Space & Cosmology:
Tackling Big Data from the Sky

Jean-Luc Starck

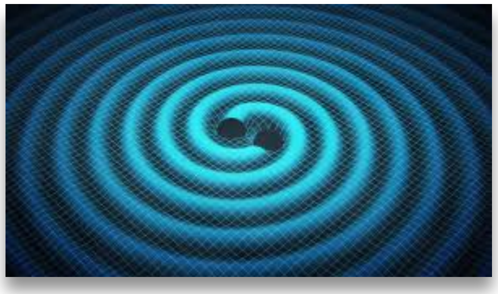
http://jstarck.cosmostat.org

Space & Cosmology

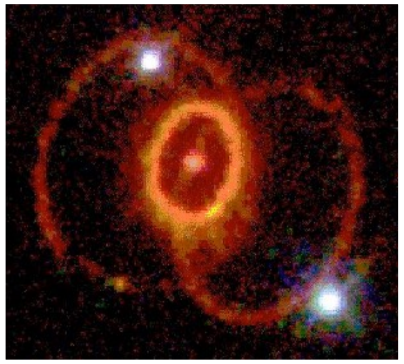
- ➡**Part 1: Introduction to Accurate Space Cosmology**
- ➡Part 2: Inverse Problems
- ➡Part 3: Euclid Weak Lensing

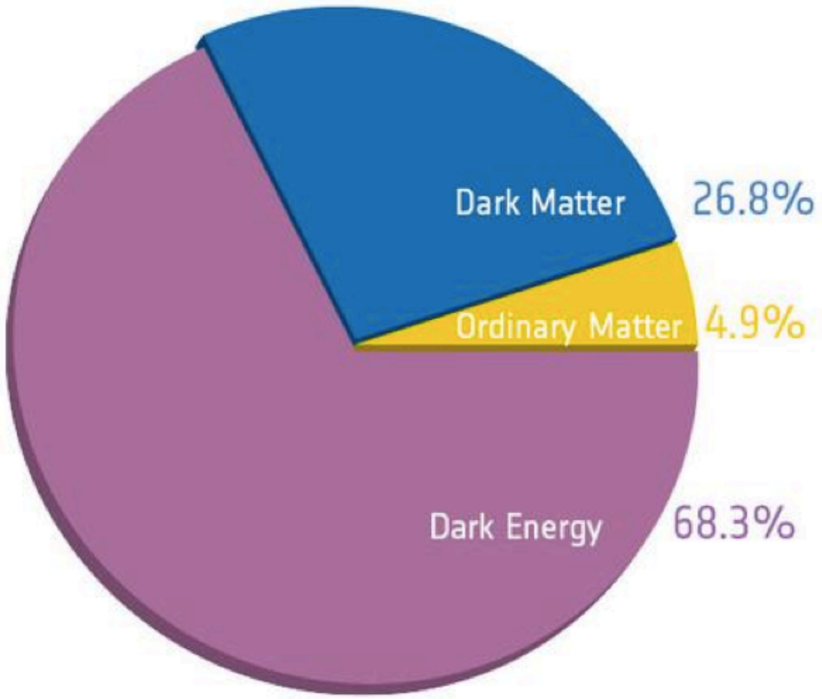
The Standard Cosmological Model

GW



Supernovae

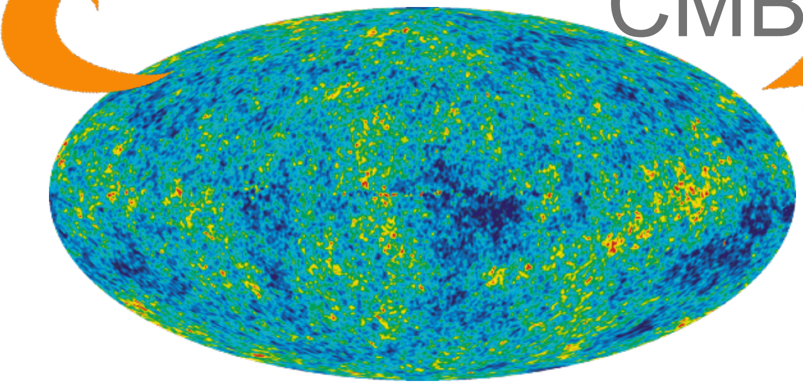


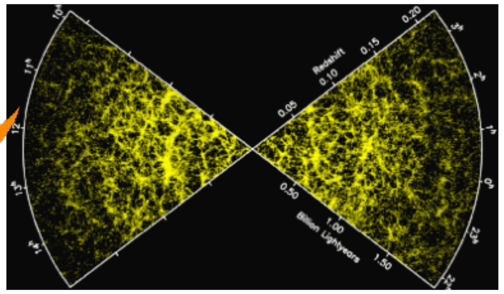


ESA/Planck

Lensing

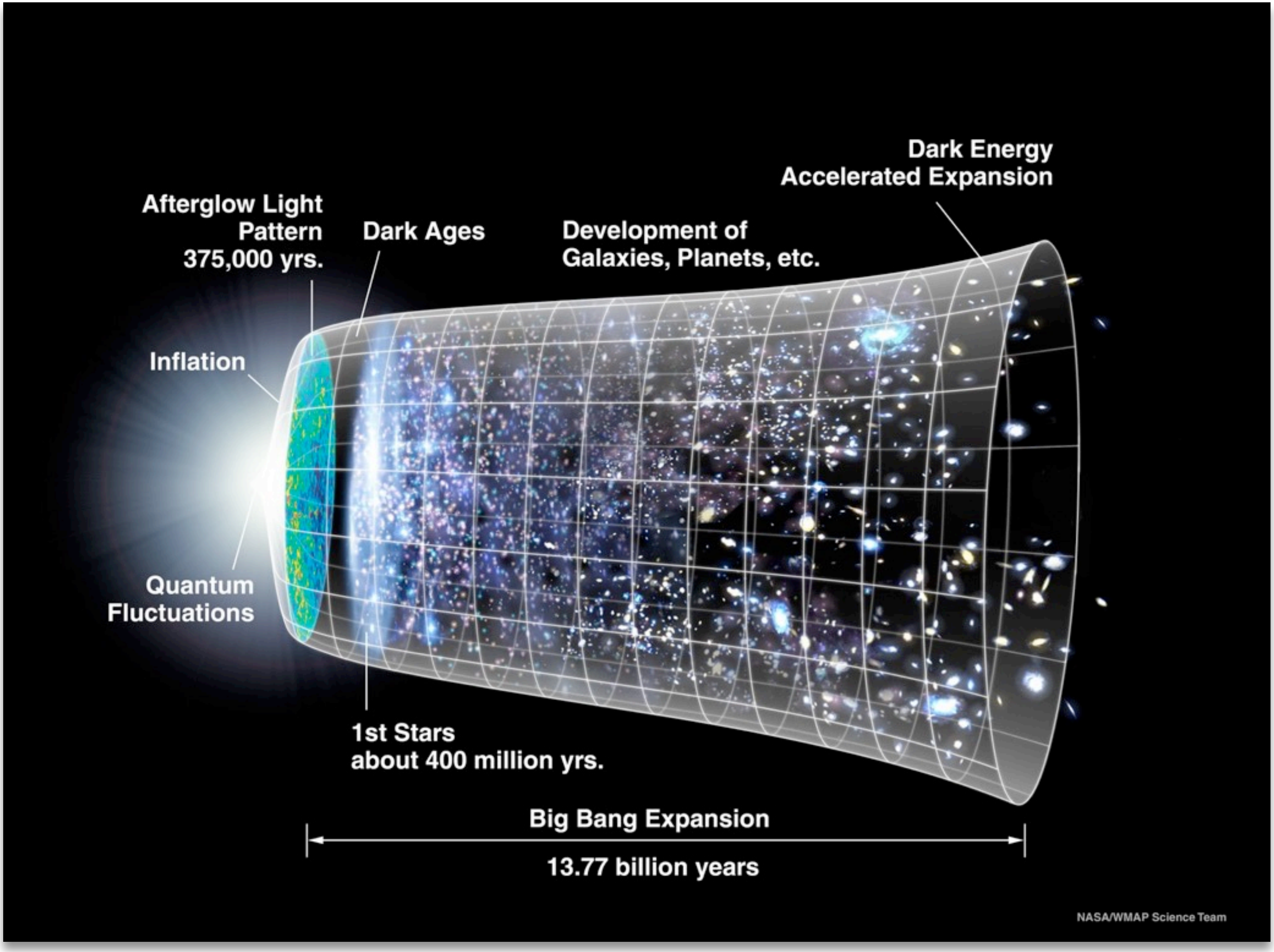
CMB





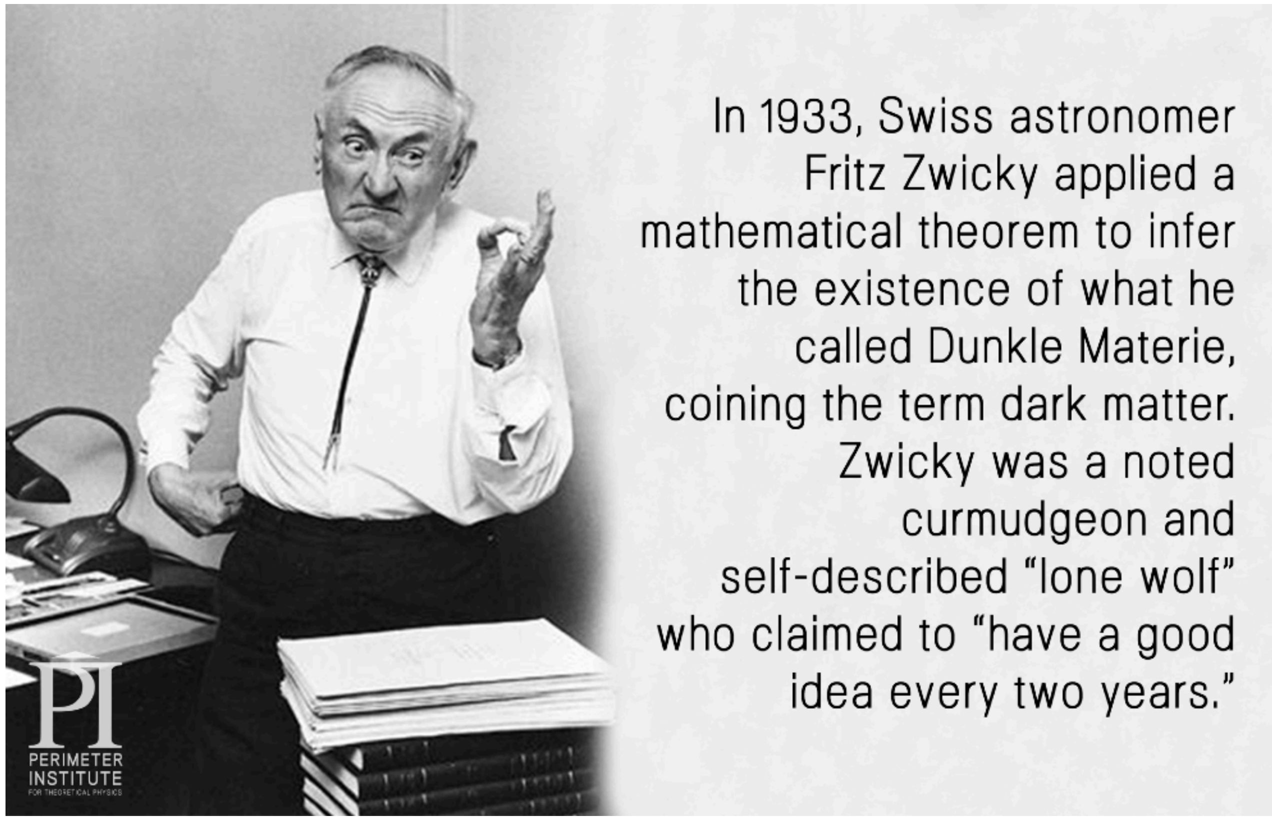
Galaxy distribution

The Standard Cosmological Model



Fritz Zwicky - 1933

Berenice Hair Cluster

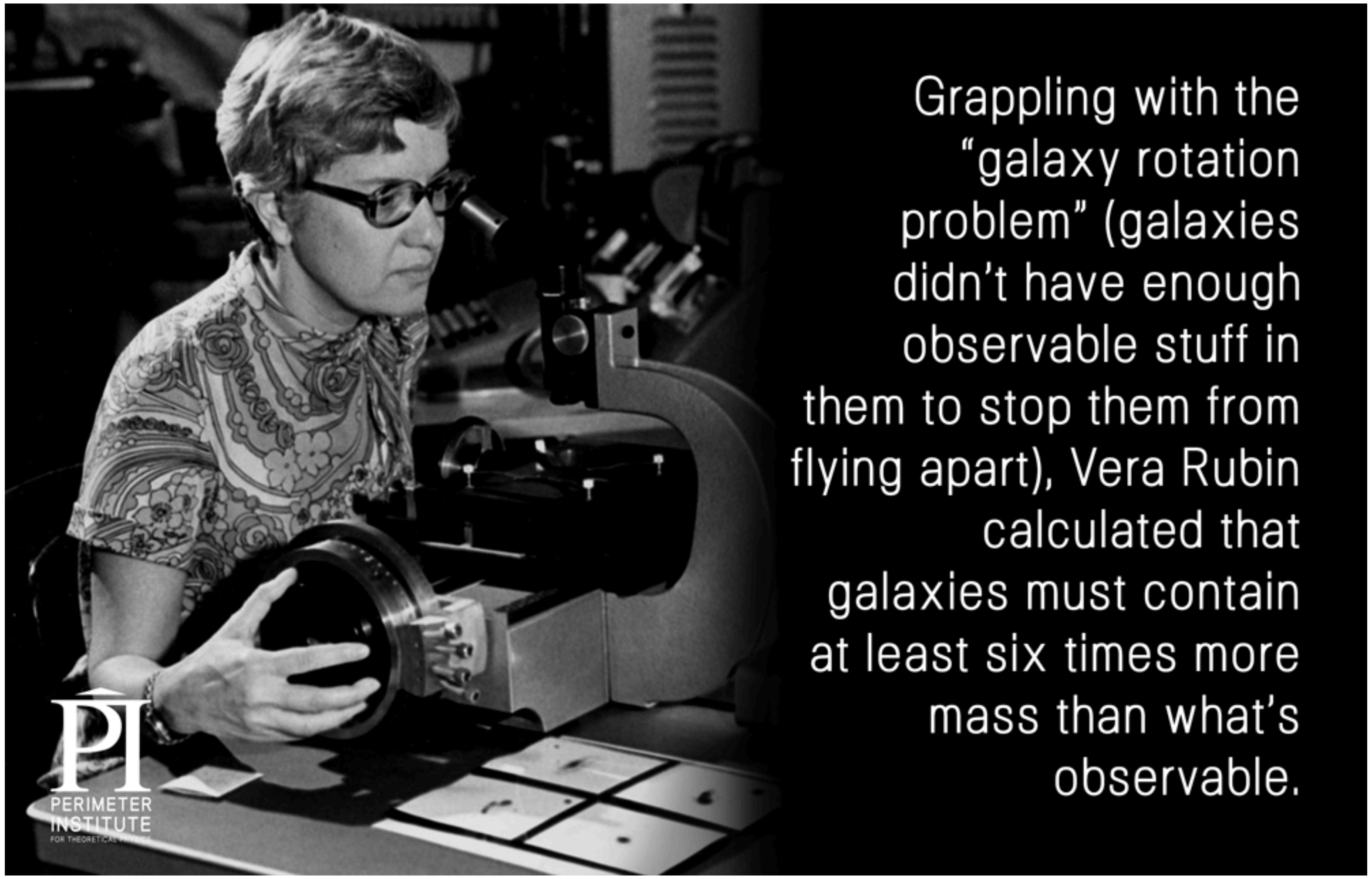


In 1933, Swiss astronomer Fritz Zwicky applied a mathematical theorem to infer the existence of what he called Dunkle Materie, coining the term dark matter. Zwicky was a noted curmudgeon and self-described “lone wolf” who claimed to “have a good idea every two years.”

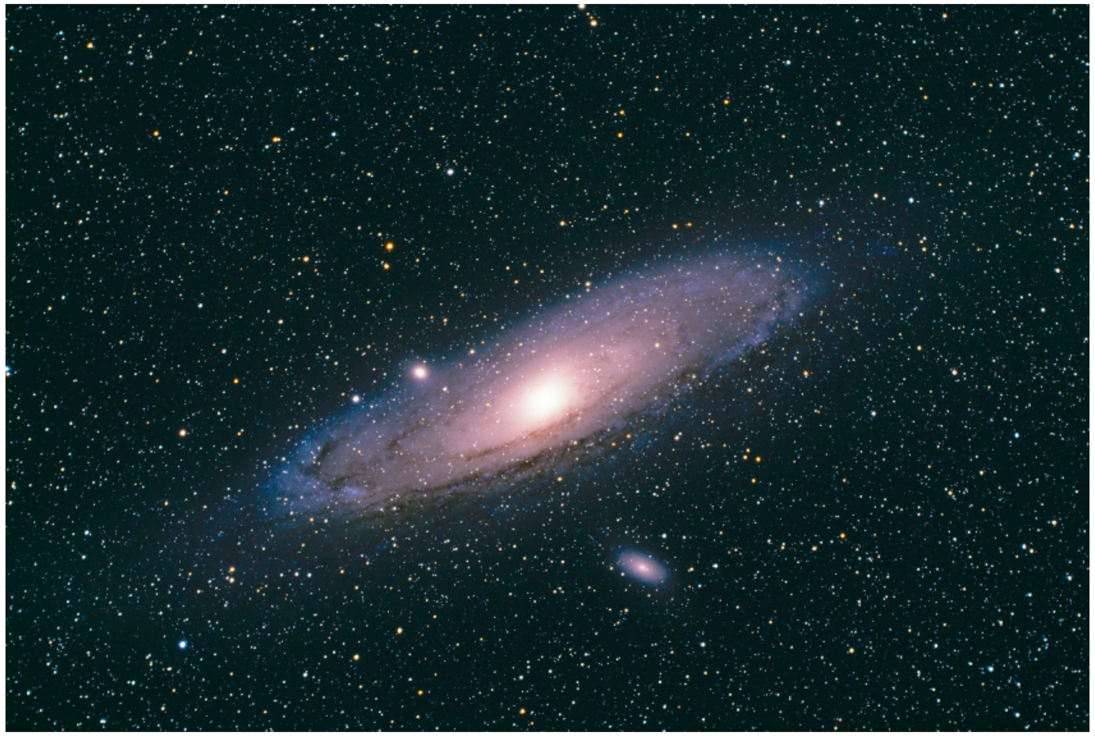
From the dispersion of the speeds of seven galaxies, he estimated that the “dynamic mass” is not compatible to the “luminous mass”, deduced from the quantity of light emitted by the cluster.

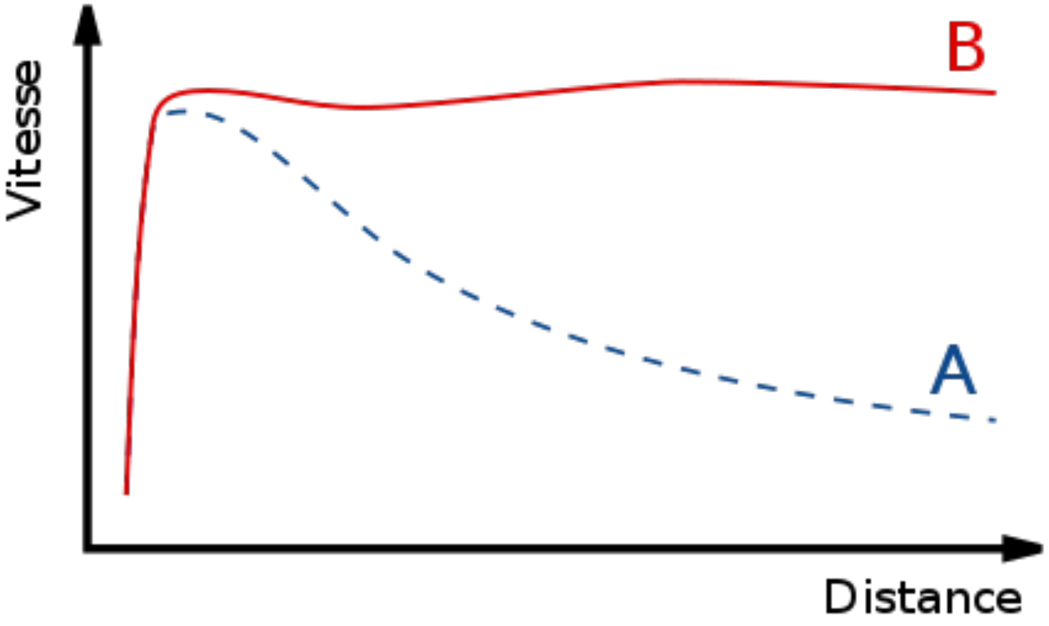
Dunkle Materie

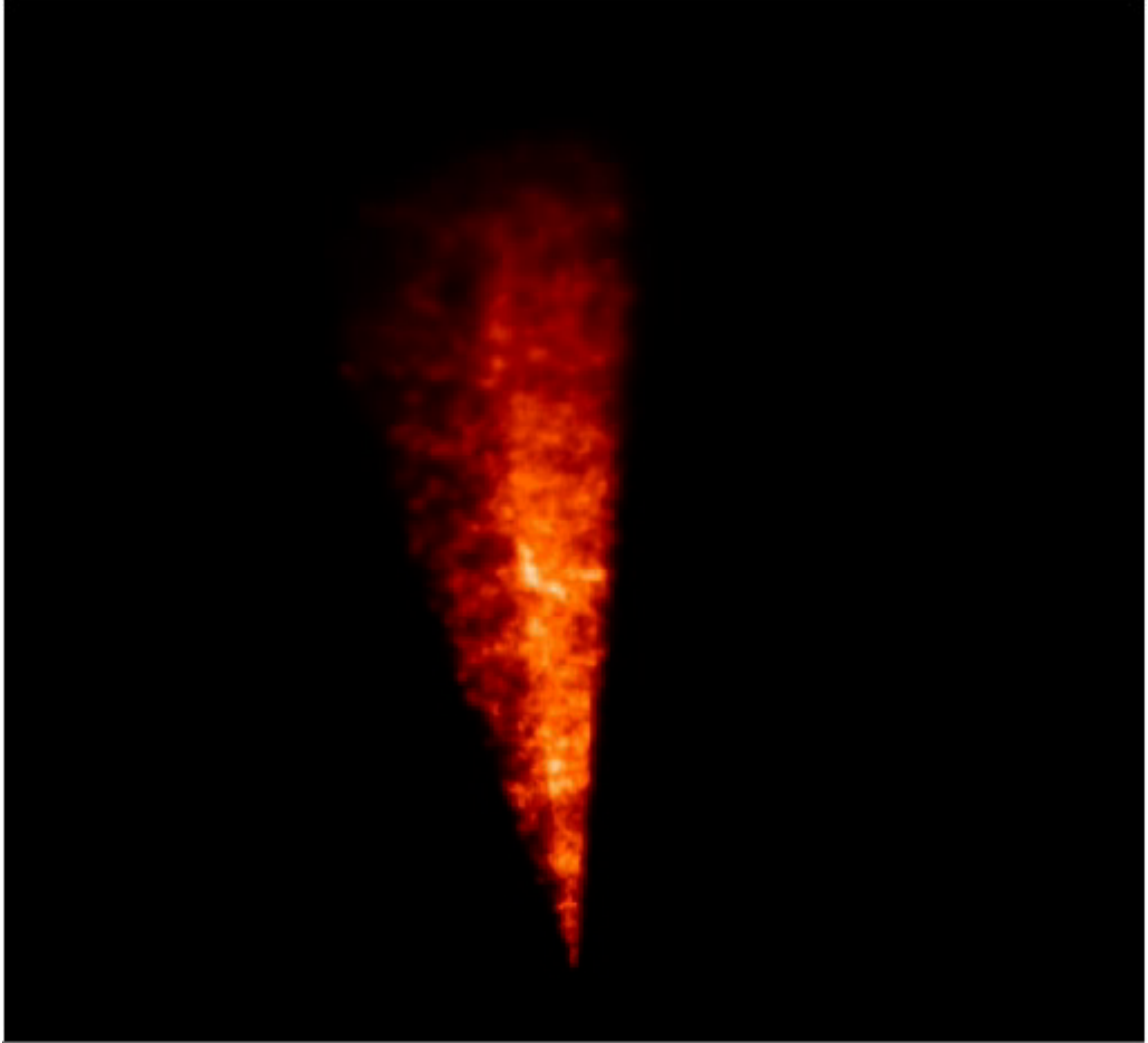
Vera Rubin - 1970

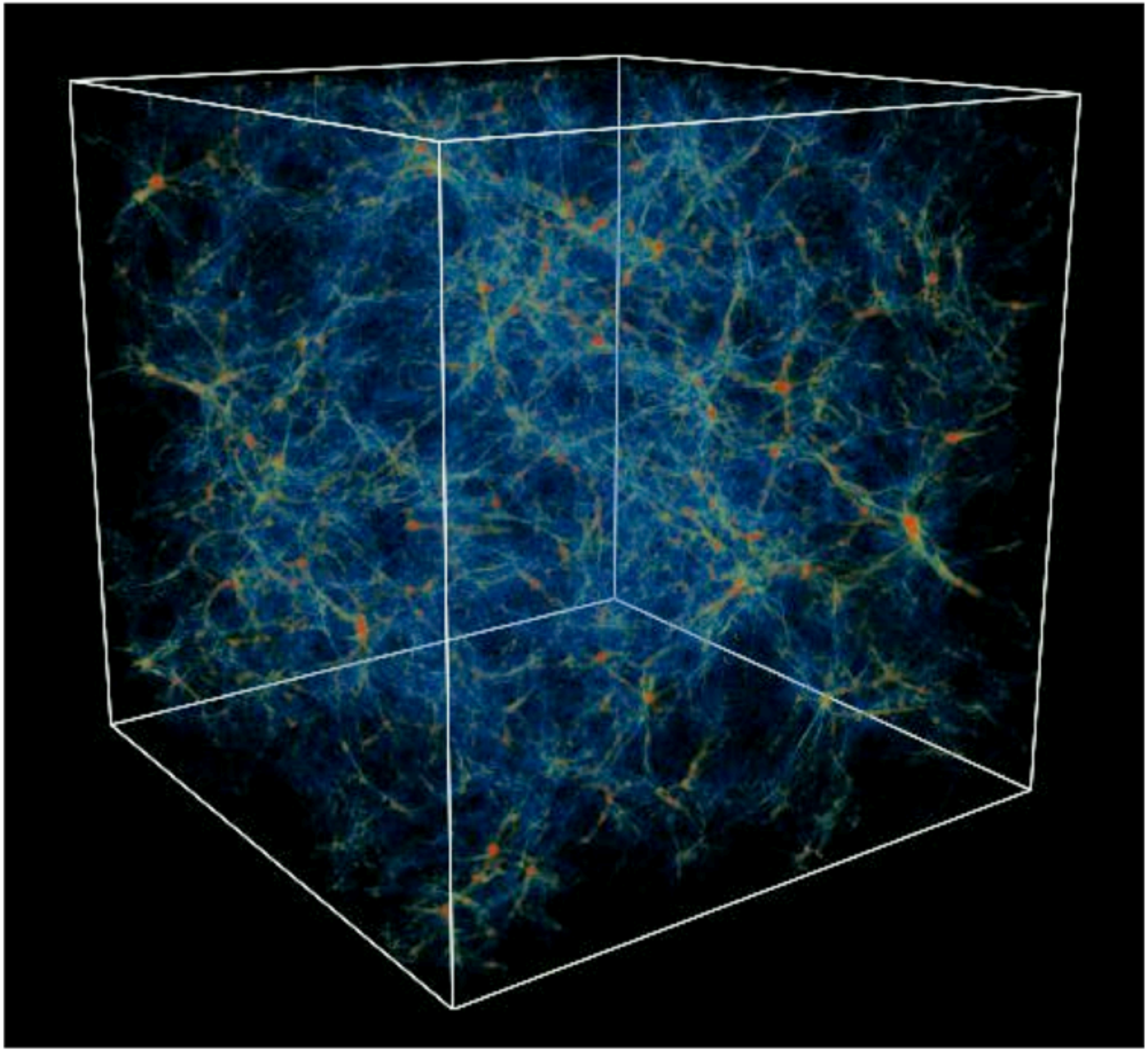


Grappling with the "galaxy rotation problem" (galaxies didn't have enough observable stuff in them to stop them from flying apart), Vera Rubin calculated that galaxies must contain at least six times more mass than what's observable.



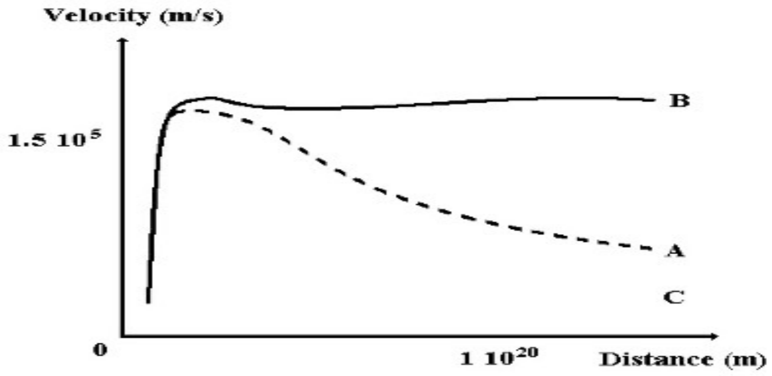




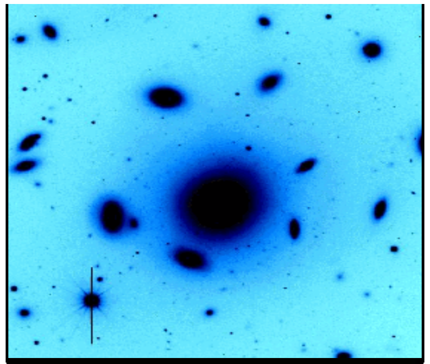


Evidence for dark matter

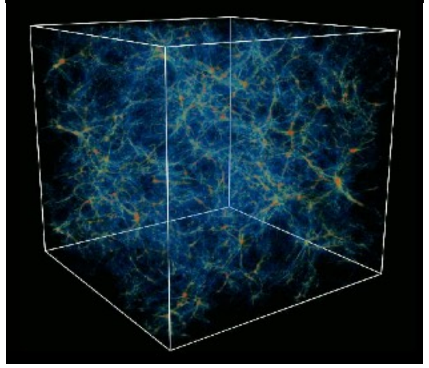
- Rotation curves of galaxies



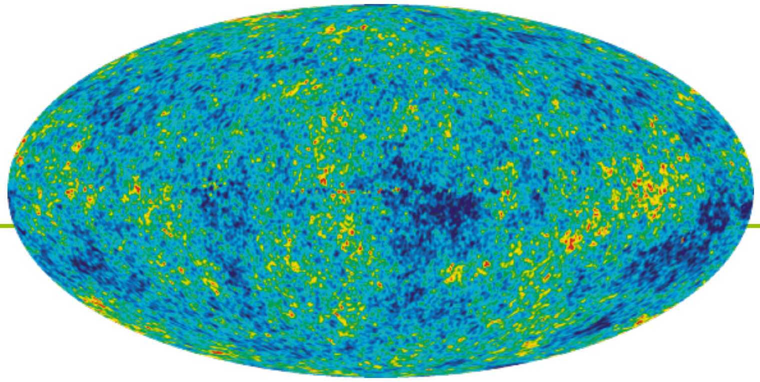
- Cluster of galaxies



- N-body simulations



- Temperature fluctuation



Two explanations

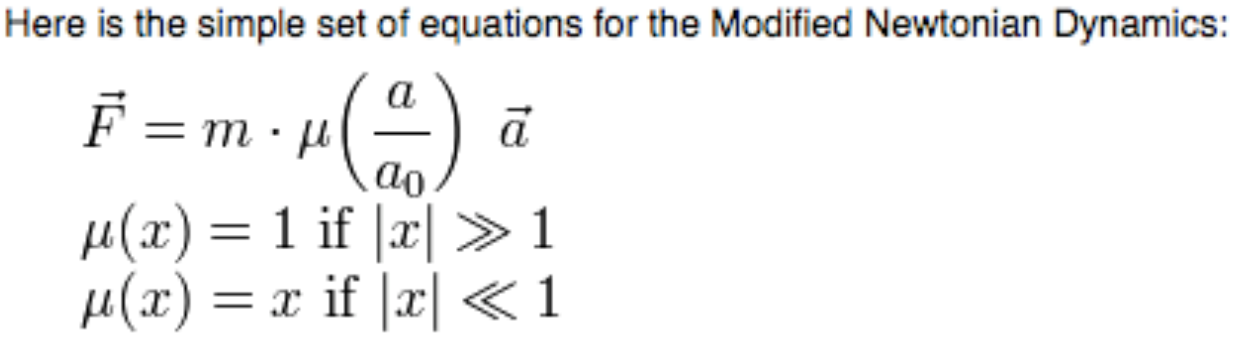
- CDM

MACHOs - ***baryonic matter***
(brown dwarfs & black holes)
(Massive Astrophysical Compact Halo Objects)
=> NOW EXCLUDED

Neutrinos - ***non baryonic matter***
=> NOW EXCLUDED

WIMPs - ***non baryonic matter*** -
(**W**eakly **I**nteractive **M**assive **P**articles)
=> hard to detect

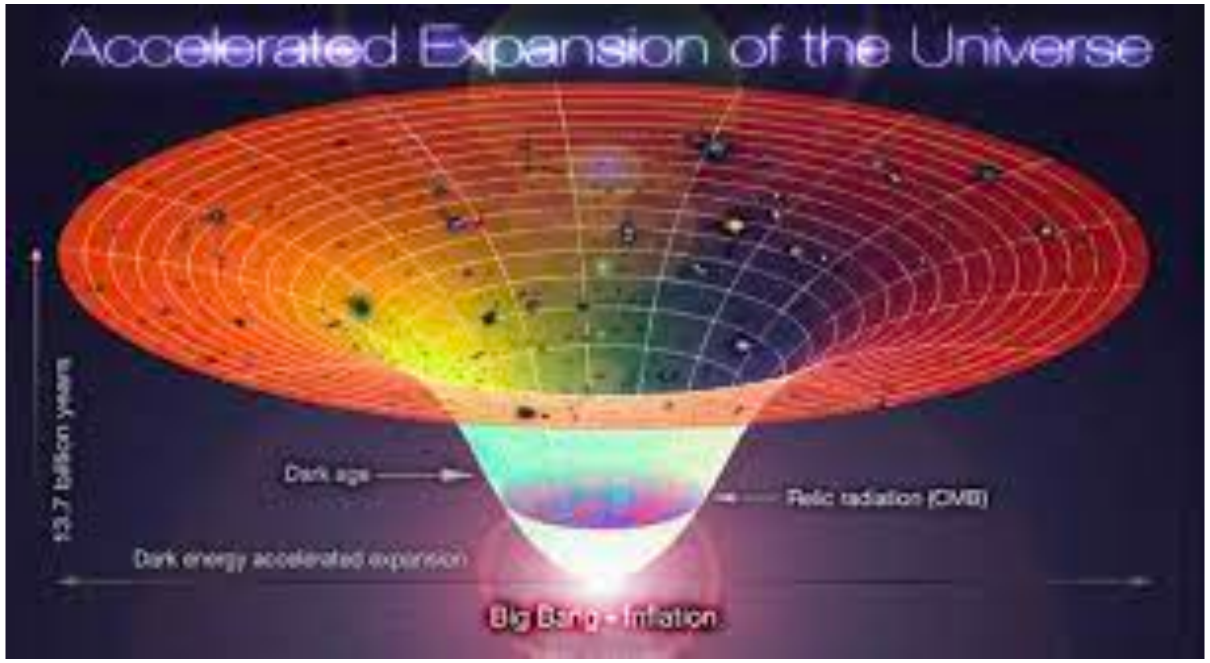
- Modify gravity (ex. MOND theory)

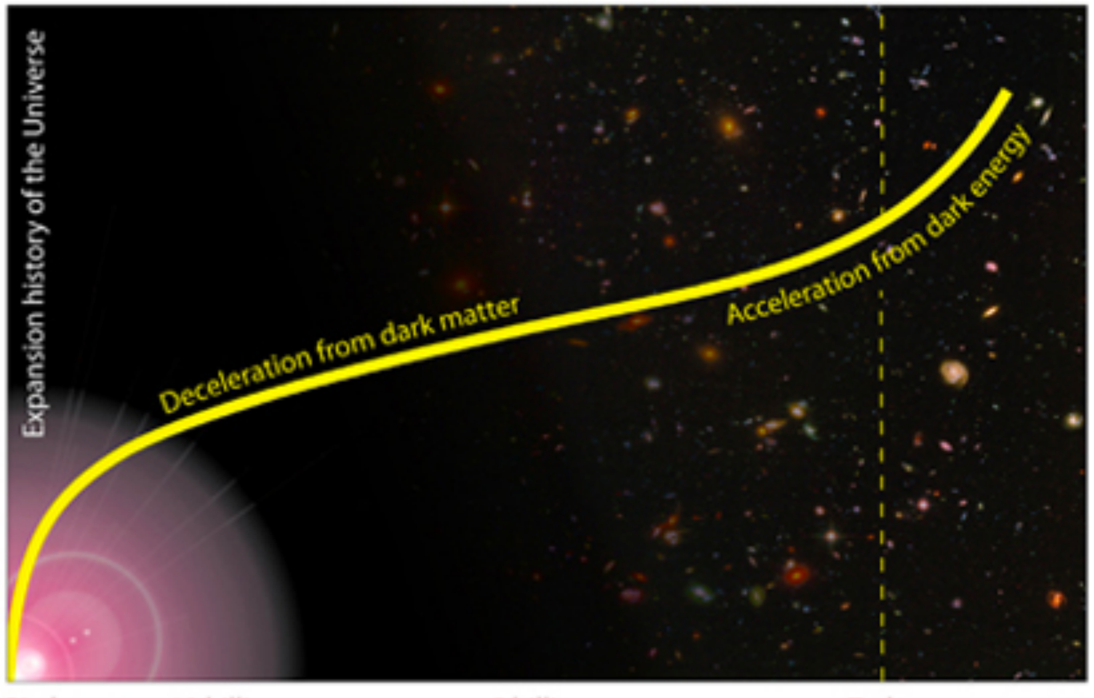


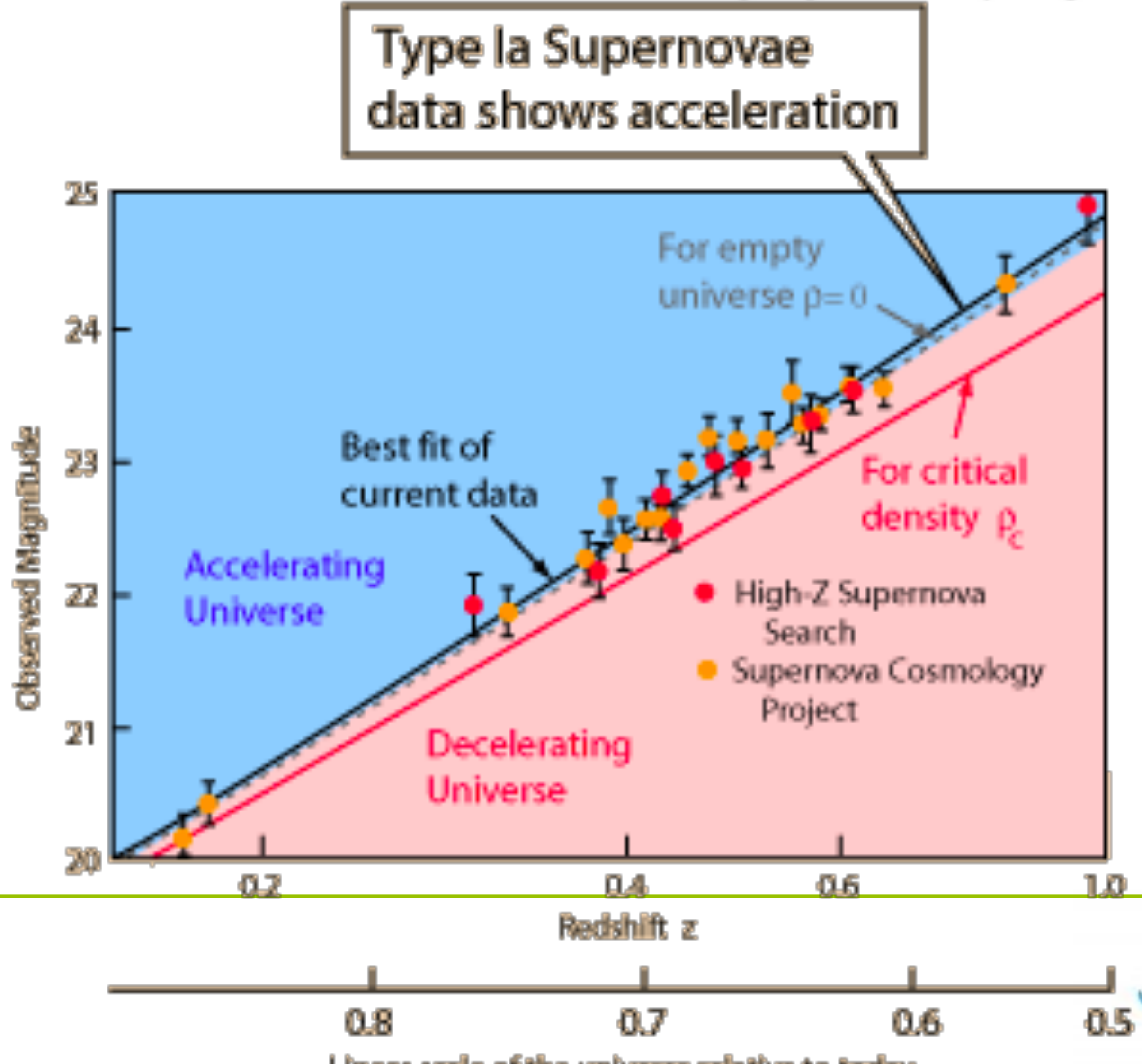
Here is the simple set of equations for the Modified Newtonian Dynamics:

$$\vec{F} = m \cdot \mu\left(\frac{a}{a_0}\right)\ \vec{a}$$
$$\mu(x) = 1 \text{ if } |x| \gg 1$$
$$\mu(x) = x \text{ if } |x| \ll 1$$

Dark Energy

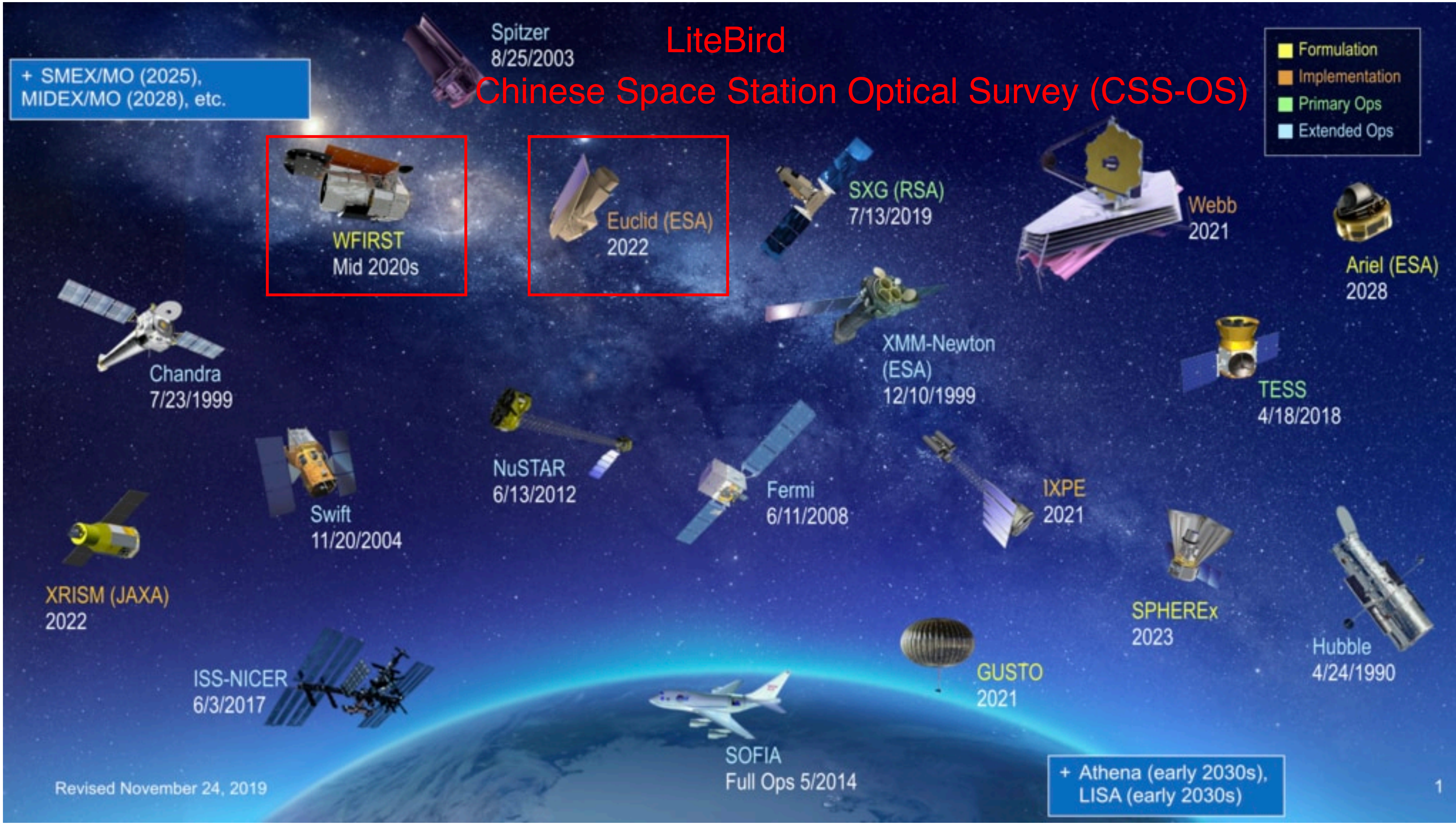

Accelerated Expansion of the Universe
13.7 billion years
Dark age
Relic radiation (CMB)
Dark energy accelerated expansion
Big Bang - Inflation


Expansion history of the Universe
Deceleration from dark matter
Acceleration from dark energy
Big bang
10 billion years ago
5 billion years ago
Today


Type Ia Supernovae
data shows acceleration
For empty
universe ρ= 0
Best fit of
current data
Accelerating
Universe
For critical
density ρc
High-Z Supernova
Search
Supernova Cosmology
Project
Decelerating
Universe
Observed Magnitude
25
24
23
22
21
20
0.2
0.4
0.6
1.0
Redshift z
0.8
0.7
0.6
0.5

LiteBird

Chinese Space Station Optical Survey (CSS-OS)



Precision Cosmology

First Source of Uncertainty: Stochastics

- New instruments, more sensitive (hardware)

- **Collect more Data** => large survey (SDSS, WMAP, Planck, KIDS, DES, etc)

==> Virtual Observatory (CDS Strasbourg)

- Data access, web services, interoperability, data model, etc

First Source of Uncertainty: Stochastics

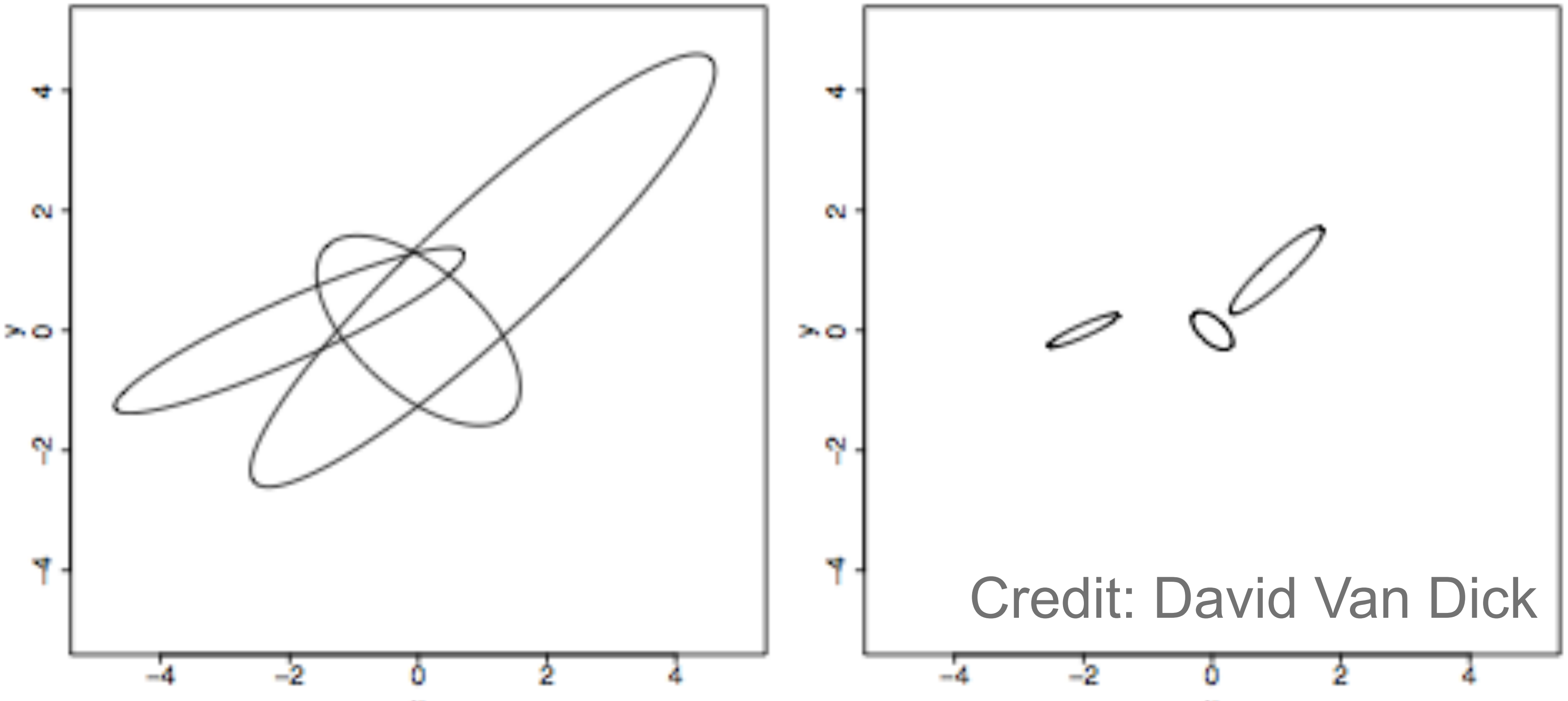
- Better statistical tools (Bayesian modeling, sparsity, BSS, machine learning, etc): beyond the second order statistics

Astrophysic + Statistics/Applied math => Astrostatistics

- **Two International organizations:**
  - ➡ **International Astrostatistics Association (IAA)**
  - ➡ **Commission on Astroinformatics and Astrostatistics within the International Astronomical Union (IAU)**
- **Two important U.S. national organizations:**
  - ➡ the Working Group in Astroinformatics and Astrostatistics within the American Astronomical Society (AAS),
  - ➡ the Interest Group in Astrostatistics within the American Statistical Association (ASA).
- **One project-level organization:** the Informatics and Statistics Science Collaboration of the Large Synoptic Survey Telescope (LSST)

- **Astrostatistics laboratories**
  - ➡USA: Penn State University, Berkeley, CMU, Cornell
  - ➡Europe:
    - Imperial Center for Inference and Cosmology (ICIC) at Imperial College
    - **CosmoStat laboratory**, CEA-Saclay

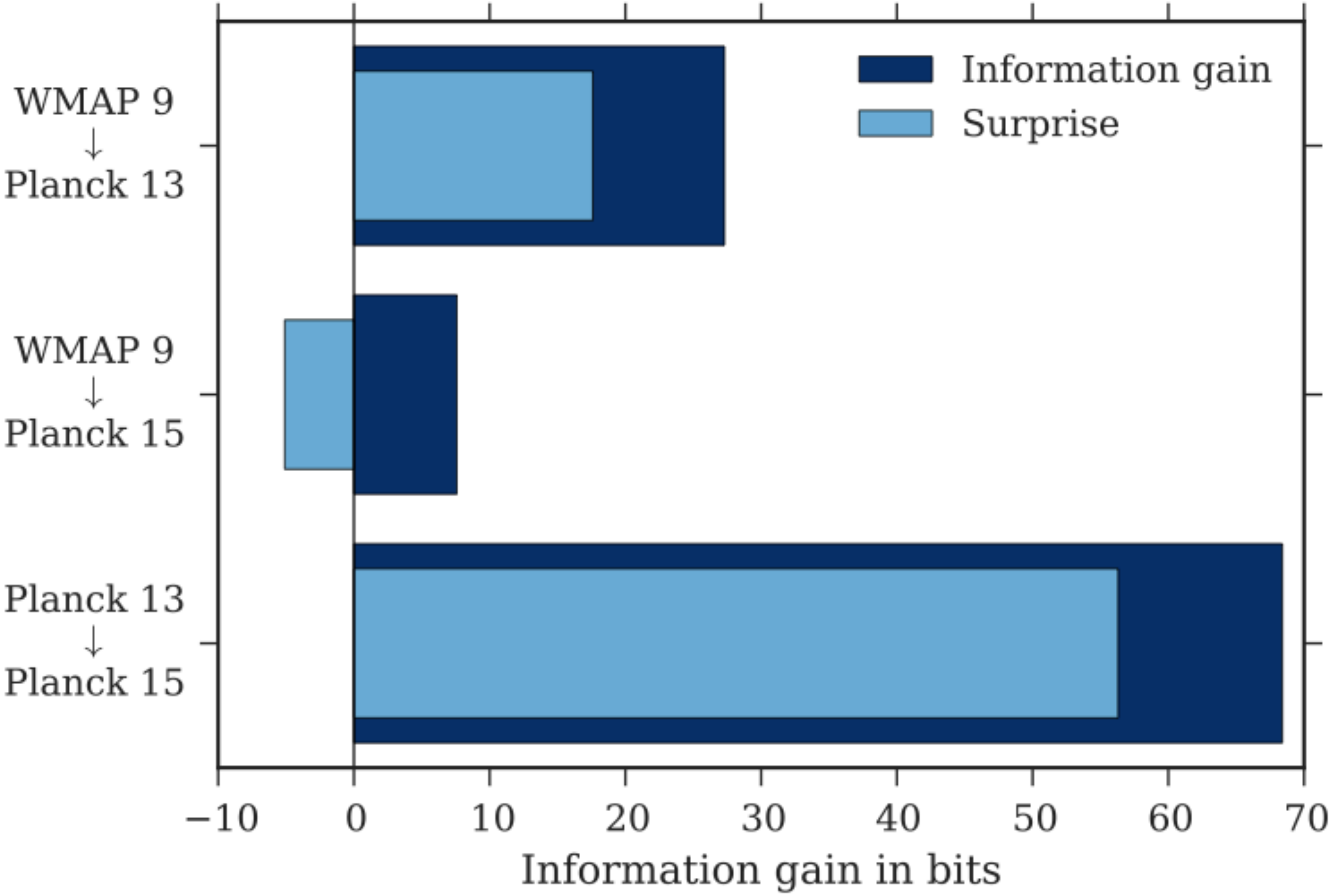
Second Source of Uncertainty: Systematics



Credit: David Van Dick

Fig: As datasets grow, systematic errors swamp statistical errors and new disparities appear.

