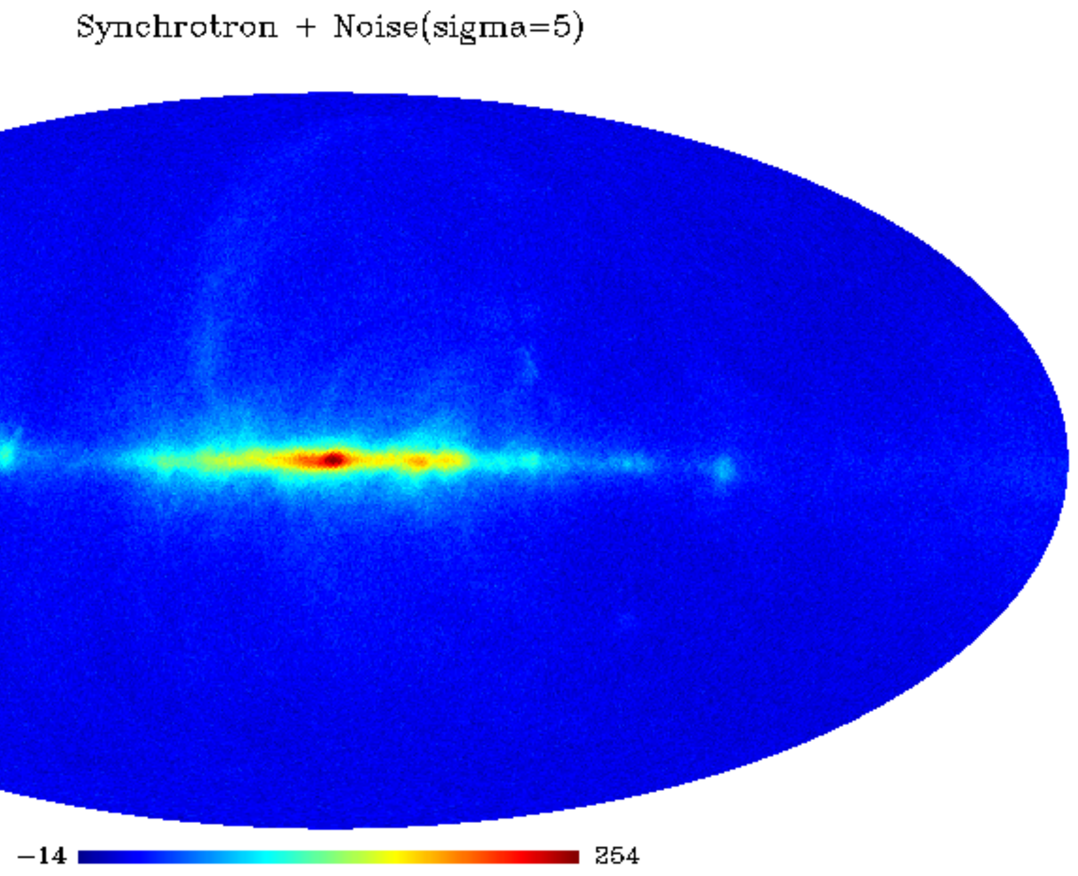




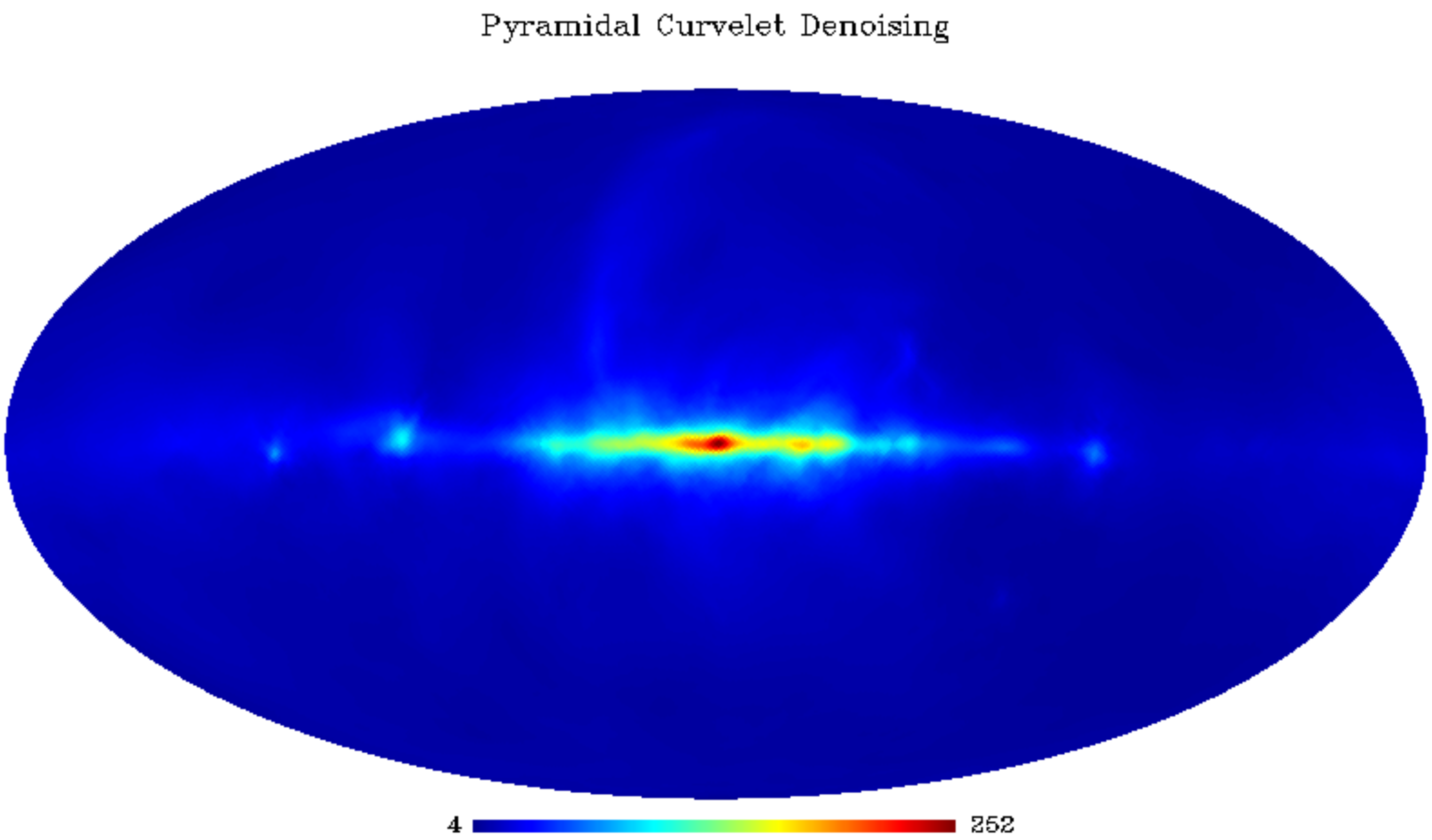
Denoising Example



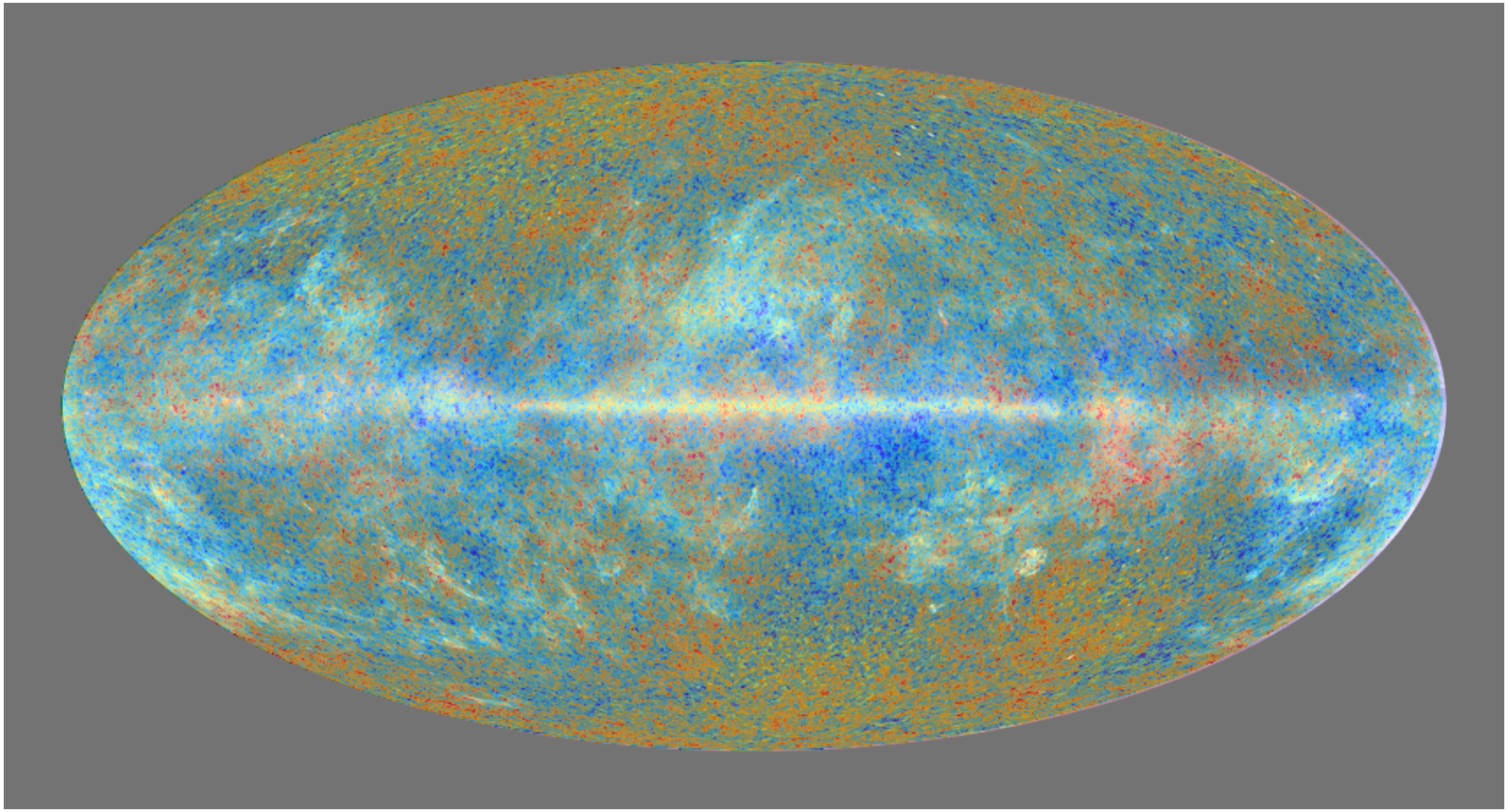
Denoising Example



Denoising Example



2nd Challenge : Blind Source Separation