Source of Uncertainty: Systematics



Seehars et al, Physical Review D, Volume 93, Issue 10, id.103507, 2016

Second Source of Uncertainty: Systematics

The need of Numerical Simulations (physics + instrument)

- to test the pipeline and its ability to measure accurately the cosmological paramters.

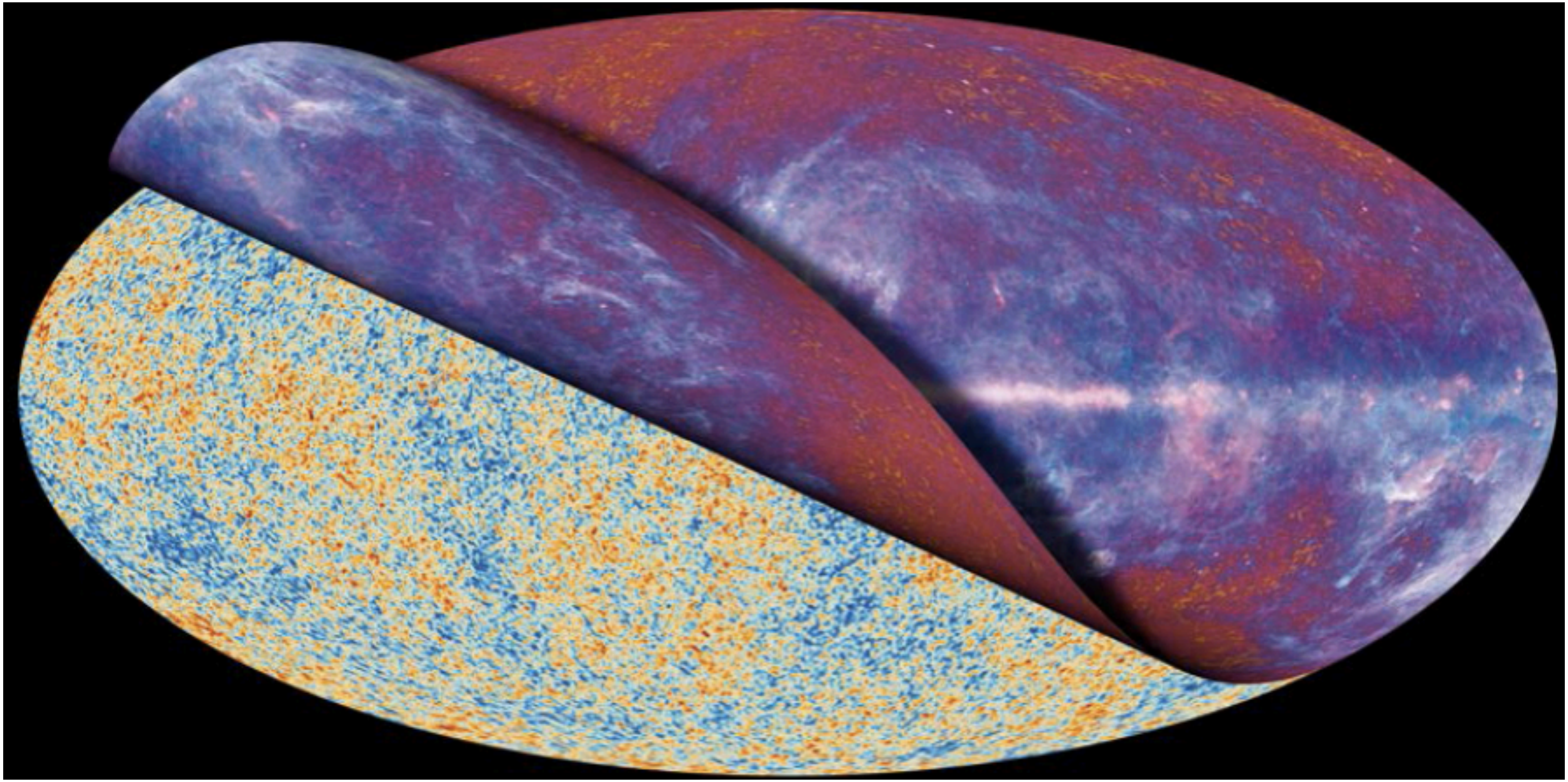
- to build the covariance matrices that are required to fit the cosmological parameters

Numerical simulations are a very important aspect of new big projets.

Foreground Removal

BICEP2:

March 2014 - Primordial Gravitationnal Wave detection claimed by BICEP2
==> it happened to be a dust signature, dust from our own galaxy !!!



Third Source of Uncertainty: Approximation

- We now perfectly how to calculate some estimators and their covariance matrices, but the **volume of data is so large** that **it is impossible** to do it, even on HPC infrastructures.

Theoretical and algorithmic work is necessary to well **control errors and biases introduced by the approximations**.

The **generalisation problem** in Deep Learning techniques can be very **challenging**.

Examples:

- Two point corrrelation functions
- Covariance matrices
- Example of ongoing work at CMU: simulate N-body simulations using machine learning.
- Approximate Bayesian Modelling (ABC) (likelihood free approach to approximating posterior where likelihood function is not specified).

Space & Cosmology

➡Part 1: Introduction to Accurate Space Cosmology

➡**Part 2: Inverse Problems**

➡Part 3: Euclid Weak Lensing

Inverse Problems

➡Part 1: Introduction to Inverse Problems

➡Part 2: From Fourier to Wavelets

➡**Part 3: Wavelet and Beyond**

➡Part 4: Sparse Regularization

➡Part 5: Application to Unmixing and Inpainting

➡Part 6: Compressed Sensing

➡Part 7: Deep Learning

Inverse Problems in Astrophysics

$$Y = HX + N$$

Examples:

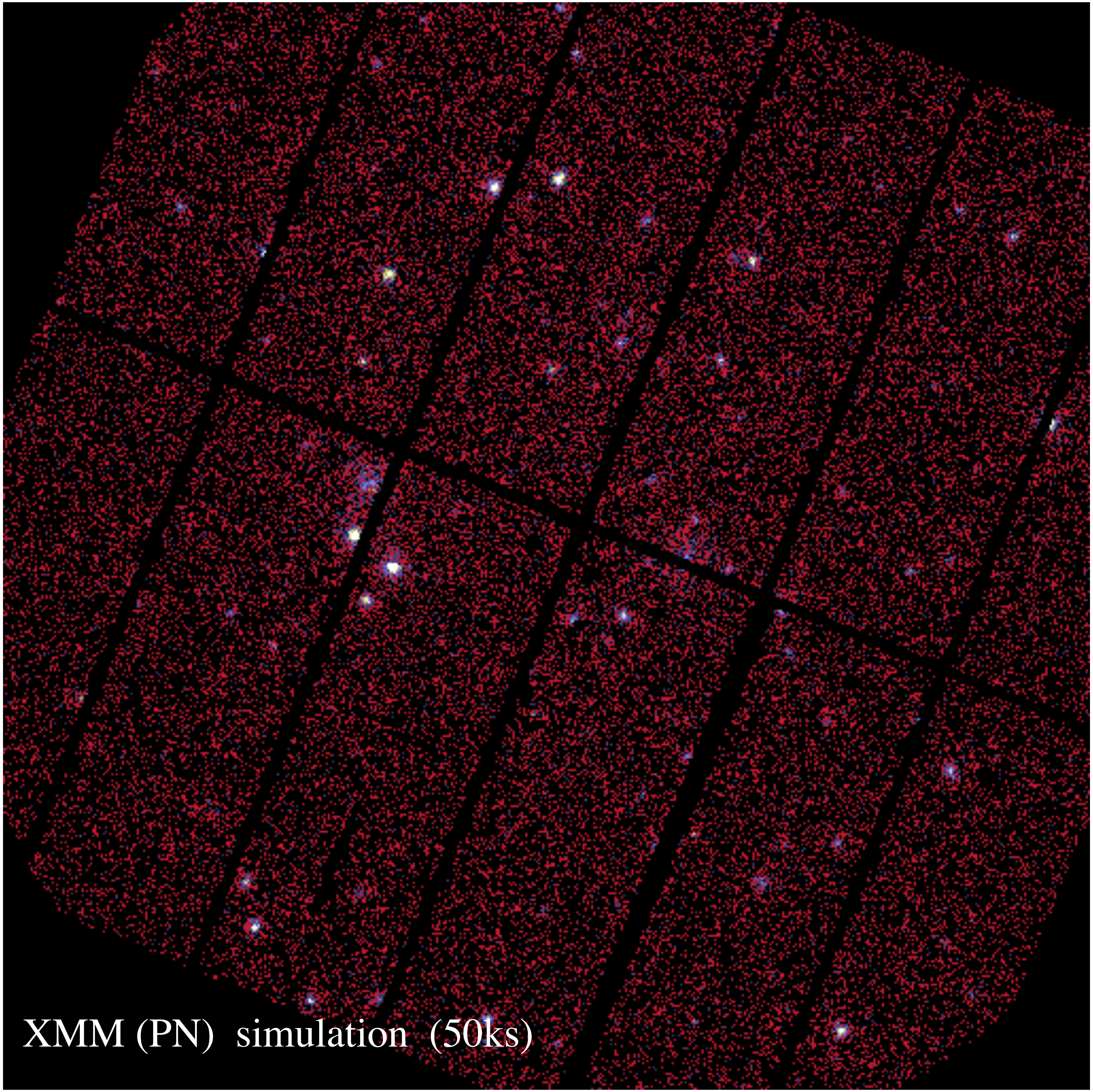
- **Denoising**
- **Deconvolution**
- **Component Separation**
- **Inpainting**
- **Blind Source Separation**
- **Compressed Sensing**

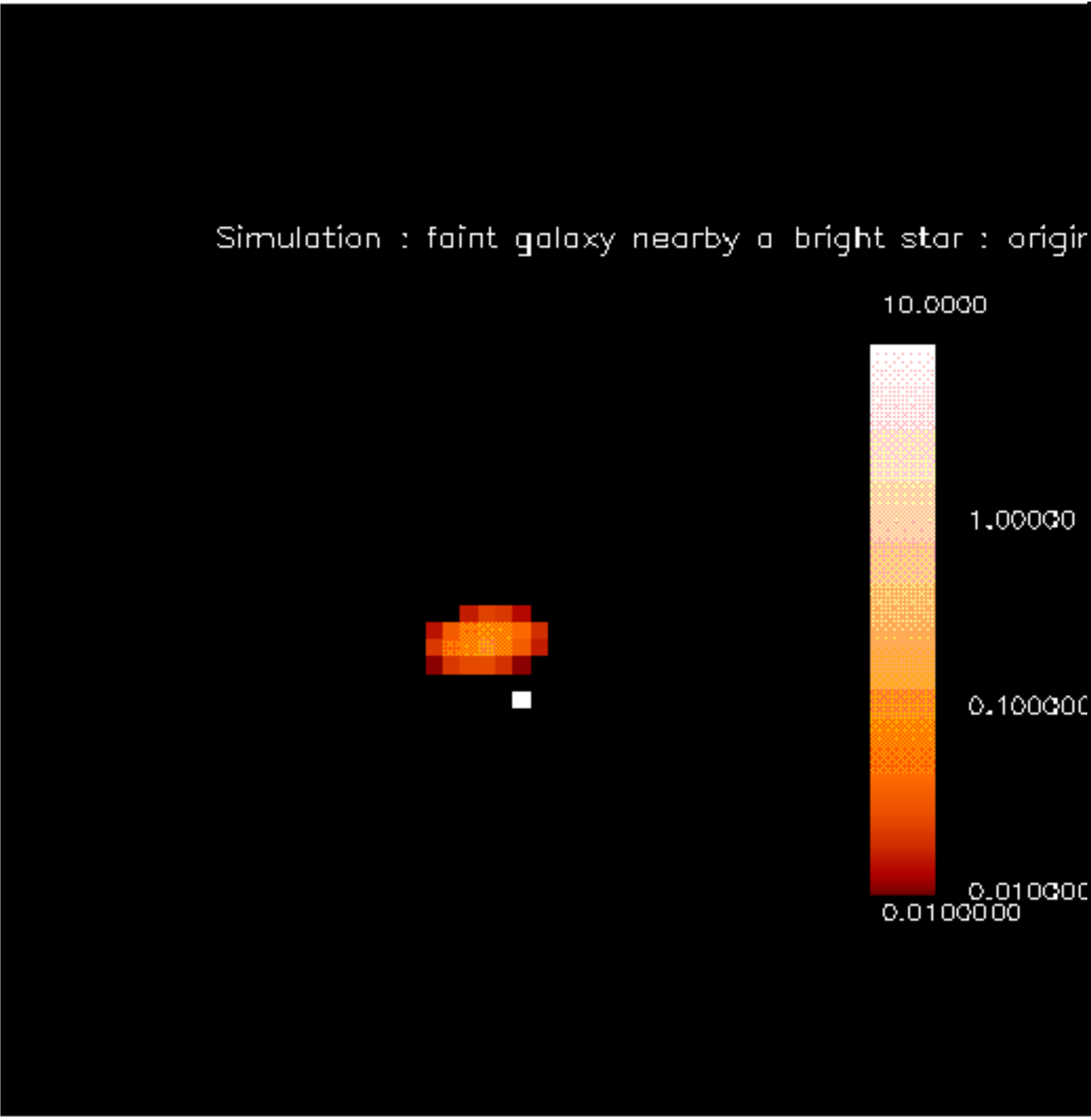
PB 1: find X knowing Y,H and the statistical properties of the noise N
Ex: Astronomical image deconvolution
Weak lensing

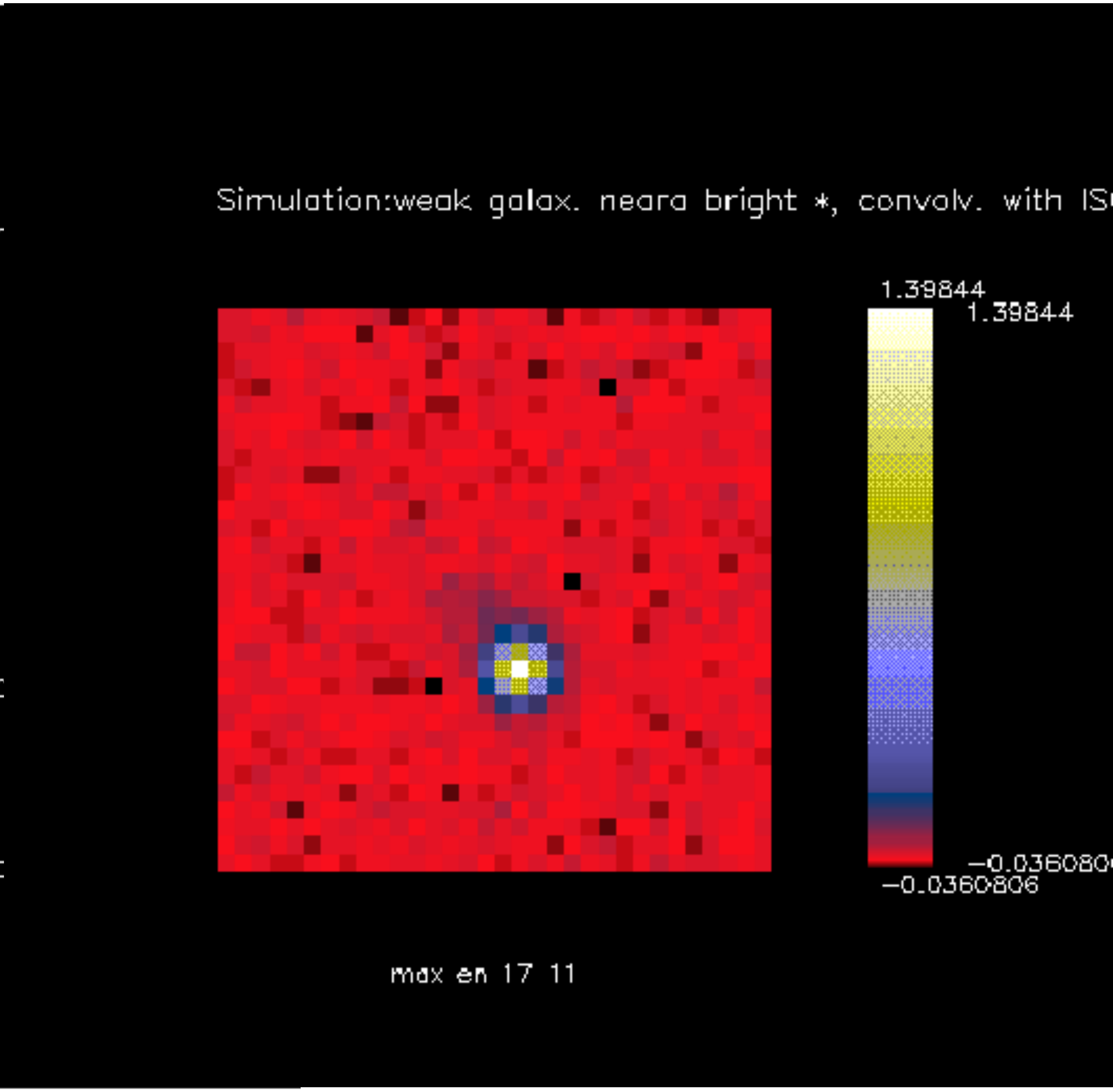
PB 2: find X and H knowing Y and the statistical properties of the noise N
Ex: Blind deconvolution

Ill posed problem, i.e. not an unique and stable solution ==> Regularization

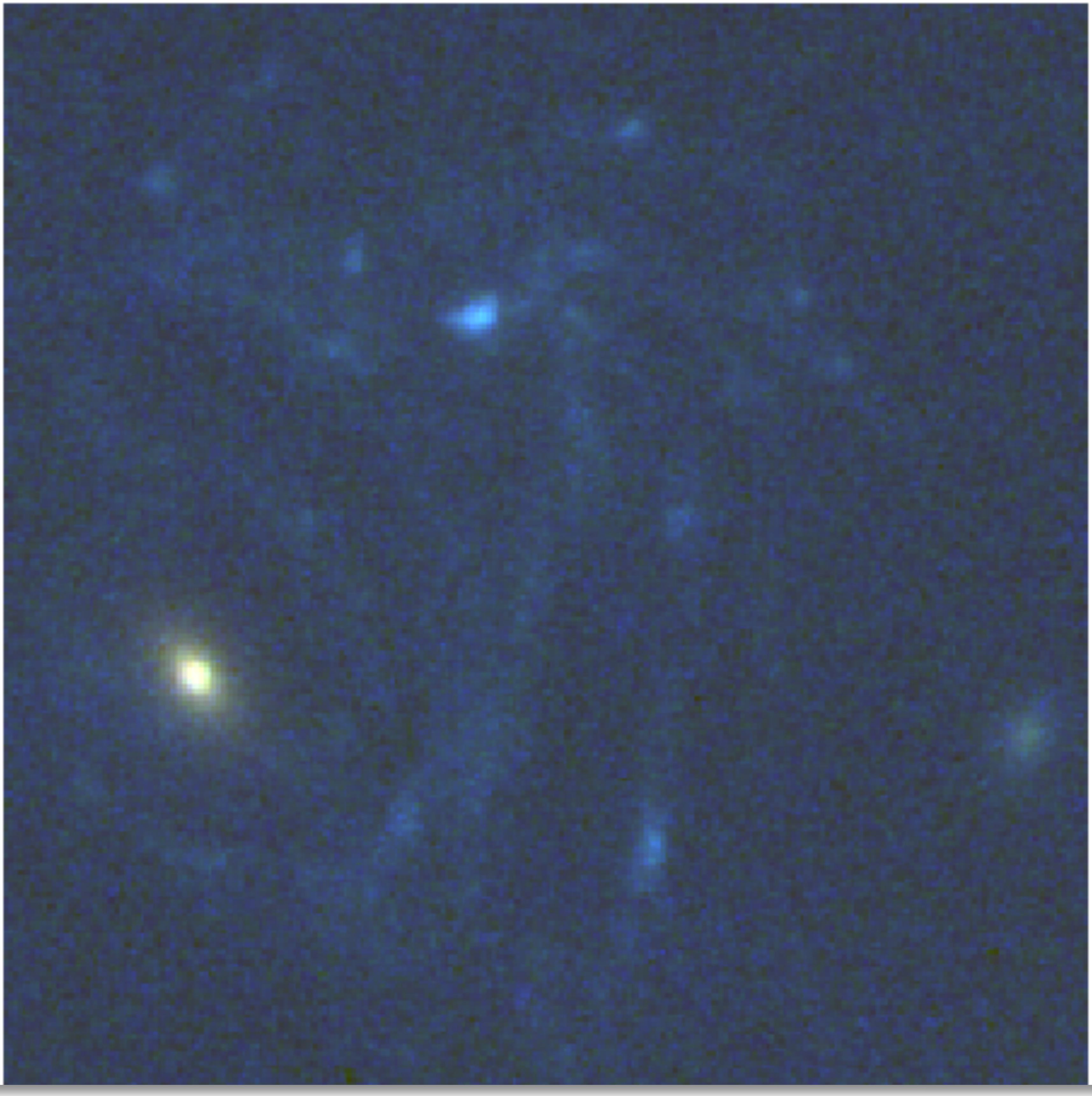
$$||Y - HX||^2$$ with some constraints on X


XMM (PN) simulation (50ks)

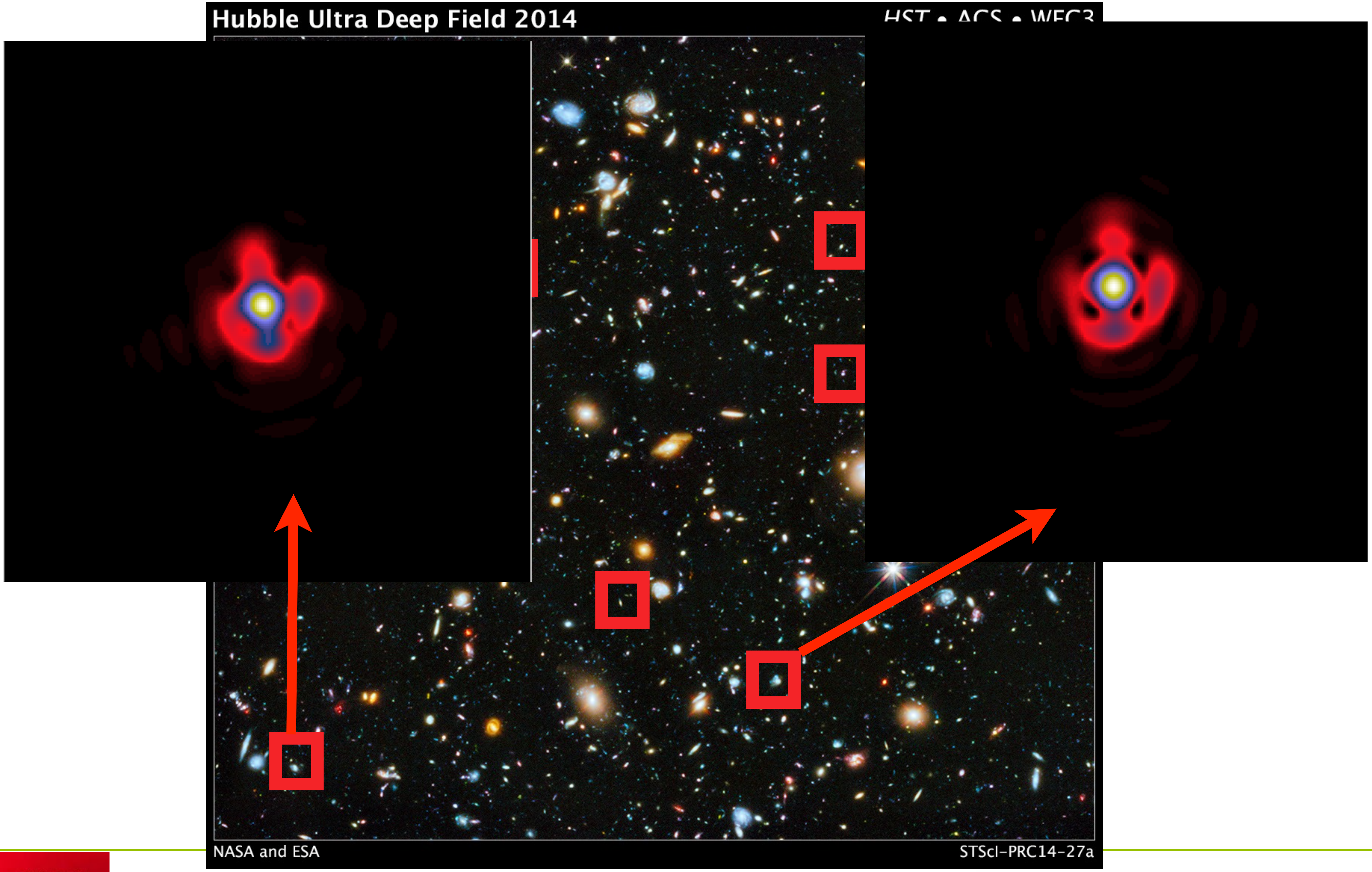

Simulation : faint galaxy nearby a bright star : origir
10.0000
1.00000
0.100000
0.010000
0.0100000


Simulation:weak galax. neara bright *, convolv. with IS
1.39844
1.39844
−0.036080
−0.0360806
max en 17 11

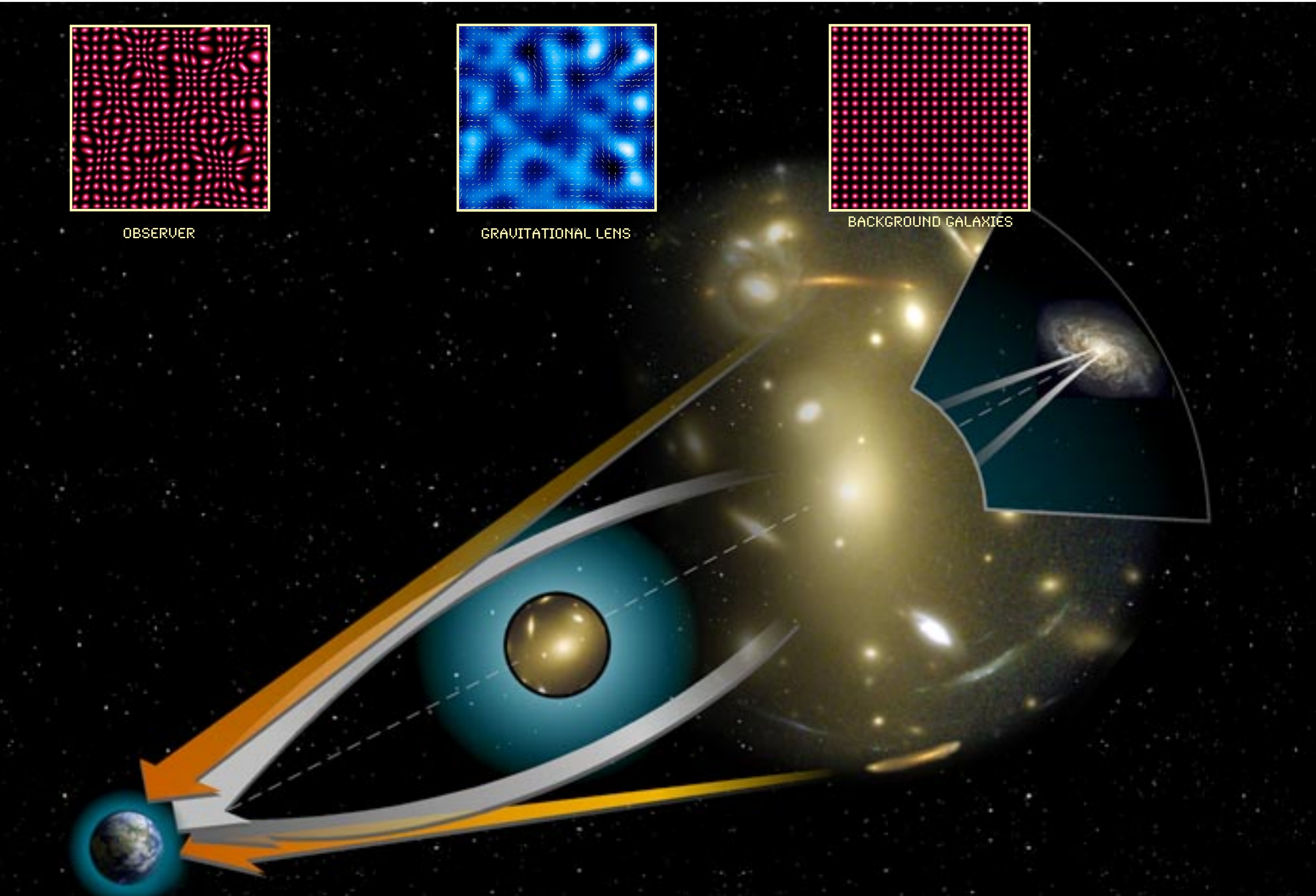
Multichannel data



Euclid Point Spread Function Field



Weak Lensing



ILL POSED INVERSE PROBLEM

$$\min_{X} ||Y - HX|| \ \ s.t. \ \ \mathcal{C}(\mathcal{X})$$

Need to add constraint

$$||Y - HX||^2$$

with some constraints on X and H

$$\min_{H,X} ||Y - HX|| \ \ s.t. \ \ \mathcal{C}(H, X)$$

If H is known then:

$$\min_{X} ||Y - HX|| \ \ s.t. \ \ \mathcal{C}(X)$$

Tikhonov Regulatization

$$Y = HX + N$$

The Tikhonov regularisation solution is: $\mathcal{C}(X) = \lambda \parallel LX \parallel^2$

$$\arg\min_X \parallel Y - HX \parallel^2 + \lambda \parallel LX \parallel^2$$

The closed-form solution of this linear inverse problem is given by:

$$\tilde{X} = (H^t H + \lambda L^t L)^{-1} H^t Y$$

Data Representation

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

$$s_i = \sum_{k=1}^{K} \underset{\substack{\uparrow \\ \text{coefficients}}}{\alpha_k} \underset{\substack{\uparrow \\ \text{basis, frame}}}{\phi_k}$$

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

The Great Father Fourier - Fourier Transforms

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

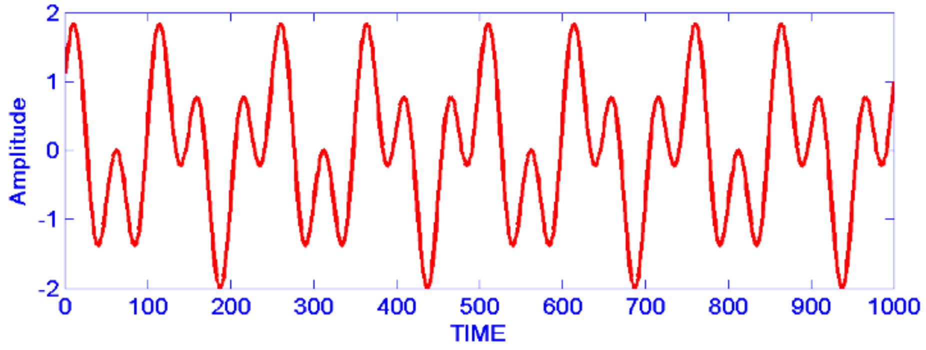
(Basis functions used are sine and cosine)

Jean-Baptiste-Joseph Fourier
(1768-1830)

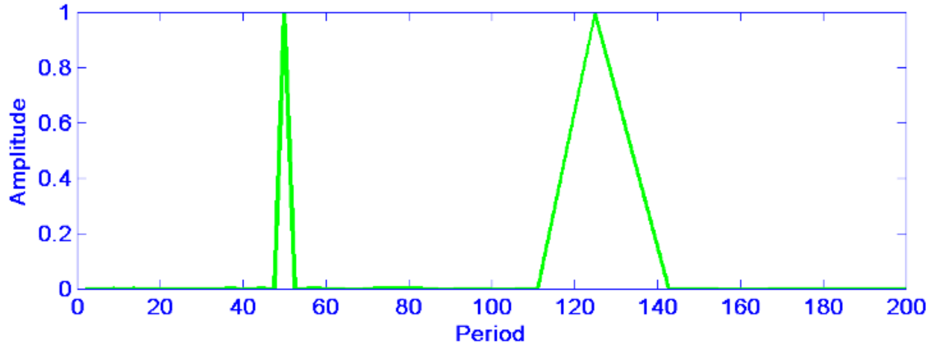
$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t} dt$$

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

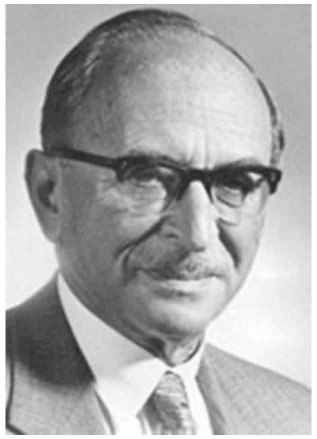
Time domain



Frequency domain



SFORT TIME FOURIER TRANSFORM (STFT)



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

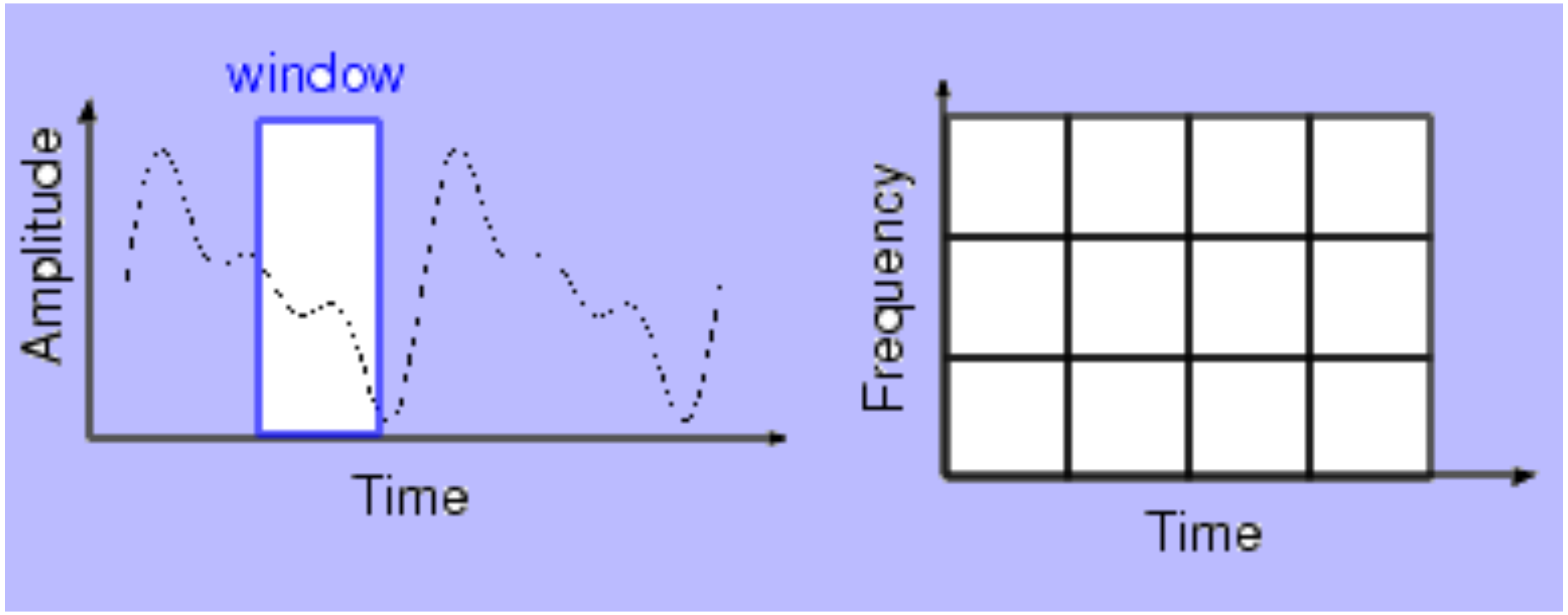
To analyze only a small section of the signal at a time -- a technique called *Windowing the Signal.*

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

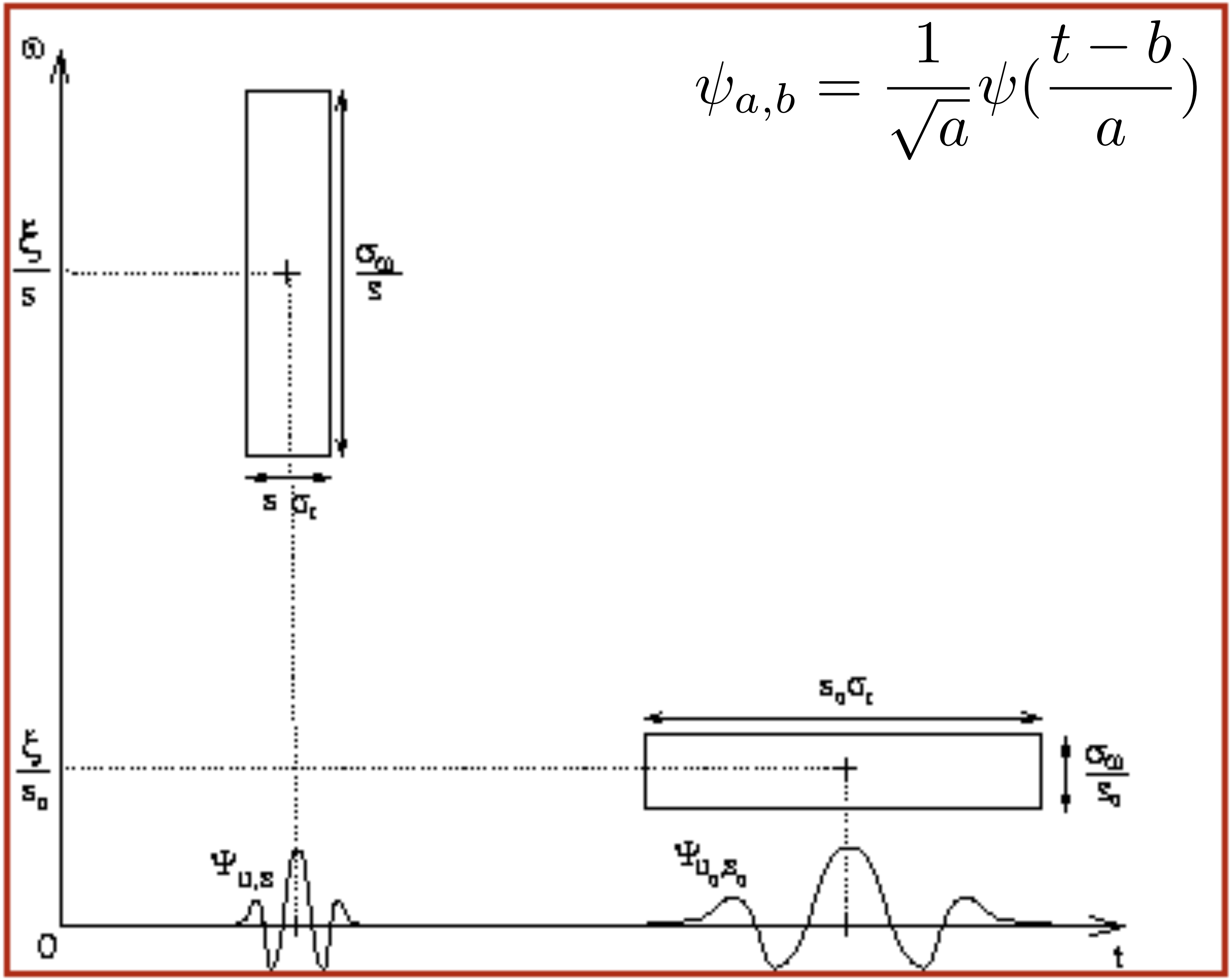
The Short Term Fourier Transform is defined by:

$$STFT(\nu, b) = \int_{\infty}^{+\infty} \exp(-j2\pi\nu t) f(t) g(t-b) dt$$

when g is a Gaussien, it corresponds to the Gabor transform.

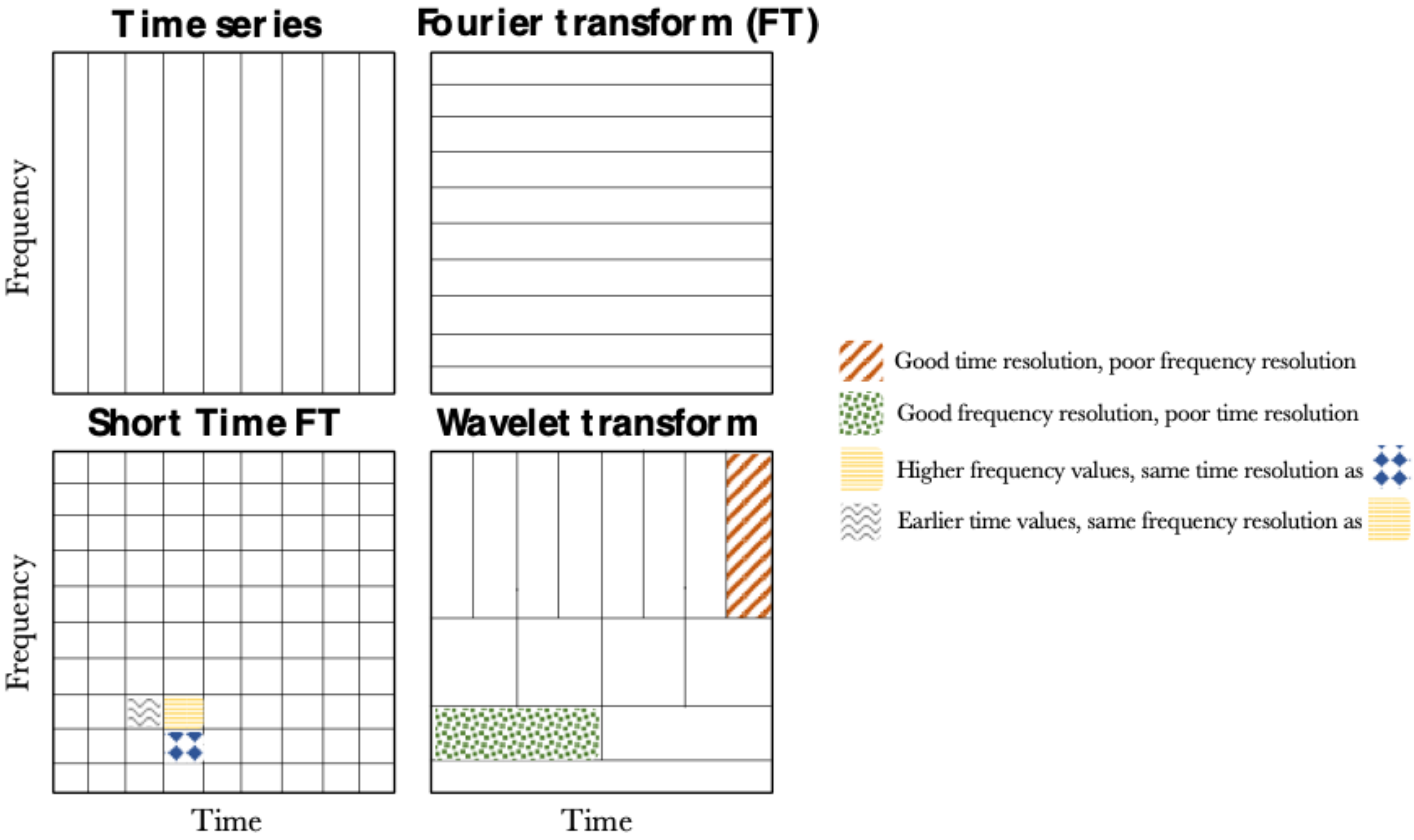


Wavelets



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

Fourier Domain Tiling



The Continuous Wavelet Transform

Jean Morlet

$$W(a,b) = K \int_{-\infty}^{+\infty} \psi^*(\frac{x-b}{a}) f(x) dx$$

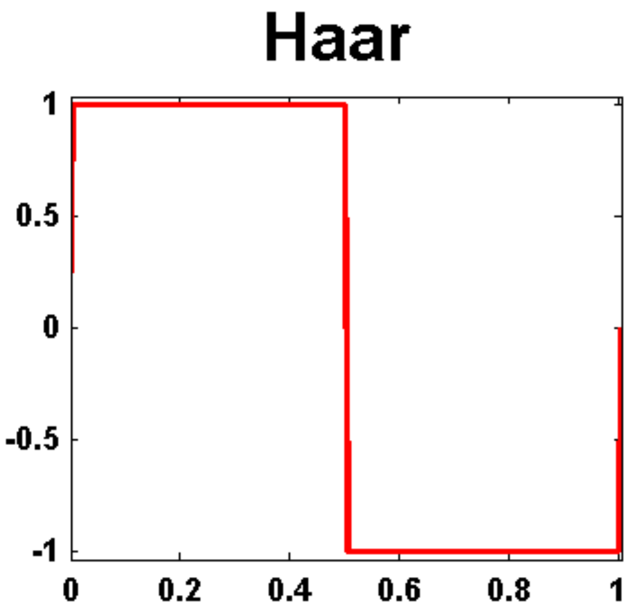
where:

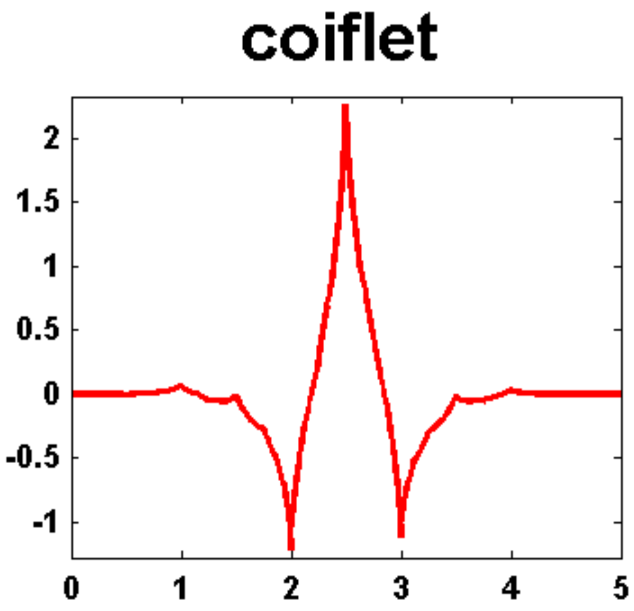
- $W(a,b)$ is the wavelet coefficient of the function $f(x)$
- $\psi(x)$ is the analyzing wavelet
- a (> 0) is the scale parameter
- b is the position parameter

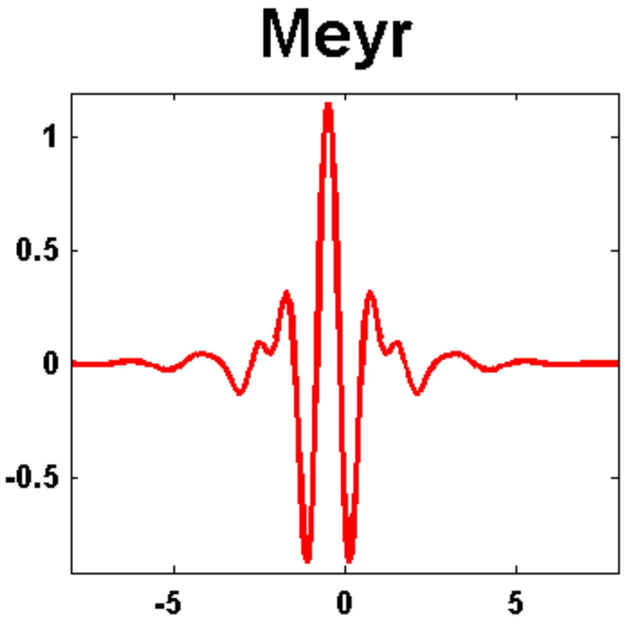
In Fourier space, we have: $\hat{W}(a,\nu) = \sqrt{a}\hat{f}(\nu)\hat{\psi}^*(a\nu)$

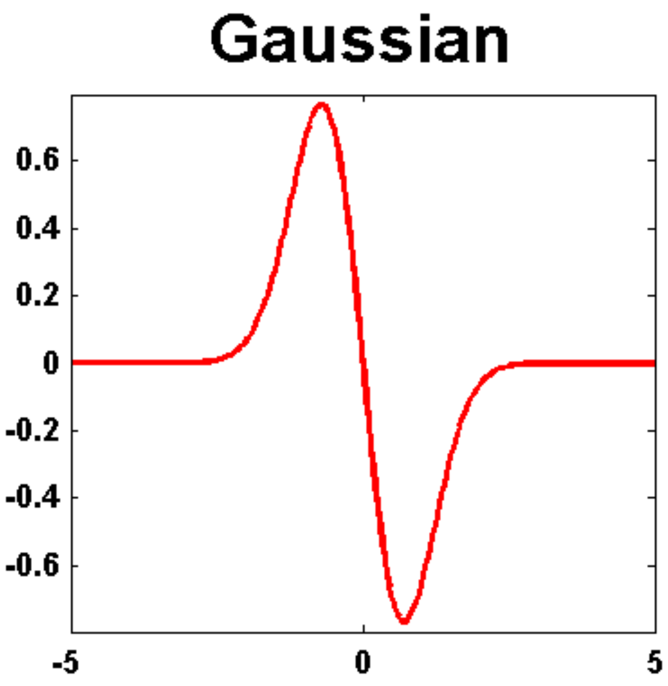
When the scale a varies, the filter $\hat{\psi}^*(a\nu)$ is only reduced or dilated while keeping the same pattern.

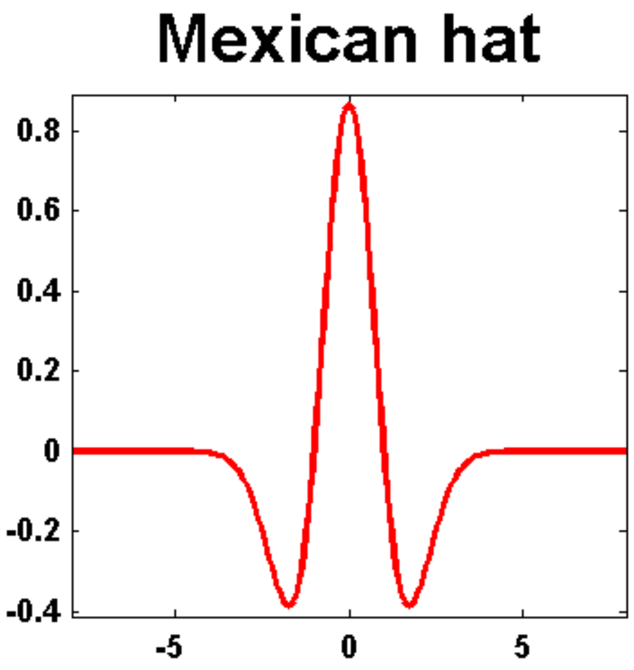
Some typical mother wavelets

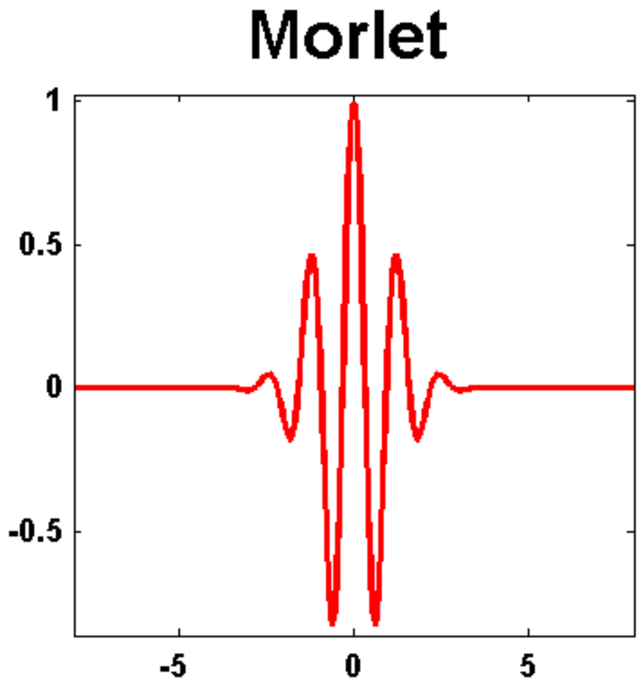




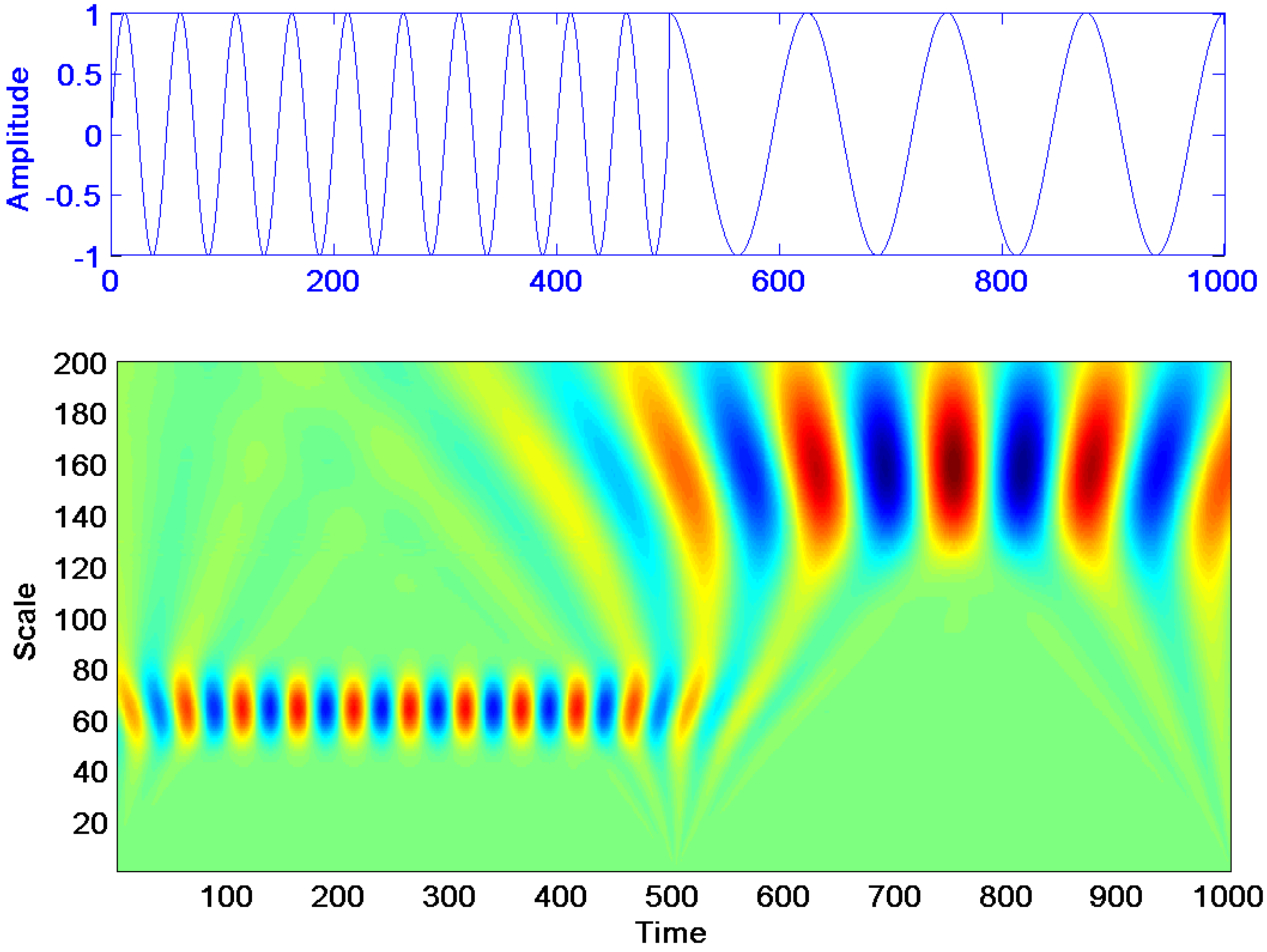








Typical picture



The Inverse Transform

The inverse transform is:

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \frac{1}{\sqrt{a}} W(a,b) \psi(\frac{x-b}{a}) \frac{dadb}{a^2}$$

where

$$C_\psi = \int_{-\infty}^{+\infty} \mid \hat{\psi}(t) \mid^2 \frac{dt}{t} < +\infty$$

Reconstruction is only possible if C_ψ is defined (admissibility condition). This condition implies $\hat{\psi}(0) = 0$, i.e. the mean of the wavelet function is 0.

Inverse Problems in Astrophysics

- ➡Part 1: Introduction to Inverse Problems
- ➡Part 2: From Fourier to Wavelets
- ➡**Part 3: Wavelet and Beyond**
- ➡Part 4: Sparse Regularization
- ➡Part 5: Application to Unmixing and Inpainting
- ➡Part 6: Compressed Sensing
- ➡Part 7: Deep Learning

Yves Meyer

A Major Breakthrough

Daubechies, 1988 and Mallat, 1989

I. Daubechies:

Compactly Supported Orthogonal and Bi-Orthogonal Wavelets

S. Mallat:

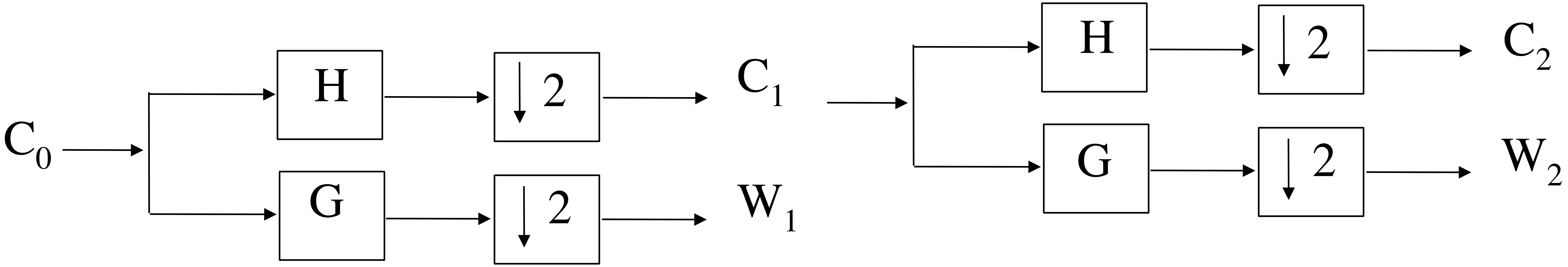
Theory of Multiresolution Signal Decomposition

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

The Orthogonal Wavelet Transform (OWT)

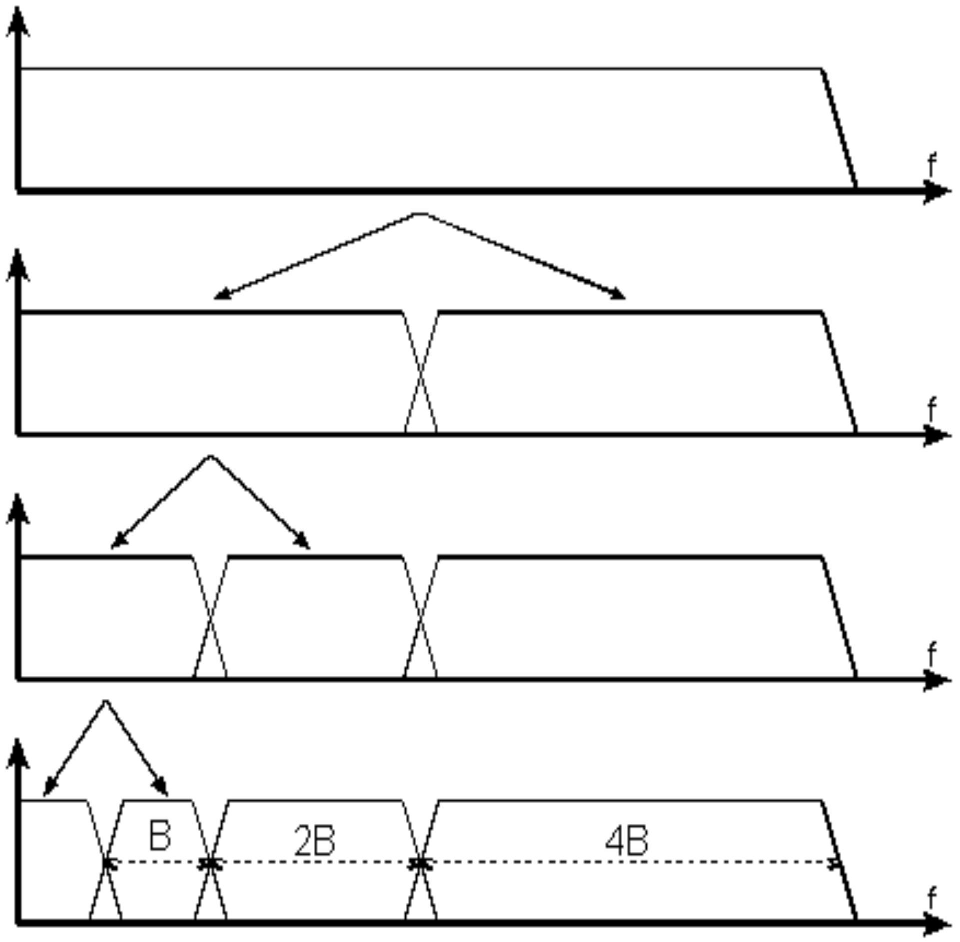
$$s_l = \sum_k c_{J,k}\phi_{J,l}(k) + \sum_k \sum_{j=1}^{J} \psi_{j,l}(k) w_{j,k}$$

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:

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

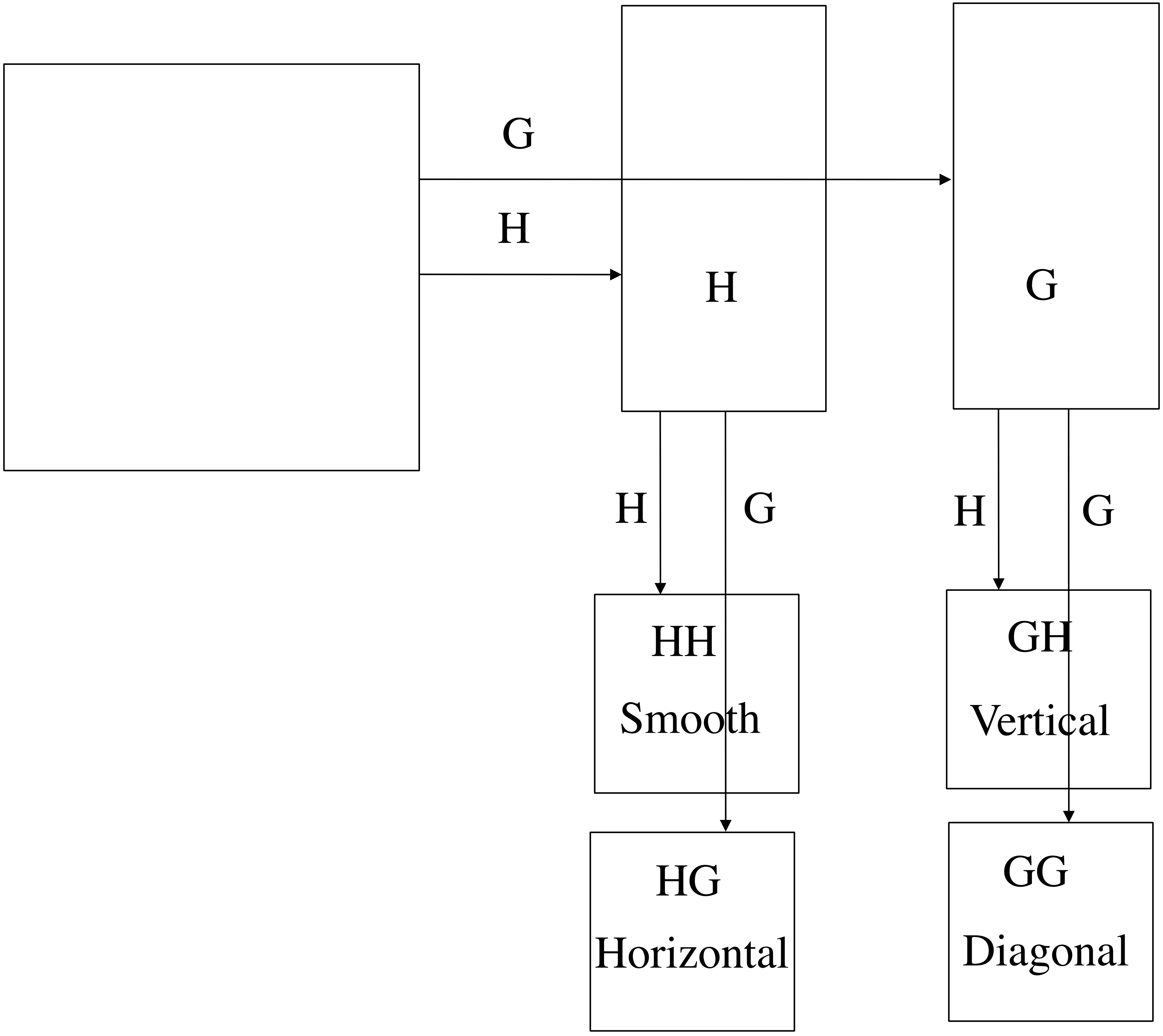
$$\breve{x} = (x_1, 0, x_2, 0, x_3, \ldots, 0, x_j, 0, \ldots, x_{n-1}, 0, x_n)$$

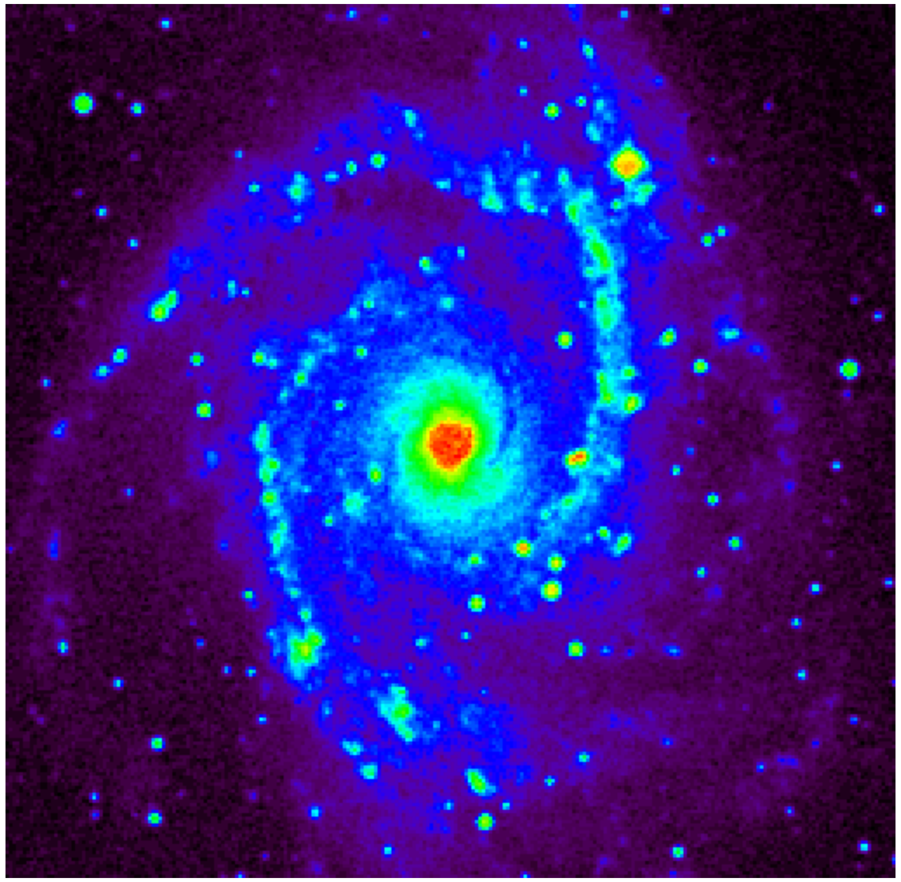
At two dimensions, we separate the variables x,y:

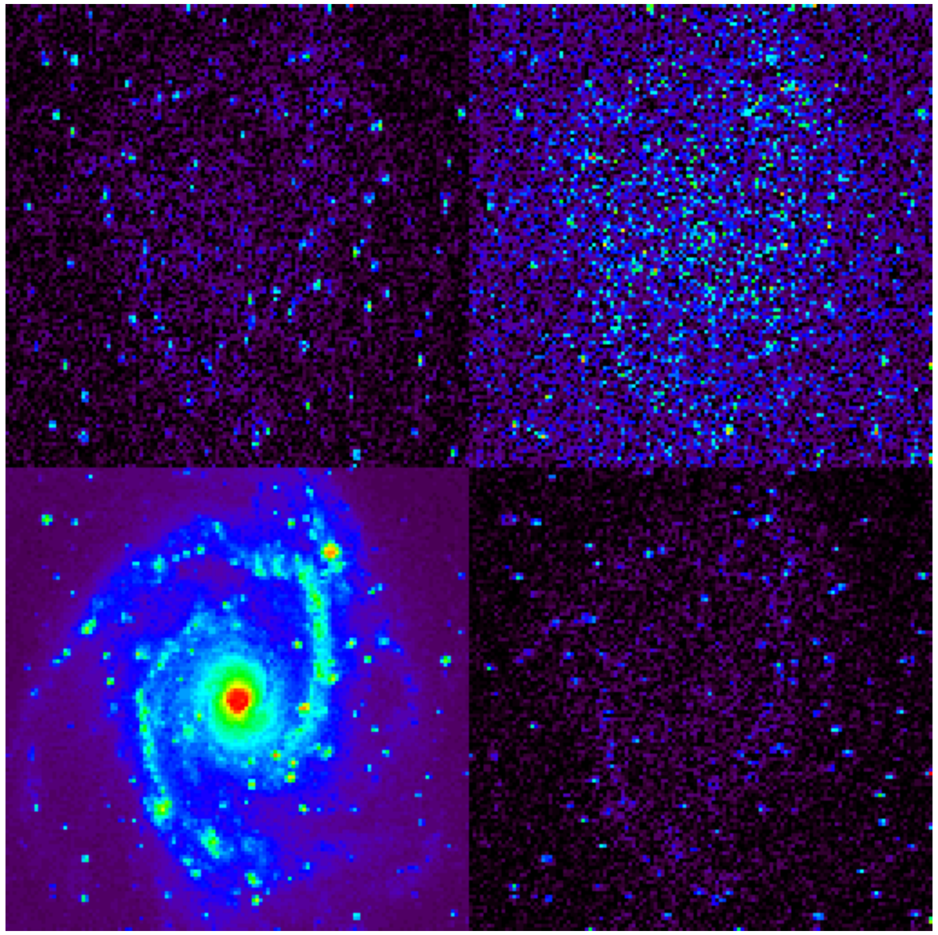
- vertical wavelet: $\psi^1(x,y) = \phi(x)\psi(y)$
- horizontal wavelet: $\psi^2(x,y) = \psi(x)\phi(y)$
- diagonal wavelet: $\psi^3(x,y) = \psi(x)\psi(y)$

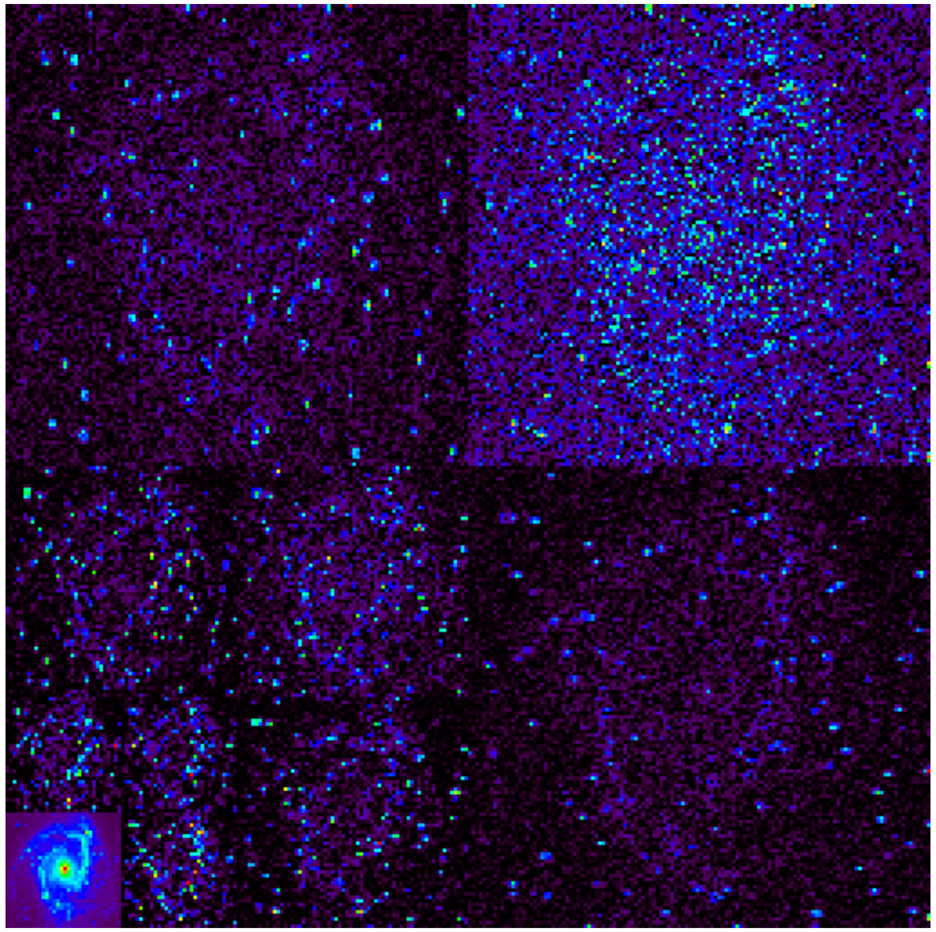
The detail signal is contained in three sub-images

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

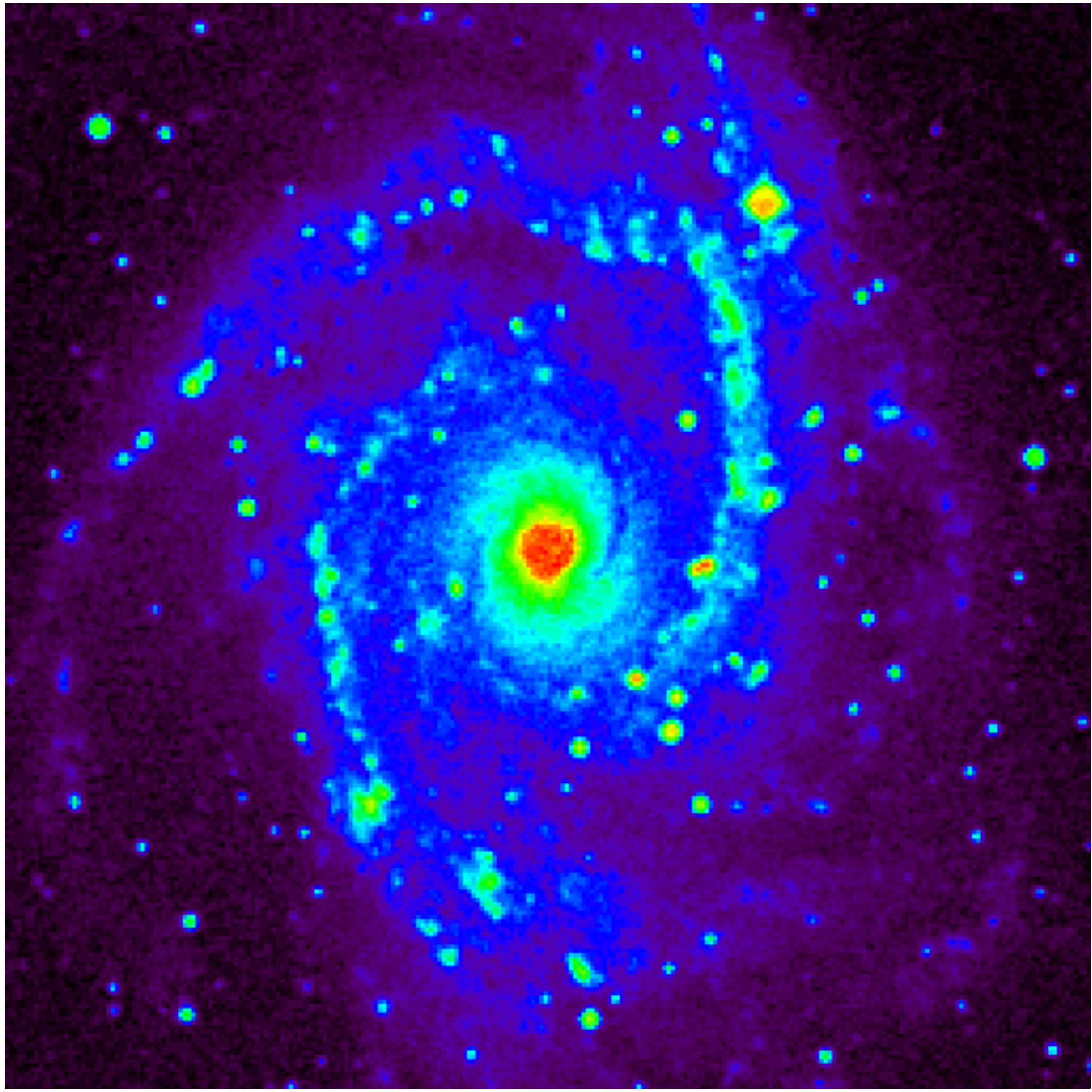

G
H
H
G
H
G
H
G
HH
Smooth
GH
Vertical
HG
Horizontal
GG
Diagonal



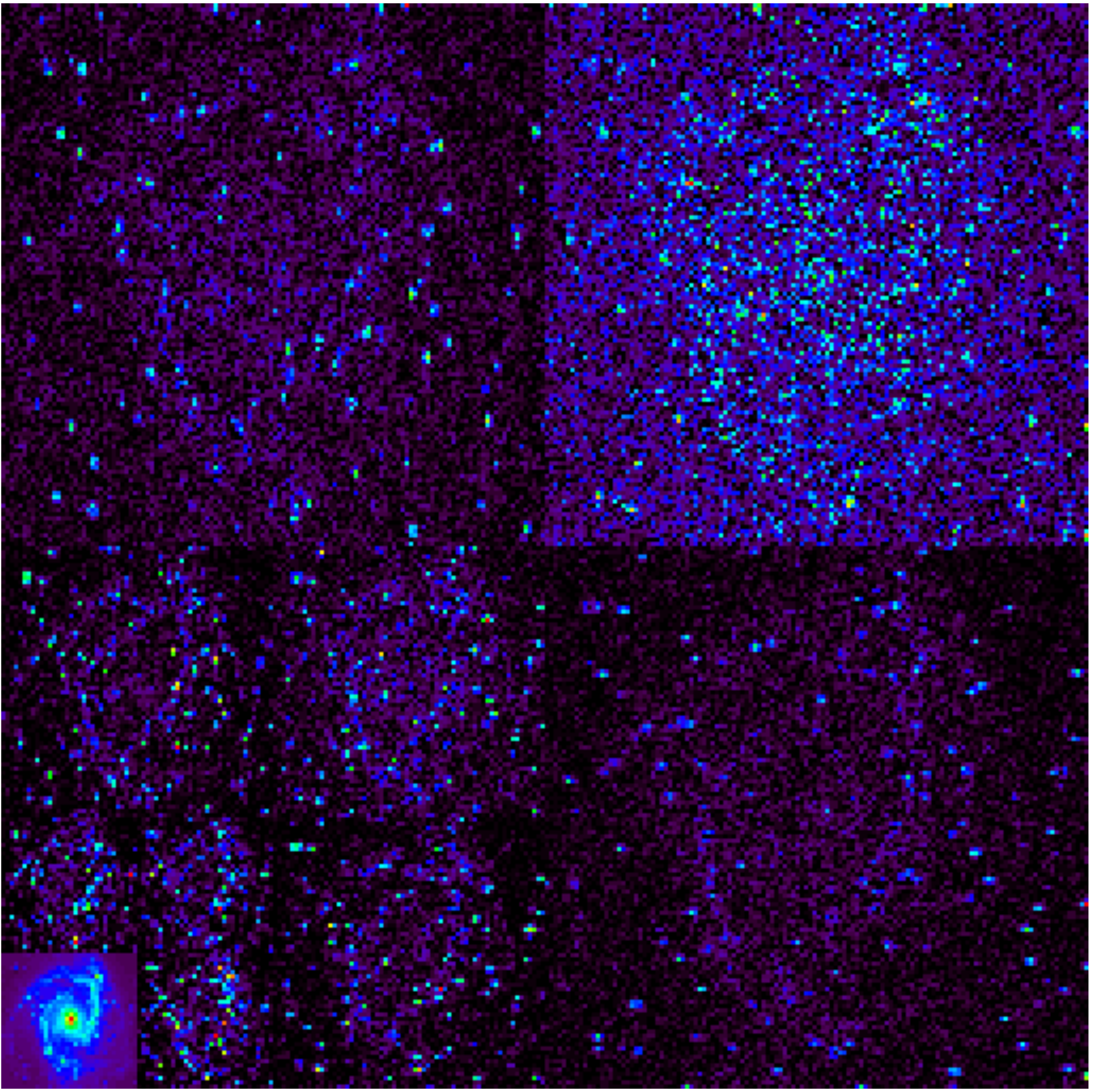


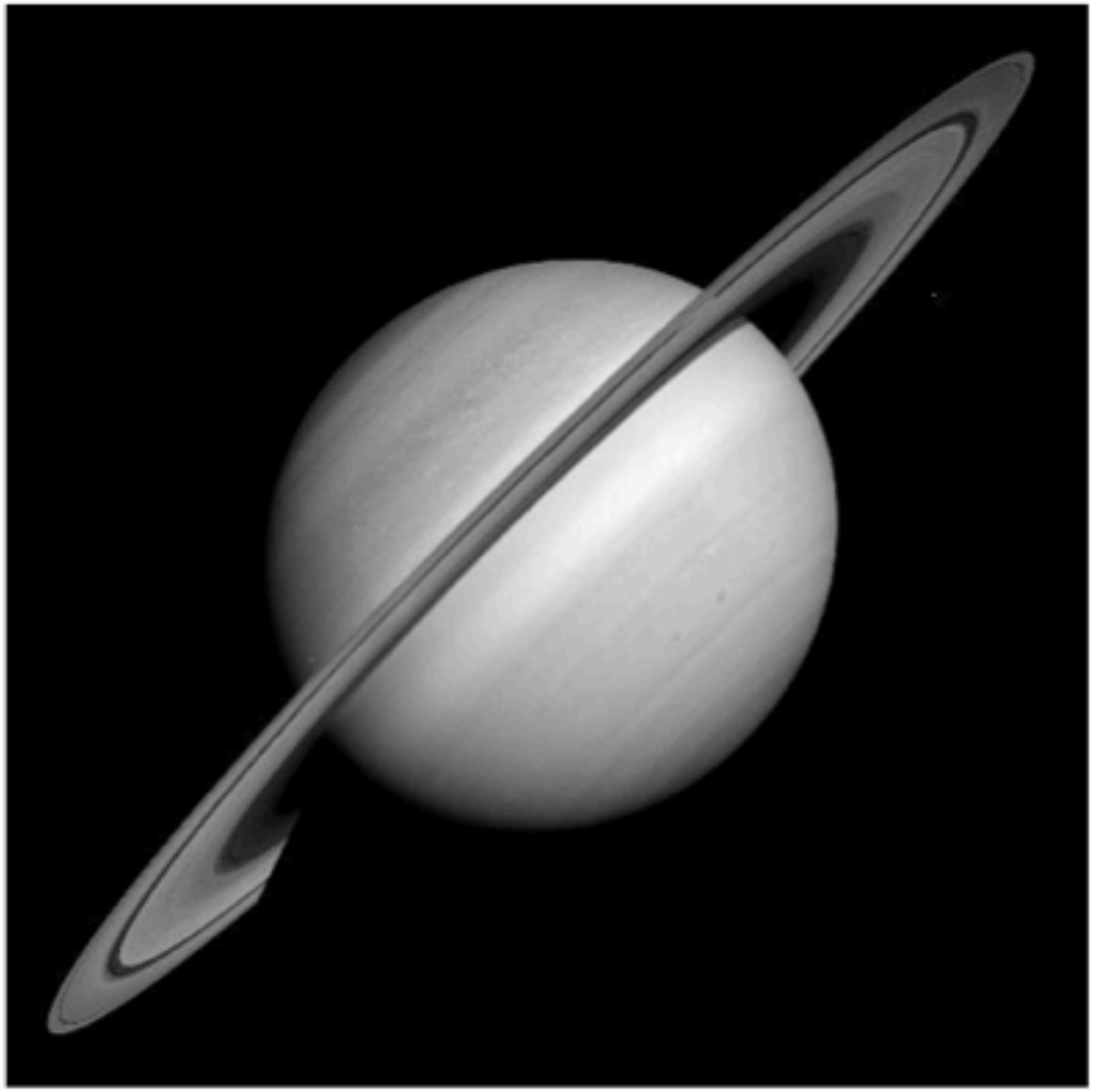


NGC2997



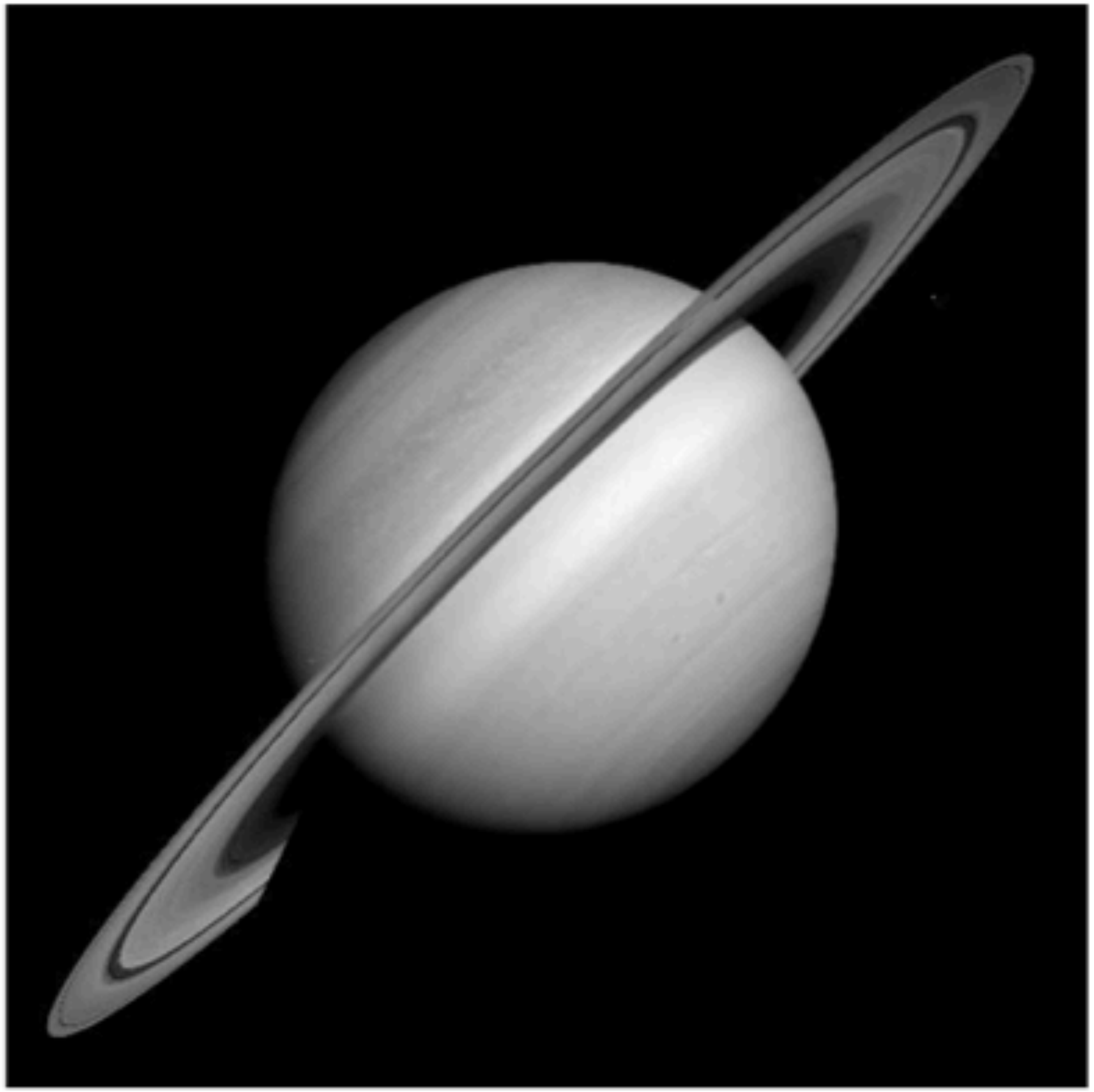
NGC2997 WT





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

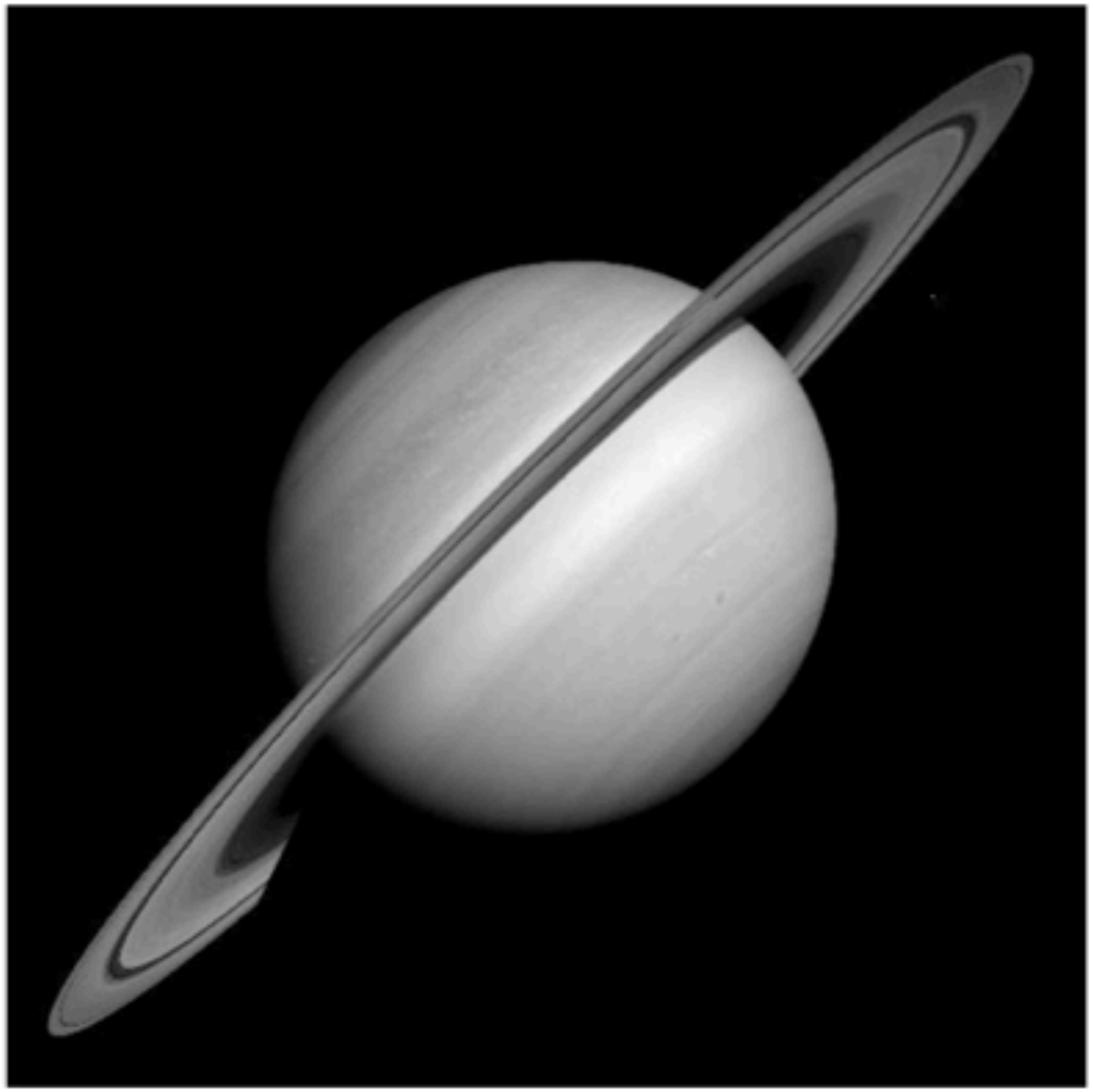
1% largest coefficients in real space
(the others are set to 0)



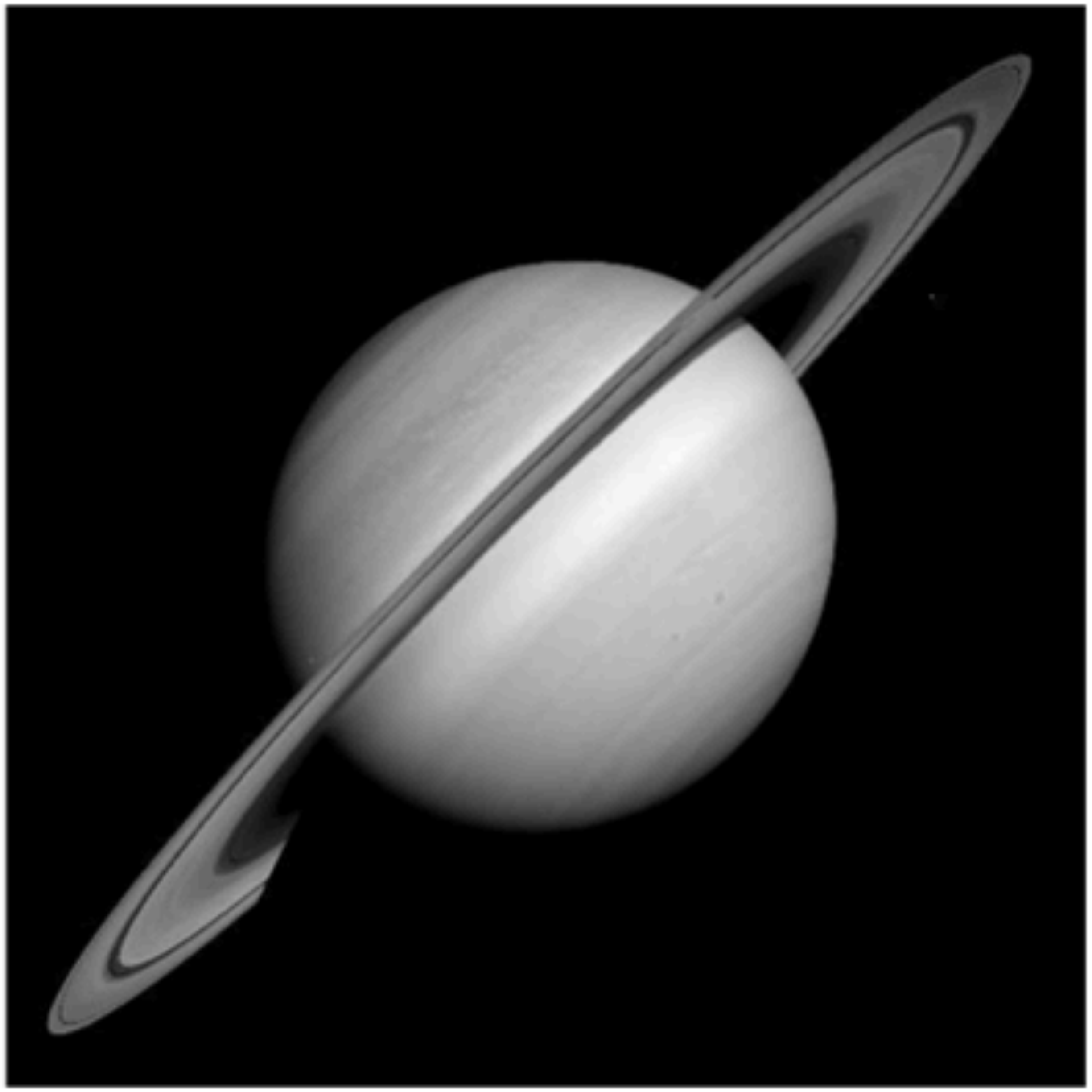
1% largest coefficients in wavelet space
(the others are set to 0)

Wavelet transform

The wavelet coefficients encode edges and large scale information.