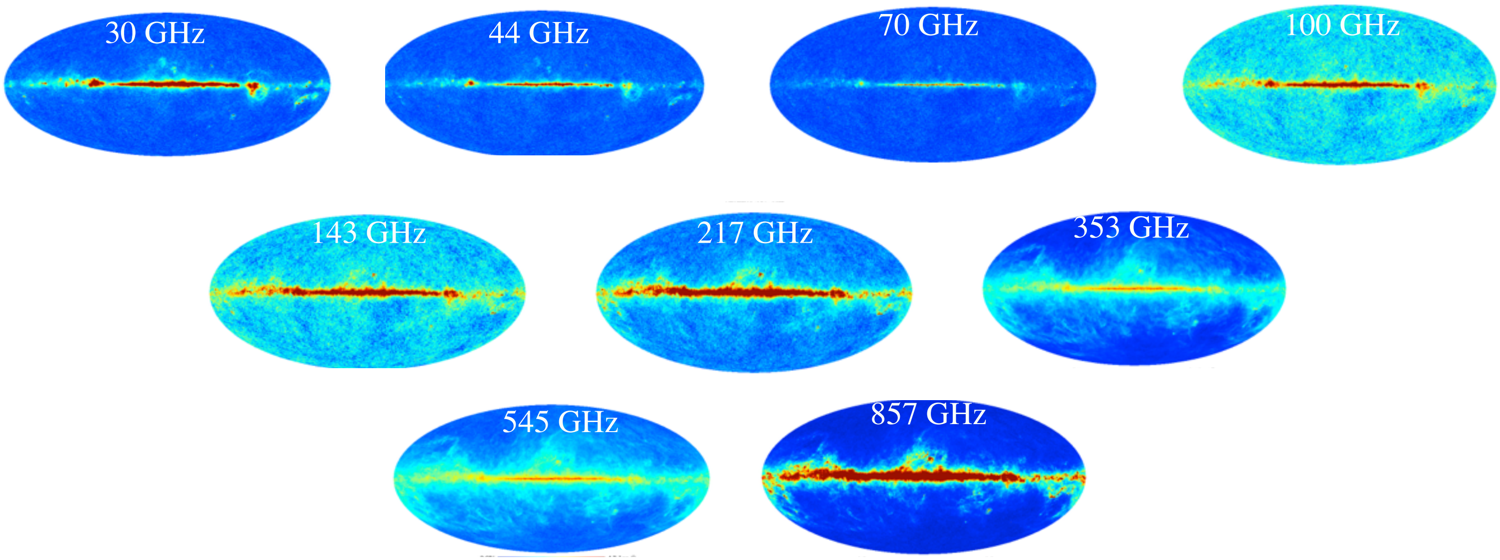
The importance of Source Separation

Extra foregrounds are