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

JPEG/JPEG 2000

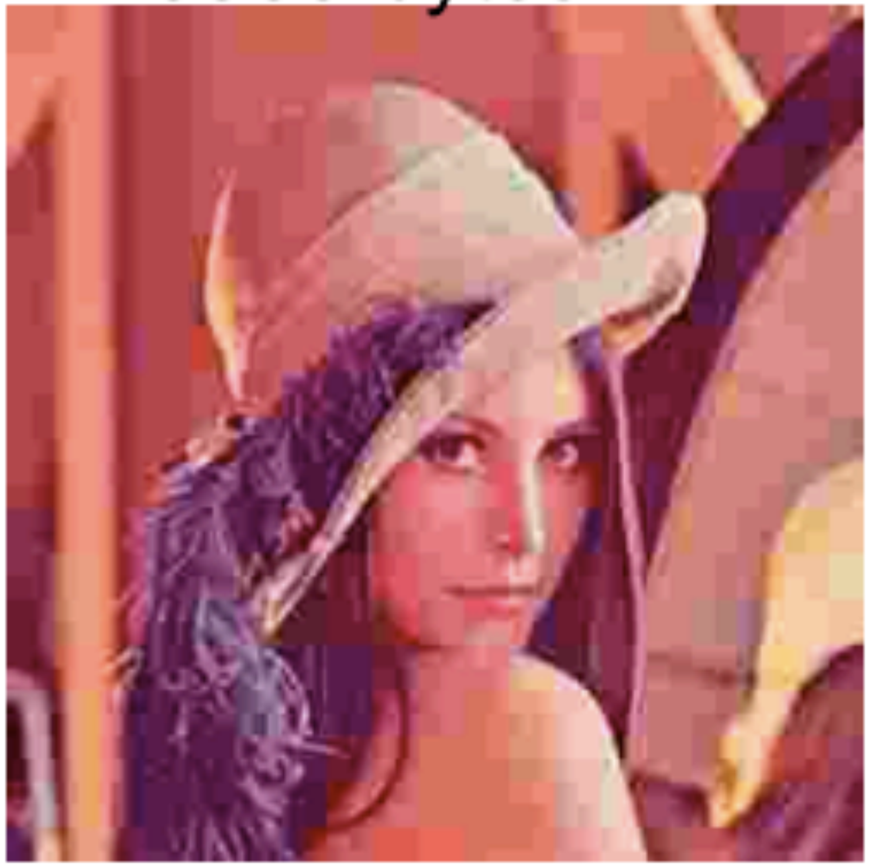
Original BMP

300x300x24

270056 bytes

JPEG 1:68

3983 bytes

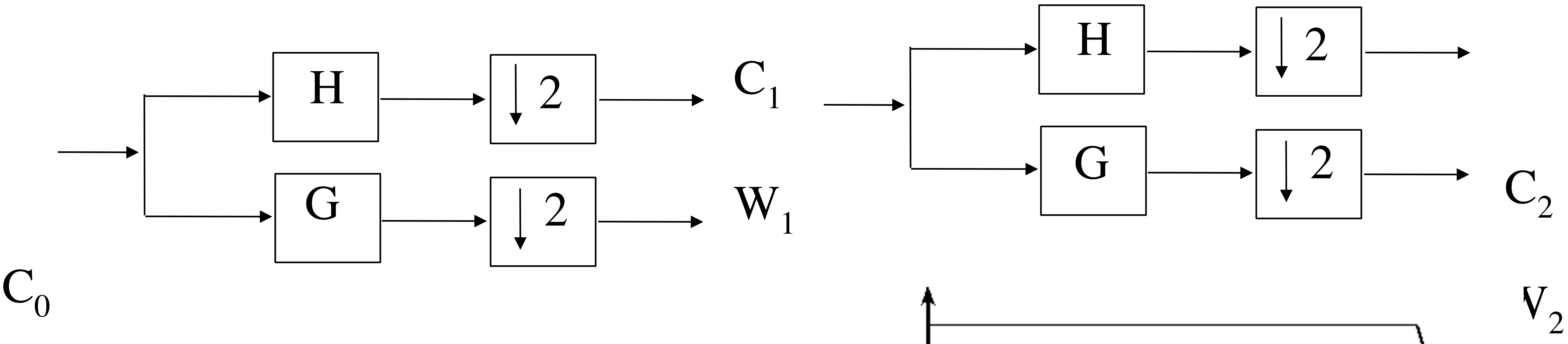


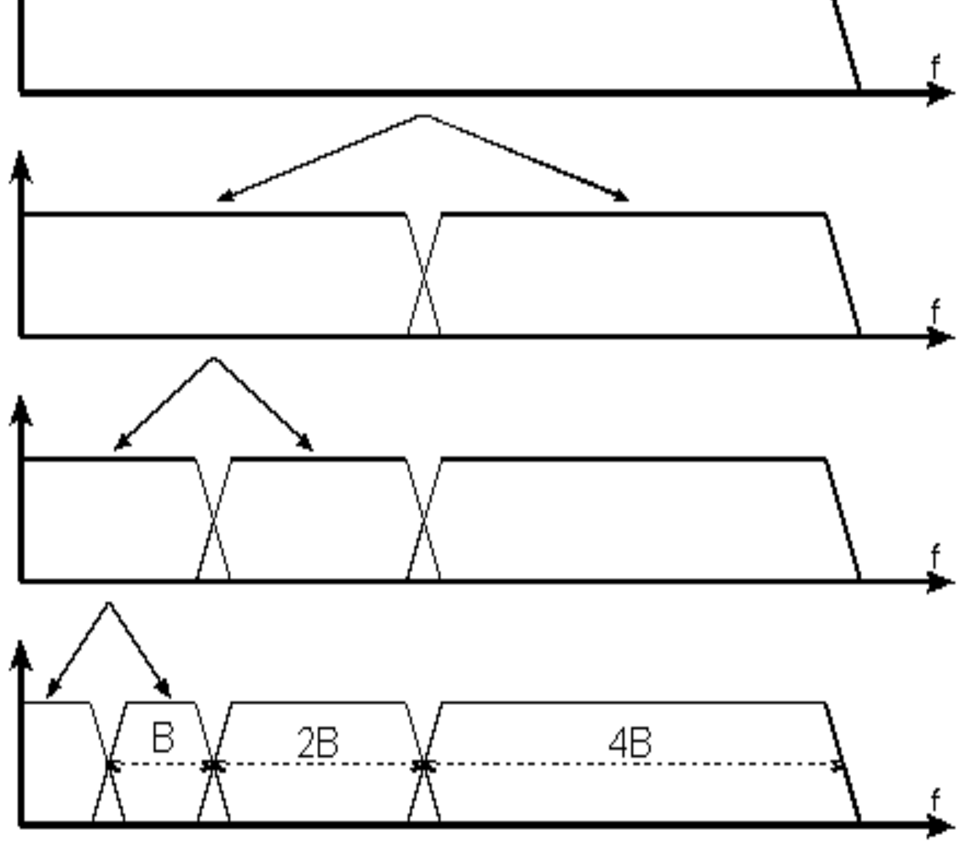
JPEG2000 1:70

3876 bytes

Translation Invariance

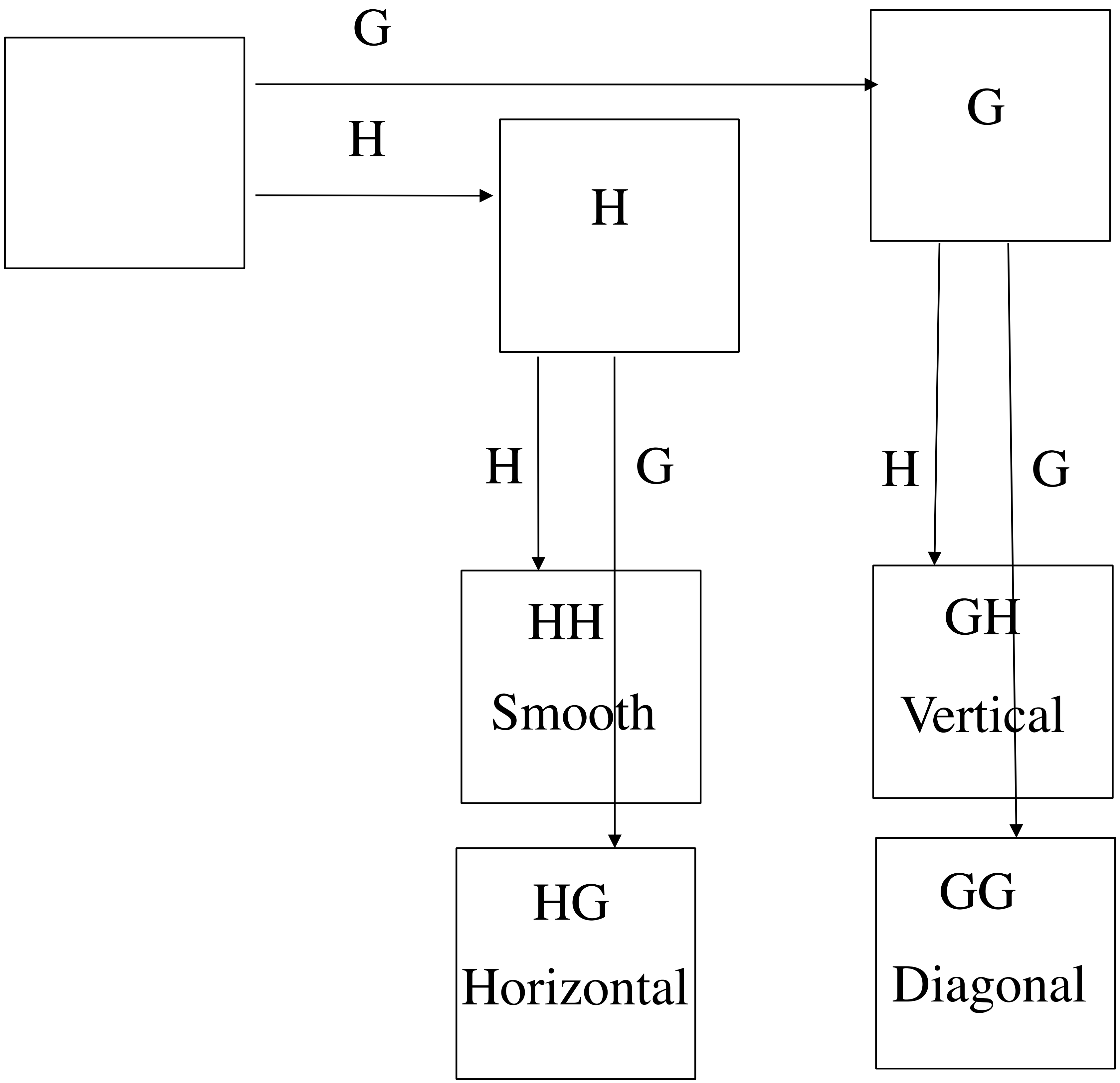
Transformation





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

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



G

H

H

G

H

G

H

G

HH

Smooth

GH

Vertical

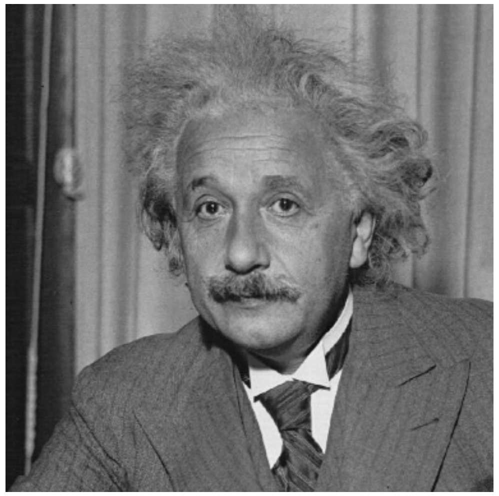
HG

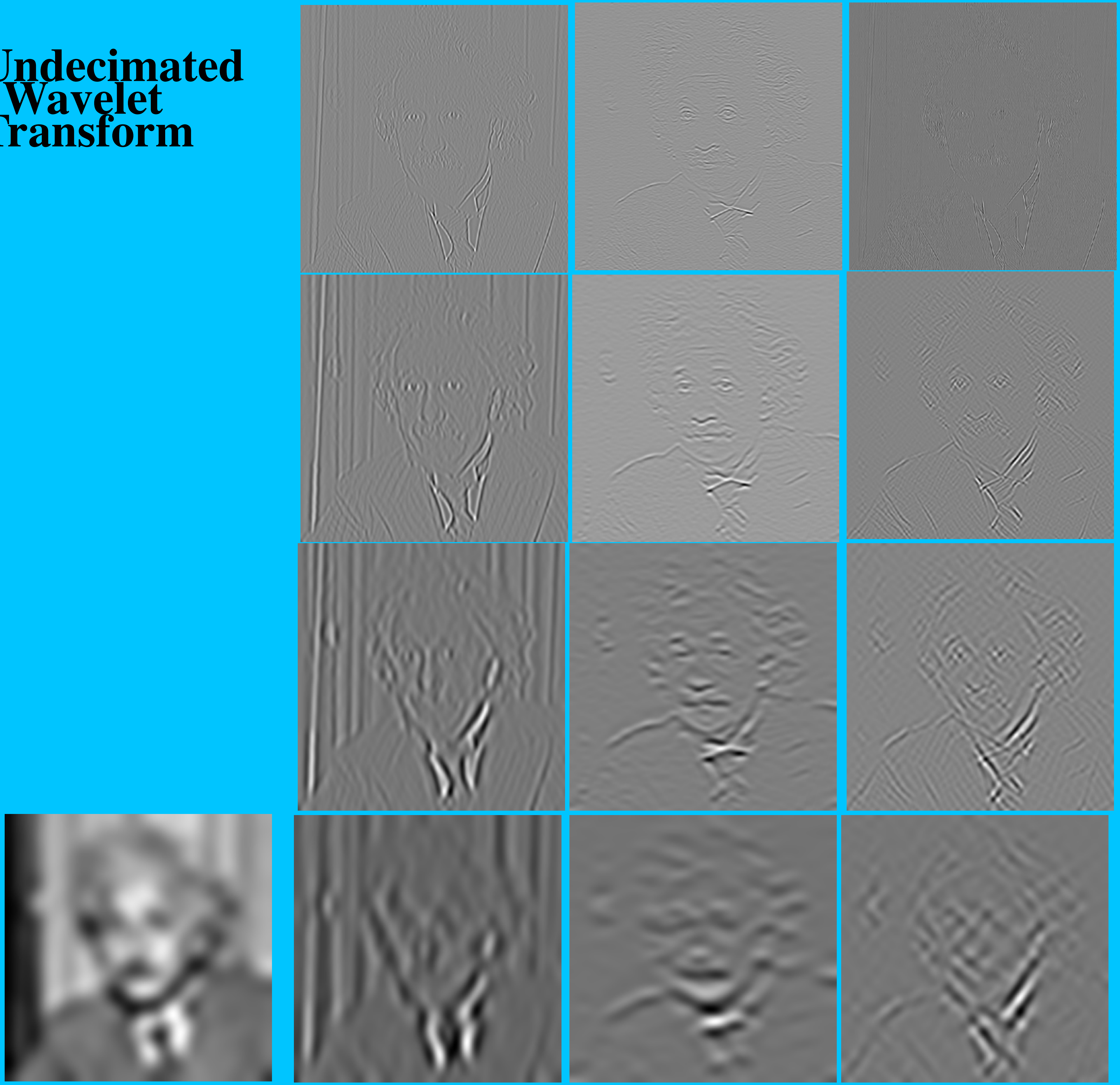
Horizontal

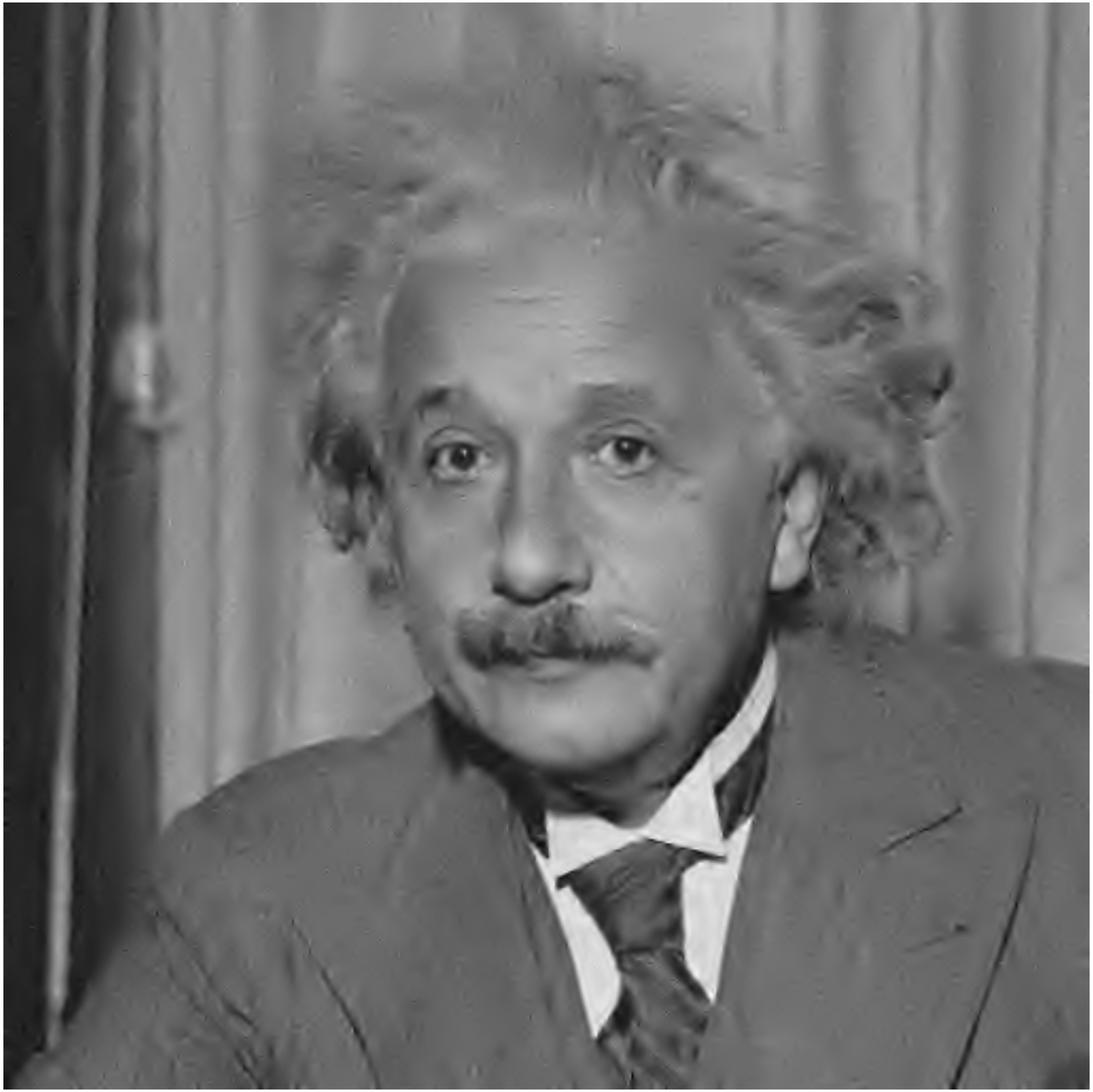
GG

Diagonal

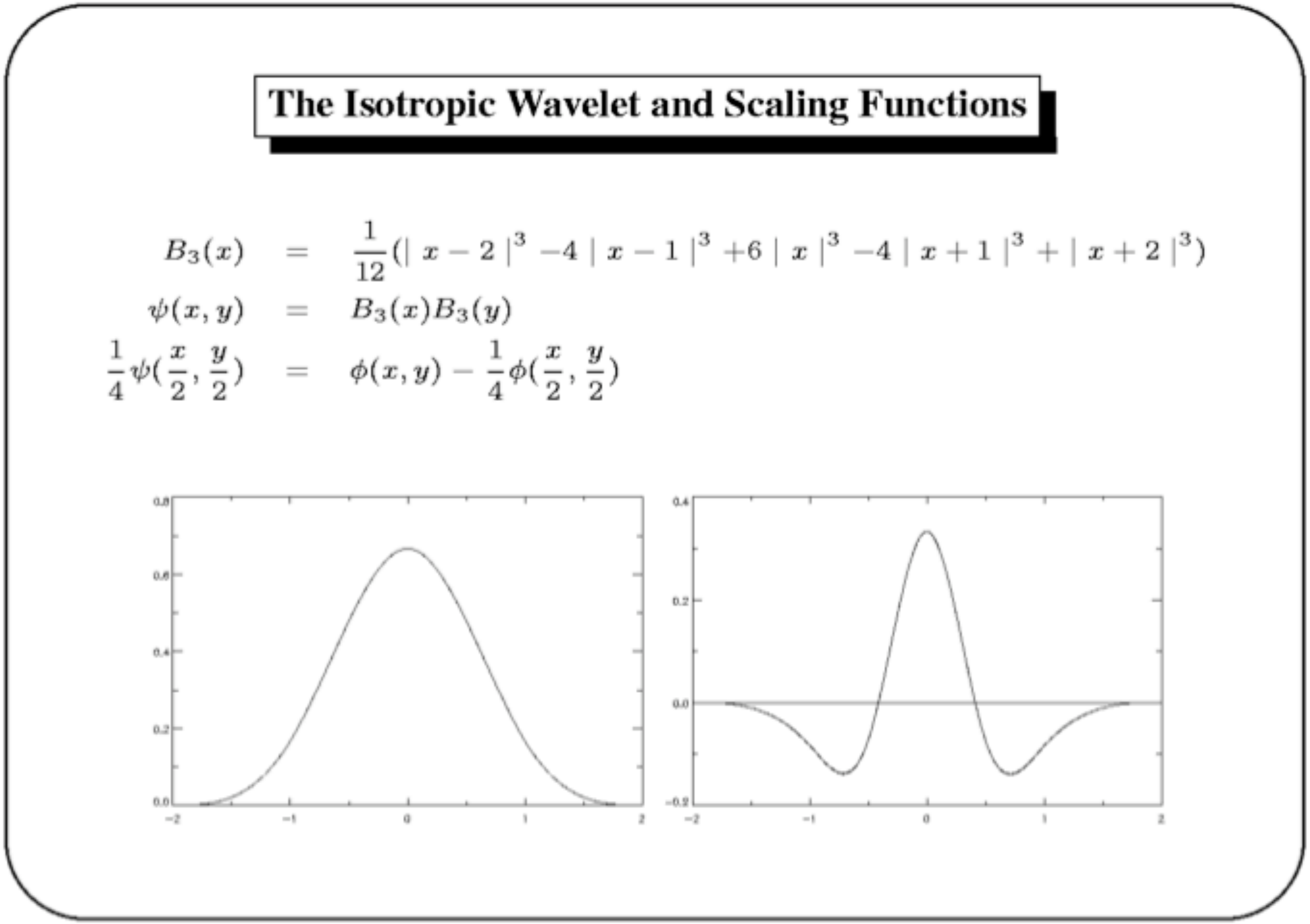
Undecimated Wavelet Transform





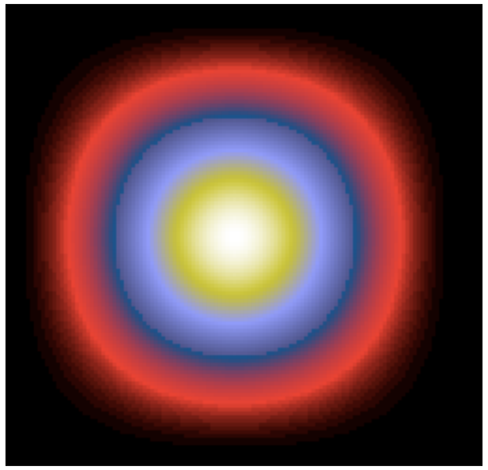


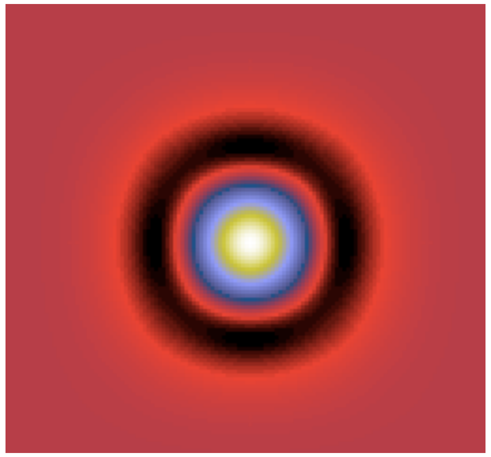
Wavelet Transform in Astronomy



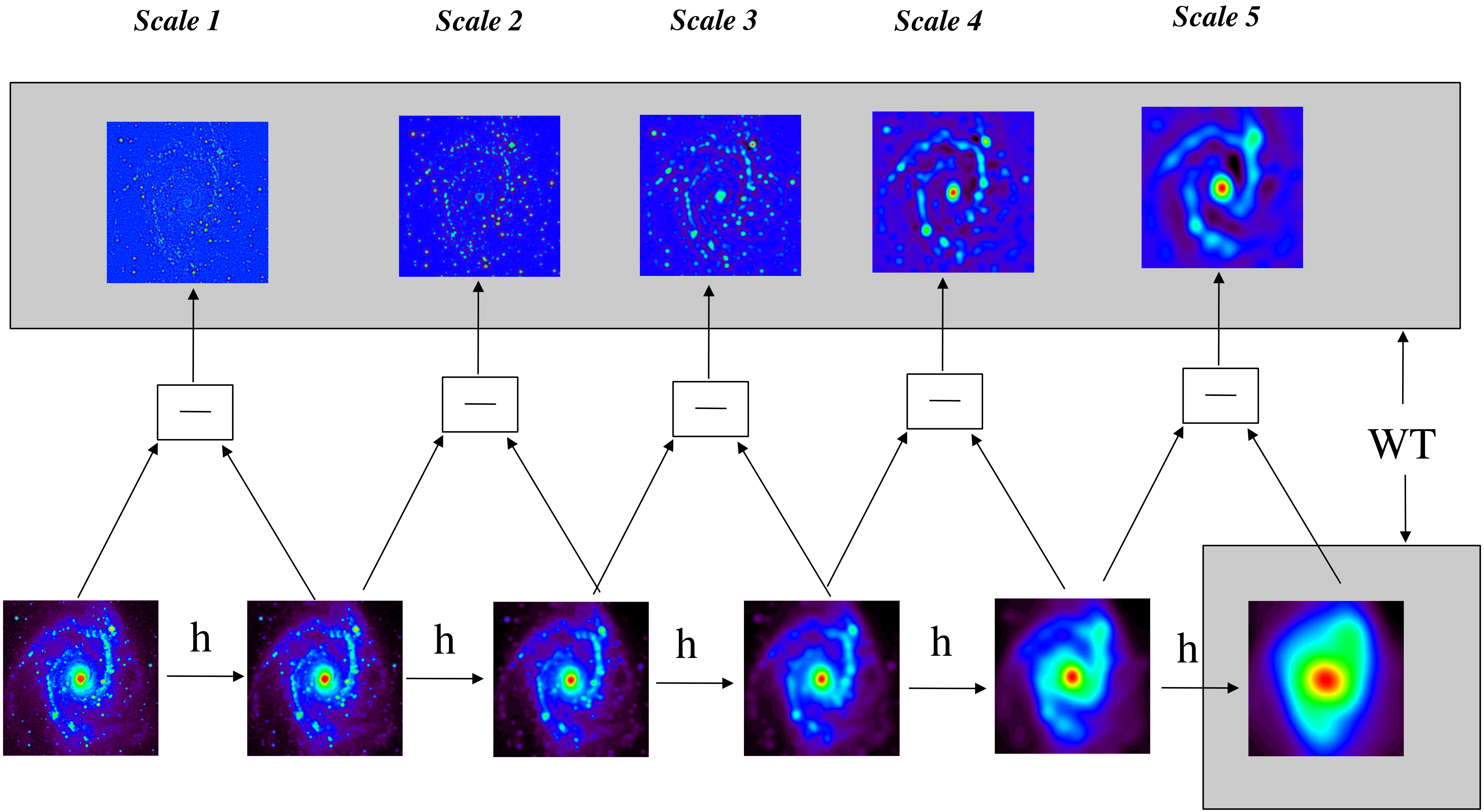
The Isotropic Wavelet and Scaling Functions

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





ISOTROPIC UNDECIMATED WAVELET TRANSFORM



Scale 1

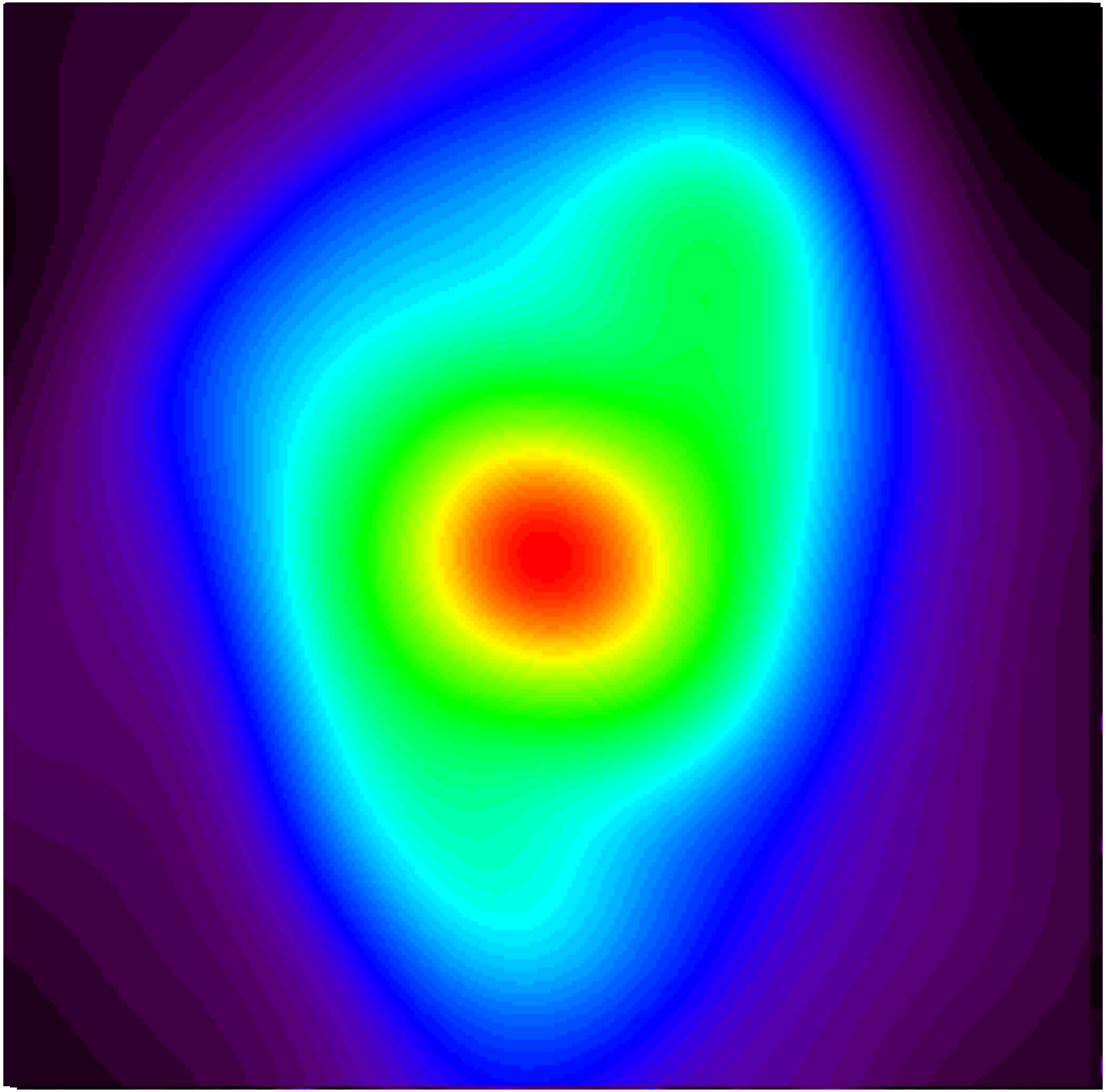
Scale 2

Scale 3

Scale 4

Scale 5

NGC2997



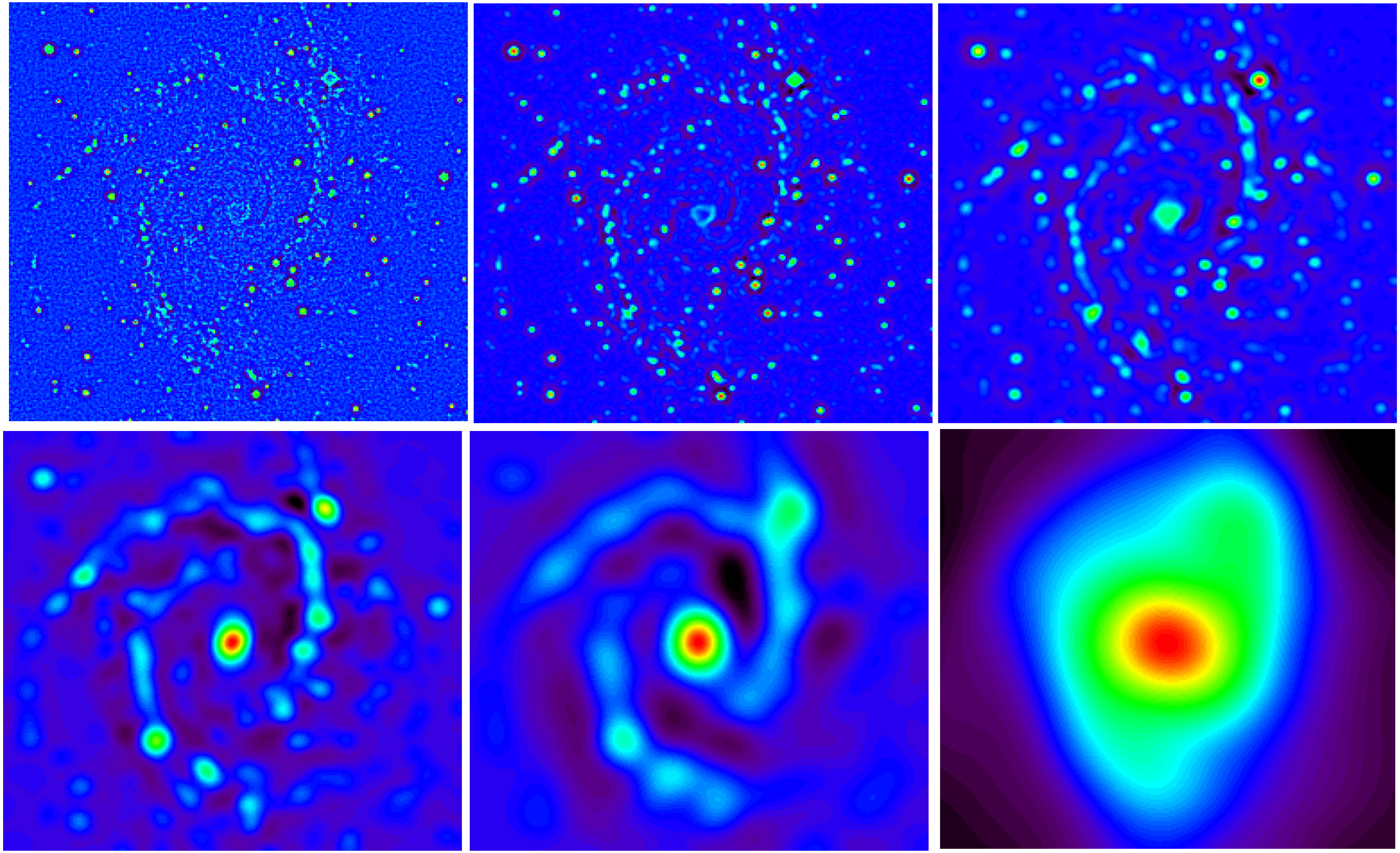
The STARLET Transform

Isotropic Undecimated Wavelet Transform (a trous algorithm)

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

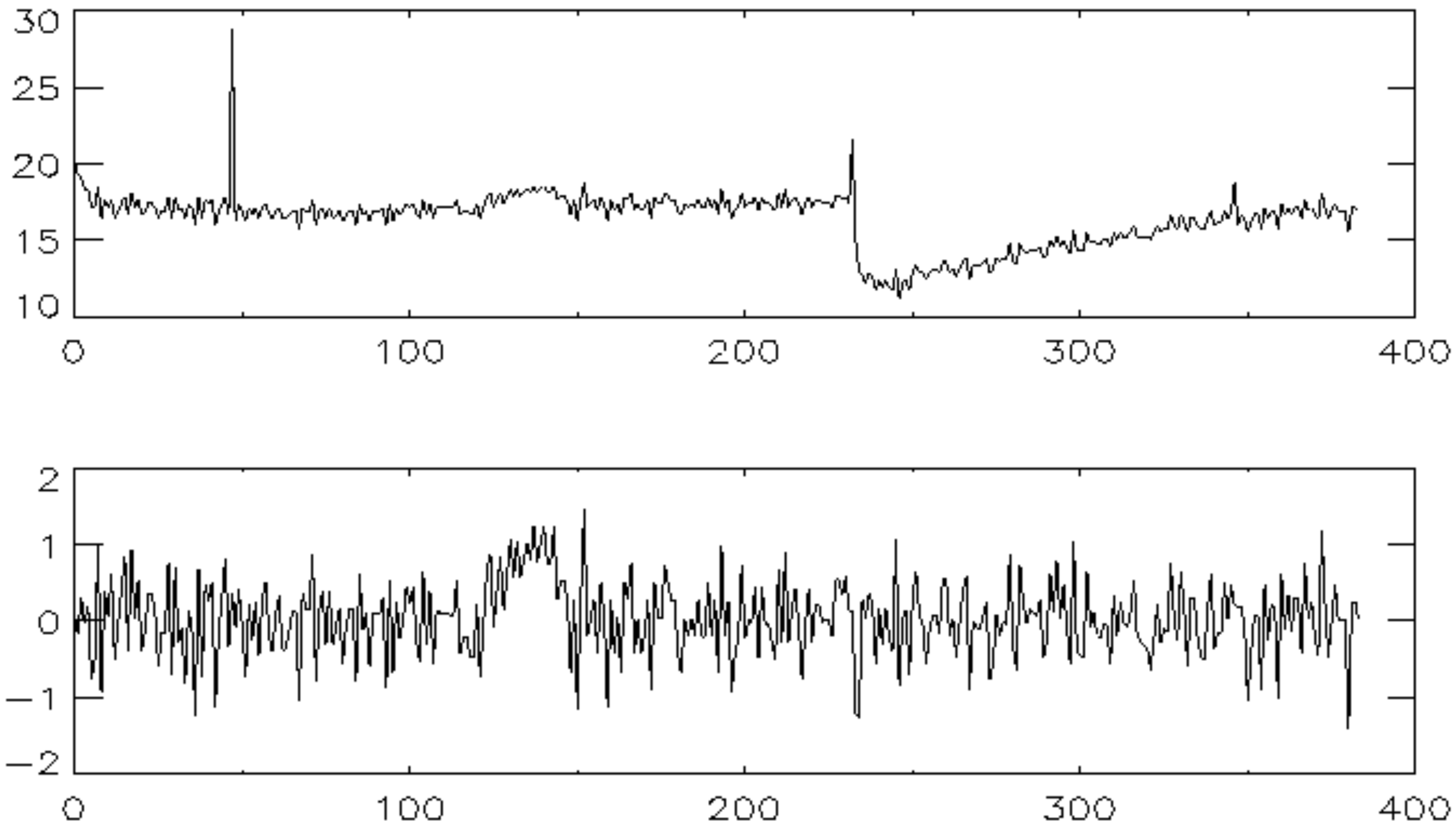
$$I(k,l) = c_{J,k,l} + \sum_{j=1}^{J} w_{j,k,l}$$

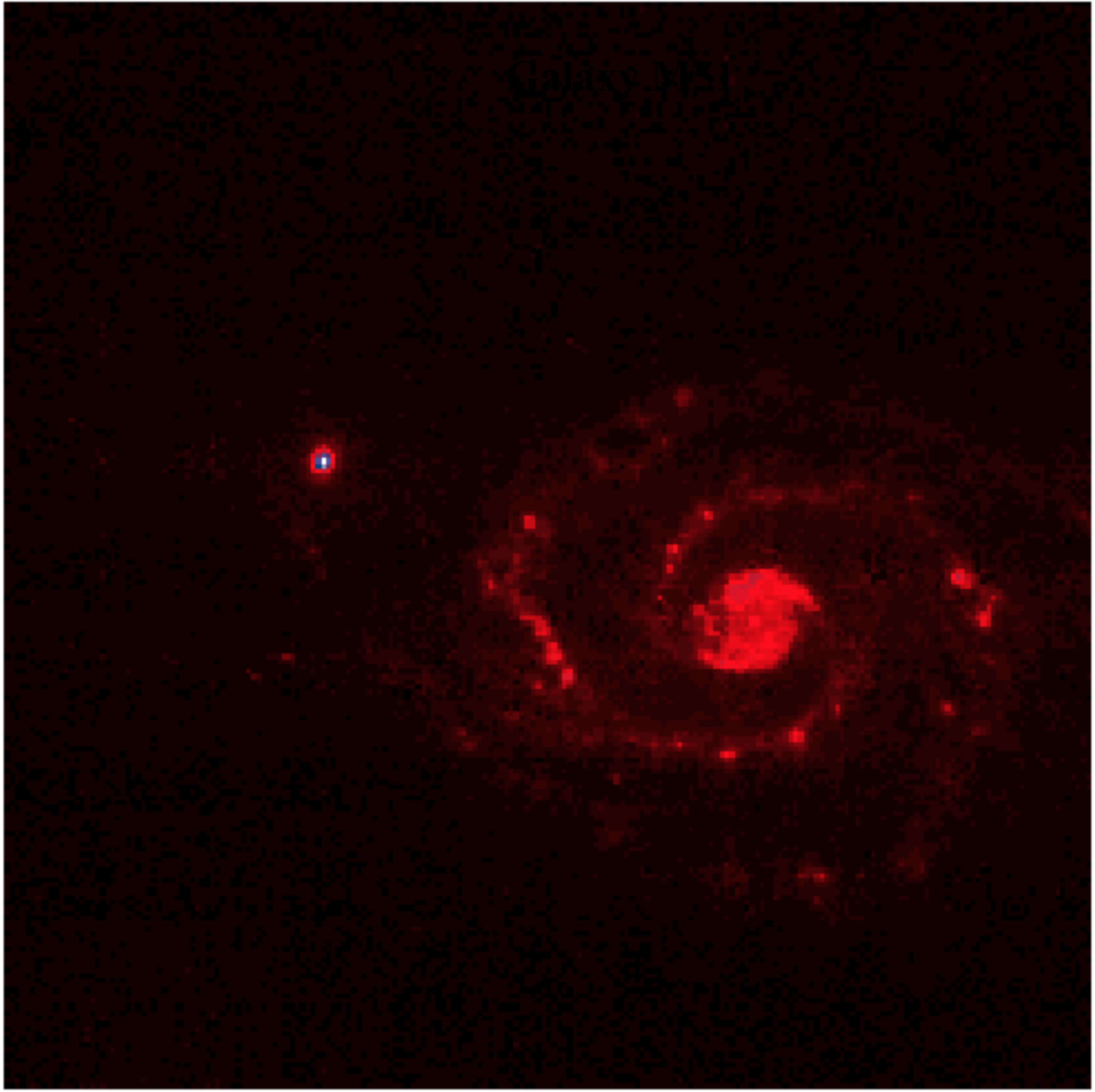
$$h = [1,4,6,4,1]/16, \quad g = \delta - h, \quad \tilde{h} = \tilde{g} = \delta$$



Infrared Space Observatory (ISO)

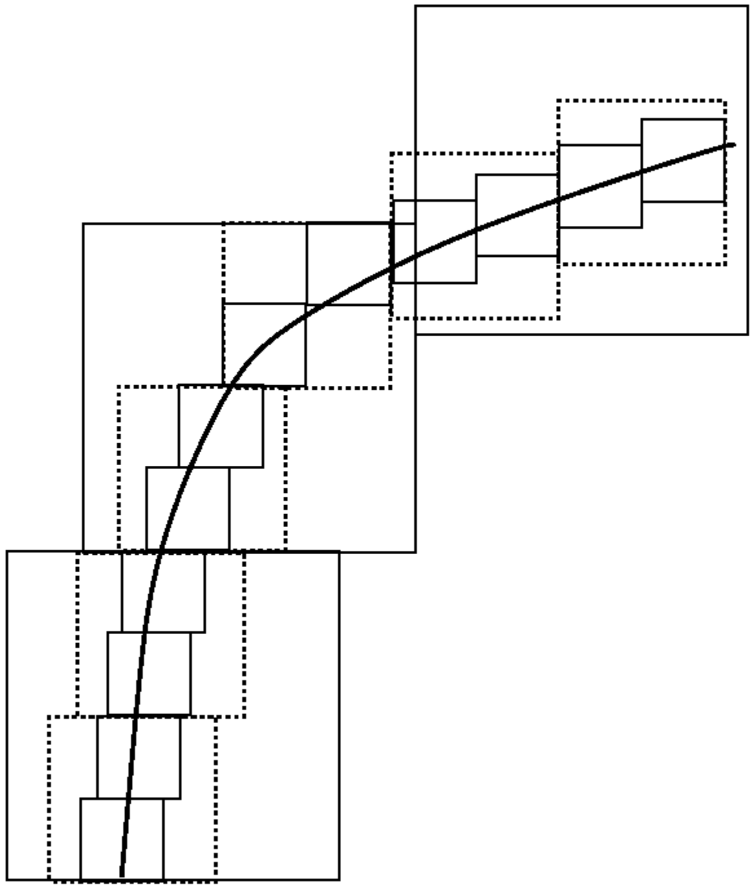
The calibration from pattern recognition consists in searching only for objects which verify given conditions. For example, finding glitches of the first type is equivalent to finding objects which are positive, strong, and with a temporal size lower than that of the sources.



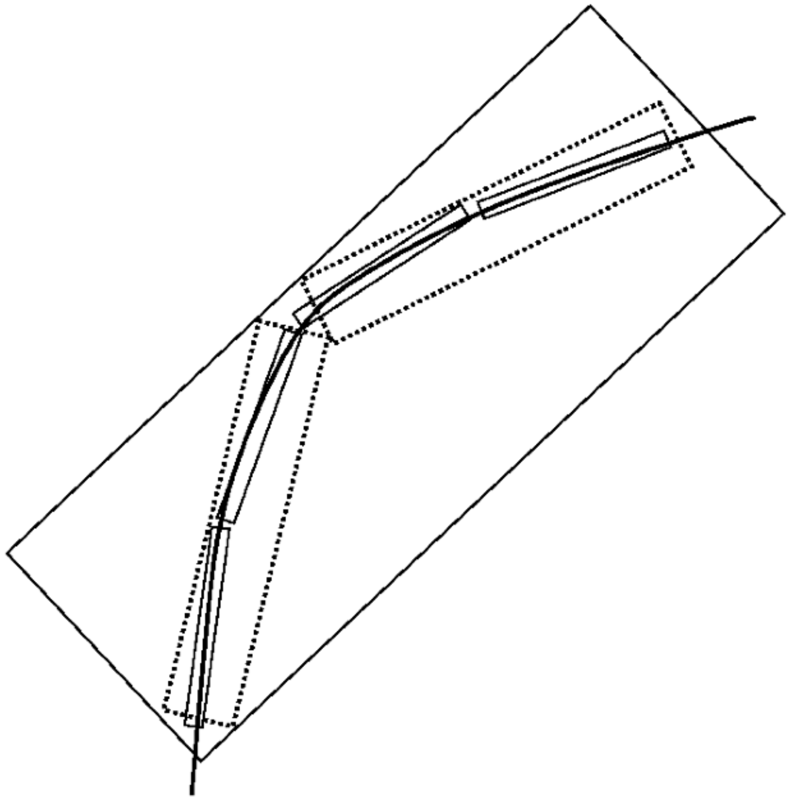

Galaxy M51

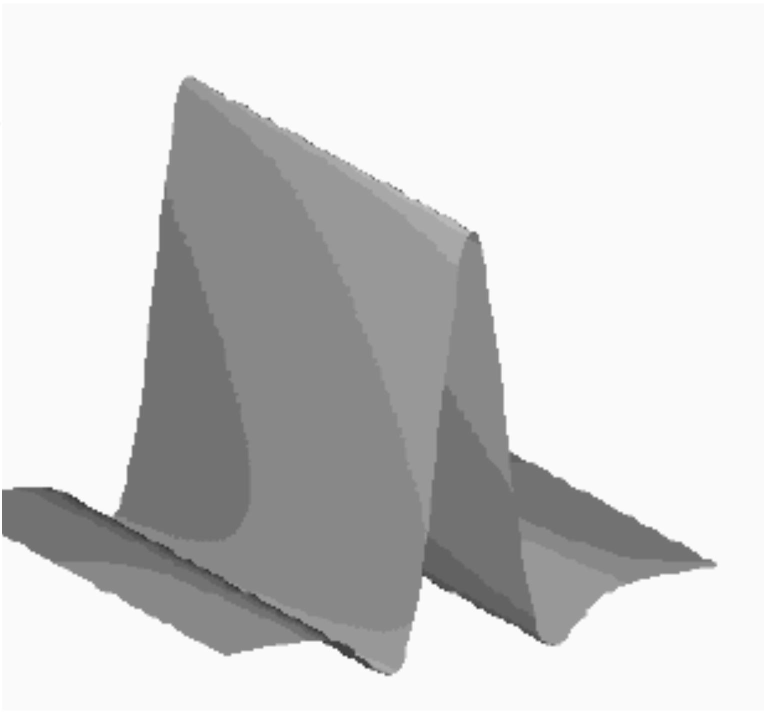
Wavelets and edges

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

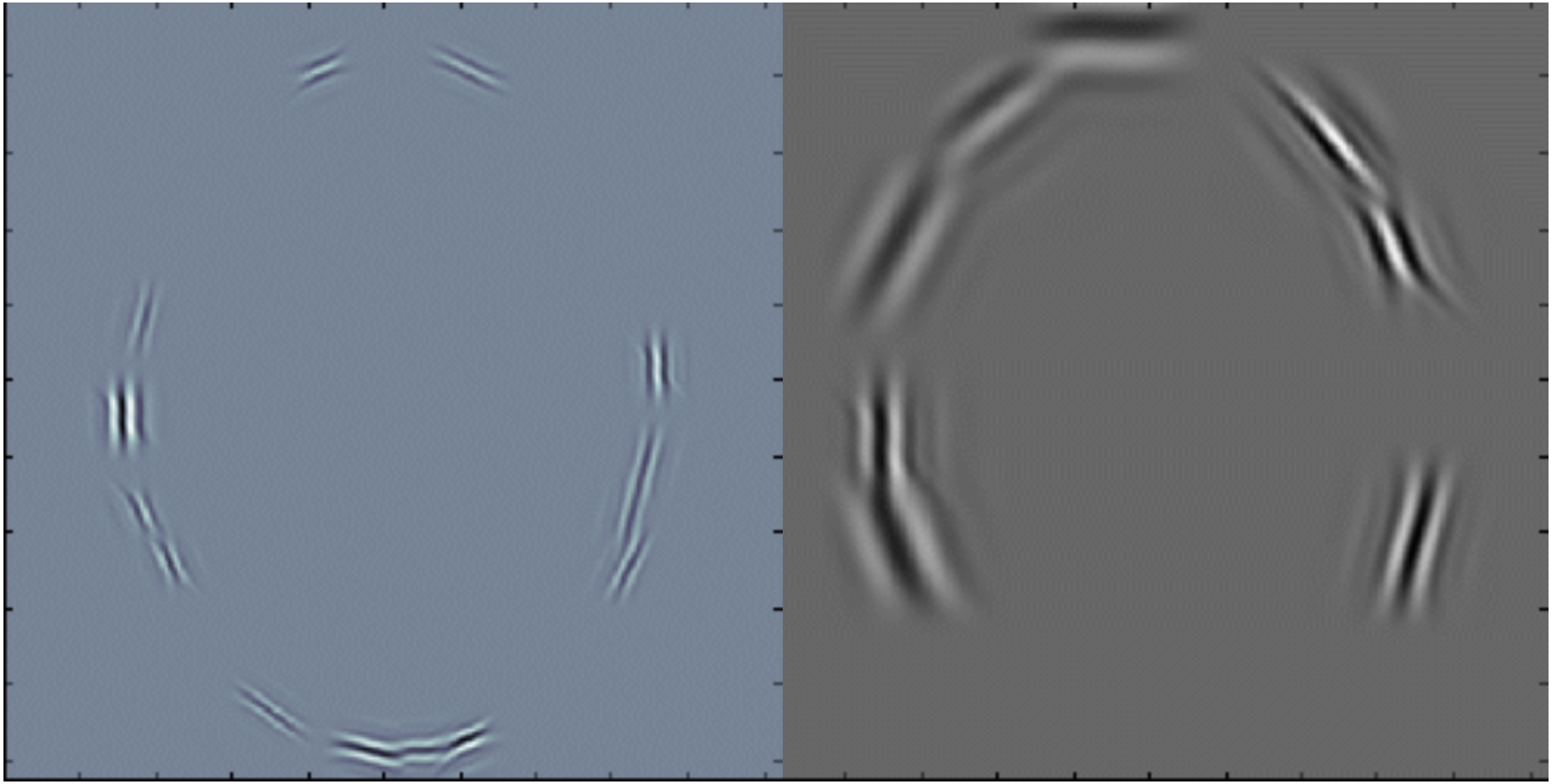


- need dictionaries of strongly anisotropic atoms :





→ ridgelets, curvelets, contourlets, bandelettes, etc.



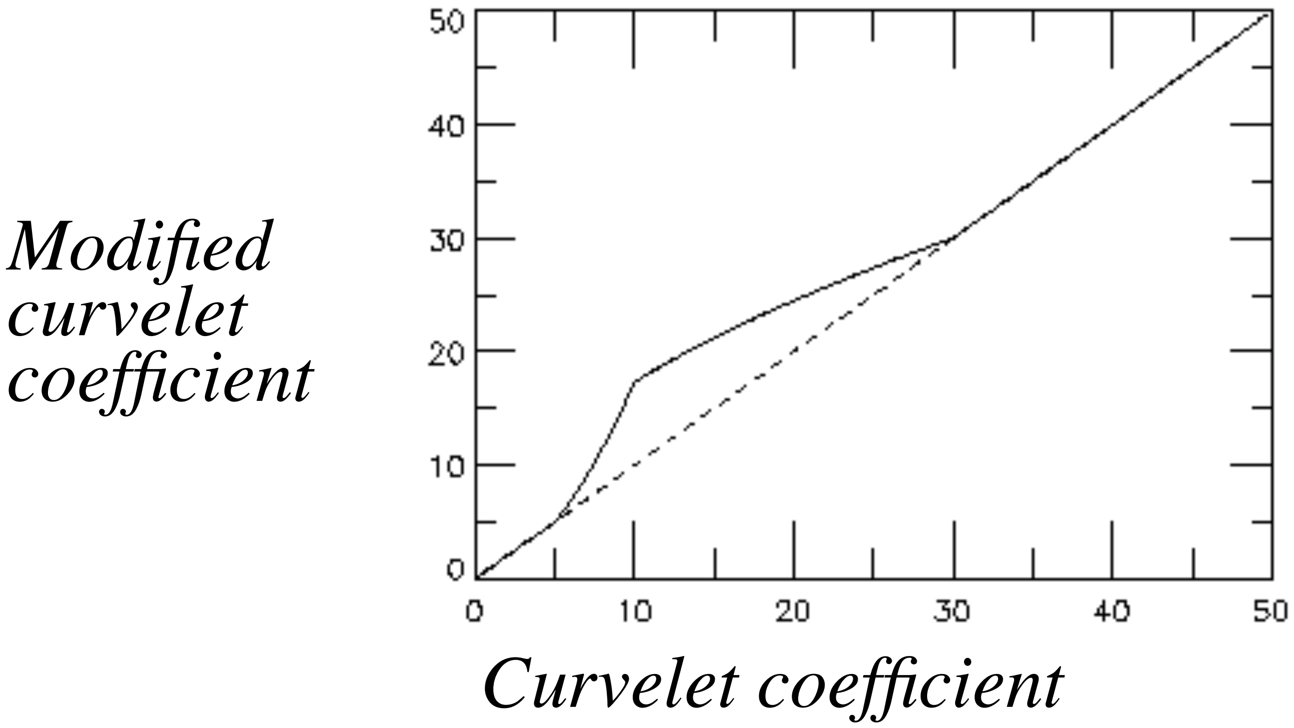
- J.L. Starck, E. Candes, and D.L. Donoho, "**The Curvelet Transform for Image Denoising**", IEEE Transactions on Image Processing , 11, 6, pp 670 -684, 2002.
- J.-L. Starck, M.K. Nguyen and F. Murtagh, "**Wavelets and Curvelets for Image Deconvolution: a Combined Approach**", Signal Processing, 83, 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

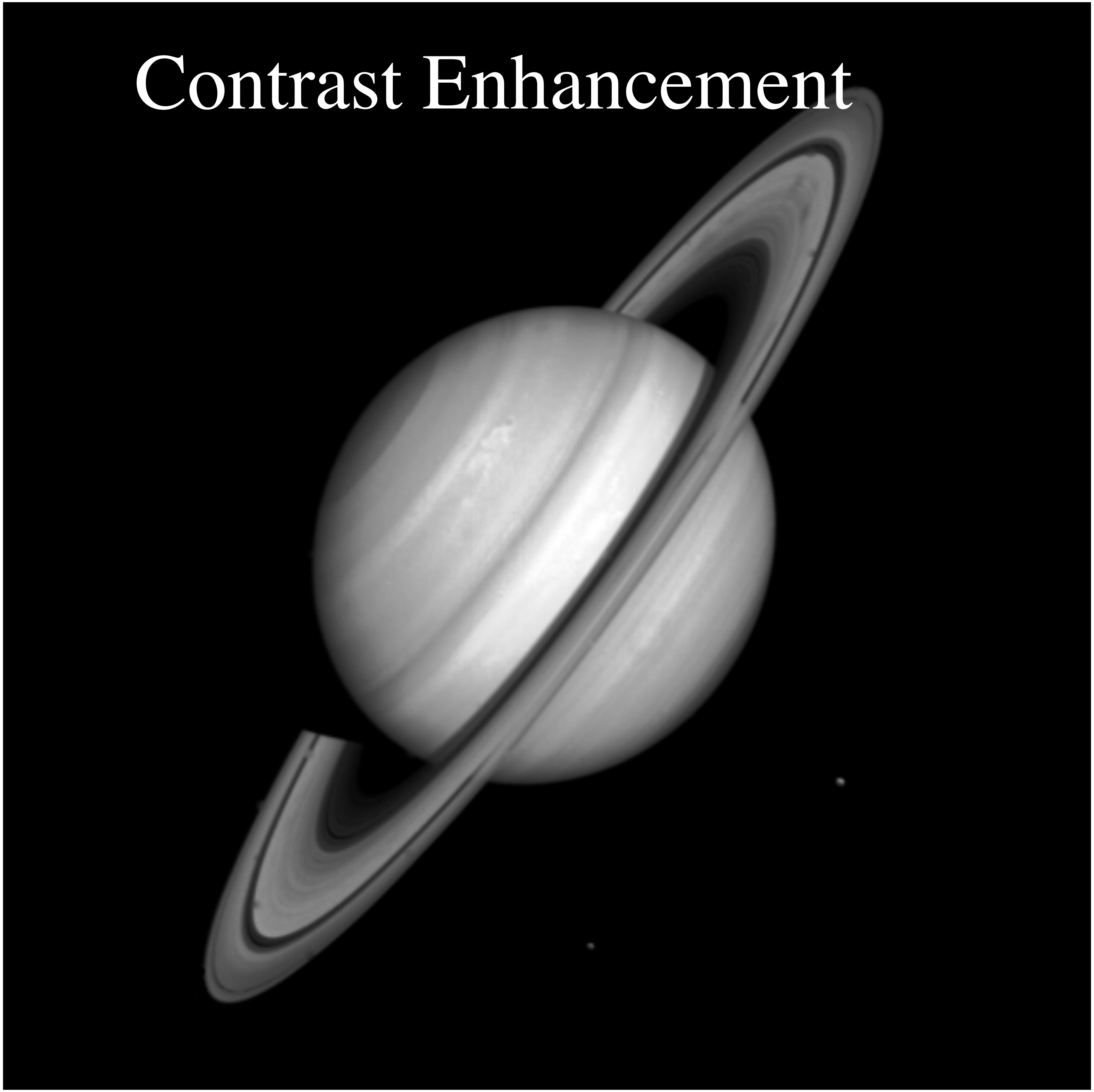
J.-L Starck, F. Murtagh, E. Candes and D.L. Donoho, ***"Gray and Color Image Contrast Enhancement by the Curvelet Transform"***,
IEEE Transaction on Image Processing, 12, 6, 2003.

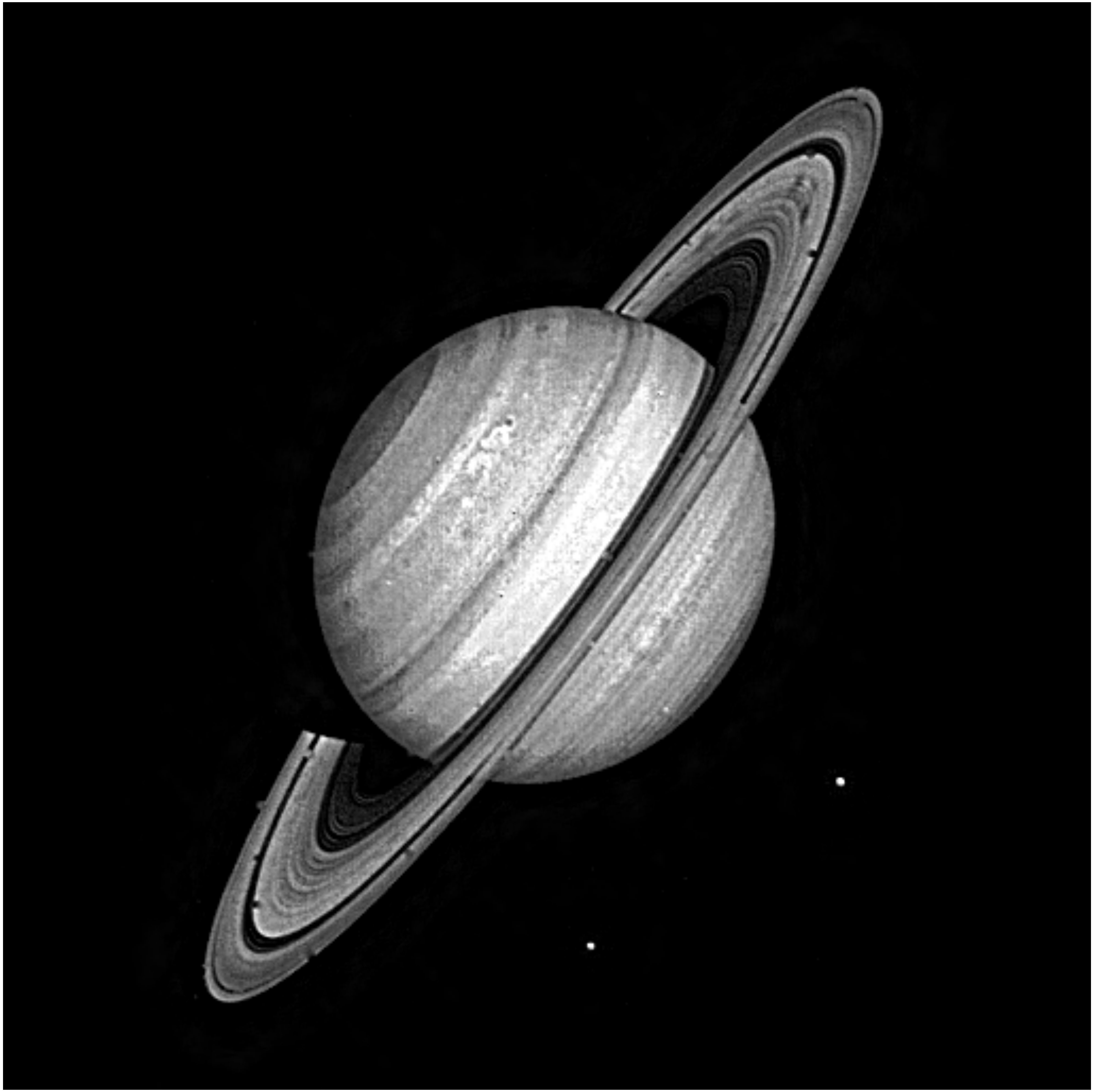
$$\tilde{I} = C_R\left(y_c\left(C_T I\right)\right)$$

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



Contrast Enhancement





1990-2010: X-let

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)

Multiscale Geometric Analysis

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

Ridgelet
Curvelet (Several implementations)
Wave Atom
Grouplet
[adaptive] Shearlet (Several implementations)

Fixed Dictionaries

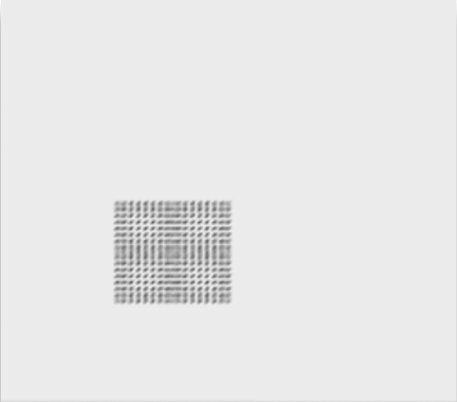
Sparsity Model : we consider a dictionary which has a fast transform/reconstruction operator:

$$X = \Phi\alpha$$

Local DCT

Stationary textures

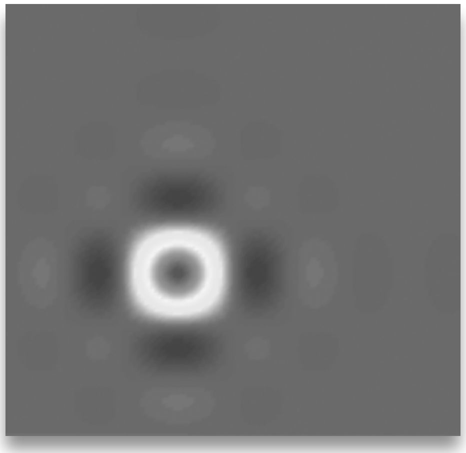
Locally oscillatory



Wavelet transform

Piecewise smooth

Isotropic structures



Curvelet transform

Piecewise smooth, edge

What is a good sparse representation for data?

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

Dictionary (basis, frame)

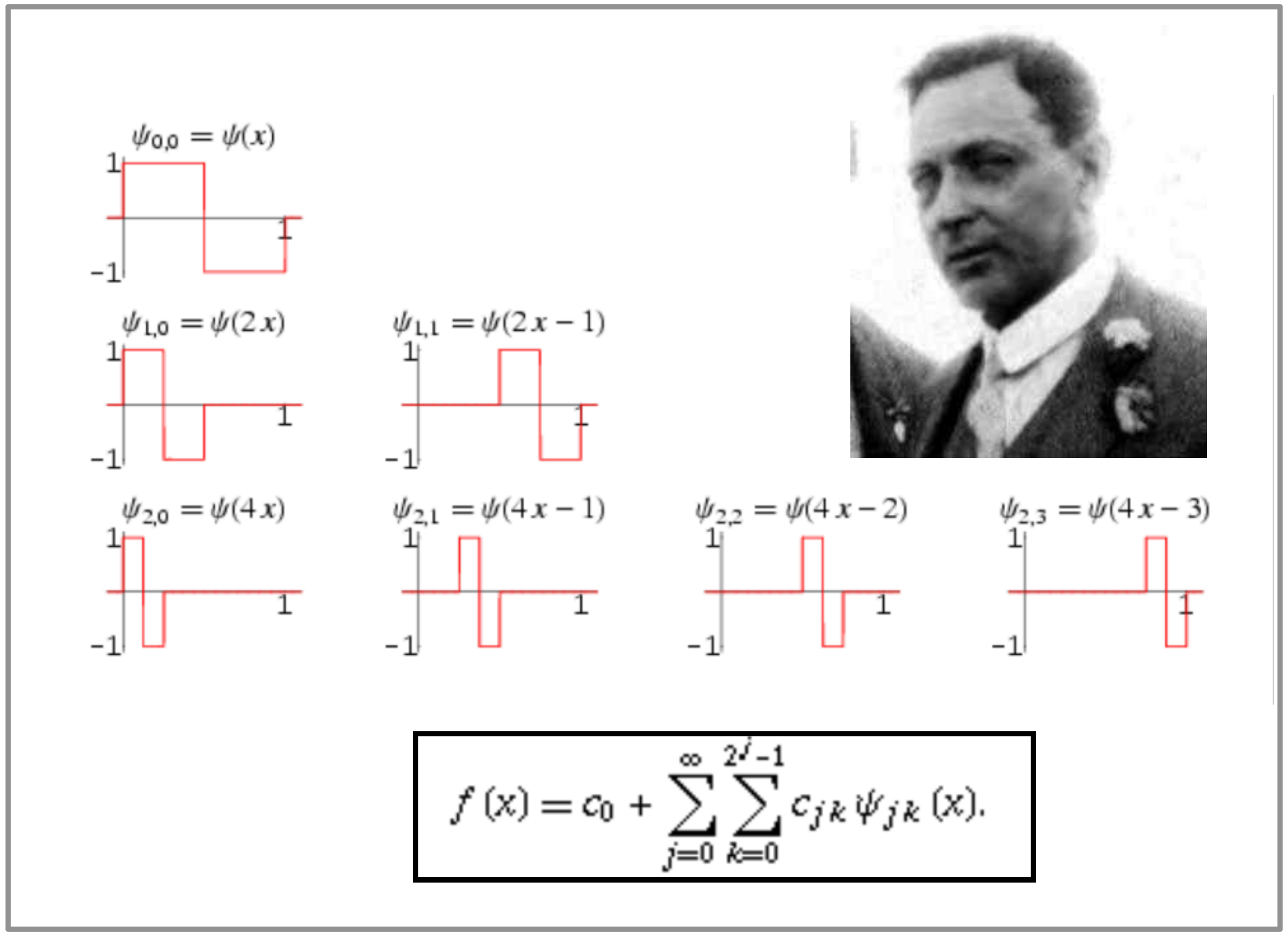
$$\Phi = \{\phi_1, \ldots, \phi_K\}$$

Atoms

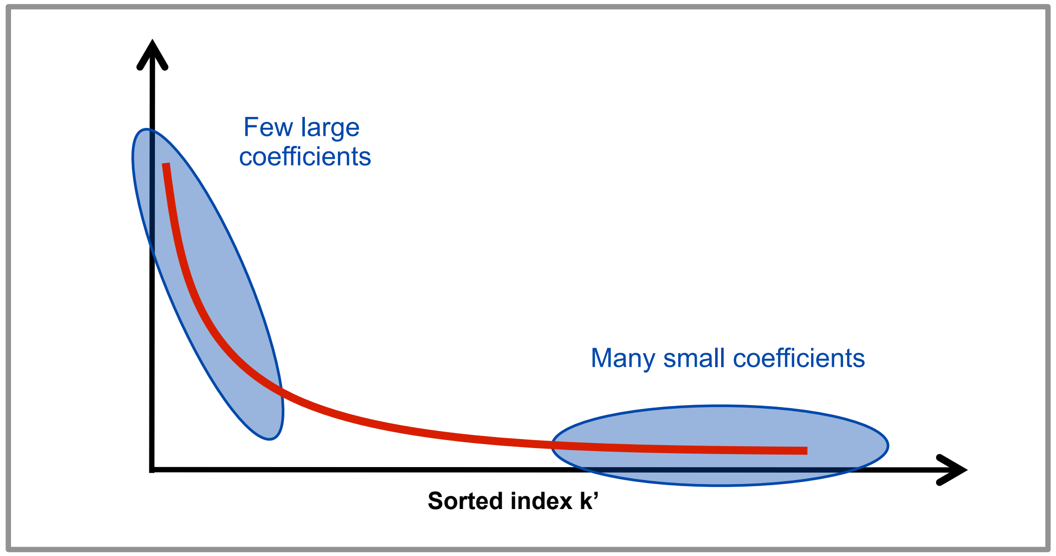
$$s = \sum_{k=1}^{K} \alpha_k \phi_k = \Phi\alpha$$

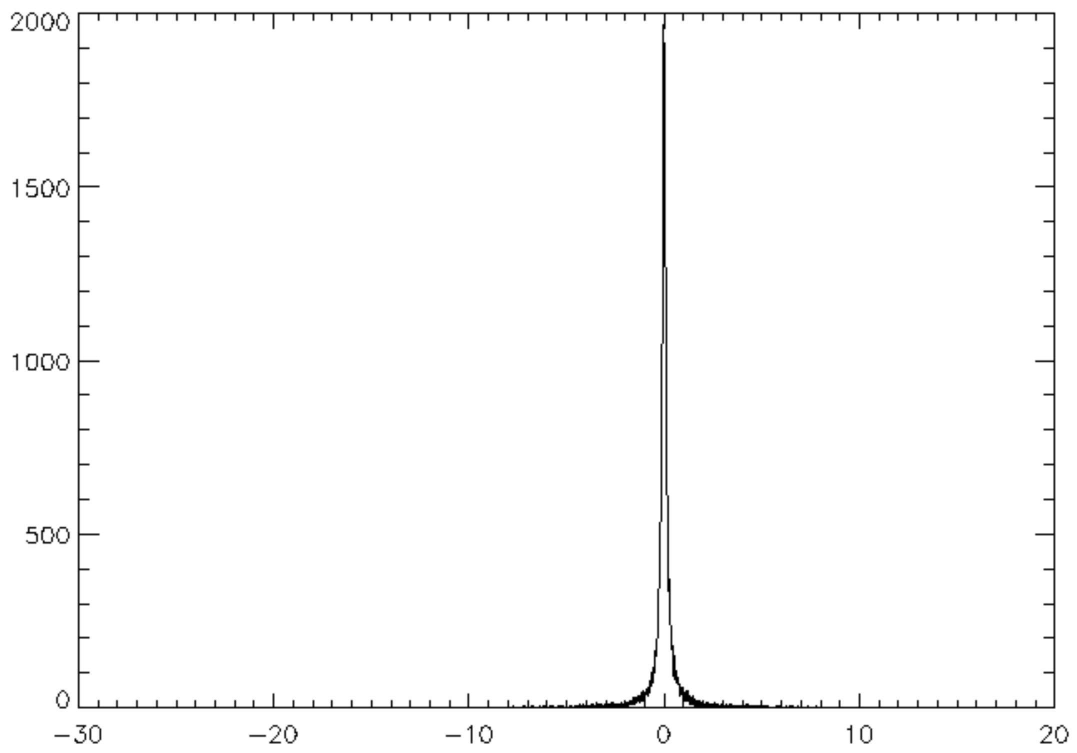
coefficients

Ex: Haar wavelet



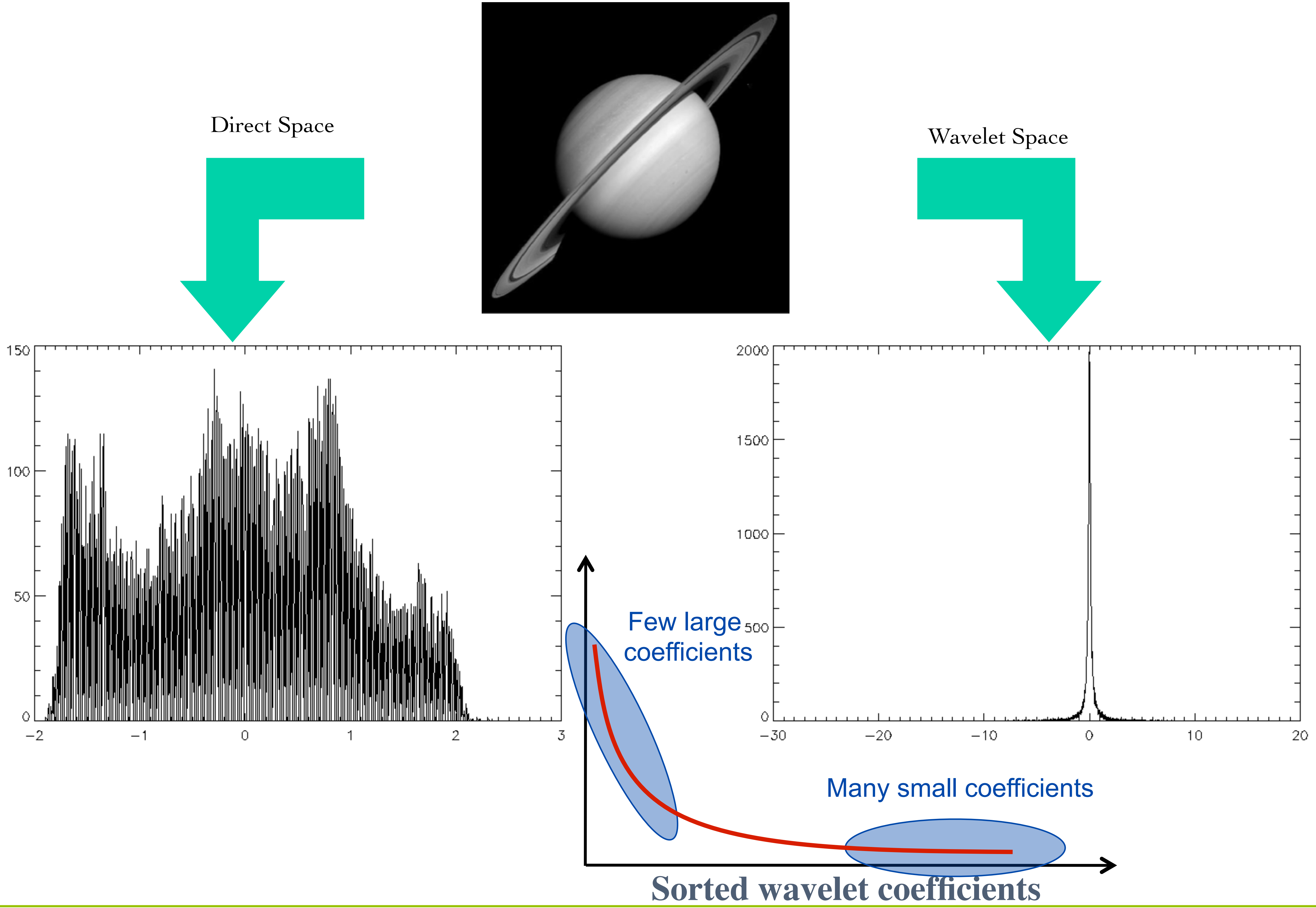
$$f(x) = c_0 + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} c_{jk}\,\psi_{jk}(x).$$





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

Sparse Representation



Direct Space

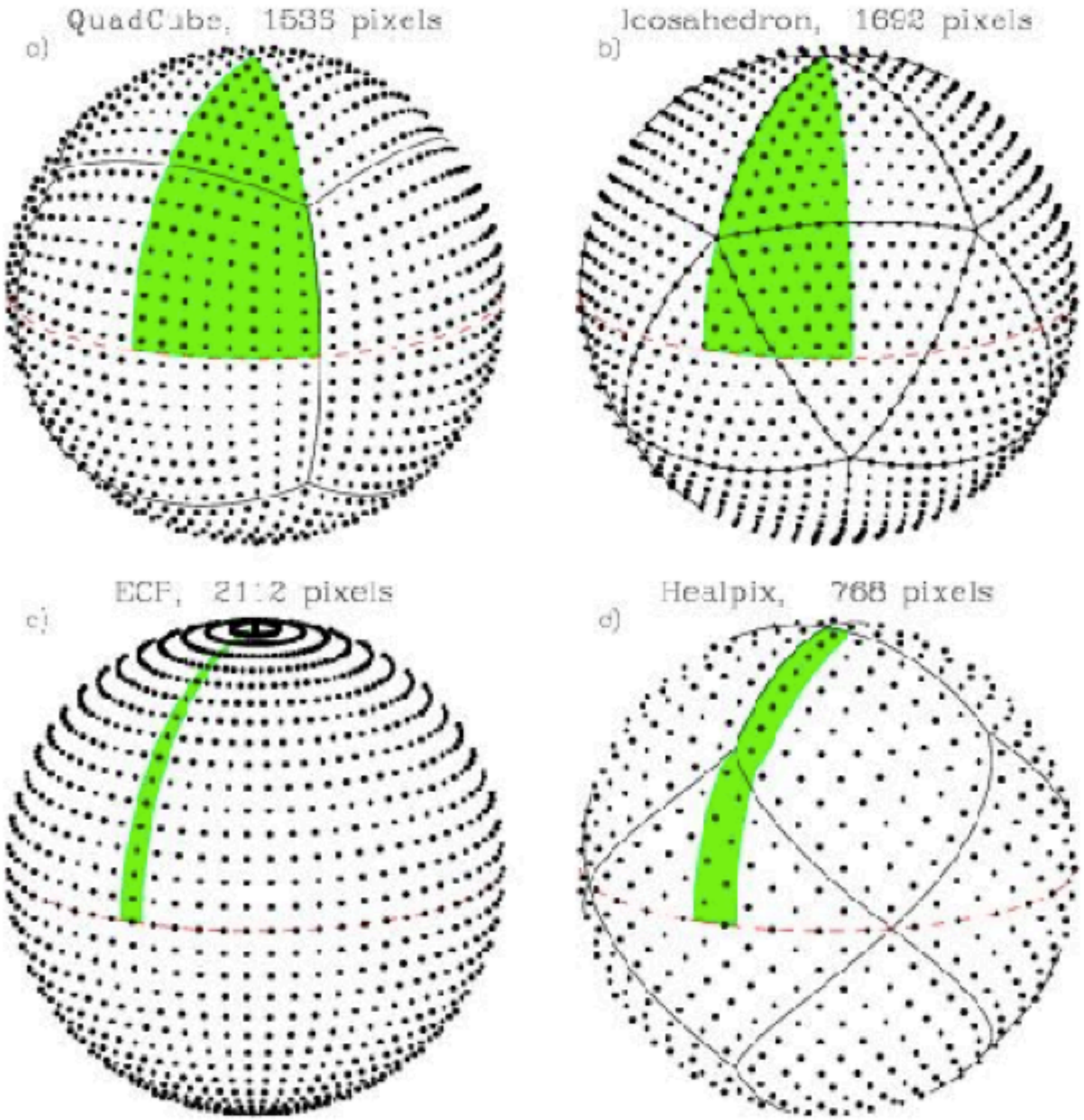
Wavelet Space

Few large coefficients

Many small coefficients

Sorted wavelet coefficients

Data on the Sphere



Healpix

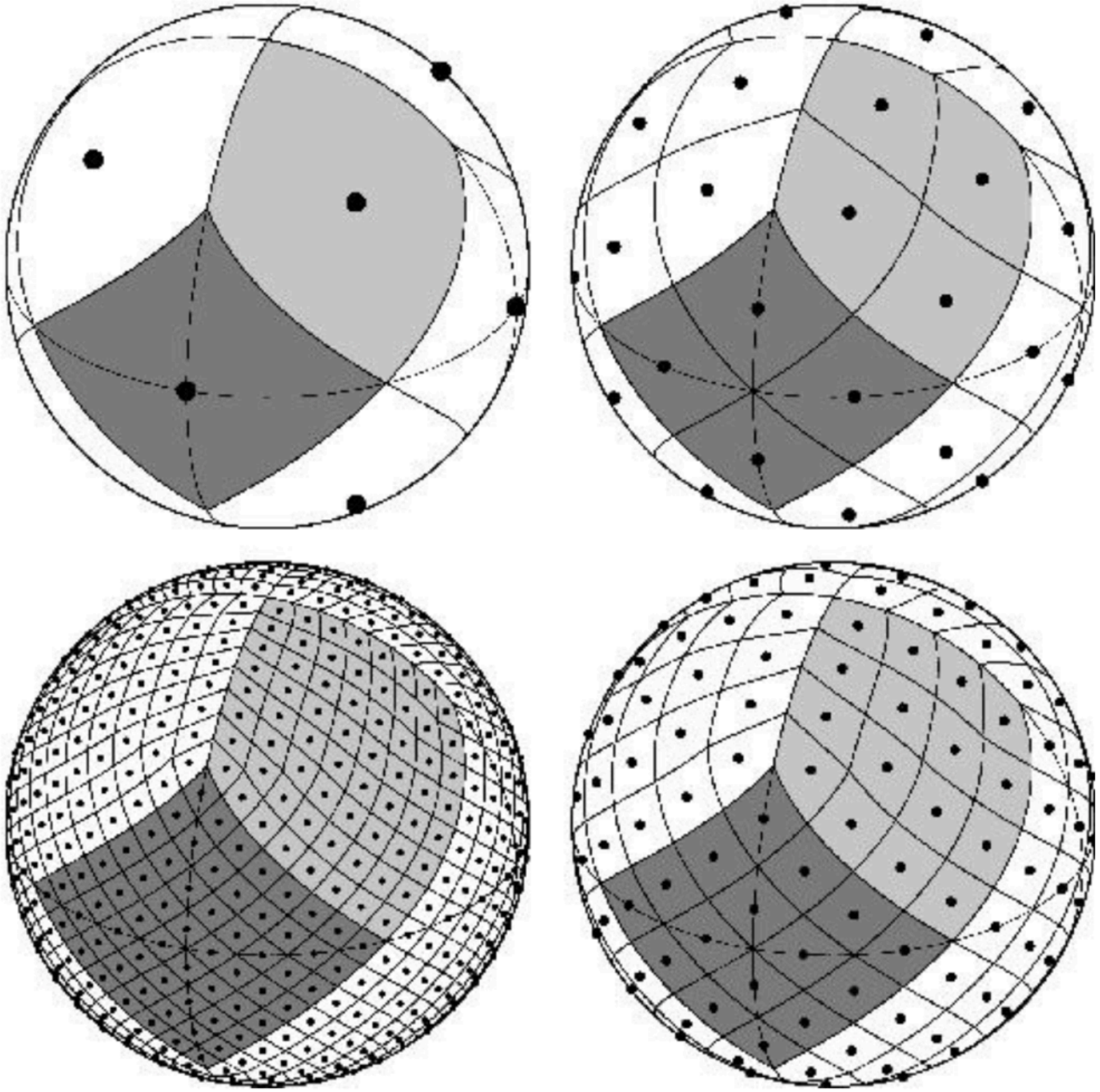
K.M. Gorski et al., 1999, astro-ph/9812350, http://www.eso.org/science/healpix

- Pixel = Rhombus
- Same Surfaces
- For a given latitude : regularly spaced
- Number of pixels:

  12 x $(N_{sides})^2$
- Included in the software:

– Anafast

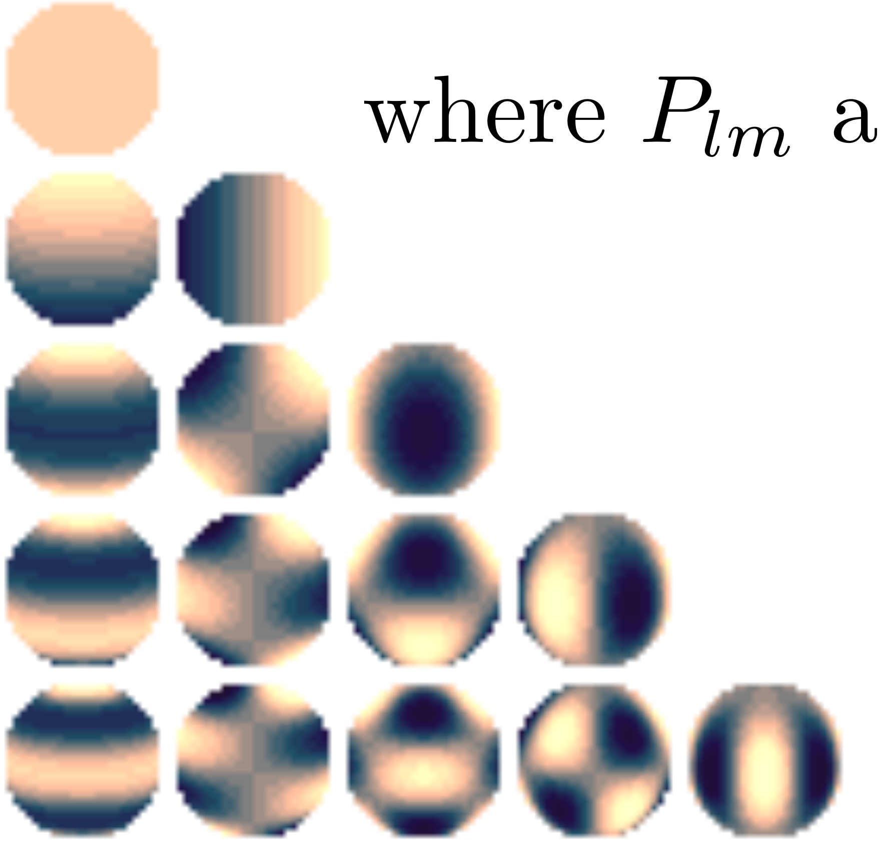
– Synfast



Spherical Harmonics

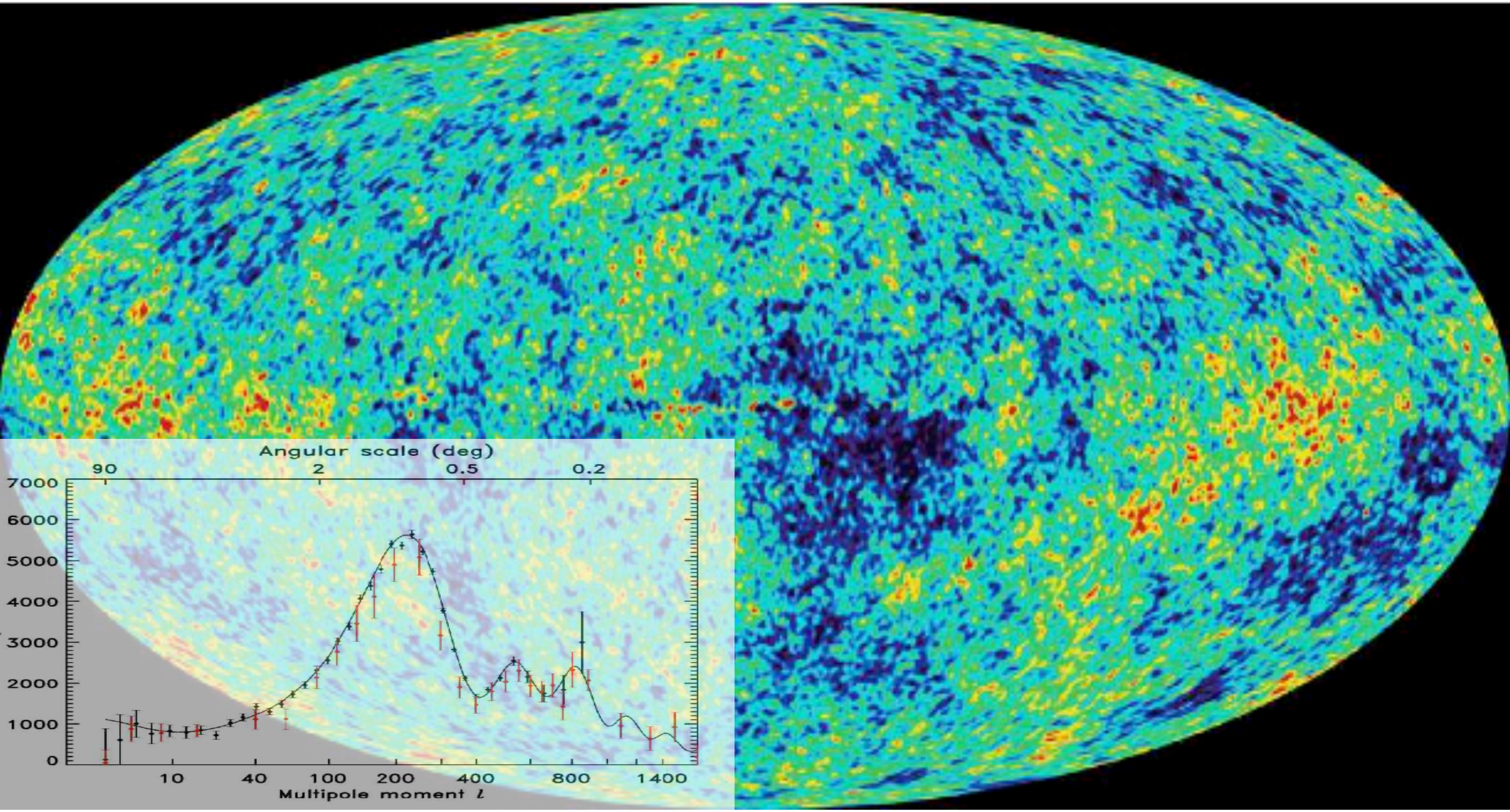
$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{-l}^{l} a_{lm} Y_{lm}(\theta, \phi)$$

$$Y_{lm}(\theta, \phi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_{lm}(\cos \phi) e^{im\theta},$$



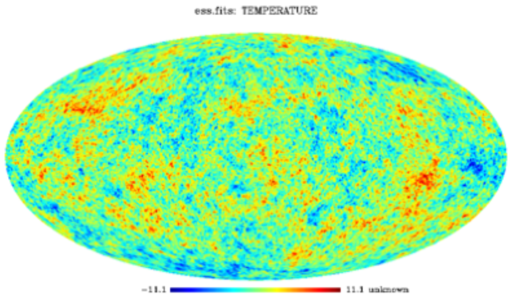
where P_{lm} are the Legendre Polynomials.

CMB & Spherical Harmonics

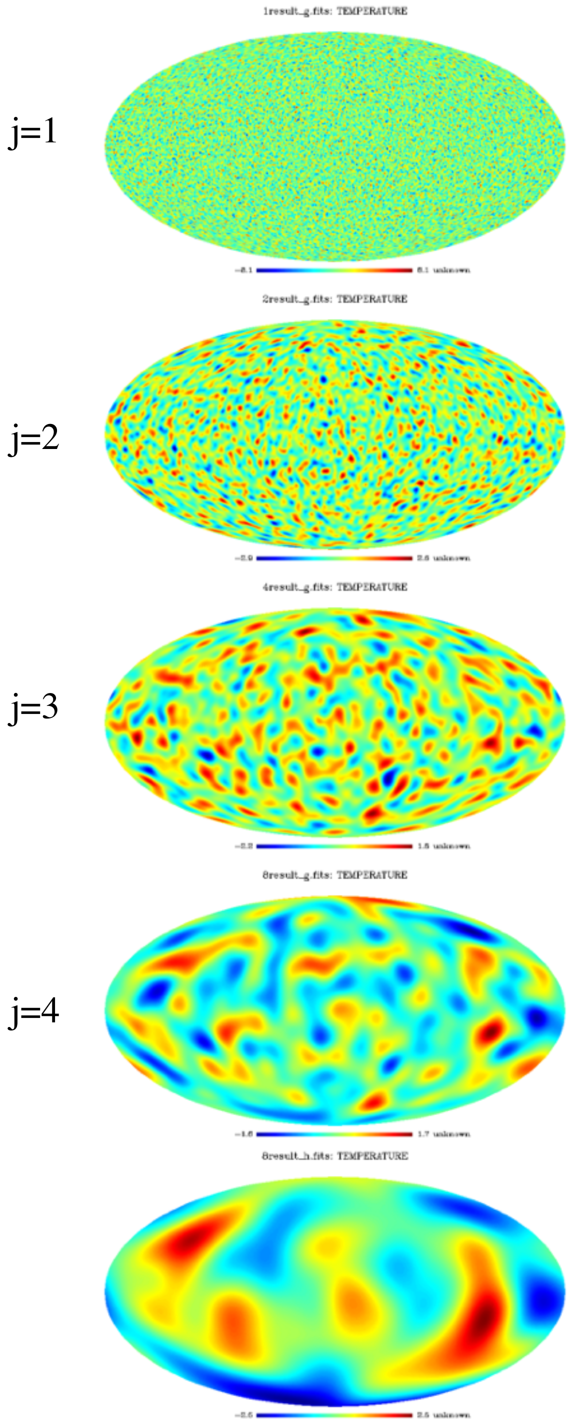


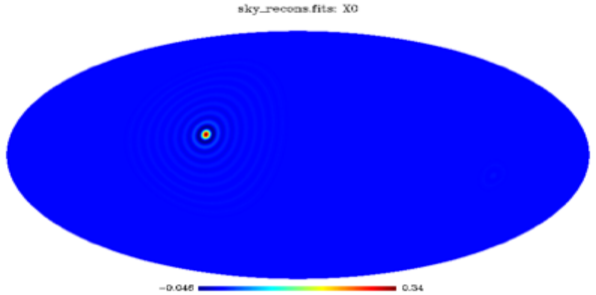
Isotropic Undecimated Wavelet on the Sphere

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



Undecimated
Wavelet Transform





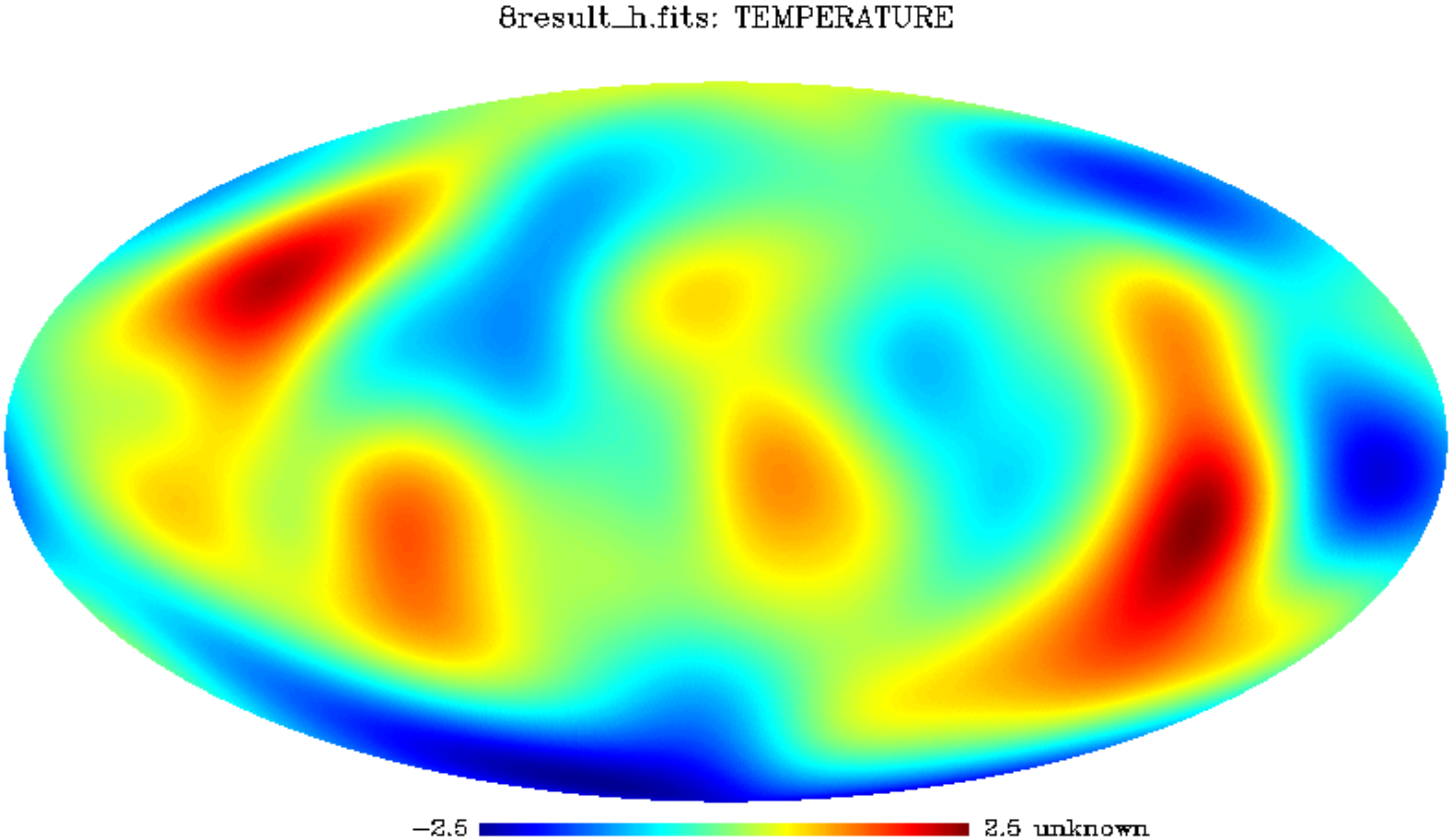
$$\hat{\psi}_{\frac{l_c}{2^j}}(l,m) = \hat{\phi}_{\frac{l_c}{2^{j-1}}}(l,m) - \hat{\phi}_{\frac{l_c}{2^j}}(l,m)$$

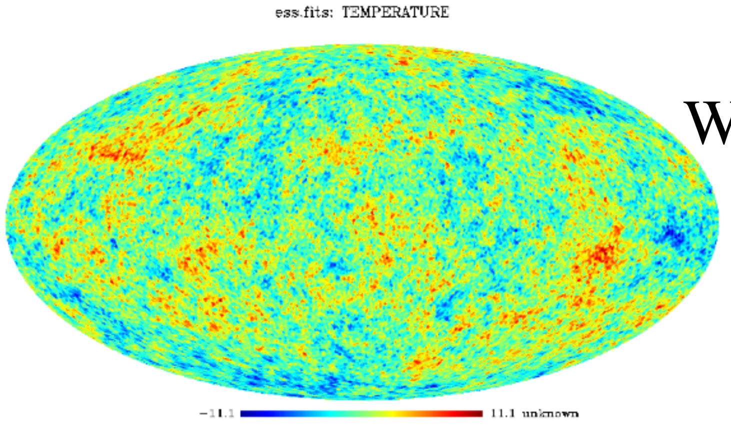
$$\hat{H}_j(l,m) \quad = \quad \begin{cases} \dfrac{\hat{\phi}_{\frac{l_c}{2^{j+1}}}(l,m)}{\hat{\phi}_{\frac{l_c}{2^j}}(l,m)} & \text{if } l < \frac{l_c}{2^{j+1}} \quad \text{and} \quad m = 0 \\ 0 & \text{otherwise} \end{cases}$$

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

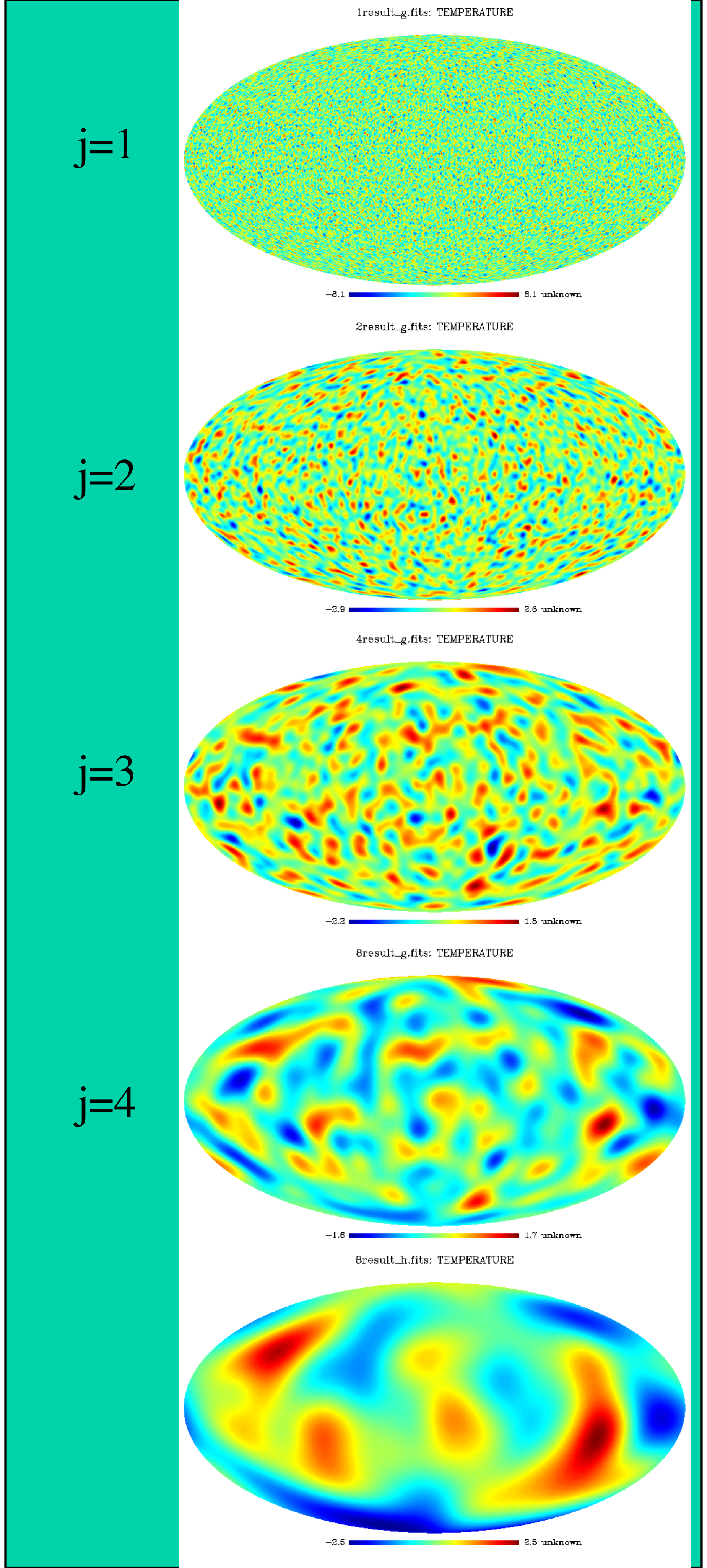
$$\hat{c}_{j+1}(l,m) = \widehat{H}_j(l,m)\hat{c}_j(l,m)$$
$$\hat{w}_{j+1}(l,m) = \widehat{G}_j(l,m)\hat{c}_j(l,m)$$

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

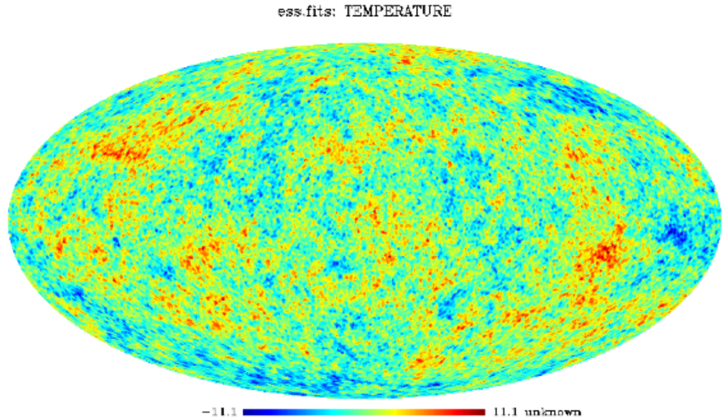

8result_h.fits: TEMPERATURE
−2.5
2.5 unknown

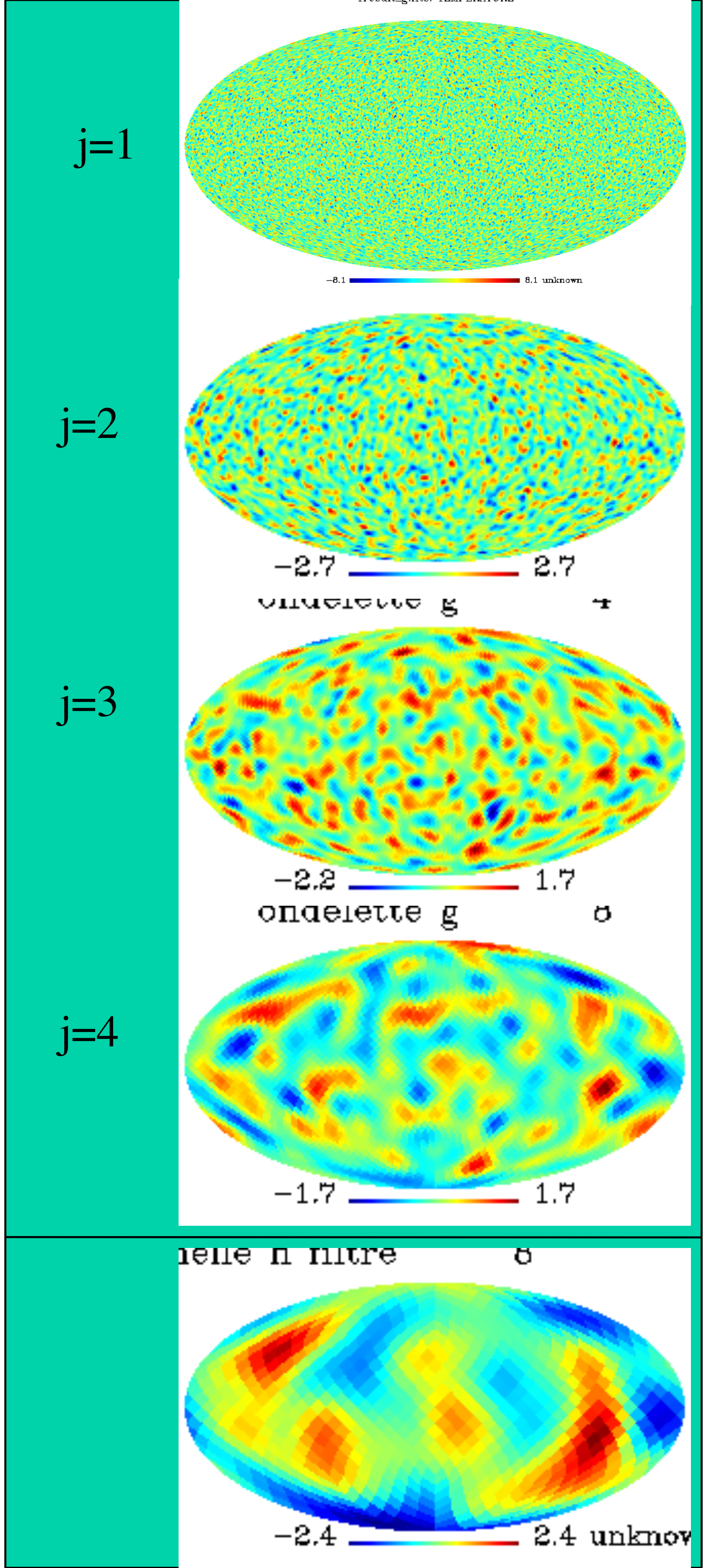


Undecimated
Wavelet Transform

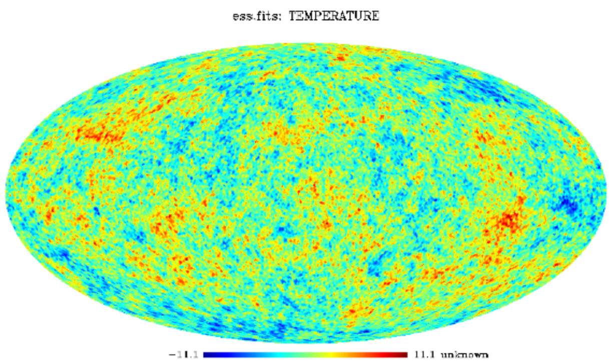


Pyramidal Wavelet Transform
On the Sphere



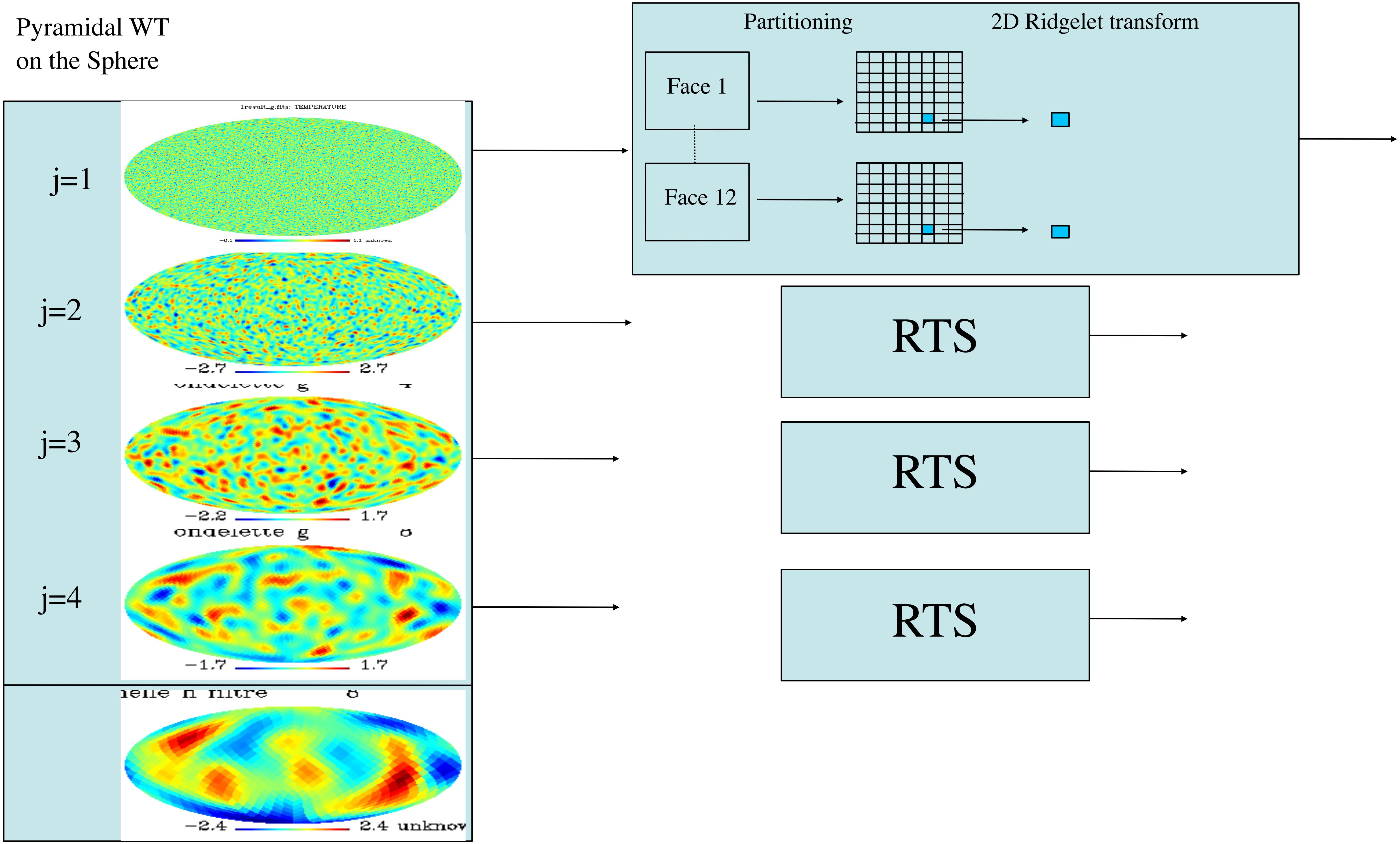


Curvelets on the Sphere



Ridgelet Transform on the Sphere (RTS)