superimposed with the CMB !!!

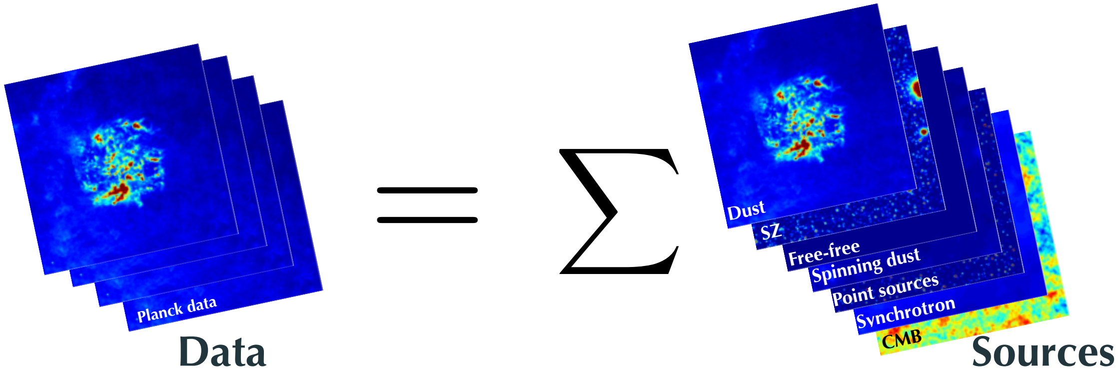
Point sources, galactic foregrounds, ... etc