Pyramidal WT
on the Sphere



j=1

j=2

j=3

j=4

Partitioning

2D Ridgelet transform

Face 1

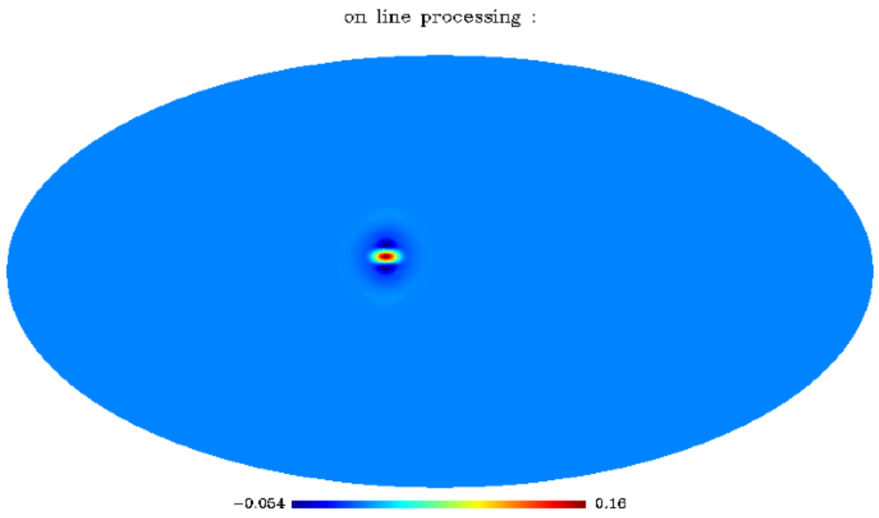
Face 12

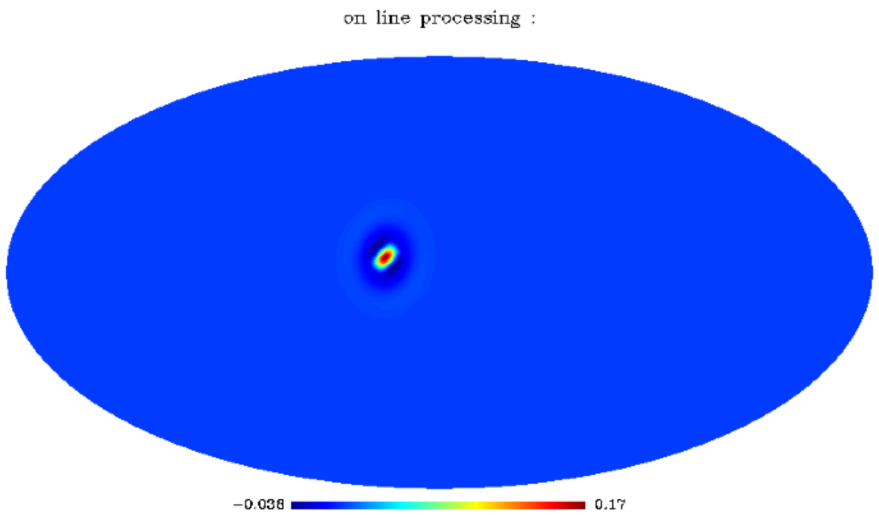
RTS

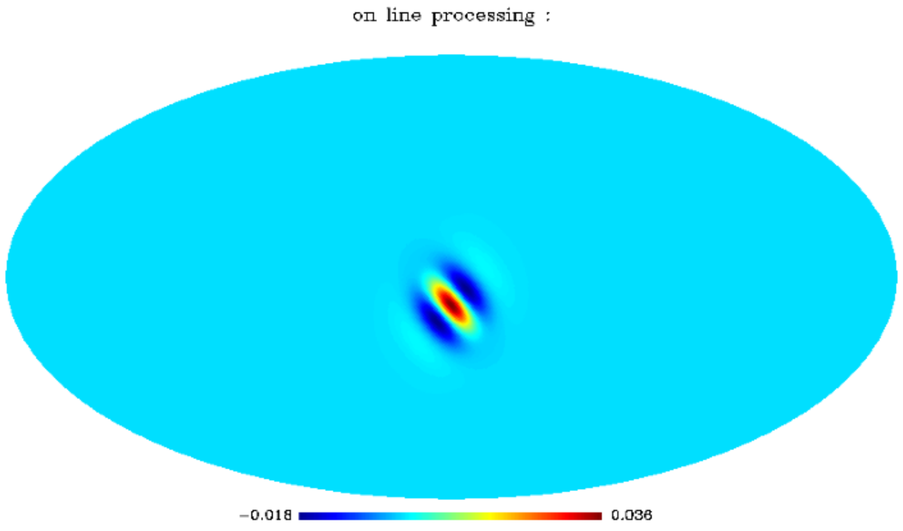
RTS

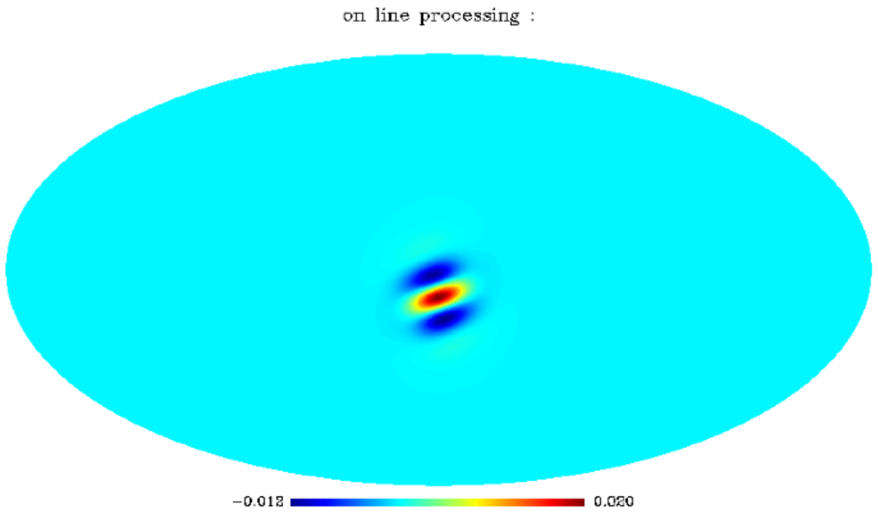
RTS

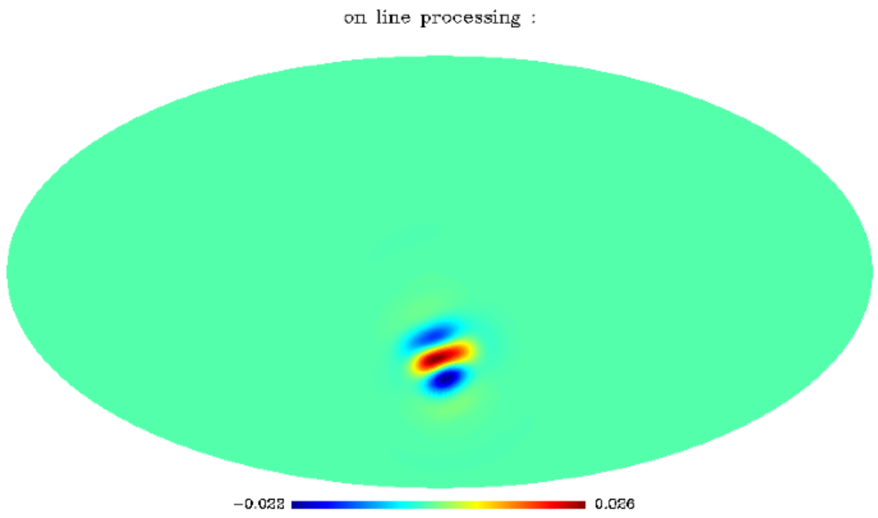
Curvelet functions on the sphere

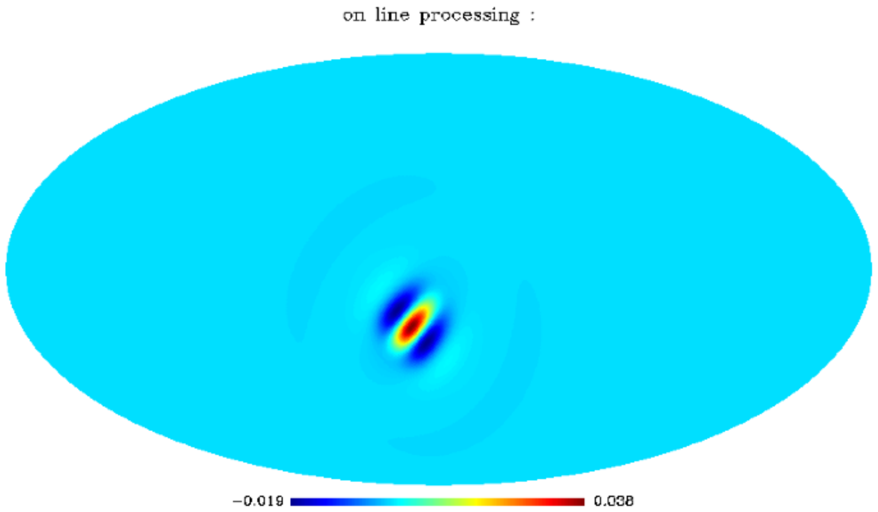






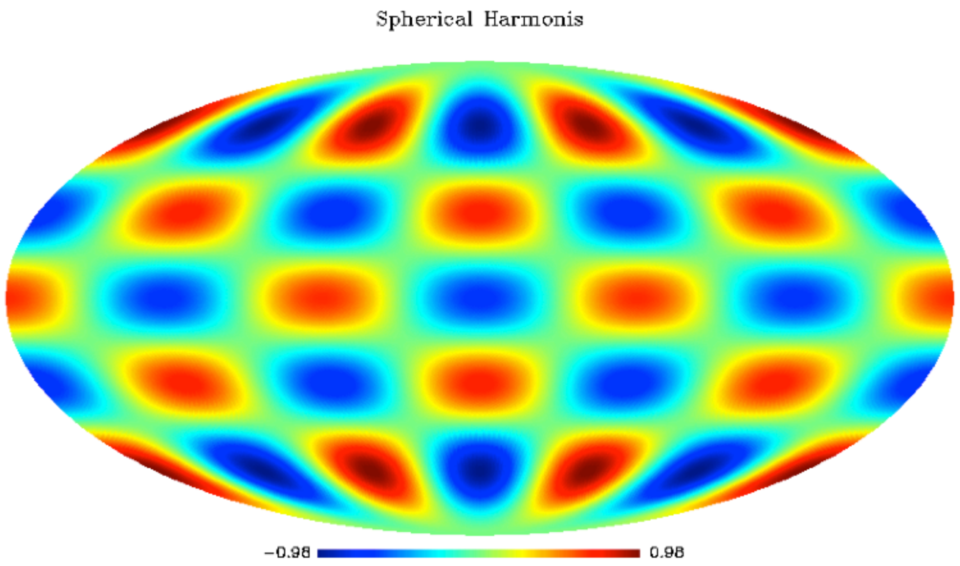




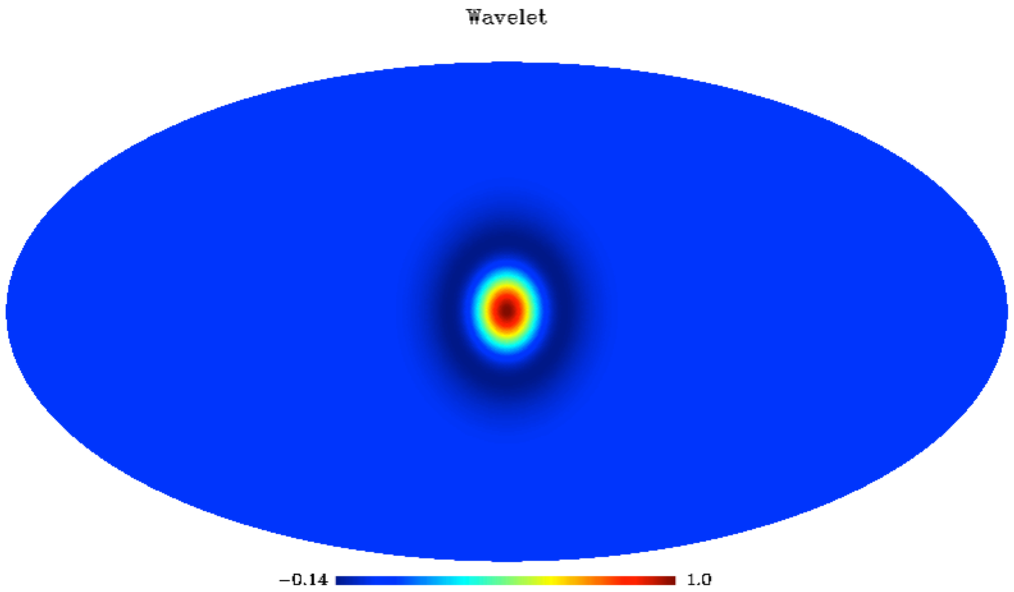


Dictionaries

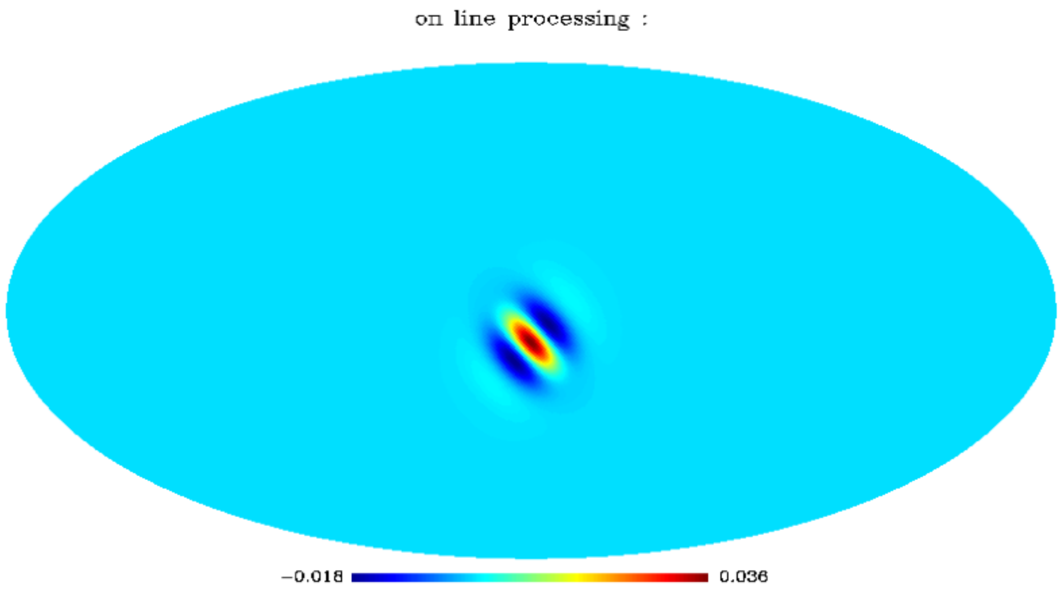
Spherical Harmonics



Wavelets



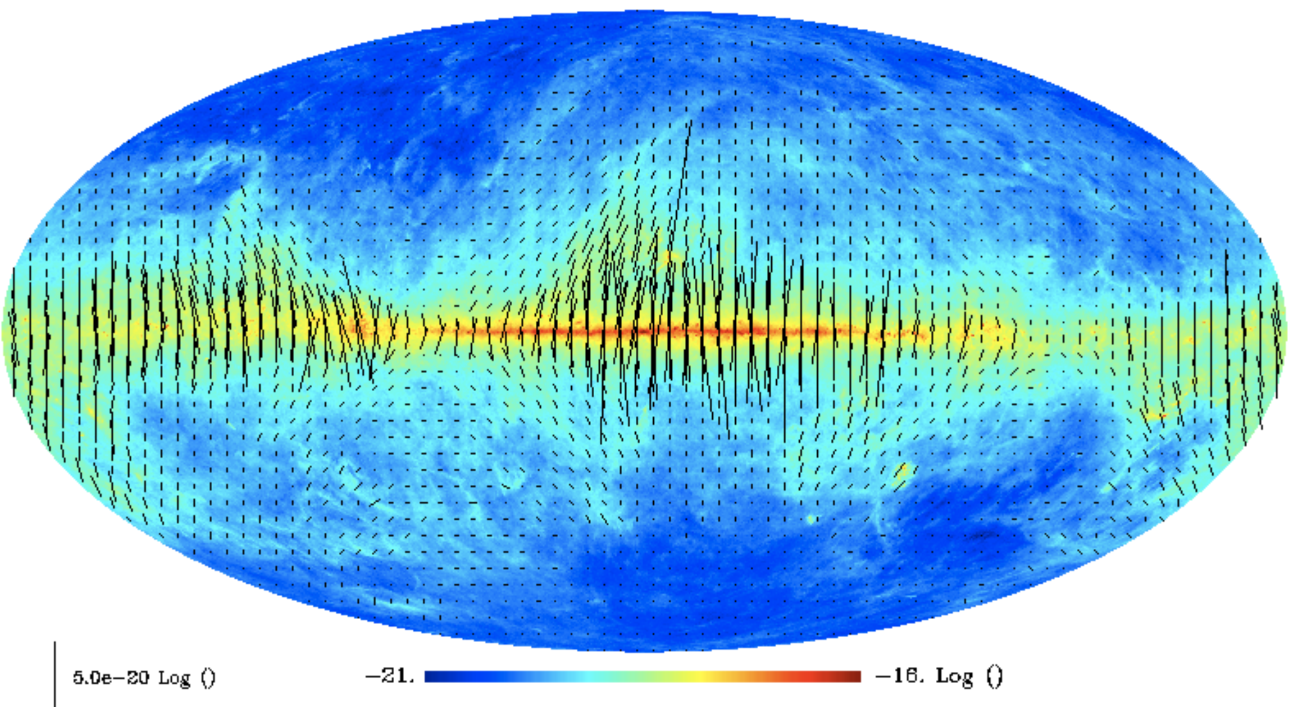
Curvelets

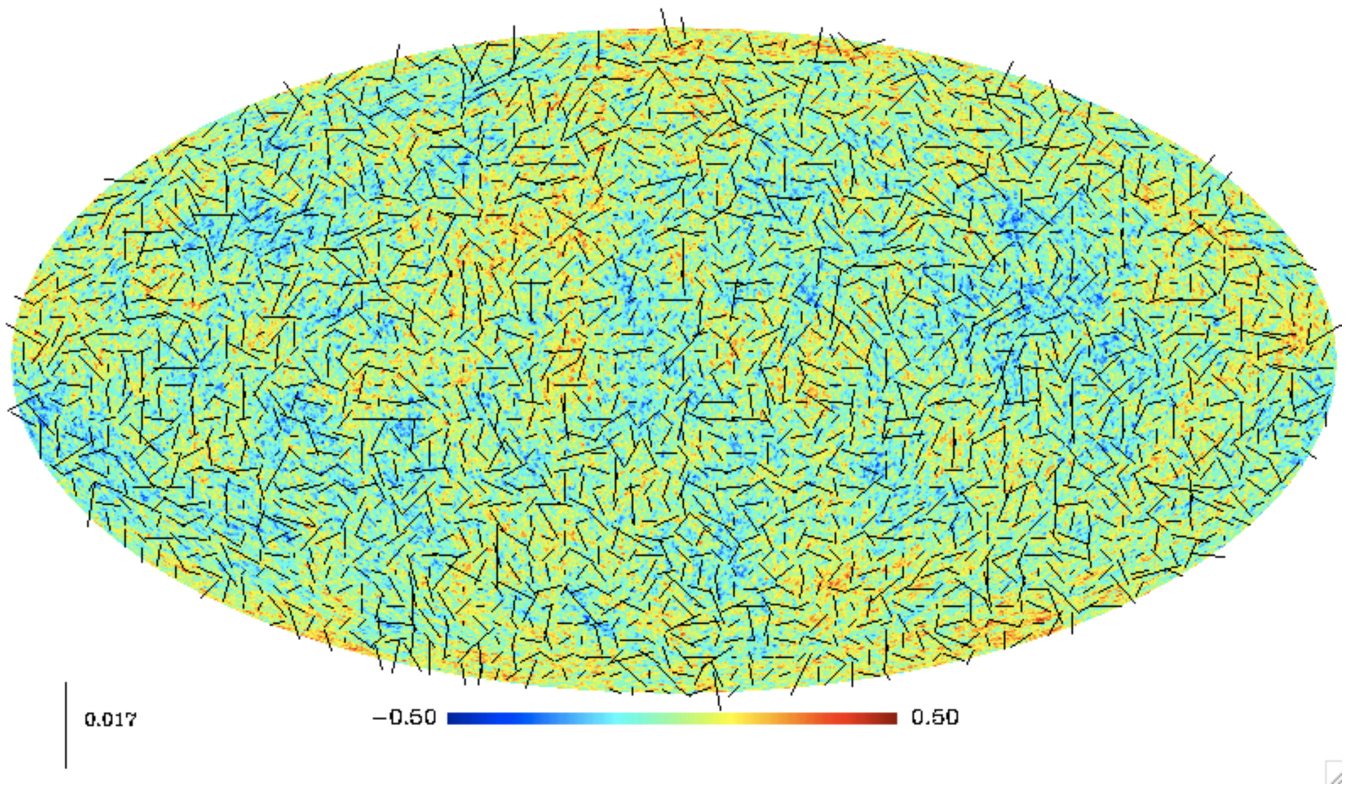


PLANCK POLARIZED DATA: T, Q, U

Magnitude $P = \sqrt{Q^2 + V^2}$

Orientation $\alpha = \arctan\left(U/Q\right)$


Dust 857 Ghz


CMB

E/B Undecimated Wavelet Transform for Polarized Data

J.-L. Starck, Y. Moudden and J. Bobin, "Polarized Wavelets and Curvelets on the Sphere", Astronomy and Astrophysics, 497, 3, pp 931--943, 2009.

$$
\begin{aligned}
E = & \quad \textstyle\sum_{\ell,m} a^E_{\ell m} Y_{\ell m} = \sum_{\ell,m} -\frac{{}_2a_{\ell m} + {}_{-2}a_{\ell m}}{2} Y_{\ell m} \\
B = & \quad \textstyle\sum_{\ell,m} a^B_{\ell m} Y_{\ell m} = \sum_{\ell,m} i \frac{{}_2a_{\ell m} - {}_{-2}a_{\ell m}}{2} Y_{\ell m}
\end{aligned}
$$

Wavelet Transform of E and B are obtained by:

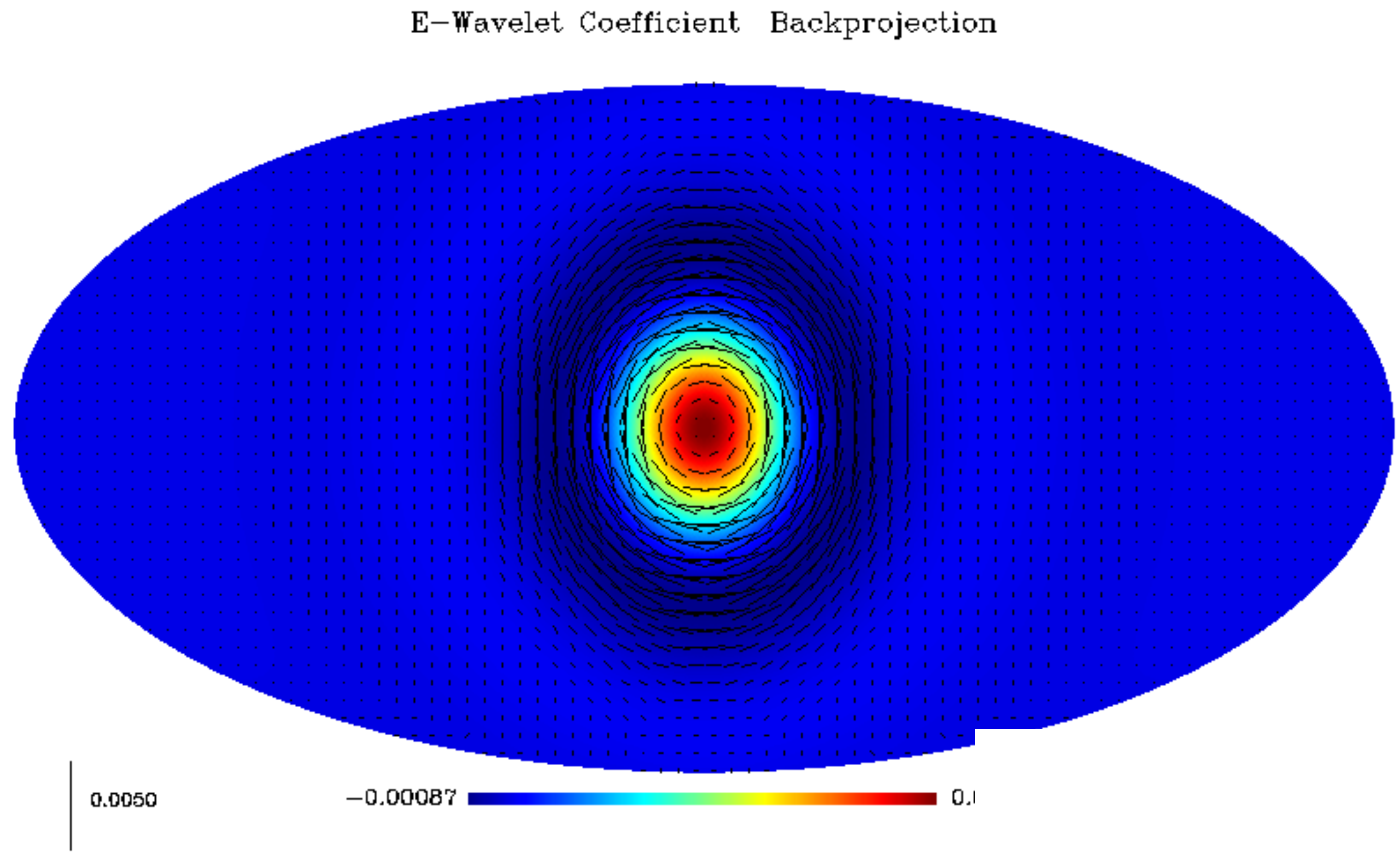
$$w_j^E =< E, \psi_j > \qquad\qquad w_j^B =< B, \psi_j >$$

Furthermore, if we use the spherical isotropic wavelet construction of (starck et al, 2006), we have

$$E(\theta, \phi) = c_J^E(\theta, \phi) + \sum_{j=1}^{J} w_j^E(\theta, \phi) \qquad\qquad B(\theta, \phi) = c_J^B(\theta, \phi) + \sum_{j=1}^{J} w_j^B(\theta, \phi)$$

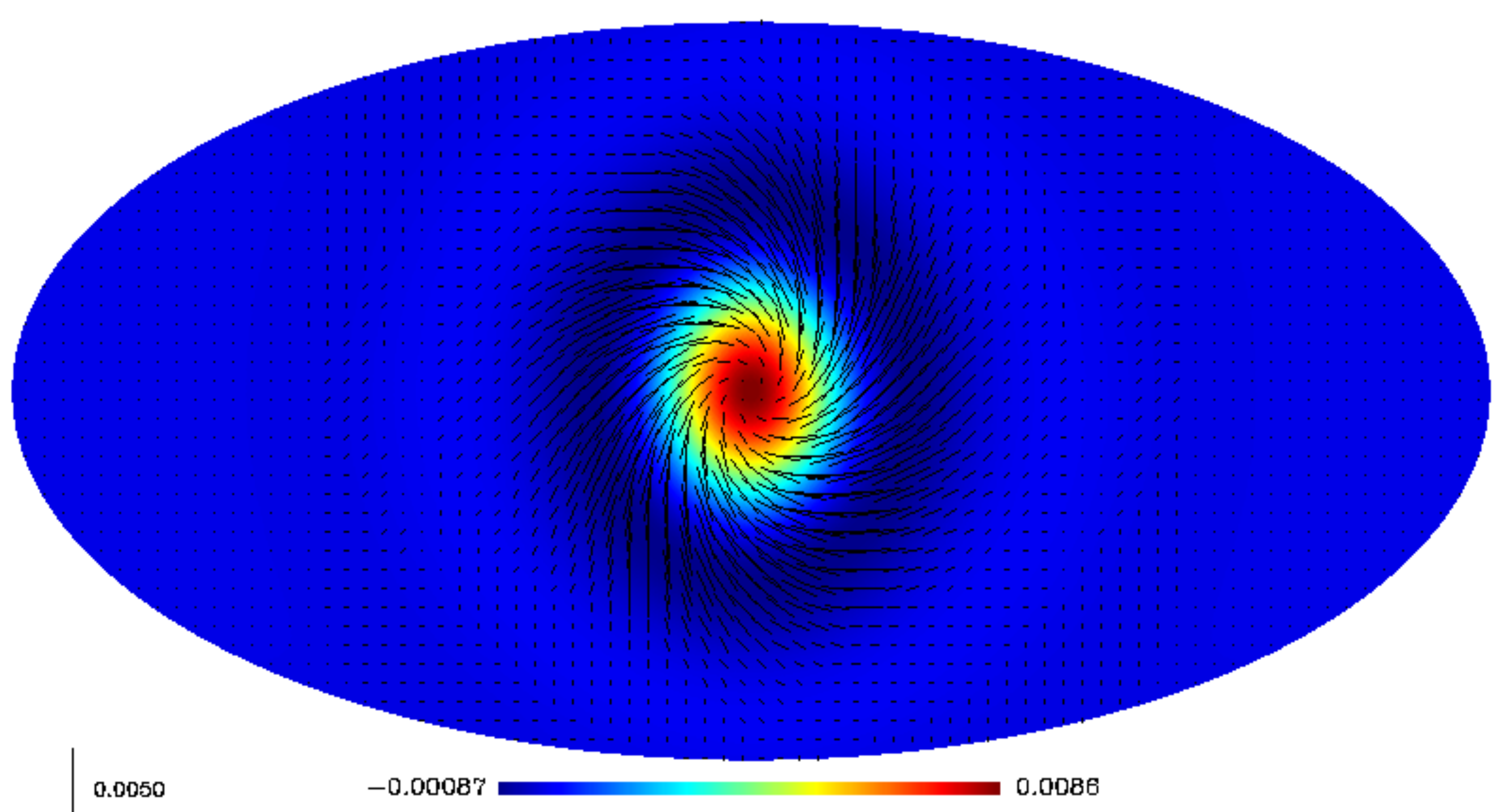
Polarized Dictionary: E/B Polarized Wavelet

J.-L. Starck, Y. Moudden and J. Bobin, "Polarized Wavelets and Curvelets on the Sphere", Astronomy and Astrophysics, 497, 3, pp 931--943, 2009.

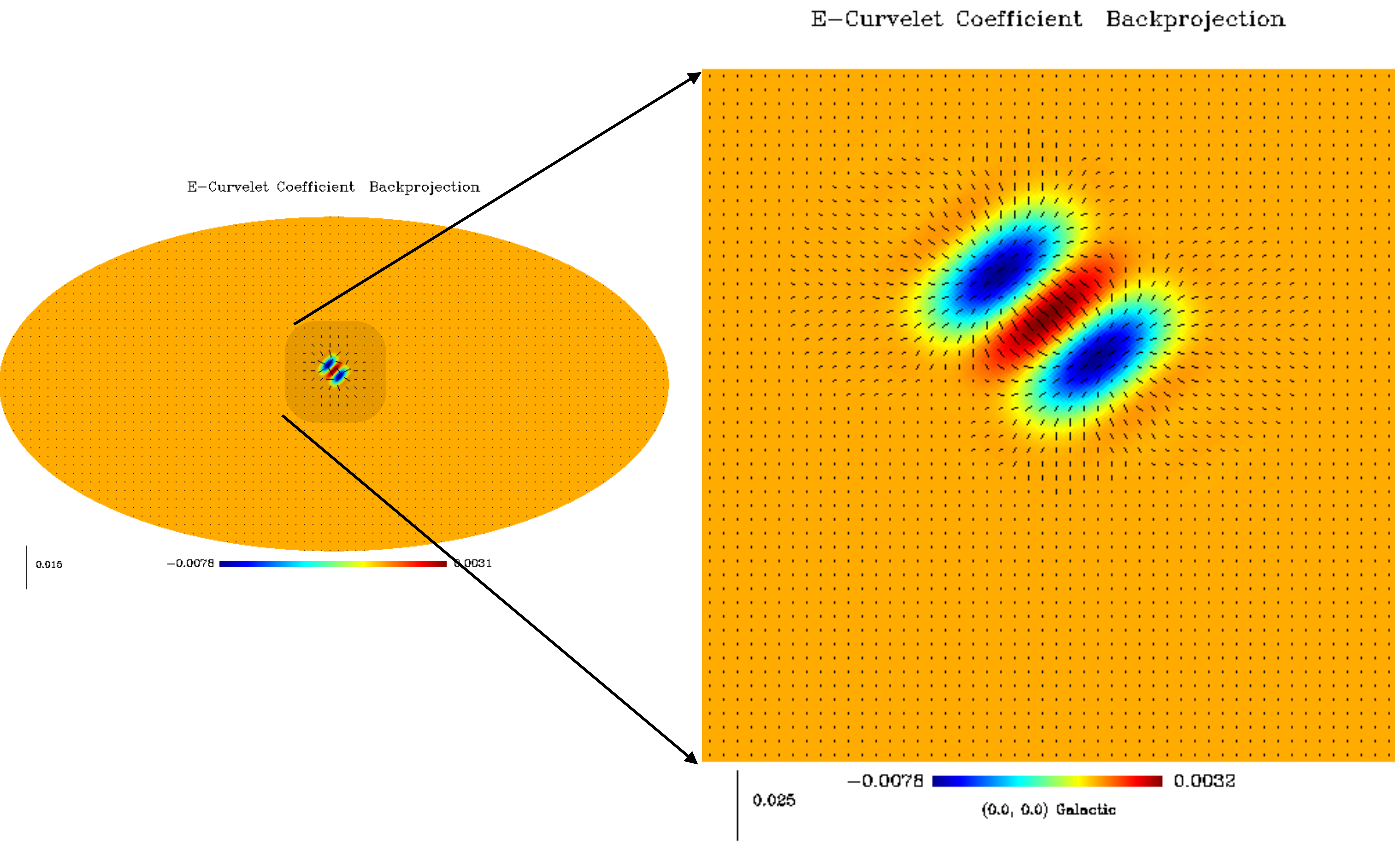

E−Wavelet Coefficient Backprojection

j=4

B−Wavelet Coefficient Backprojection



Curvelet and E/B Mode Decomposition



Inverse Problems in Astrophysics

- ➡Part 1: Introduction to Inverse Problems
- ➡Part 2: From Fourier to Wavelets
- ➡Part 3: Wavelet and Beyond
- ➡**Part 4: Sparse Regularization**
- ➡Part 5: Application to Unmixing and Inpainting
- ➡Part 6: Compressed Sensing
- ➡Part 7: Deep Learning

Example: DENOISING

$$\arg\min_{X} \parallel Y - HX \parallel^2 + \lambda \parallel LX \parallel^2$$

Denoising using a sparsity model (H=Id)

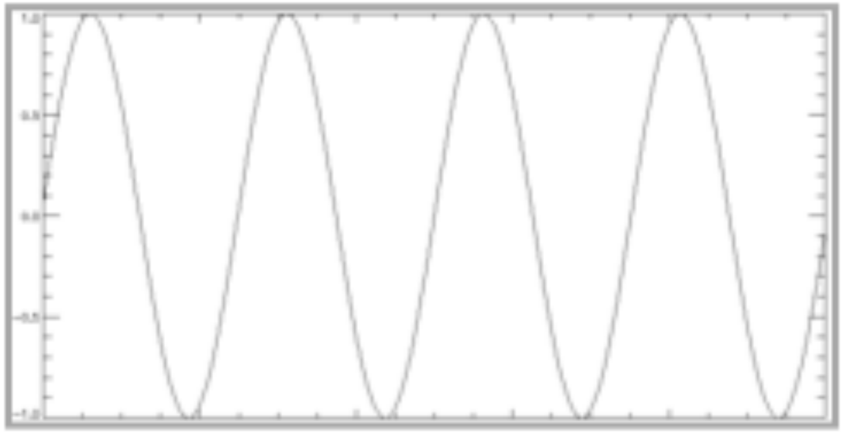
$$H = Id$$

$$Y = X + N$$

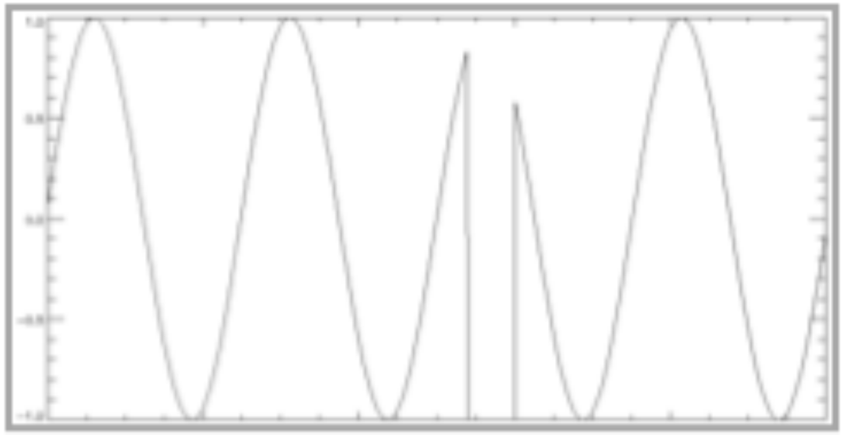
$$X = \Phi\alpha$$

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

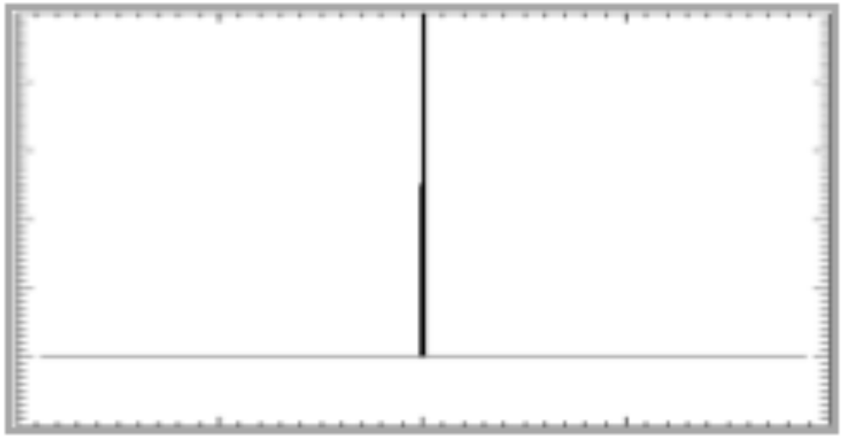
Minimizing the l_0 norm



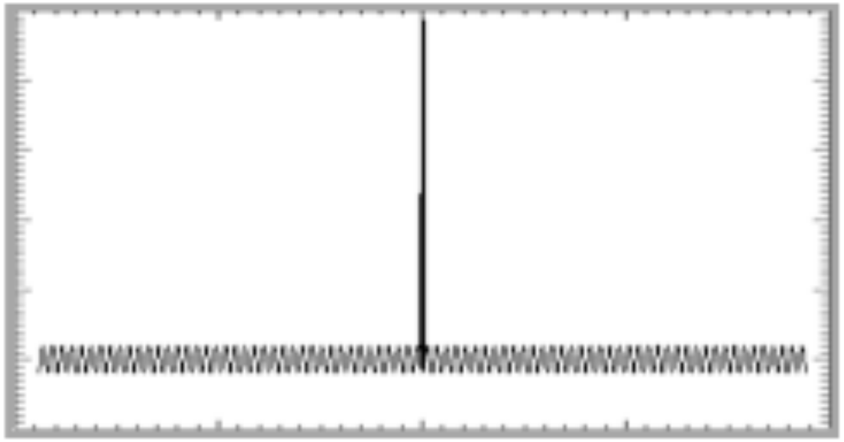
Sine curve



Truncated sine curve



TF of a sine curve



TF of a truncated sine curve

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

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 $s = \phi\alpha$

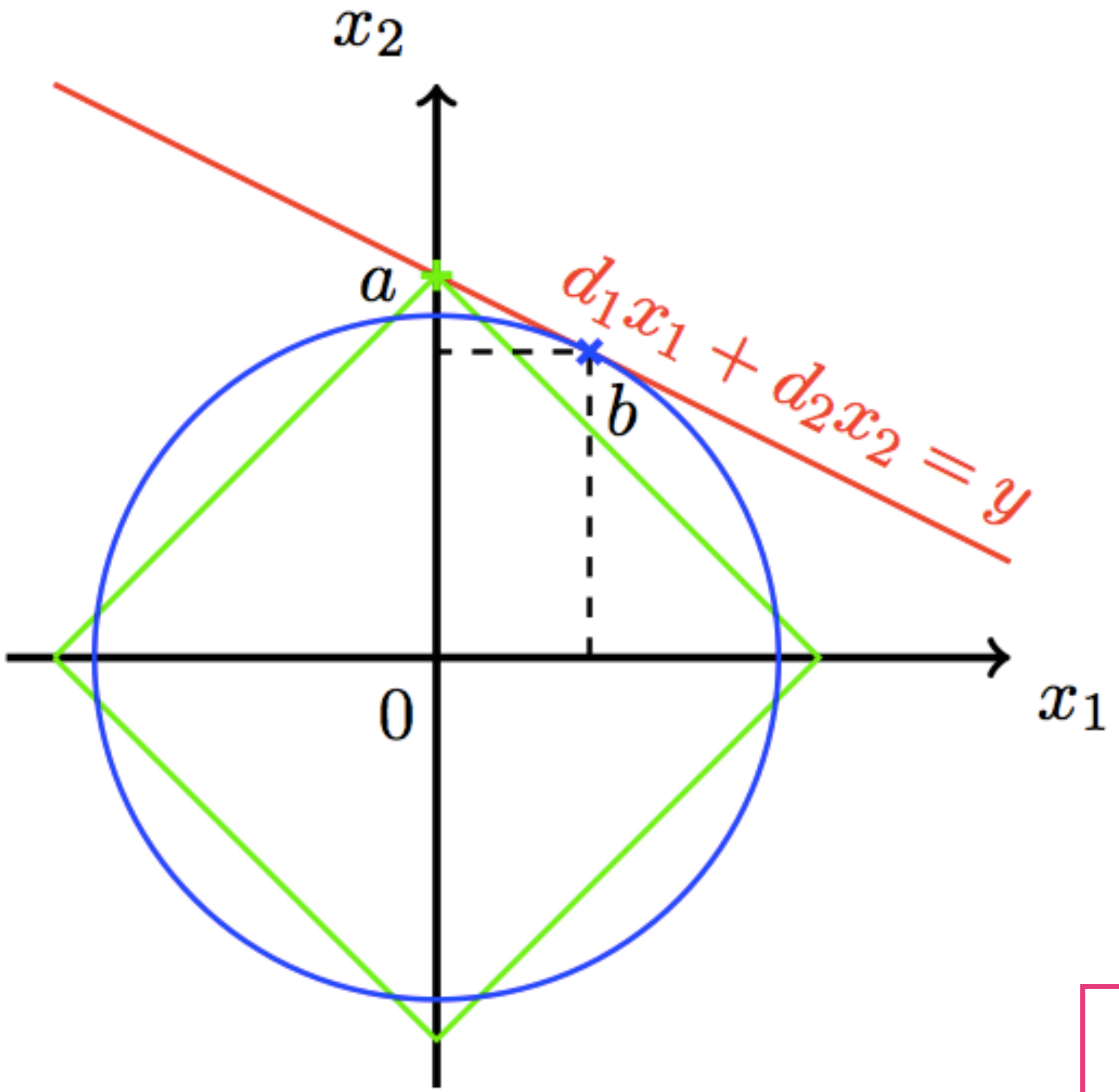
It can be seen as a kind of convexification of (P0).

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.

L1 Norm & Sparsity

$$\|X\|_p = (\sum_i | X_i |^p)^{\frac{1}{p}}$$



$$\boxed{p < 2}$$

Minimization: P=0

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

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

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

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

$$\boxed{\lambda = \frac{t^2}{2}}$$

1st iteration solution:

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

Exact for Φ orthonormal.

Minimization: P=1

$$\tilde{\alpha} \in \arg\min_{\alpha} \frac{1}{2} \parallel Y - \Phi\alpha \parallel^2 + \lambda \parallel \alpha \parallel_1$$

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

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

$$\mathrm{SoftThresh}_t(x) = \mathrm{sgn}(x)\max(0, |x| - t)$$

1st iteration solution:

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

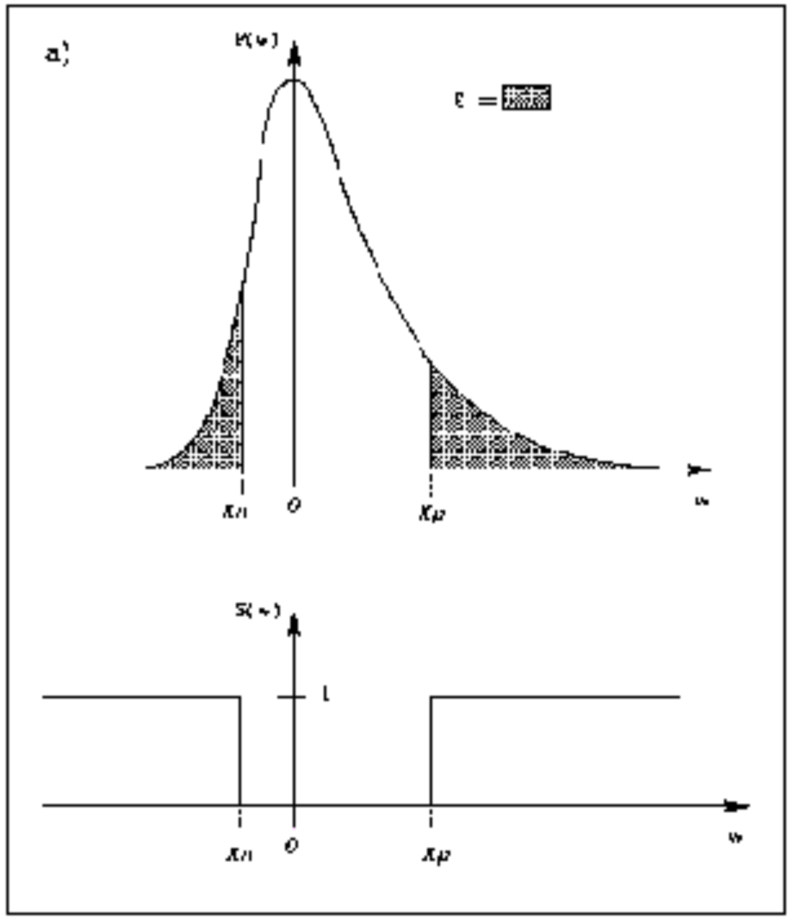
Exact for Φ orthonormal.

Regularization Parameter & Noise Modeling

DETECTION of Active Coefficients:

For a positive coefficient: $P = Prob(x > \alpha_i)$

For a negative coefficient: $P = Prob(x < \alpha_i)$



Hypothesis testing: given a threshold **e**:

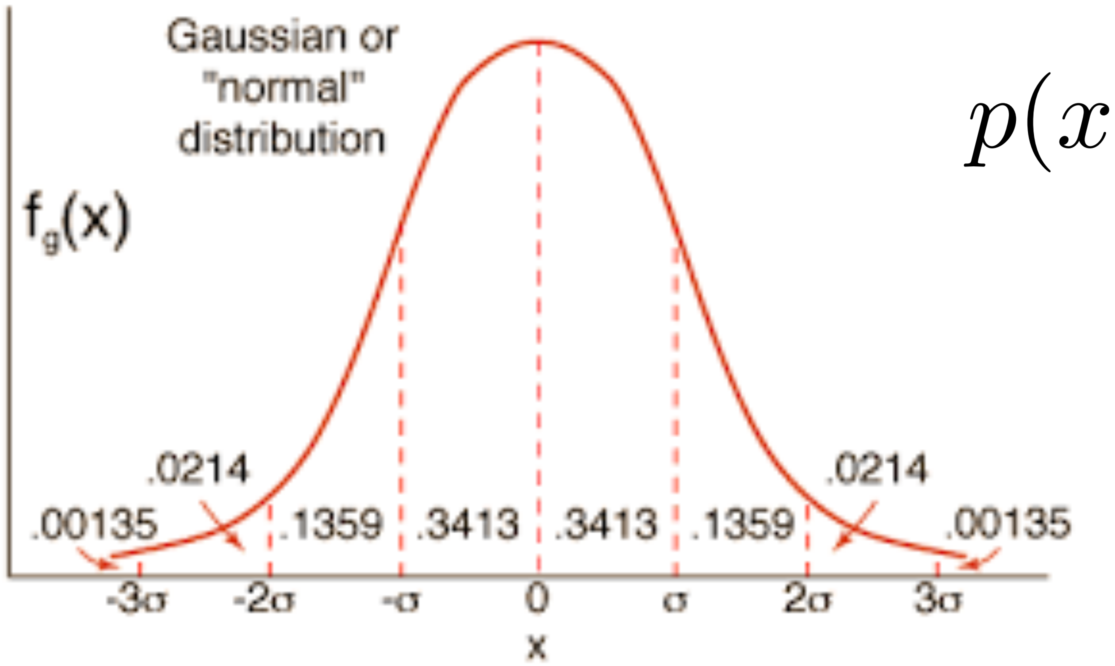
H0: if P > **e**, the coefficient could be due to the noise.

H1: if P < **e**, the coefficient cannot be due to the noise,

and a **significant coefficient** is detected.

Gaussian Noise

Since we apply a linear operator on the data, the noise on each coefficient is also Gaussian:



$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2}{2\sigma^2}}$$

It suffixes to compare the coefficients α to $\boxed{t = k\sigma}$.

if $\mid \alpha_i \mid \geq k\sigma$ then α_i is **significant** or **active**.

if $\mid \alpha_i \mid < k\sigma$ then α_i is **NOT significant**.

Regularization = DETECTION STRATEGY $\lambda = t^2 = (k\sigma)^2$

More **sophisticated tests** exist: False Discovery Rate (FDR), block sparsity, etc

Non Gaussian Noise

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.

Variance Stabilisation

Poisson Noise

If the noise in the data Y is Poisson, the transform

$$t(Y) = 2\sqrt{Y + \frac{3}{8}}$$

acts as if the data arose from a Gaussian white noise model (Anscombe, 1948), with $\sigma = 1$, under the assumption that the mean value of Y is large.

Poisson Noise + Gaussian Noise

The generalization of the variance stabilizing is (Murtagh, Starck, Bijaoui, 1995):

$$t(Y) = \frac{2}{\alpha}\sqrt{\alpha Y + \frac{3}{8}\alpha^2 + \sigma^2 - \alpha g}$$

where α is the gain of the detector, and g and σ are the mean and the standard deviation of the read-out noise.

Multiscale Variance Stabilization

B. Zhang, M.J. Fadili and J.-L. Starck, "Wavelets, Ridgelets and Curvelets for Poisson Noise Removal" , IEEE Transactions on Image Processing , Vol 17, No 7, pp 1093--1108, 2008

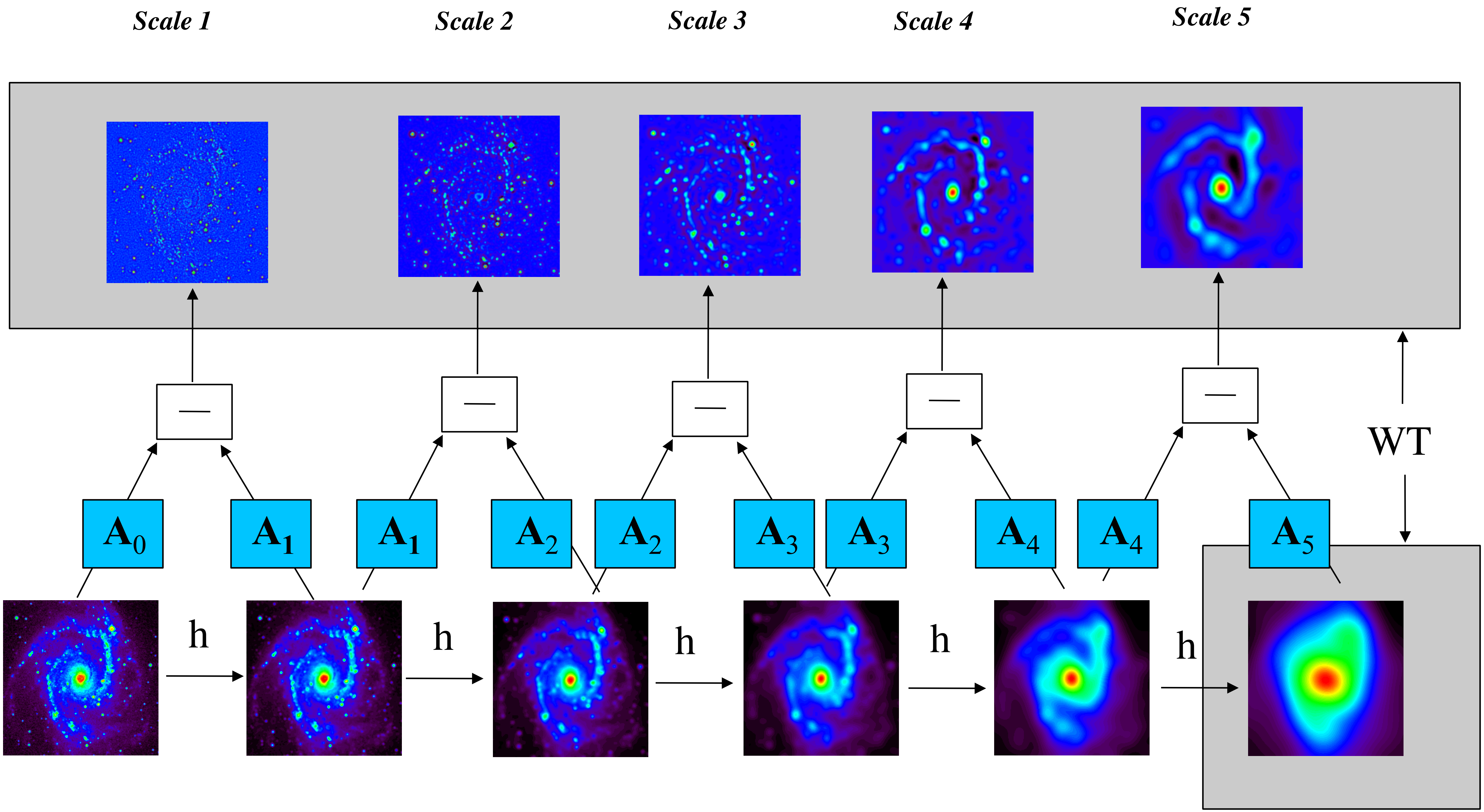
IUWT $$\begin{cases} a_j &= \bar{h}^{\uparrow j-1} \star a_{j-1} \\ w_j &= a_{j-1} - a_j \end{cases}$$

$$\Longrightarrow \begin{array}{c}\text{MSVST}\\+\\\text{IUWT}\end{array} \quad \begin{cases} a_j &= \bar{h}^{\uparrow j-1} \star a_{j-1} \\ w_j &= \mathcal{A}_{j-1}(a_{j-1}) - \mathcal{A}_j(a_j) \end{cases}$$

$$\mathcal{A}_j(a_j) = b^{(j)} \sqrt{a_j + c^{(j)}}$$

$$c^{(j)} = \frac{7\tau_2^{(j)}}{8\tau_1^{(j)}} - \frac{\tau_3^{(j)}}{2\tau_2^{(j)}} \quad , \quad b^{(j)} = 2\sqrt{\frac{\tau_1^{(j)}}{\tau_2^{(j)}}} \qquad \tau_k^{(j)} = \sum_i \left(h^{(j)}[i]\right)^k$$

ISOTROPIC UNDECIMATED WAVELET TRANSFORM

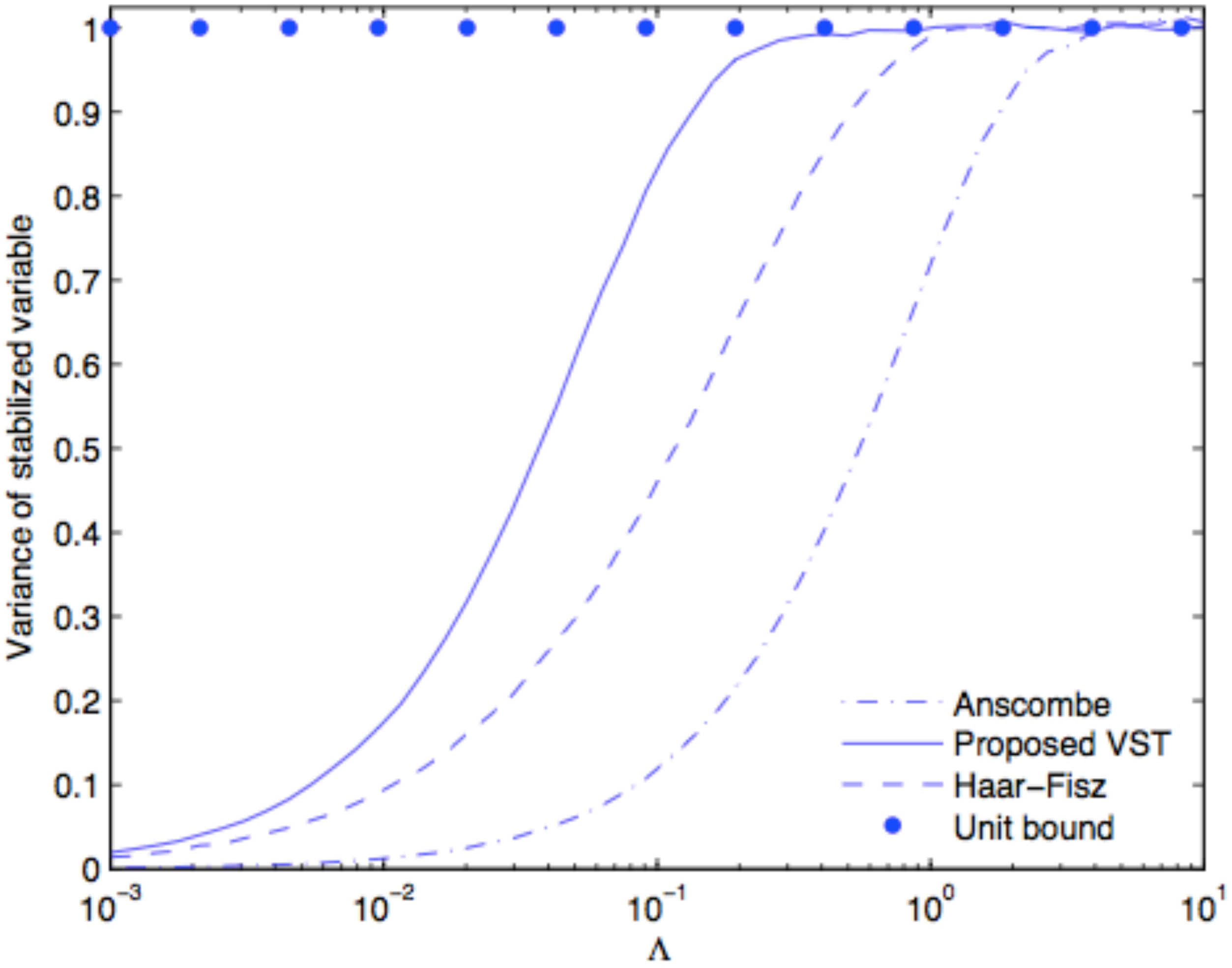


J.-L. Starck, M.J. Fadili, S. Digel , B. Zhang and J. Chiang, "Source Detection Using a 3D Sparse Representation: Application to the Fermi Gamma-ray Space Telescope ", Astronomy and Astrophysics , 504, 2, pp.641-652, 2009.

J. Schmitt, J.L. Starck, J.M. Casandjian, M.J. Fadili, I. Grenier, "Poisson Denoising on the Sphere: Application to the Fermi Gamma Ray Space Telescope", Astronomy and Astrophysics, 517, A26, **2010.**

Multiscale VST

For multiscale VST, the convergence toward the asymptotic behavior can be much more rapid for the new VST than Anscombe or Fisz.



➔ A Poisson process can be reasonably considered stabilized for:

➔ $\lambda > 10$ using **Anscombe**,

➔ $\lambda \gtrsim 1$ using **Fisz** and for

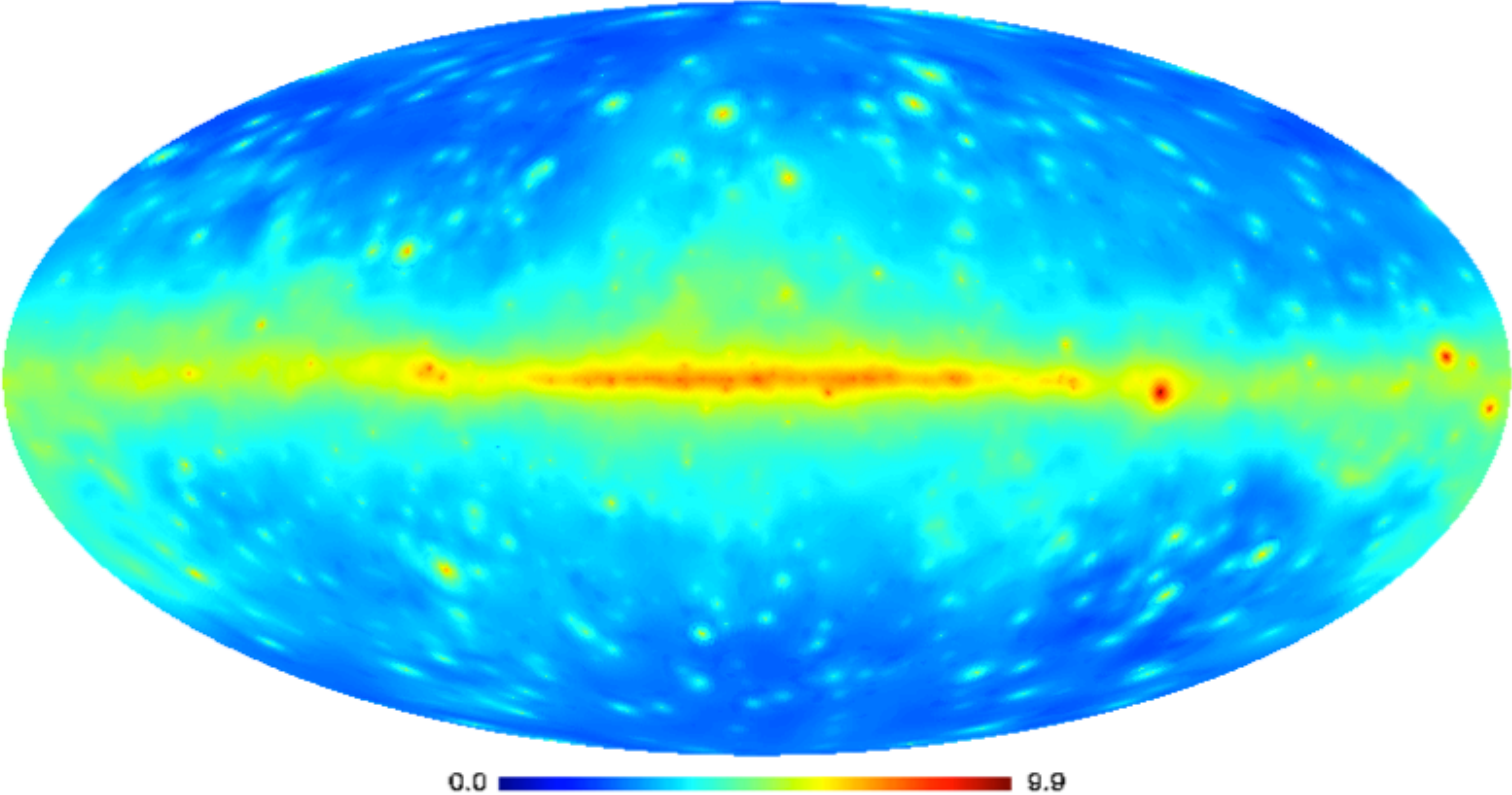
➔ $\lambda > 0.1$ using the multiscale **VST**

Poisson noise, wavelets, and sources detection

J.-L. Starck, M.J. Fadili, S. Digel , B. Zhang and J. Chiang, "Source Detection Using a 3D Sparse Representation: Application to the Fermi Gamma-ray Space Telescope ", Astronomy and Astrophysics , 504, 2, pp.641-652, 2009.

J. Schmitt, J.L. Starck, J.M. Casandjian, M.J. Fadili, I. Grenier, "Poisson Denoising on the Sphere: Application to the Fermi Gamma Ray Space Telescope", Astronomy and Astrophysics, A&A, 2010.

J. Schmitt, J.L. Starck, J.M. Casandjian, M.J. Fadili, I. Grenier, "Poisson Denoising on the Sphere: Application to the Fermi Gamma Ray Space Telescope", **Astronomy and Astrophysics, A&A, 2010.**

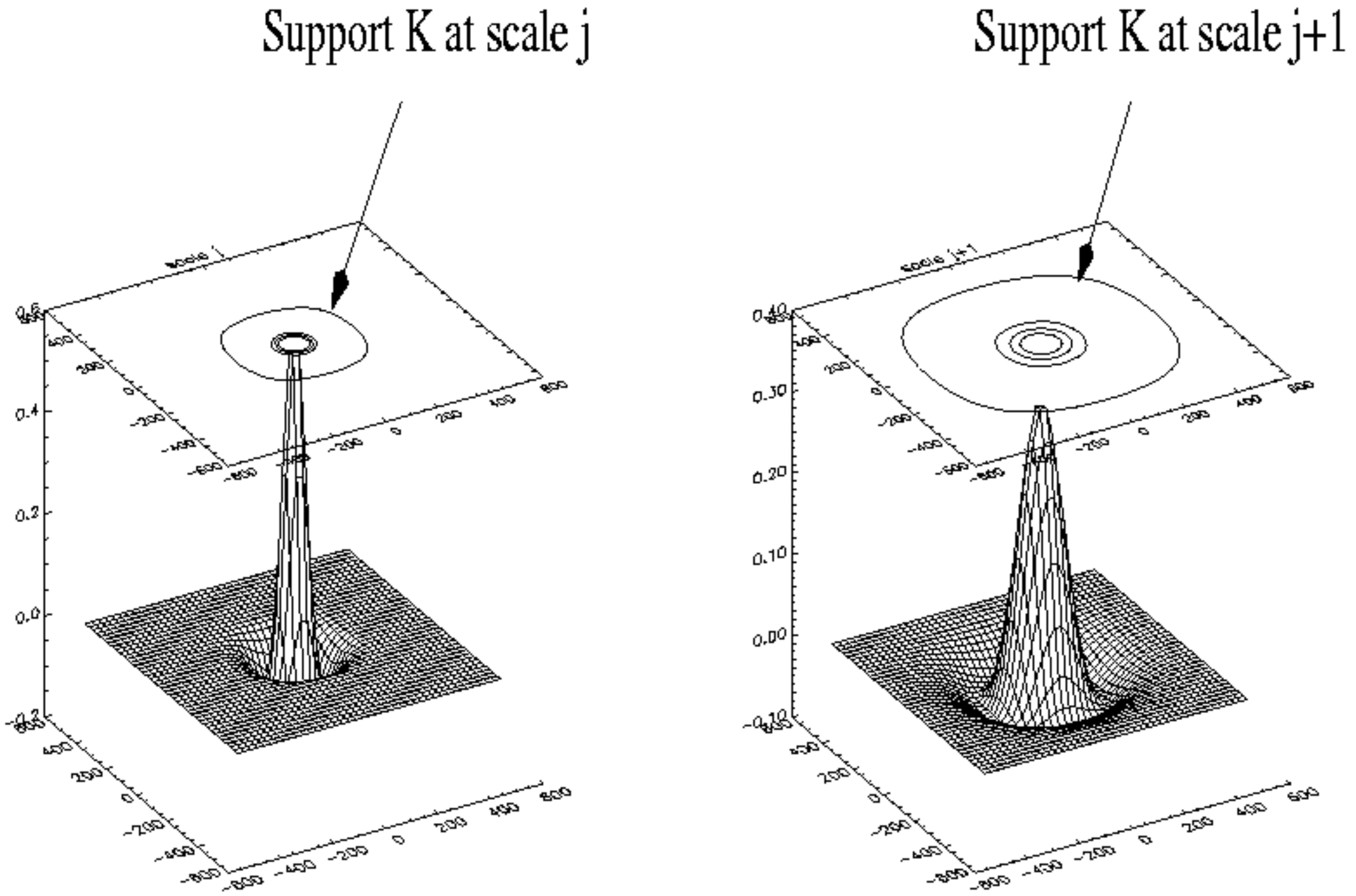


Poisson Noise with Few Photons

There are N events at positions $(k_1, \ldots, k_N)$ which contribute to the calculation of a wavelet coefficient w_j a location l (in 3D, $l = (x, y, z)$):

$$w_j(l) = \langle s, \psi_{j,l} \rangle = \sum_{i=1}^{N} \psi \left(\frac{k_i - l}{2^j} \right)$$

If a wavelet coefficient w_j is due to the noise, it can be consider as a realisation of the sum N independent random variables with the same distribution law which is the law of the wavelet function.



Support K at scale j

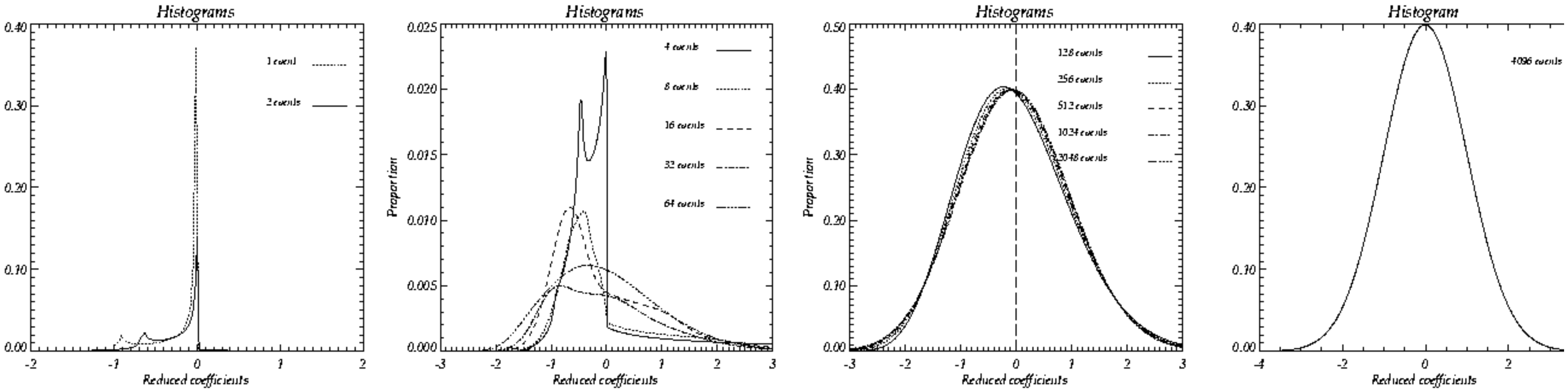
Support K at scale j+1

Support K of ψ at two consecutive scales j and $j+1$.

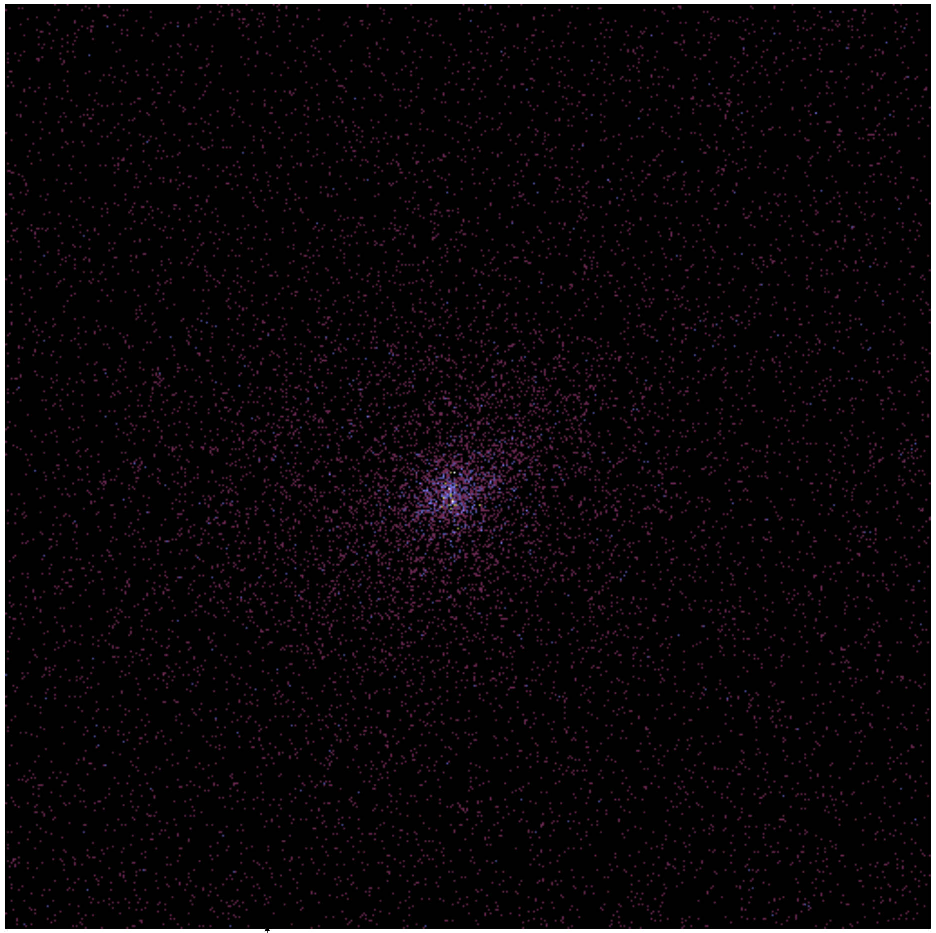
Autoconvolution histograms

As we consider independant events, the distribution law of a coefficient W_n related to N events is given by N autoconvolutions of H_1

$$H_n = H_1 \otimes H_1 \otimes \ldots \otimes H_1$$

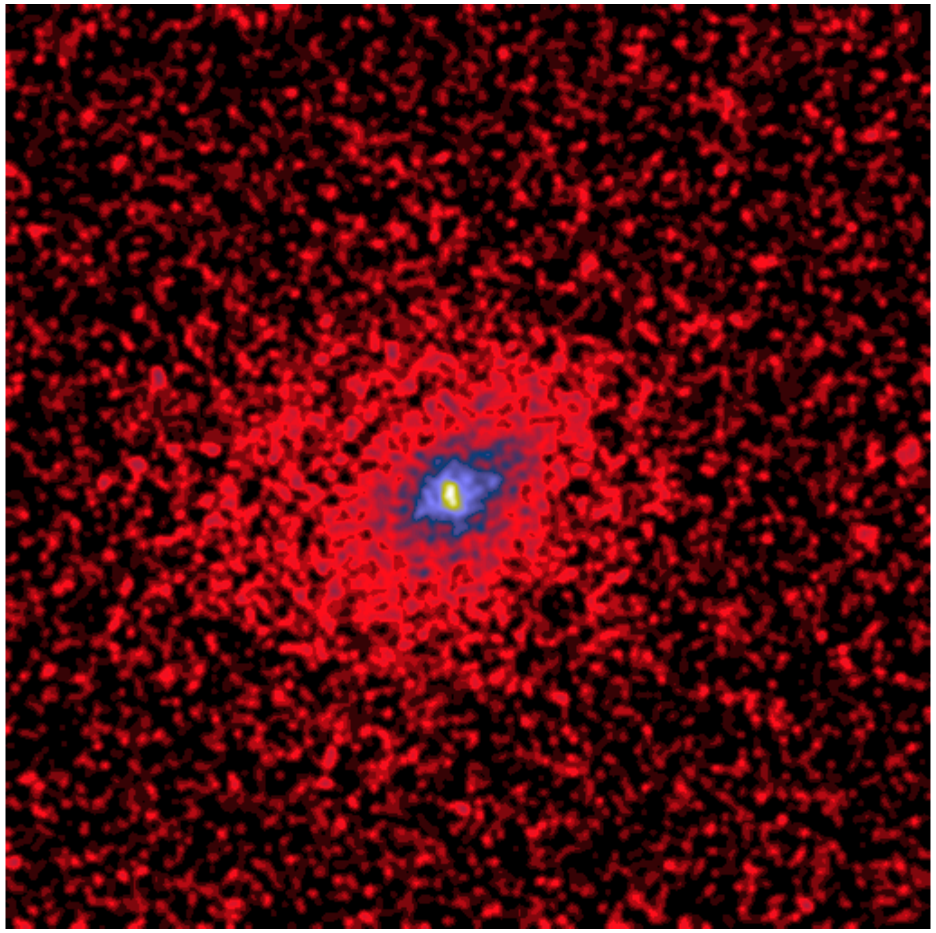


Autoconvolution histograms for the wavelet associated with a B_3 spline scaling function for one and 2 events (first), 4 to 64 events (second), 128 to 2048 (third), and 4096 (right).

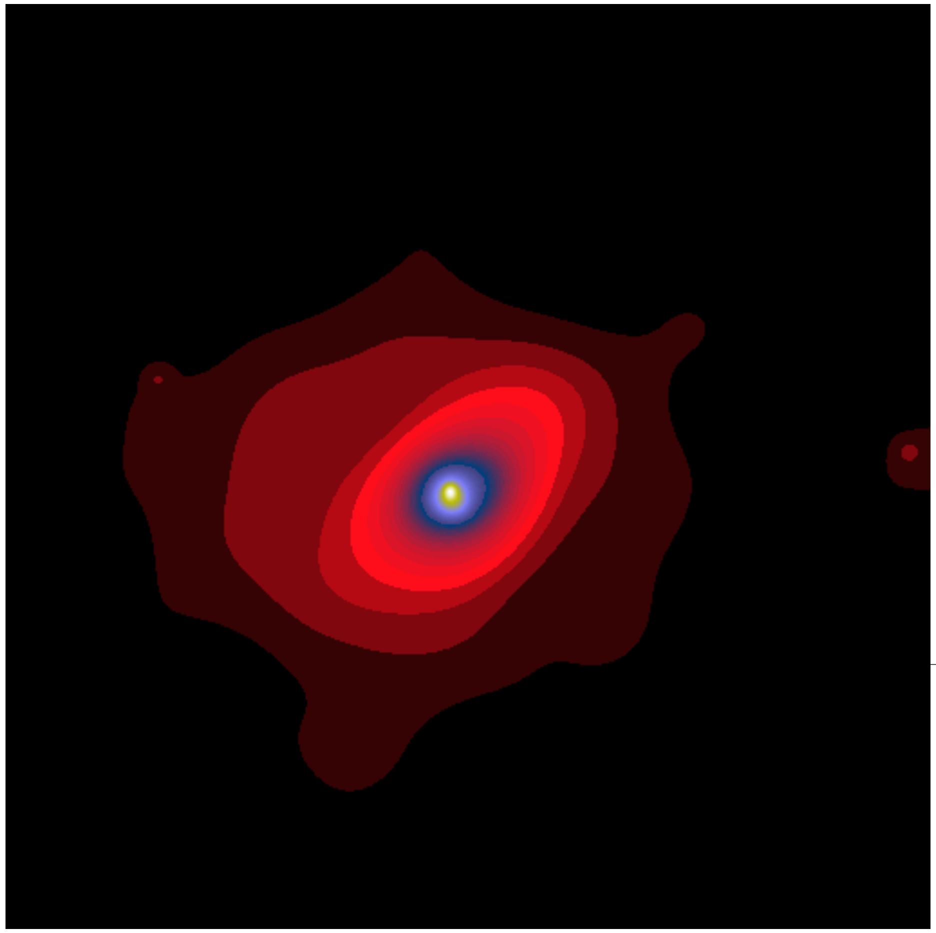


ROSAT A2390

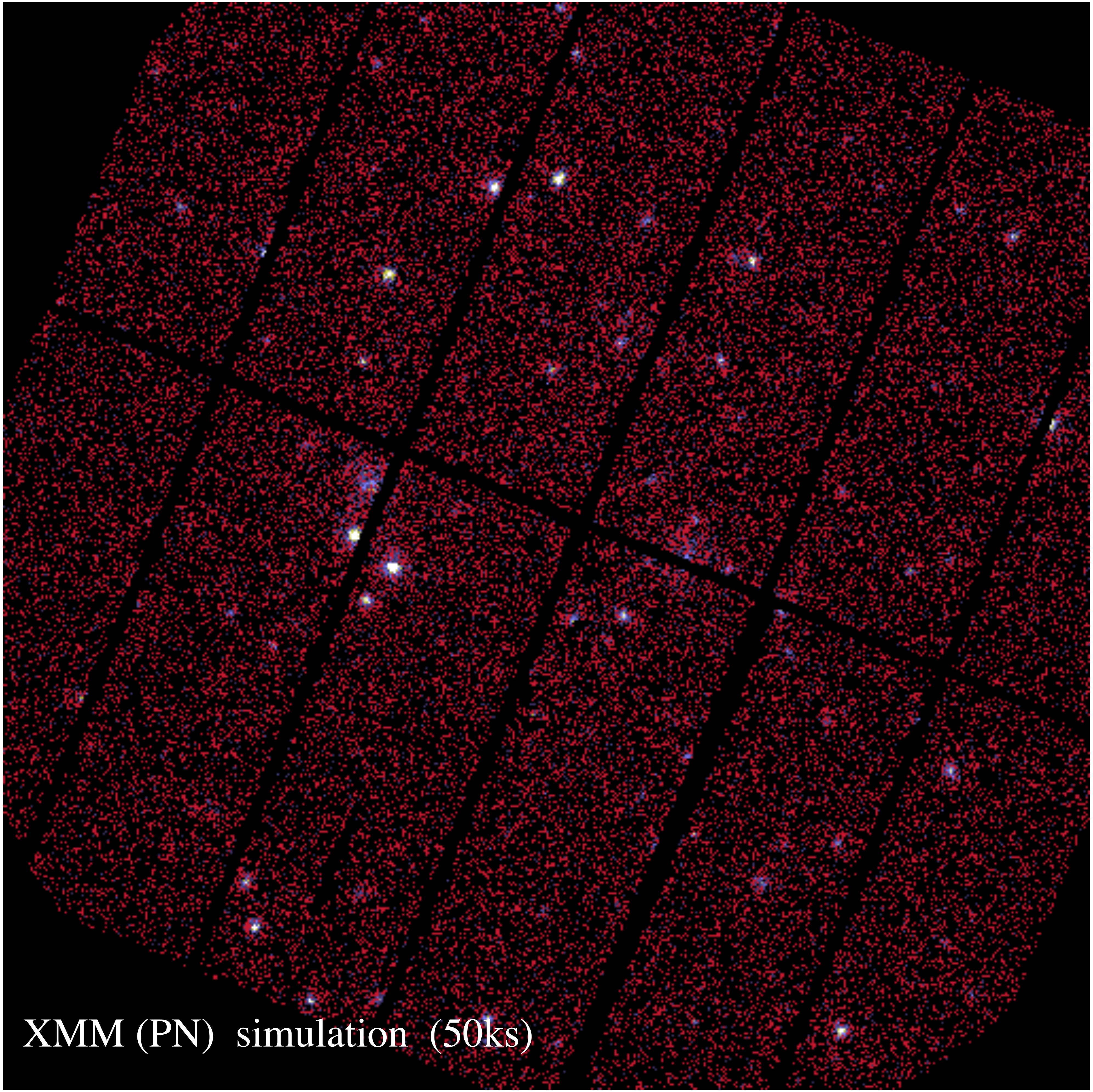
FILTERING



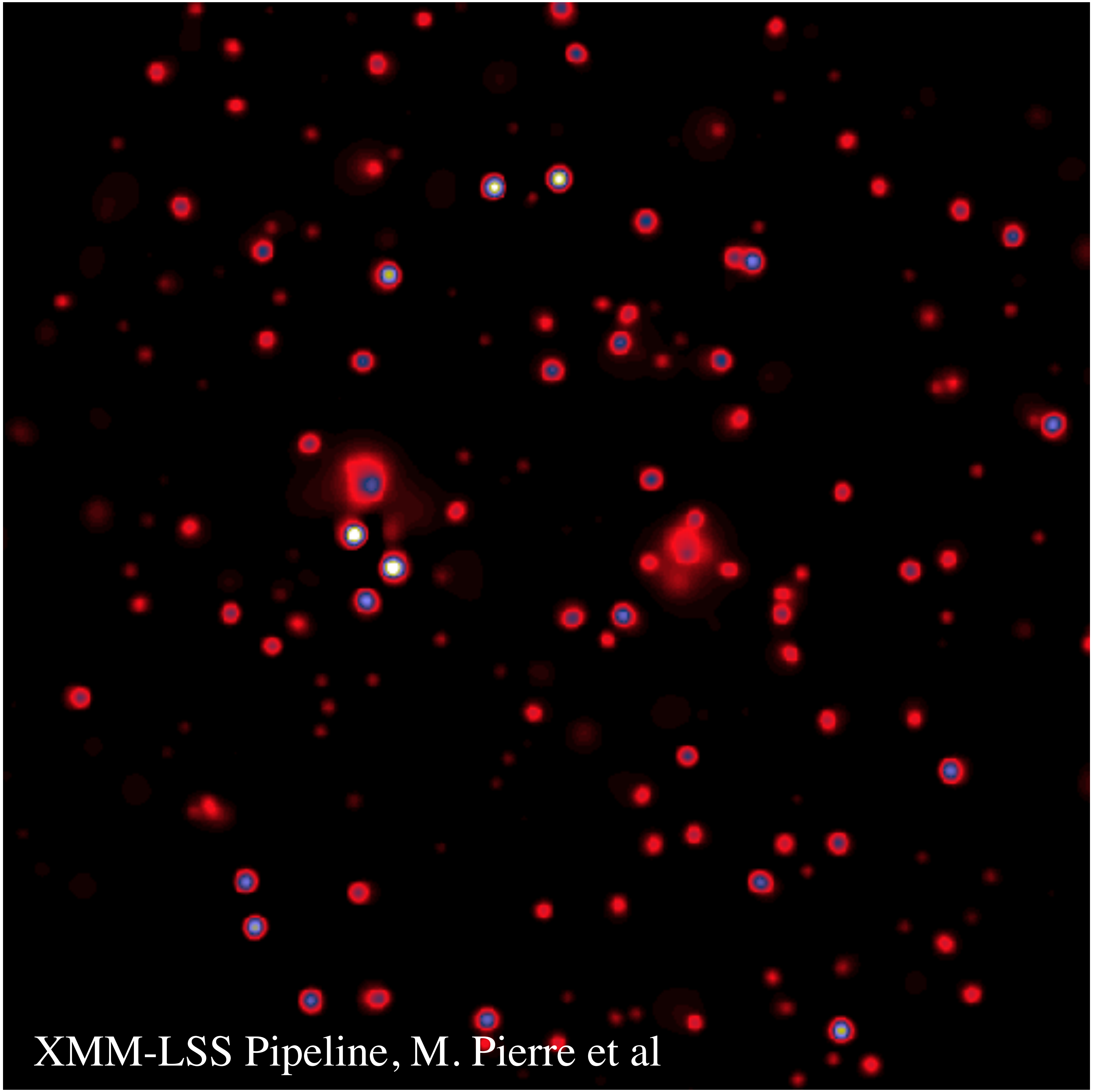
Gaussian Filtering



Wavelet Filtering



XMM (PN) simulation (50ks)



XMM-LSS Pipeline, M. Pierre et al

Generalization to Iterative Soft/Hard Thresholding Algorithm

Iterative thresholding methods were proposed initially in [F. and Nowak, 2003], [Daubechies, Defrise, De Mol, 2003], [Starck, Candès, Donoho, 2003]:

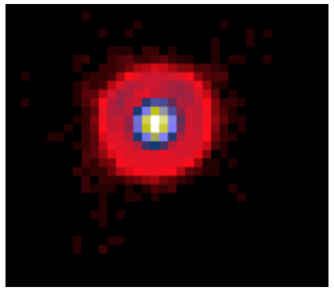
$$\alpha^{(n+1)} = \mathrm{HardThresh}_{\lambda}\left(\alpha^{(n)} + \Phi^T H^T \left(Y - H\Phi\alpha^{(n)}\right)\right)$$
$$\alpha^{(n+1)} = \mathrm{SoftThresh}_{\lambda}\left(\alpha^{(n)} + \Phi^T H^T \left(Y - H\Phi\alpha^{(n)}\right)\right)$$

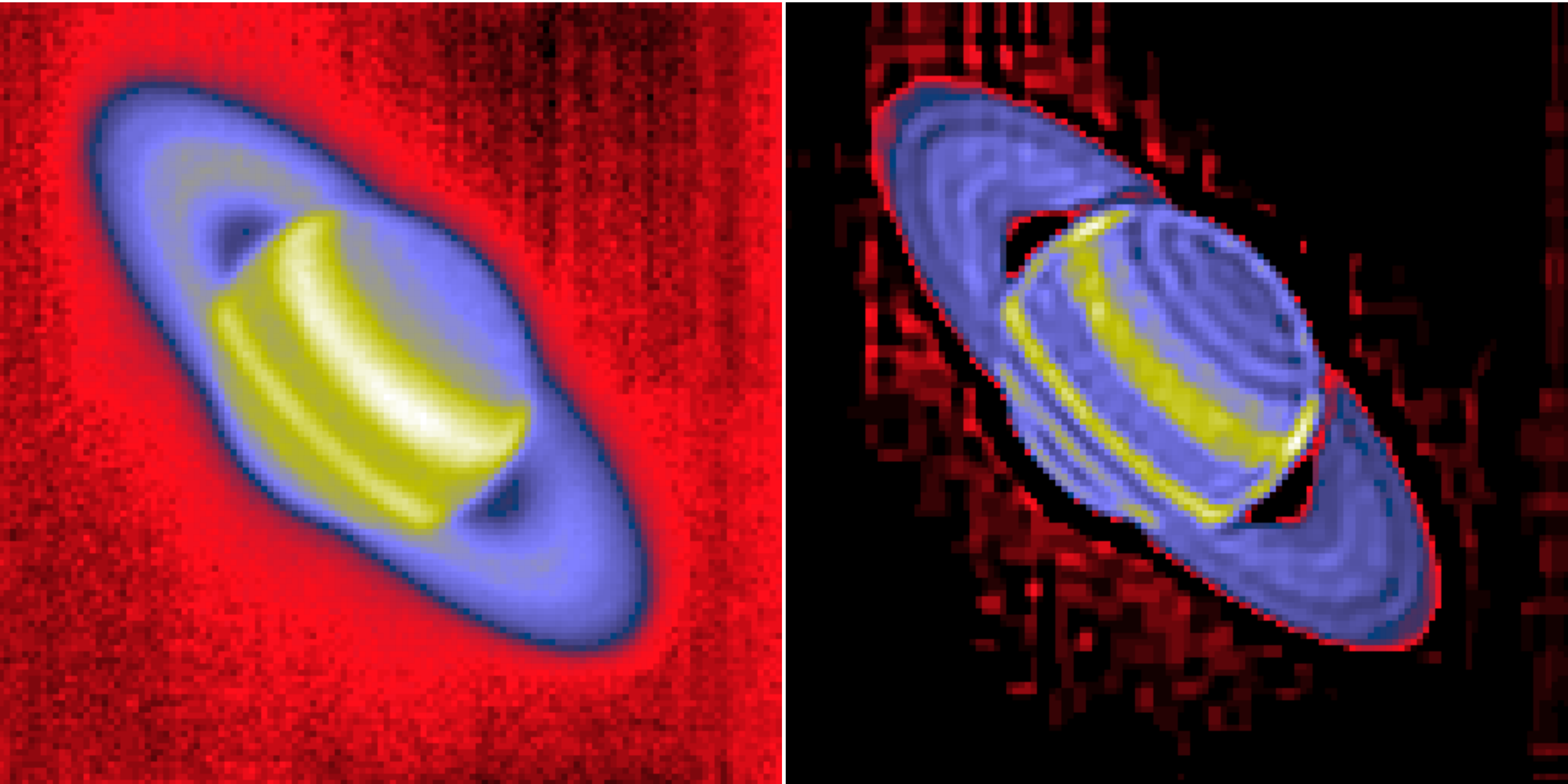
IST can be seen as a generalization of projected gradient descent In the framework of **proximal theory** [Moreau 62], we have [Combettes and Wajs, 2005]:

$$\alpha^{n+1} = \mathrm{prox}_{\mathcal{C}}(\alpha^n + \mu\Phi^t H^t (Y - H\Phi\alpha^n)).$$

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.





Polarized Data Denoising

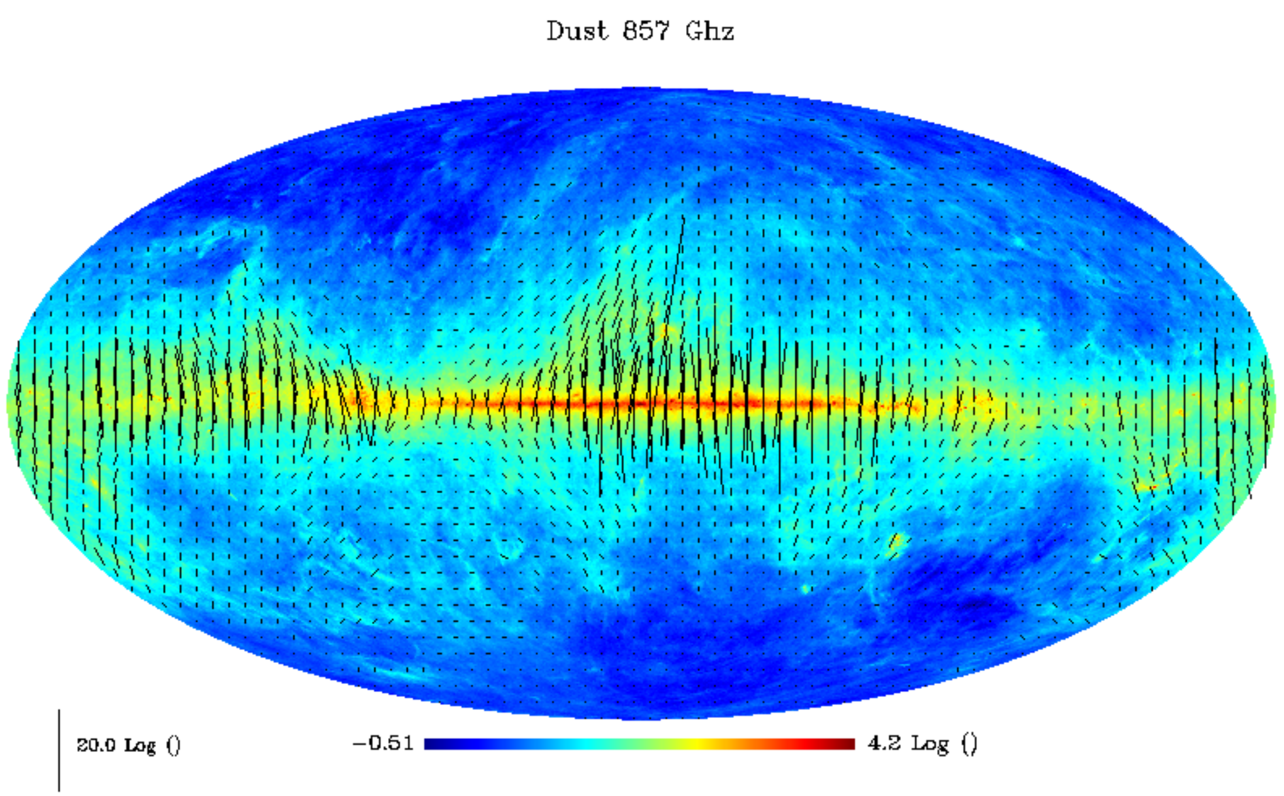
$$Q(\theta,\phi) = \sum_{l,m} c^E_{J,l,m} Z^+_{l,m} + i c^B_{J,l,m} Z^-_{l,m} + \sum_j \sum_{l,m} \tilde{w}^E_{j,l,m} Z^+_{l,m} + i \tilde{w}^B_{j,l,m} Z^-_{l,m}$$
$$U(\theta,\phi) = \sum_{l,m} c^B_{J,l,m} Z^+_{l,m} - i c^E_{J,l,m} Z^-_{l,m} + \sum_j \sum_{l,m} \tilde{w}^B_{j,l,m} Z^+_{l,m} - i \tilde{w}^E_{j,l,m} Z^-_{l,m}$$

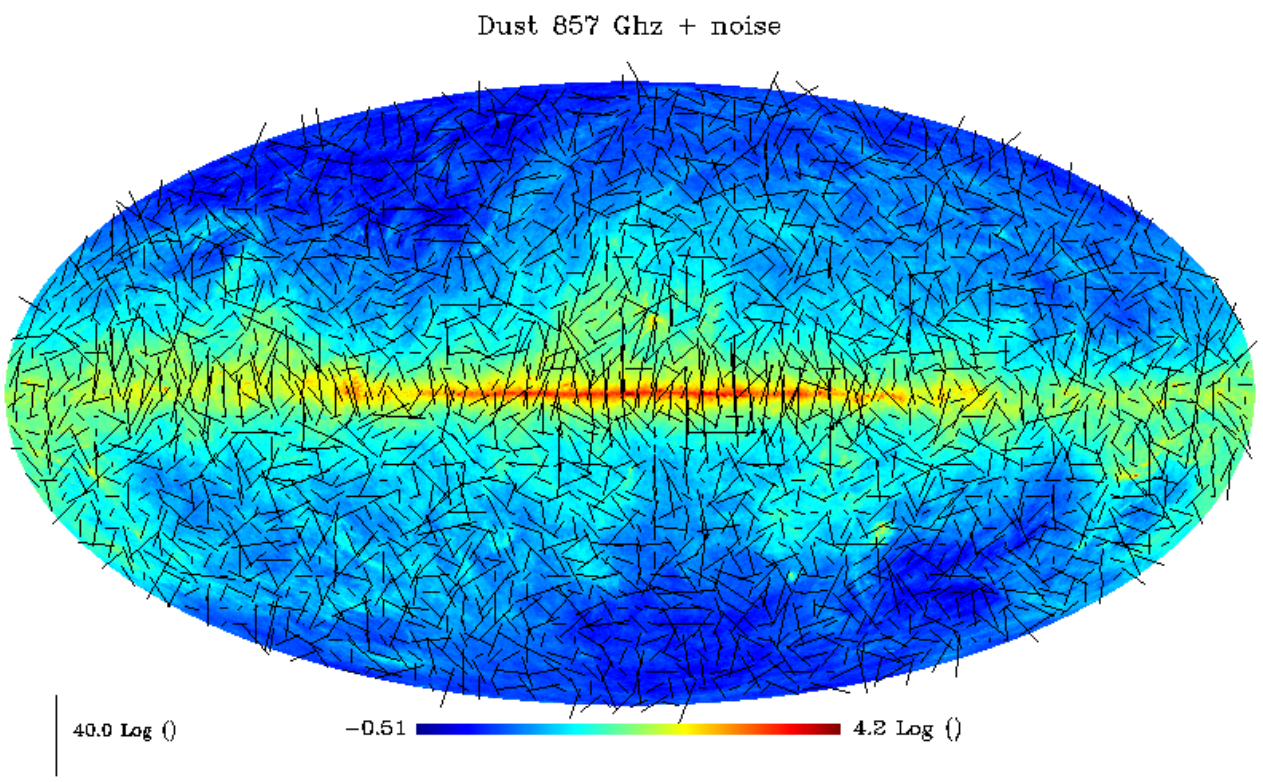
Where

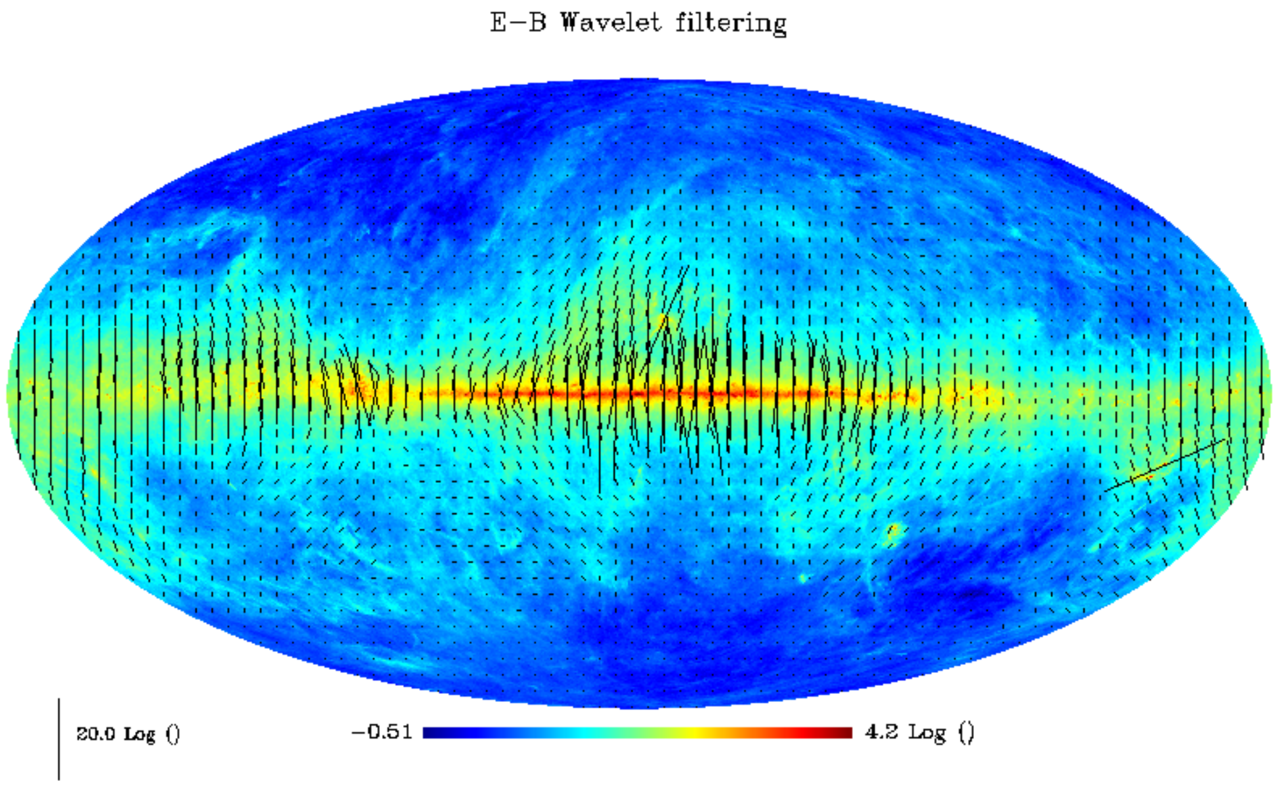
$$\tilde{w}^E_{j,k} = \delta(w^E_{j,k})$$
$$\tilde{w}^B_{j,k} = \delta(w^B_{j,k})$$

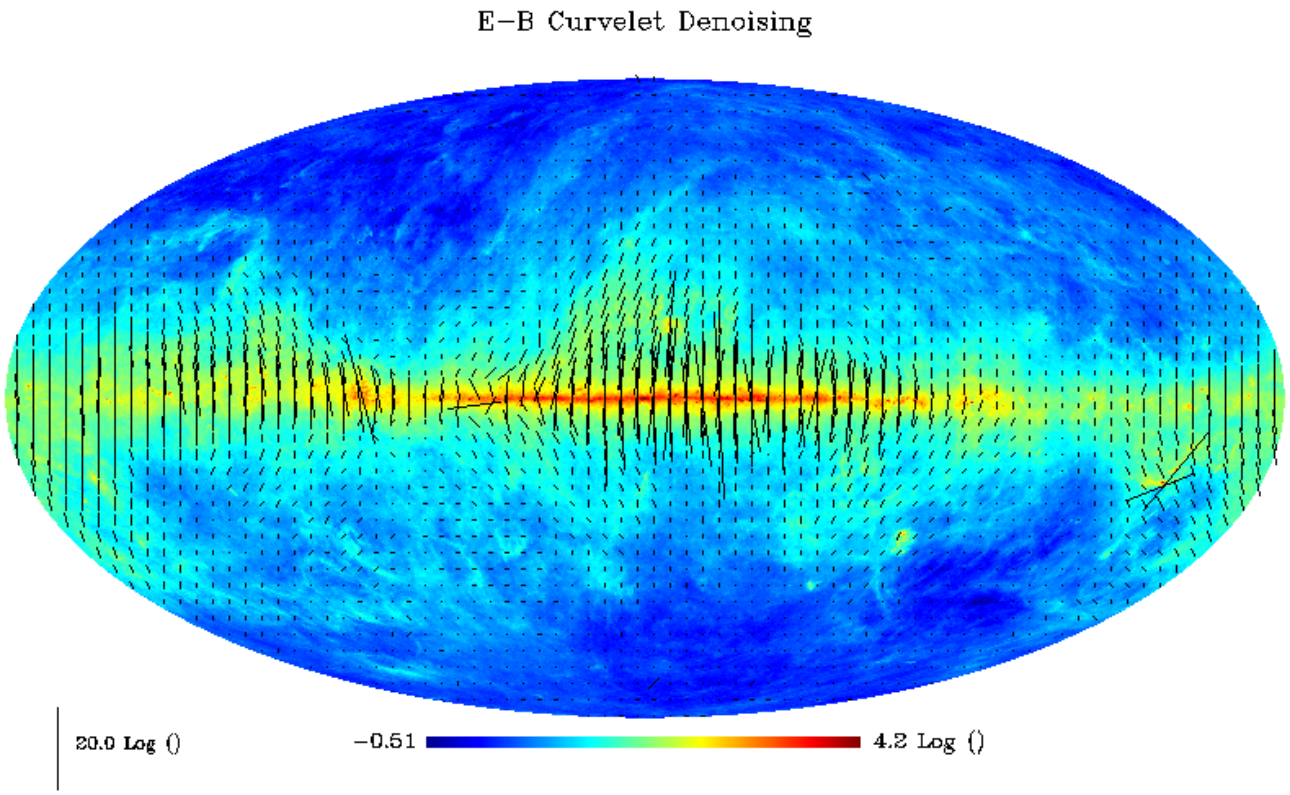
Hard thresholding corresponds to the following non linear operation:

$$\tilde{w}_{j,k} = \begin{cases} w_{j,k} & \text{if } \mid w_{j,k} \mid \geq T_j \\ 0 & \text{otherwise} \end{cases}$$