Blind Source Separation

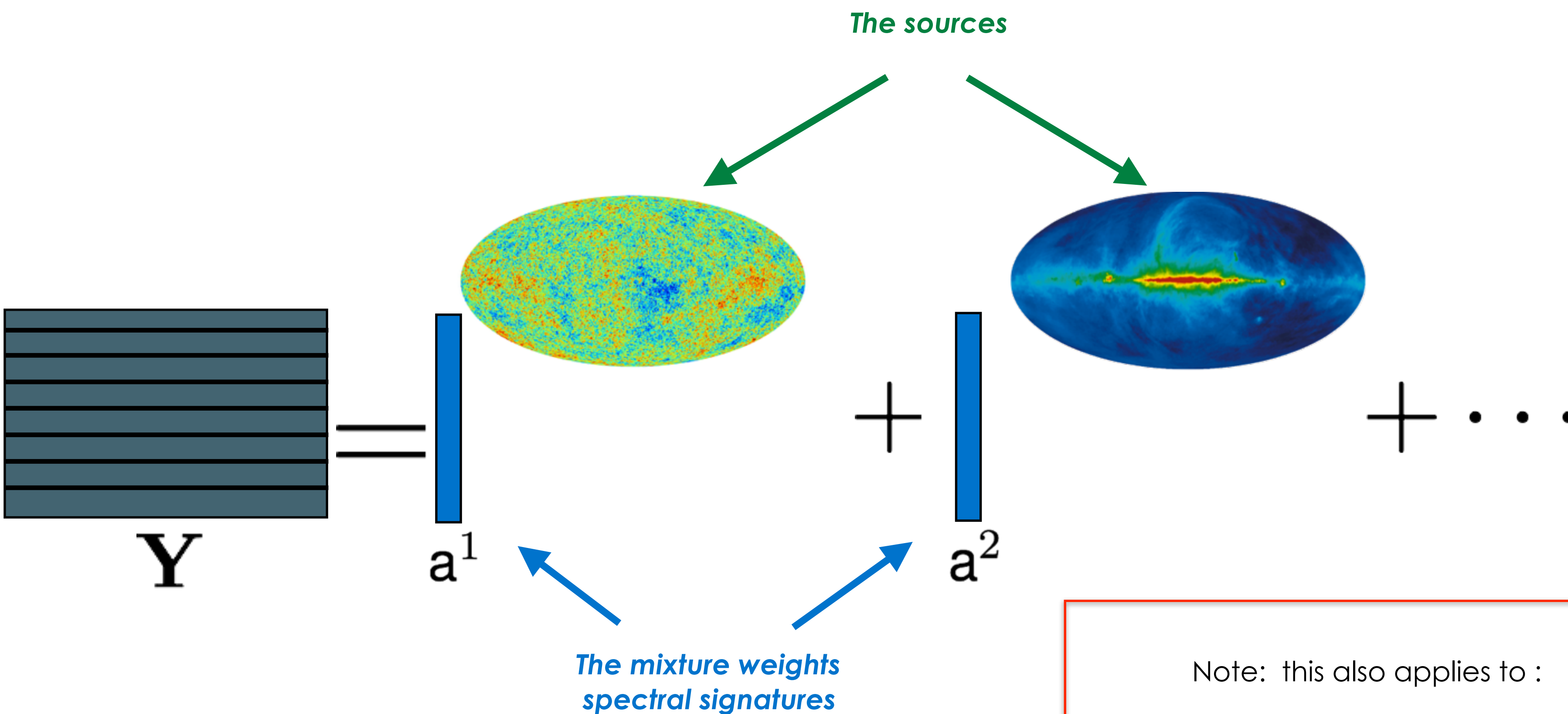
- Planck: 9 channels with different resolution and noise properties



- Each channel is a superposition of emissions



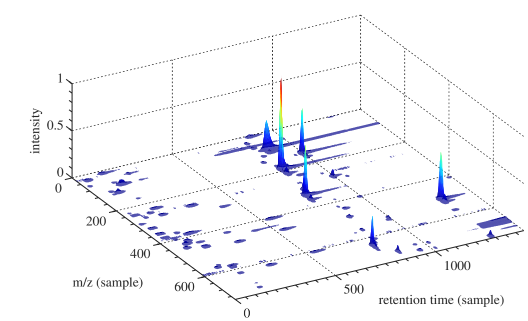
Forward model: Linear Mixture



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

Blind Source Separation:
Estimation both A and X from Y only
Matrix Factorization Problem

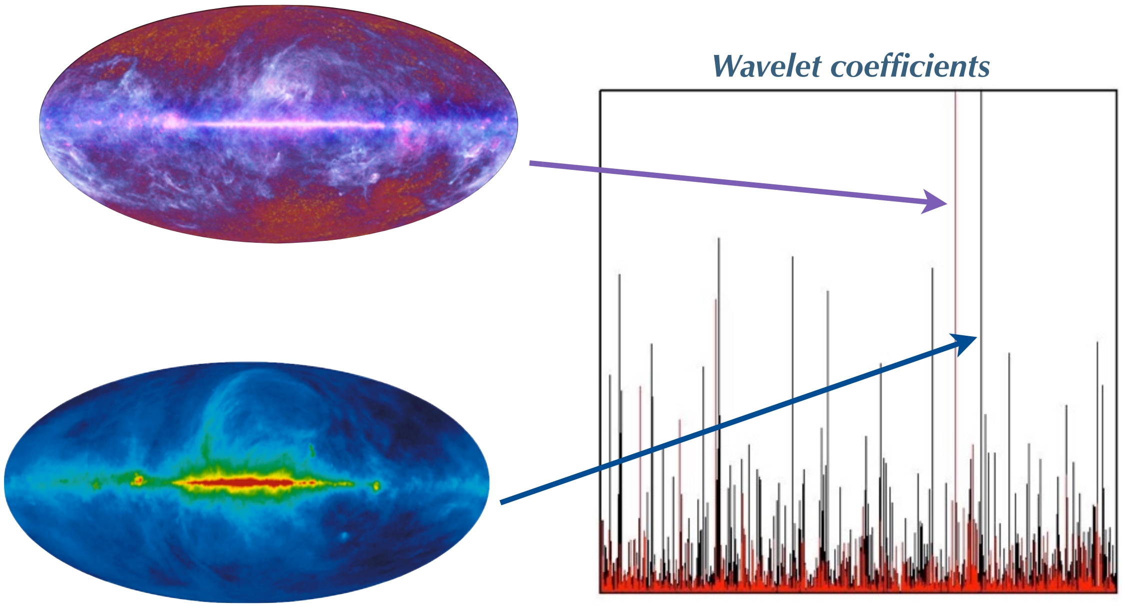
Note: this also applies to :



Multi-measurement data
EEG, ECG, NMR, MS, LC-MS, etc.

Separation principle: Morphological diversity

Different sparsity patterns



Morphological Diversity +
Electromagnetic spectra of CMB, SZ…

- Inverse problem: $$\underset{\mathbf{A}\in\mathcal{C},\mathbf{X}}{\arg\min}\ \underbrace{\frac{1}{2}\|\mathbf{Y}-\mathbf{A}\mathbf{X}\|_F^2}_{\text{Data Fidelity Term}} + \underbrace{\|\boldsymbol{\Lambda}\odot\mathbf{X}\boldsymbol{\Phi}\|_p}_{\text{Weighted Sparse Prior}}$$

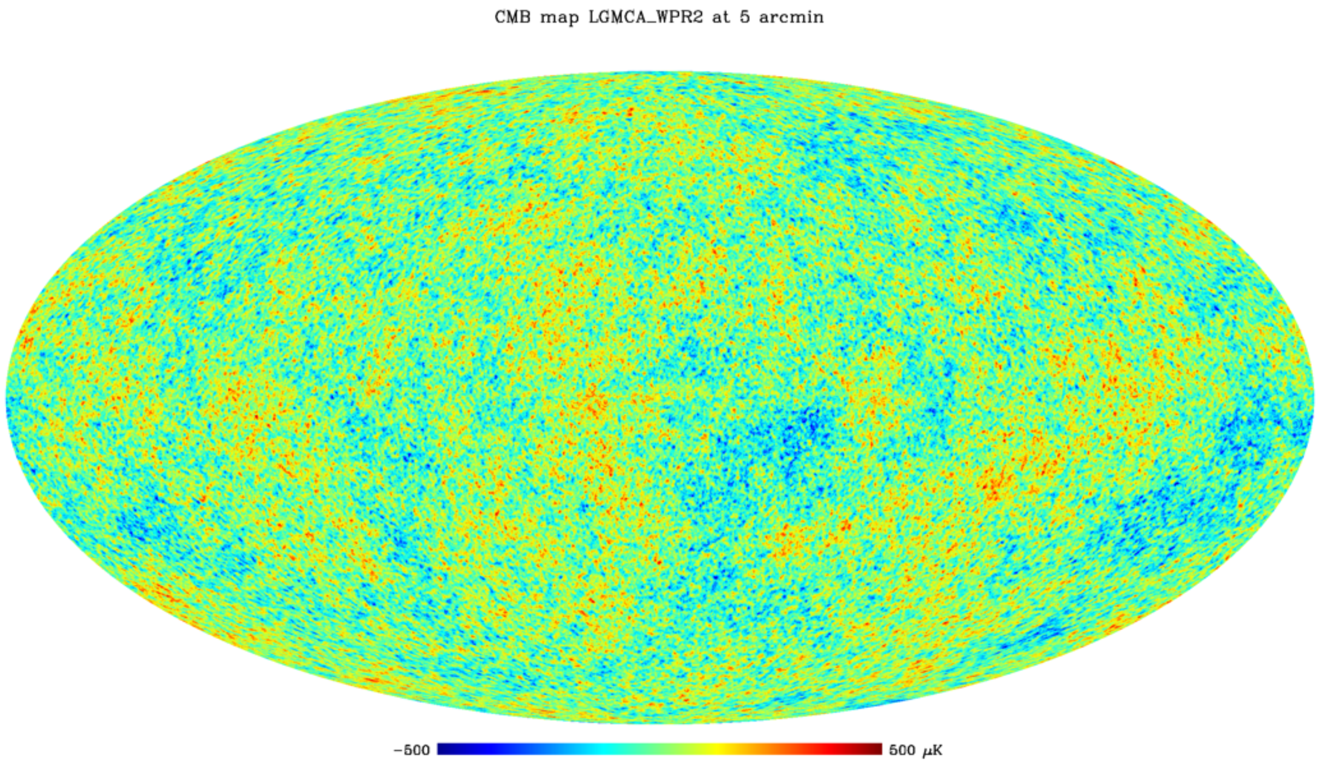
- GMCA is a variant of projected Alternating Least Squares (pALS), with sparse regularization

- Automatic and robust choice of thresholds

- Fast Algorithm

Bobin 2007, IEEE Trans. Sig. Proc. 13
Bobin 2008, IEEE Trans. Imag. Proc. 16

CMB Map W-PR2 by LGMCA

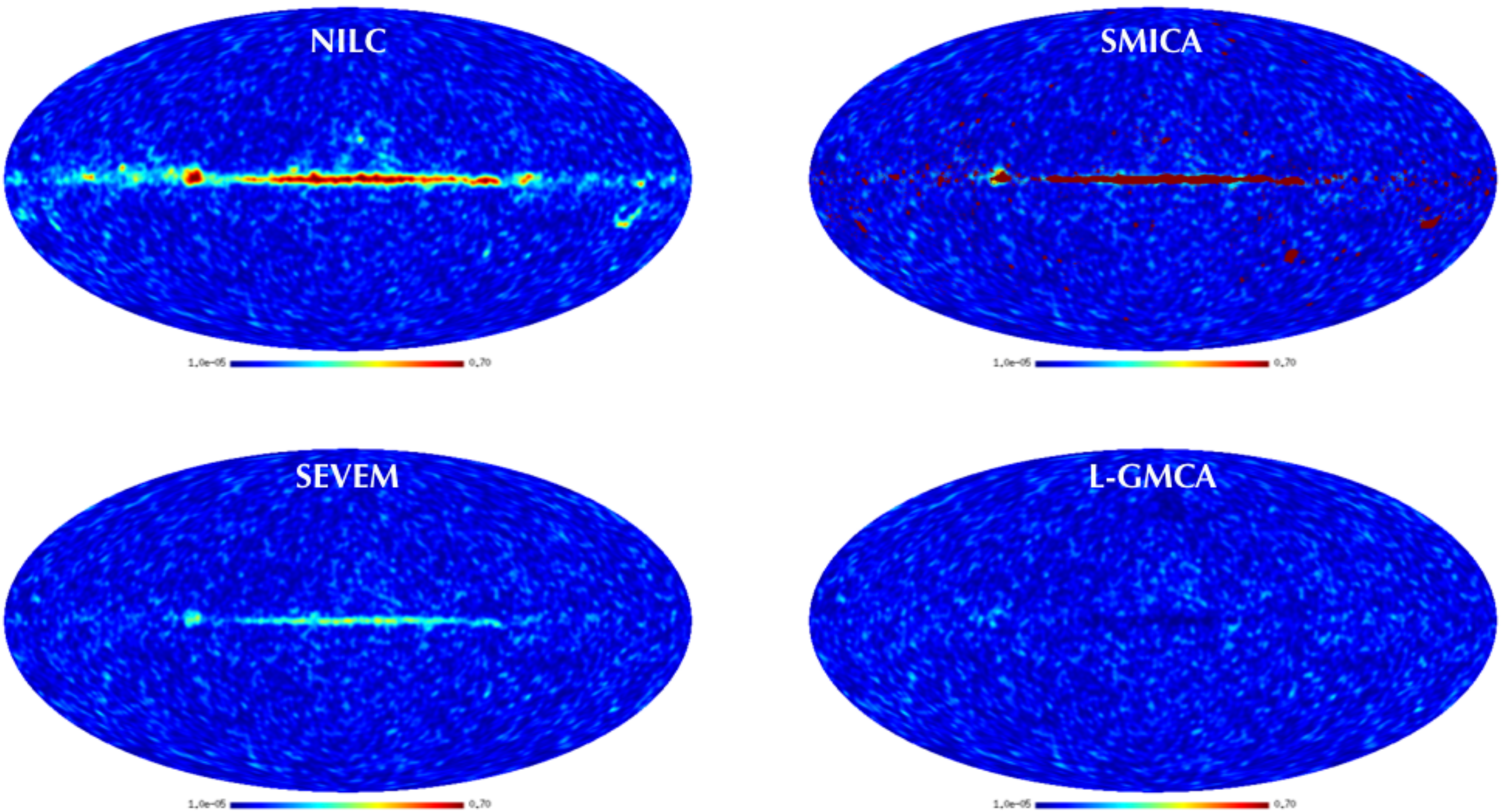


Bobin J., Sureau F., Starck J-L, Rassat A. and Paykari P., Joint Planck and WMAP CMB map reconstruction, A&A, 563, 2014

Bobin J. , Sureau F., Starck J.-L. , Polarized cosmic microwave background map recovery with sparse component separation A&A 583, 2015

Bobin J., Sureau F., Starck, CMB reconstruction from WMAP and Planck PR2 data, A&A, 591, 2016

Comparison with Planck CMB Maps



Measuring an excess of CMB power with the quality map