Dust 857 Ghz


Dust 857 Ghz + noise


E−B Wavelet filtering

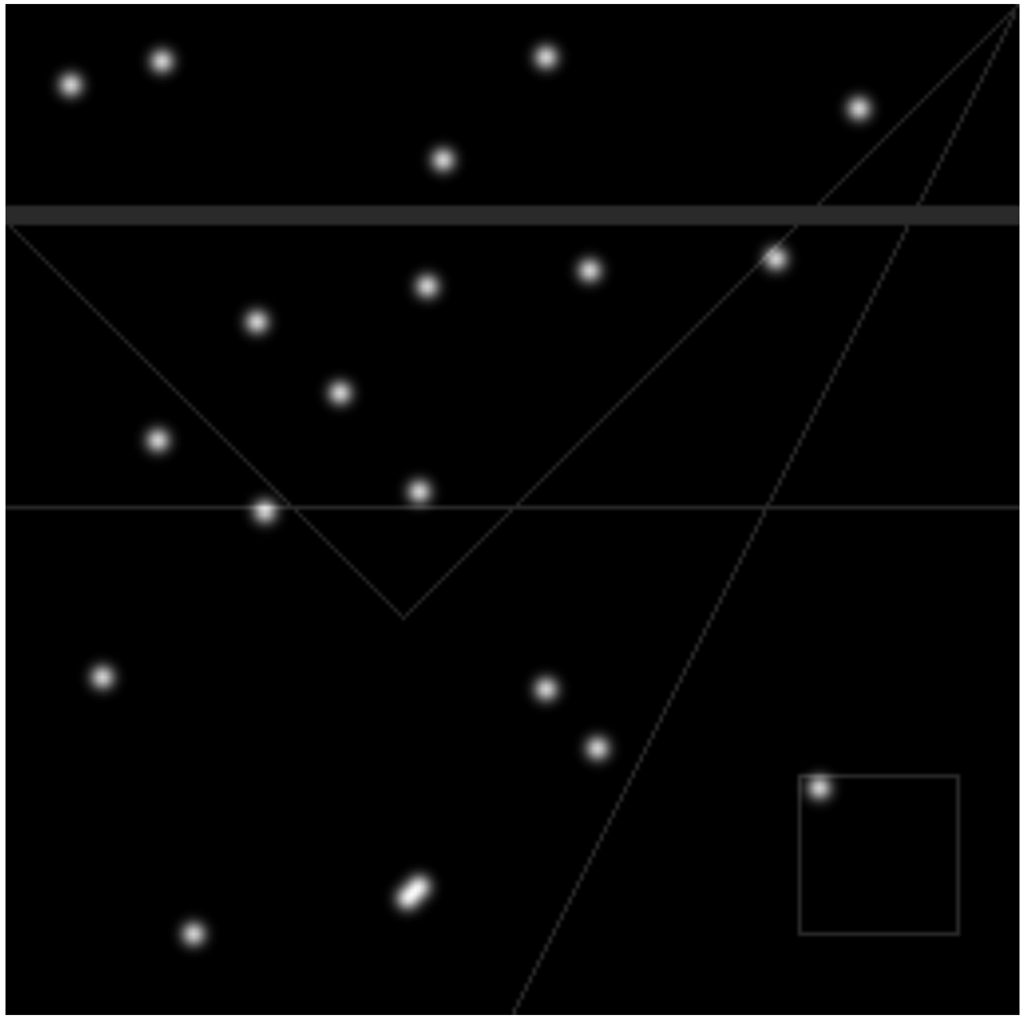

E−B Curvelet Denoising

Inverse Problems in Astrophysics

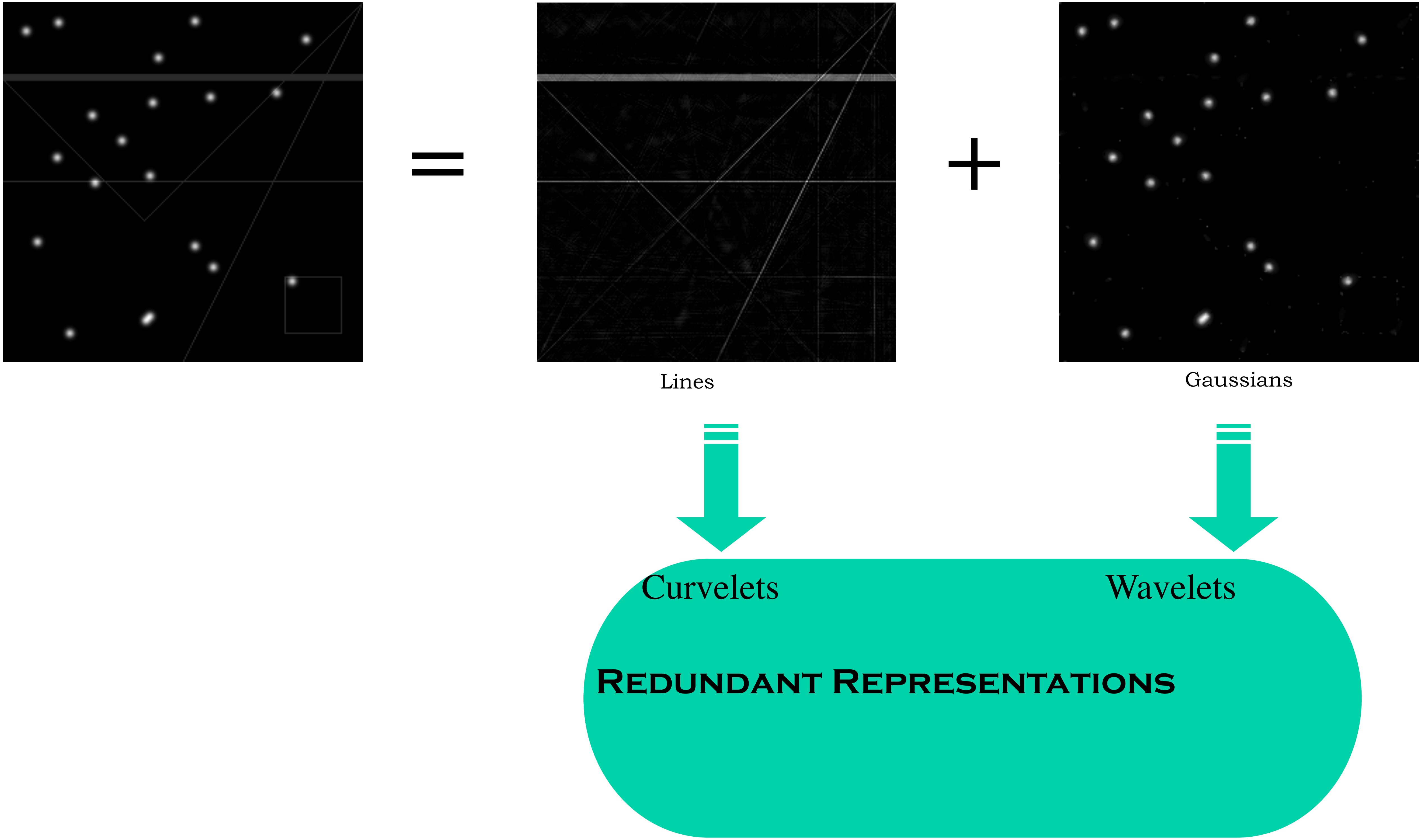
- ➡Part 1: Introduction to Inverse Problems
- ➡Part 2: From Fourier to Wavelets
- ➡Part 3: Wavelet and Beyond
- ➡Part 4: Sparse Regularization
- ➡**Part 5: Application to Unmixing and Inpainting**
- ➡Part 6: Compressed Sensing
- ➡Part 7: Deep Learning

A difficult issue

Is there any representation that well represents the following image ?



Going further



Lines

Gaussians

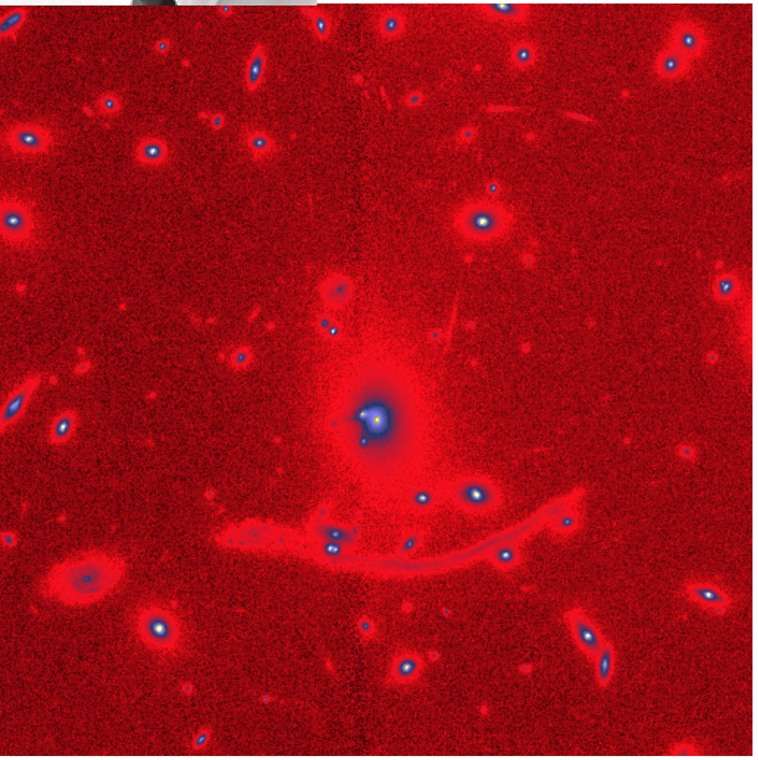
Curvelets

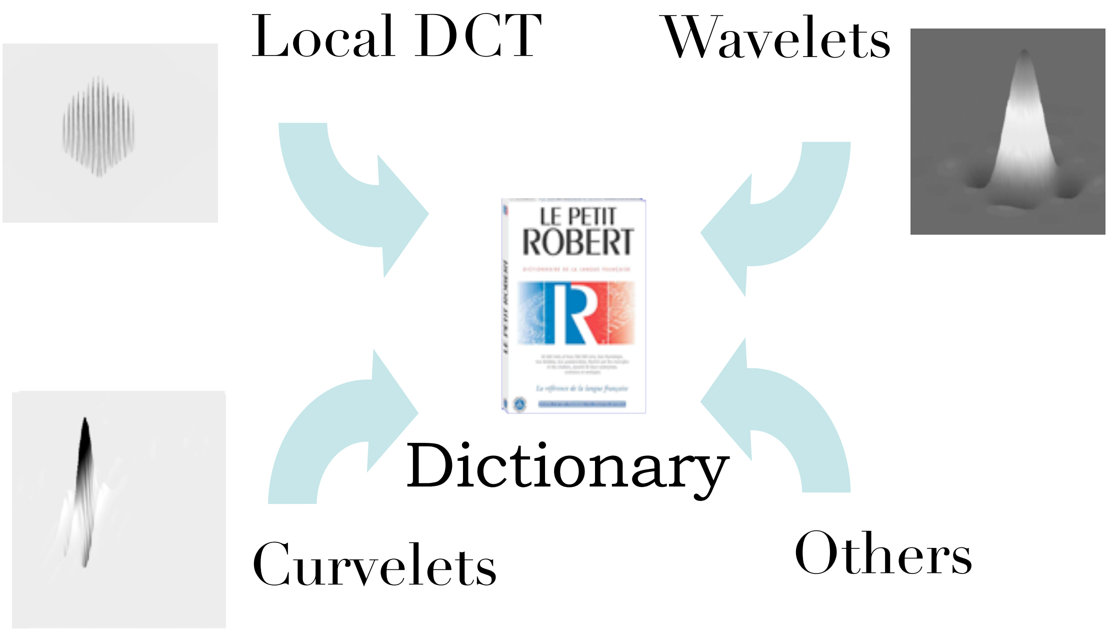
Wavelets

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.*





Local DCT

Wavelets

Dictionary

Curvelets

Others

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

Model:

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

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

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.*

Sparsity Model: we consider a signal as a sum of K components s_k, ,
each of them being sparse in a given dictionary :

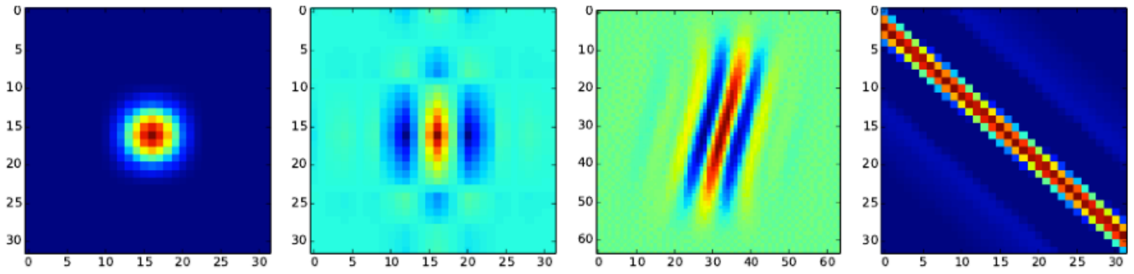
$$Y = X_1 + X_2$$

X_1 can be well approximated with few coefficients in a given domain.
X_2 can be well approximated with few coefficients in **another** domain.

$$min_{X_1,X_2} \parallel Y - (X_1 + X_2) \parallel^2 + C_1(X_1) + C_2(X_2)$$

$$C_1(X_1) = \parallel \Phi_1 X_1 \parallel_1$$

$$C_2(X_2) = \parallel \Phi_2 X_2 \parallel_1$$



(A) Starlet (B) Bi-orthogonal (C) Curvelet (D) Ridgelet

Morphological Component Analysis (MCA)

$$\min_{X_1,\ldots,X_L} \parallel Y - \sum_{k=1}^{L} X_k \parallel^2 + \lambda \sum_{k=1}^{L} \parallel \Phi_k^t X_k \parallel_p$$

. Initialize all X_k to zero

. Iterate j=1,...,Niter

- Iterate k=1,..,L

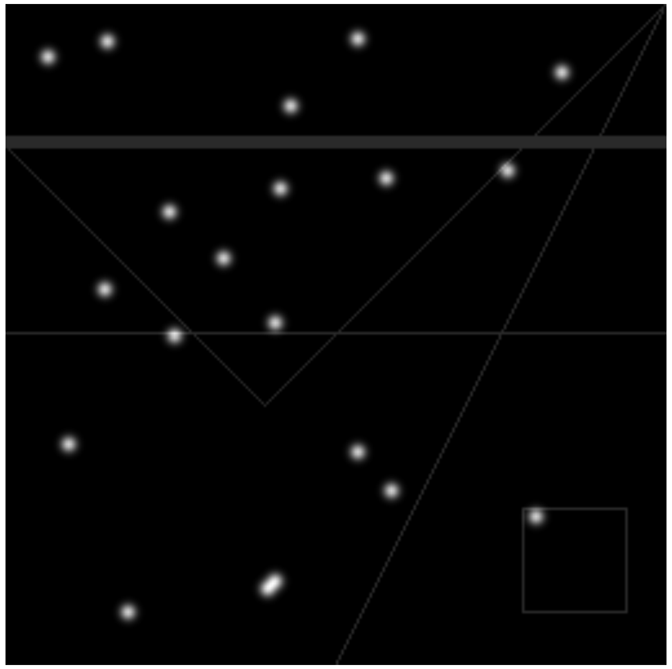
    - Update the kth part of the current solution by fixing all other parts and minimizing:

$$\min_{X_k} \parallel Y - \sum_{i=1, i \neq k}^{L} X_i - X_k \parallel^2 + \lambda_j \parallel \Phi_k^t X_k \parallel_p$$

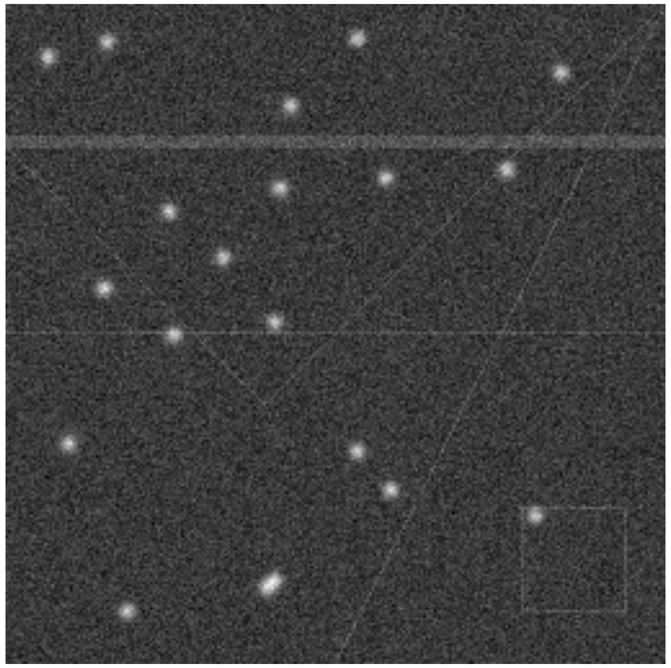
Which is obtained by a simple **hard**/soft thresholding of : $Z = Y - \sum_{i=1, i \neq k}^{L} X_i$

- Decrease the threshold λ_j

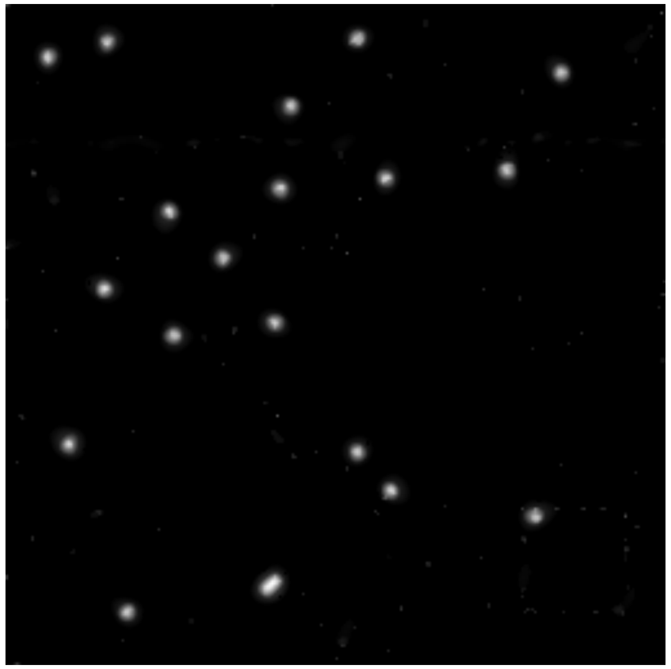
$$MIN_{s_1,s_2}\left(\|Ws_1\|_p + \|Cs_2\|_p\right)$$ subject to $$\|s-(s_1+s_2)\|_2^2 < \varepsilon$$



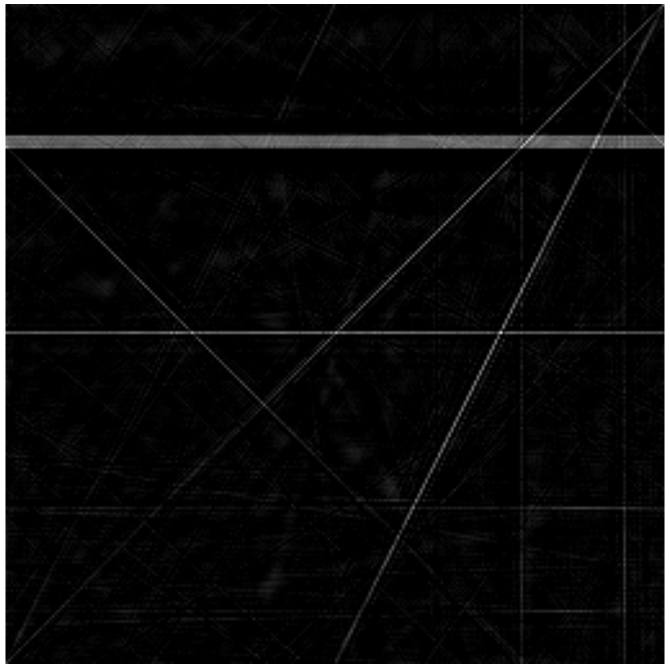
a) Simulated image (gaussians+lines)



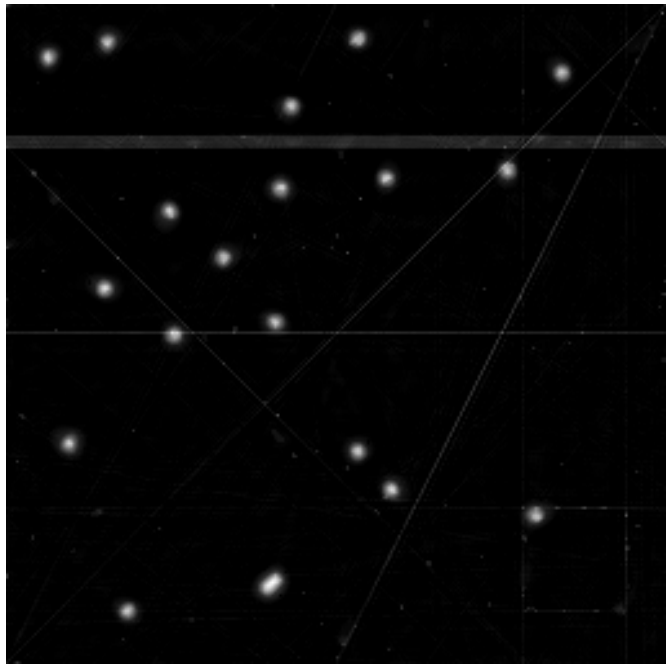
b) Simulated image + noise



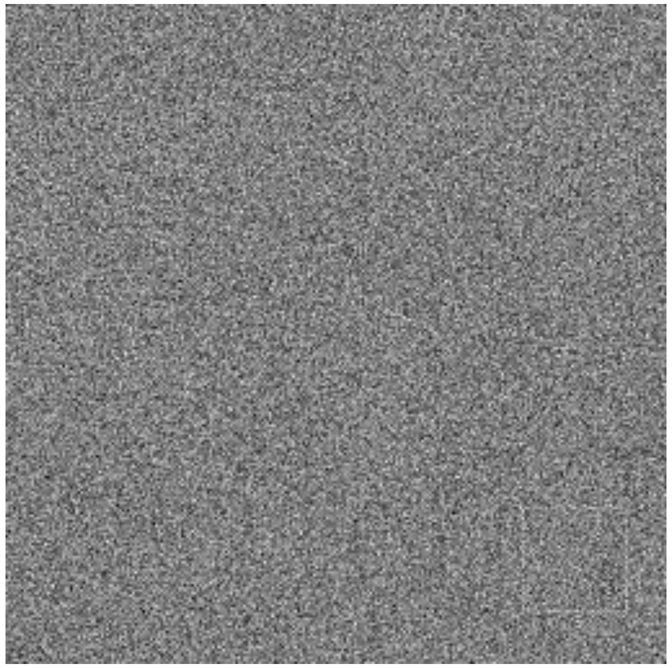
c) A trous algorithm



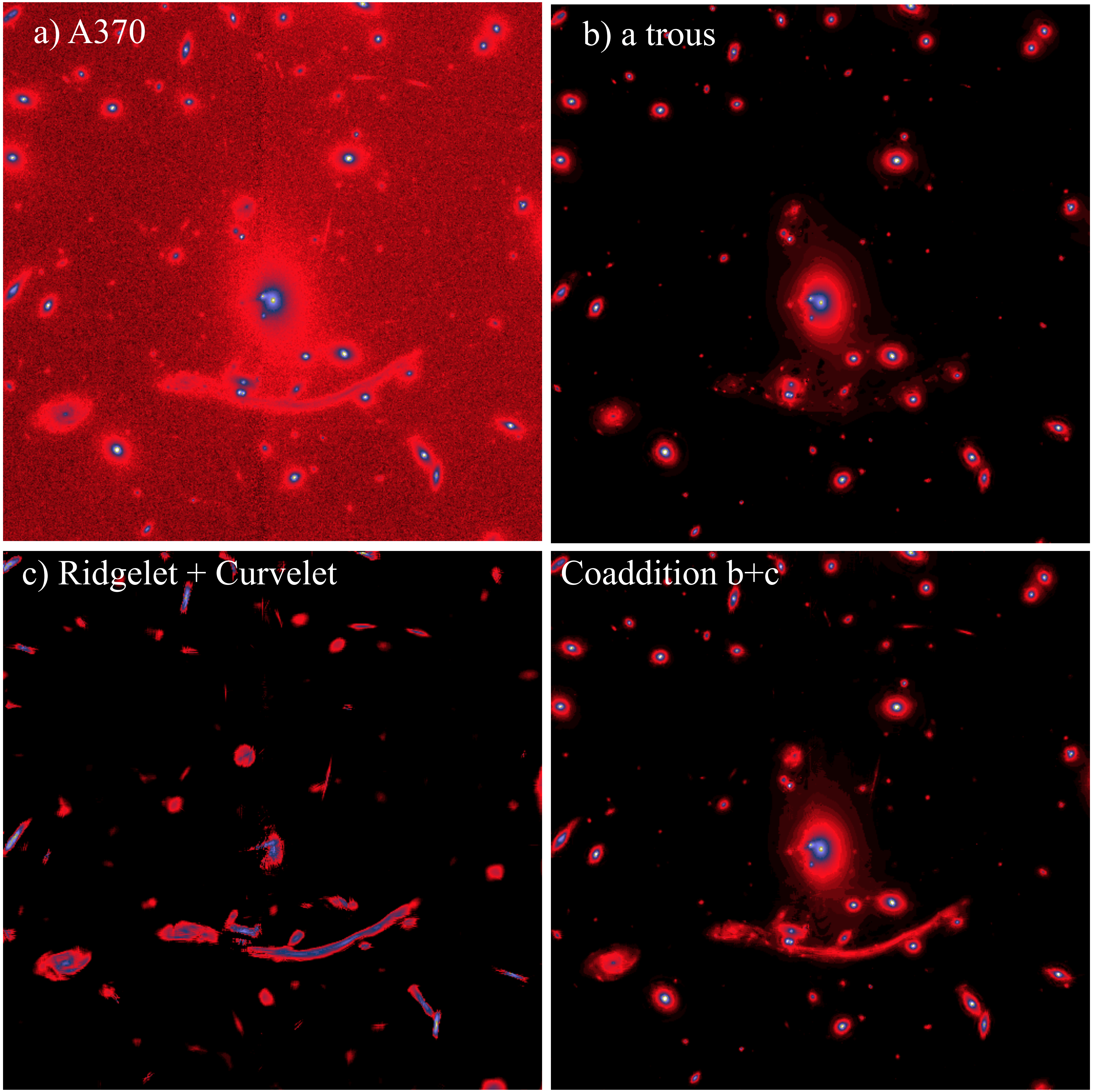
d) Curvelet transform

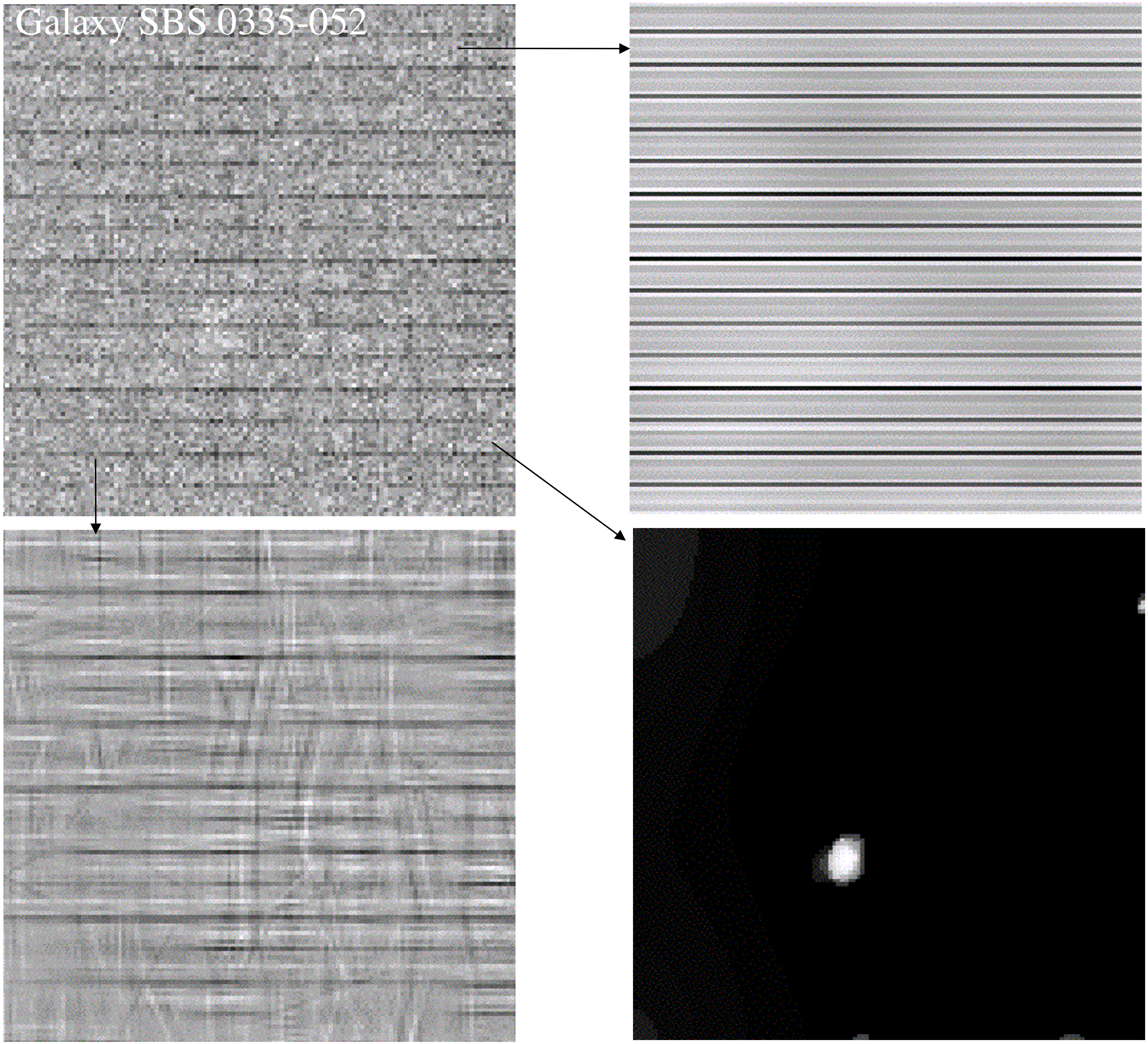


e) coaddition c+d

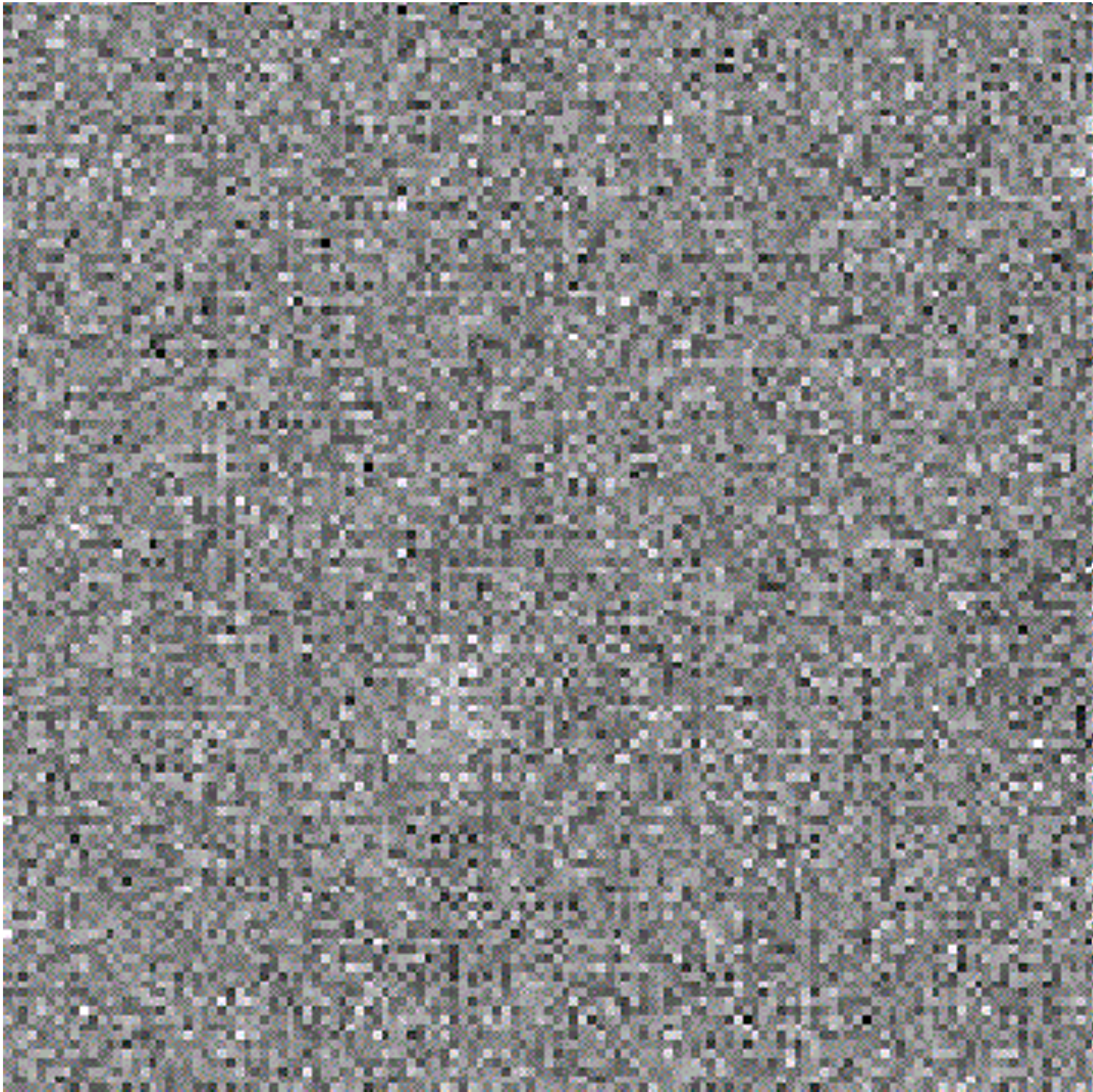


f) residual = e-b

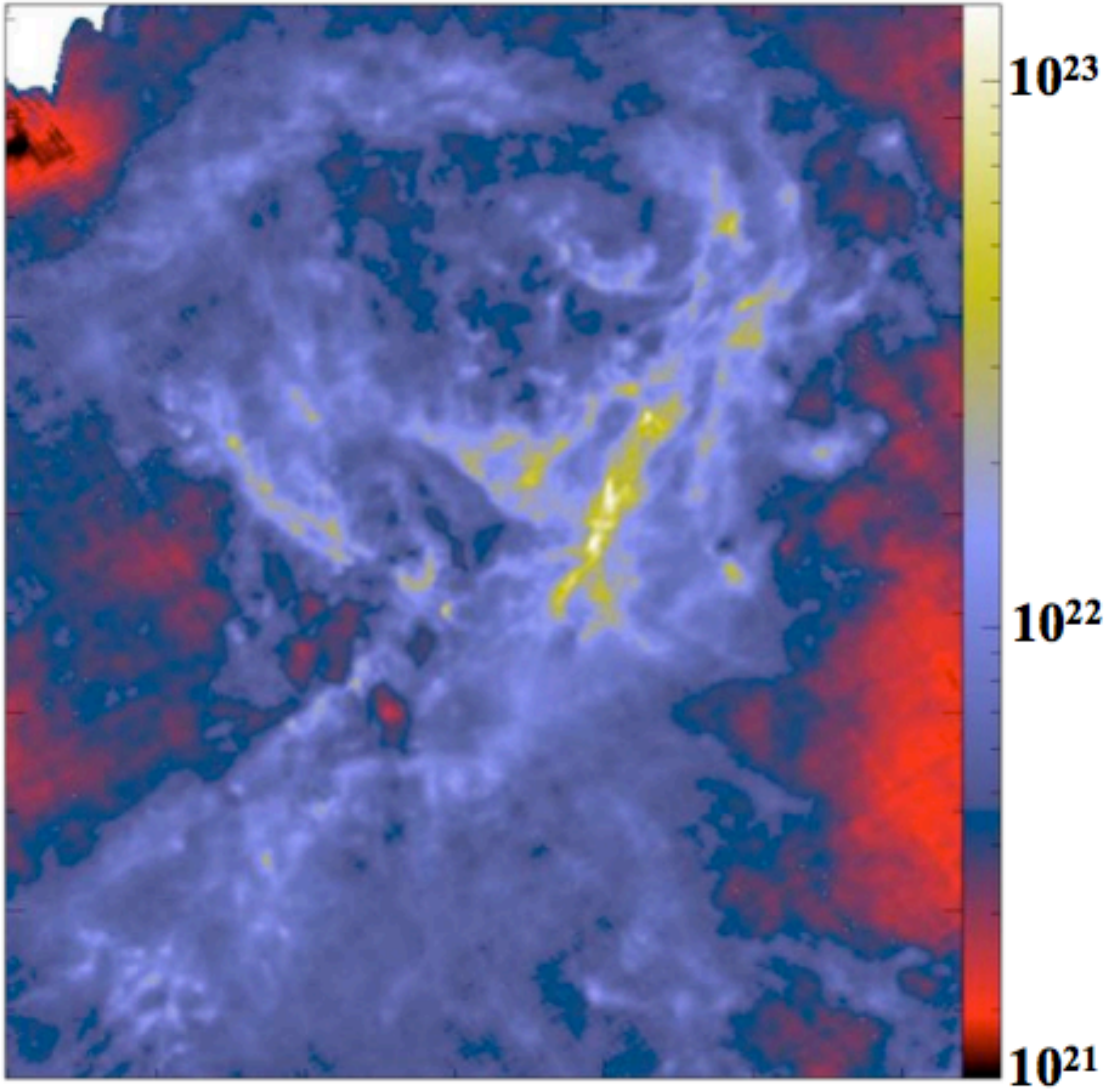

a) A370
b) a trous
c) Ridgelet + Curvelet
Coaddition b+c


Galaxy SBS 0335-052

Galaxy SBS 0335-052
10 micron
GEMINI-OSCIR



Herschel (SPIRE+PACS) Column density map (H_2/cm^2)



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

Dense cores form primarily in filaments

Morphological Component Analysis:

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

***Herschel* Column density map**

=

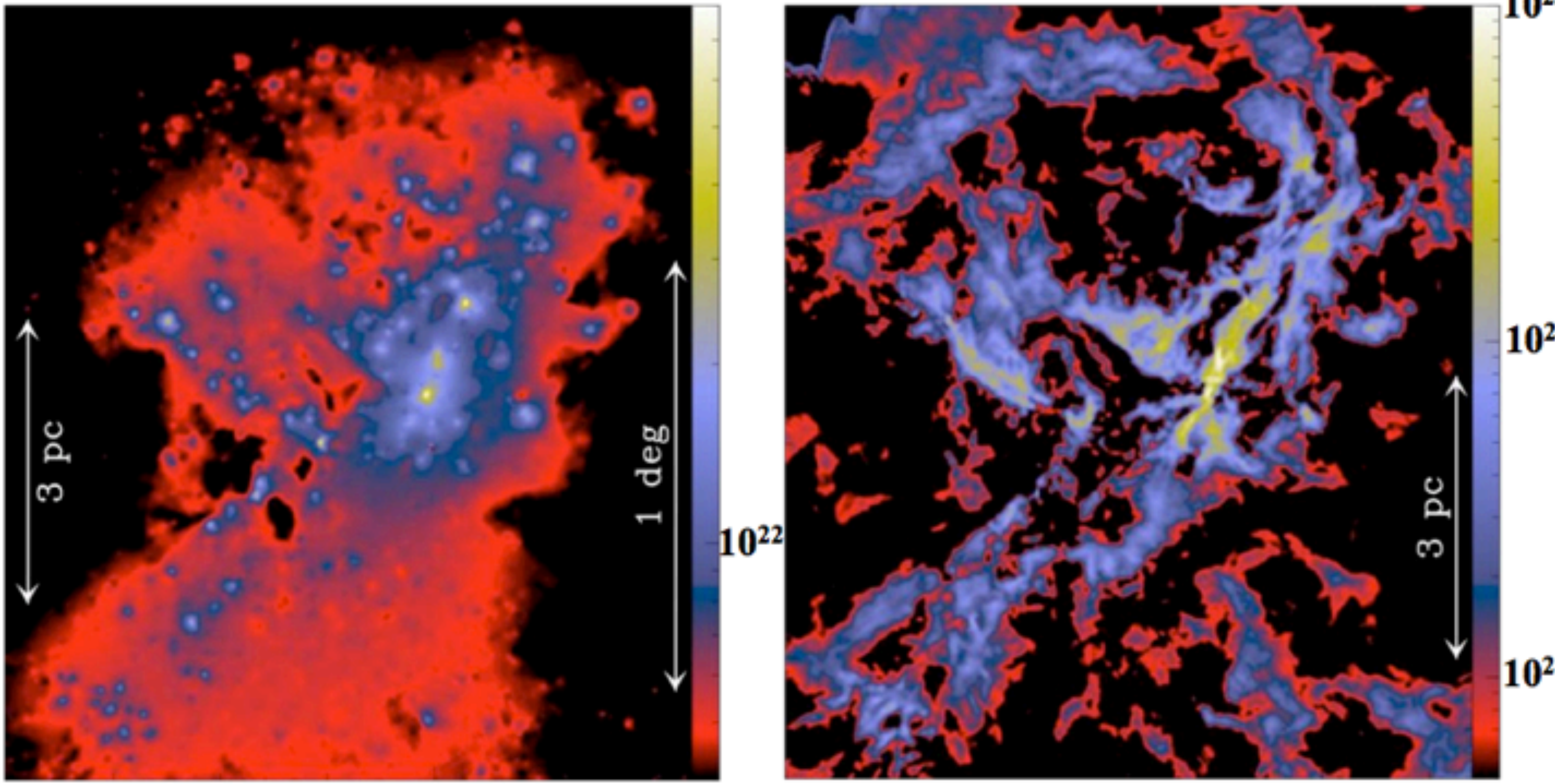
Cores

Wavelet component (H_2/cm^2)

+

Filaments

Curvelet component (H_2/cm^2)

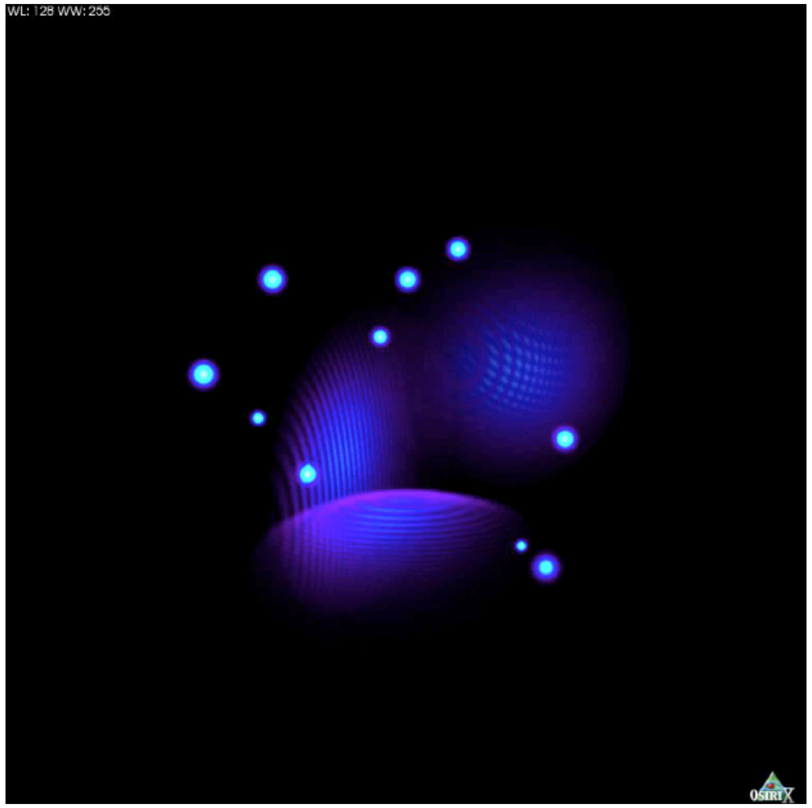


A. Menshchikov, Ph.André. P. Didelon, et al, “Filamentary structures and compact objects in the Aquila and Polaris clouds observed by Herschel”, A&A, 518, id.L103, 2010.

3D Morphological Component Analysis

A. Woiselle

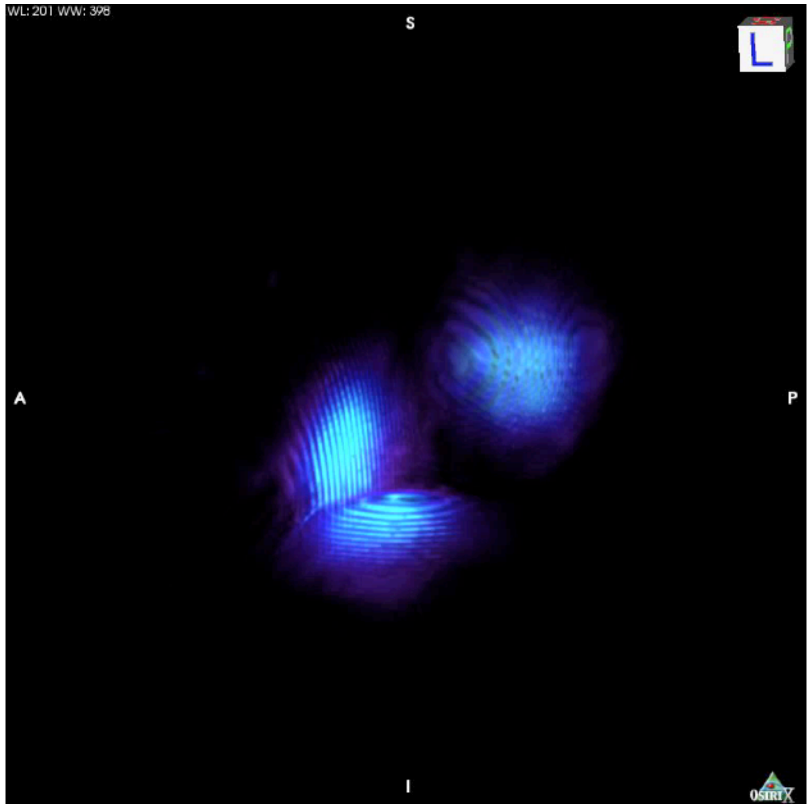
Original (3D shells + Gaussians)



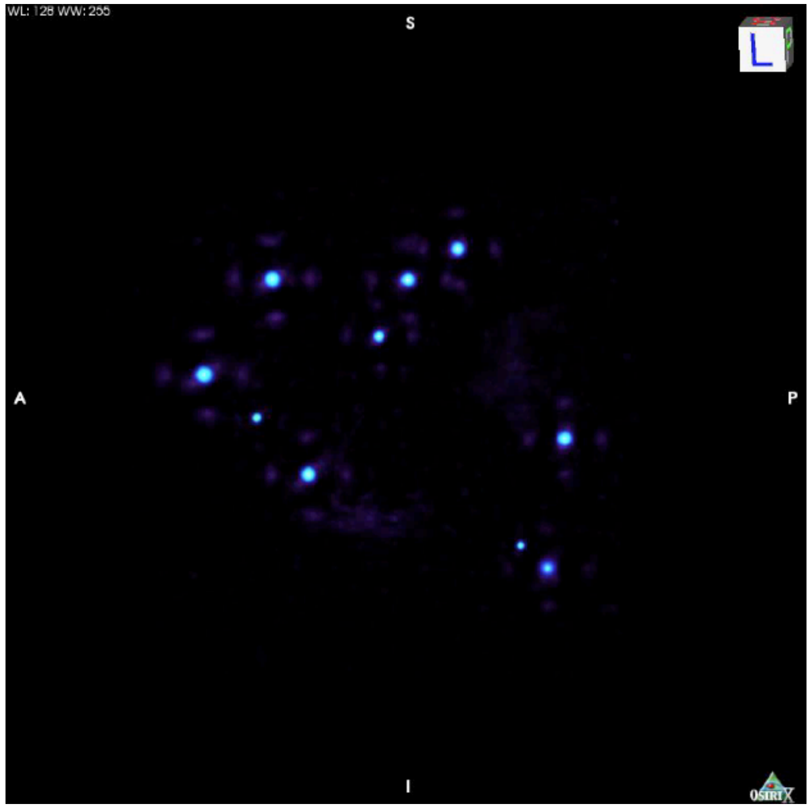
Dictionary

RidCurvelets + 3D UDWT.

Shells



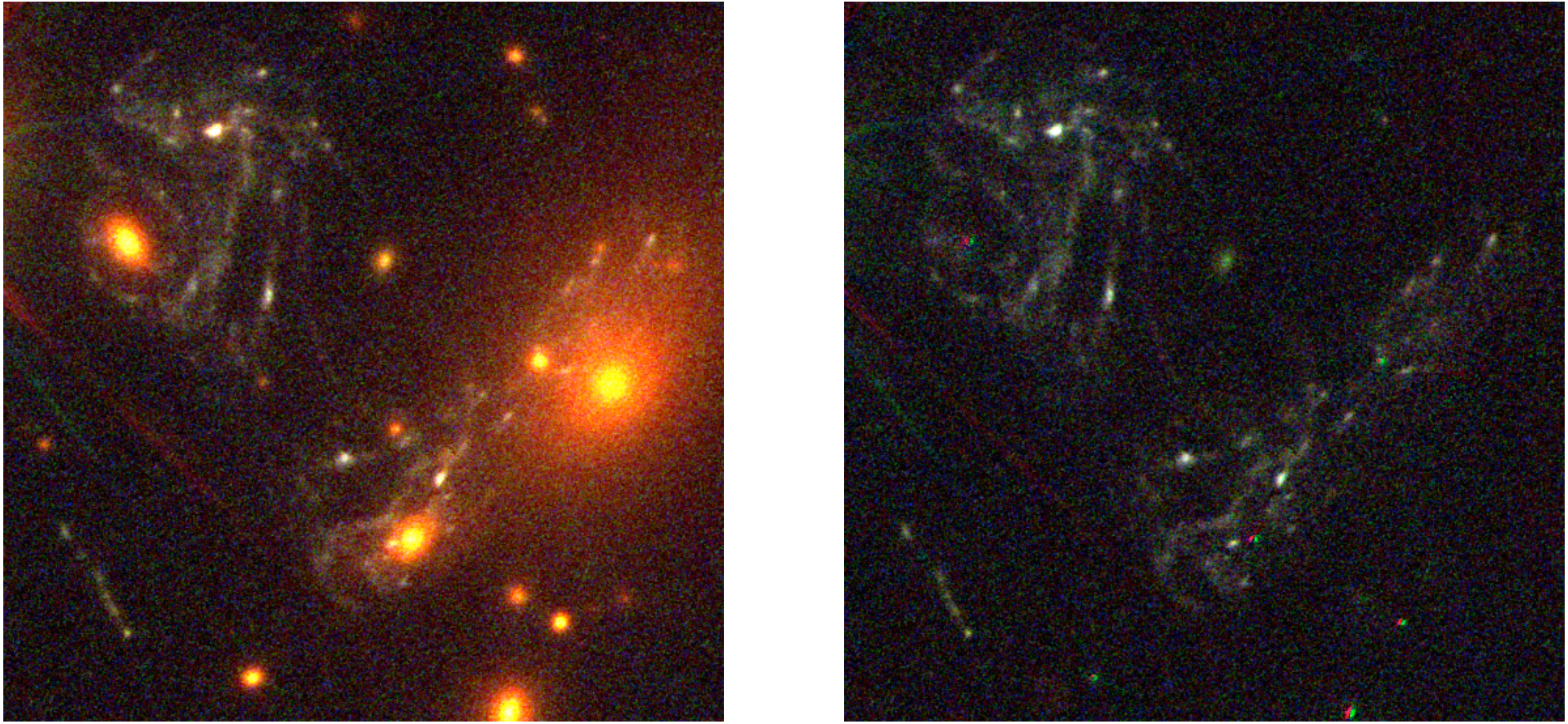
Gaussians



- 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.

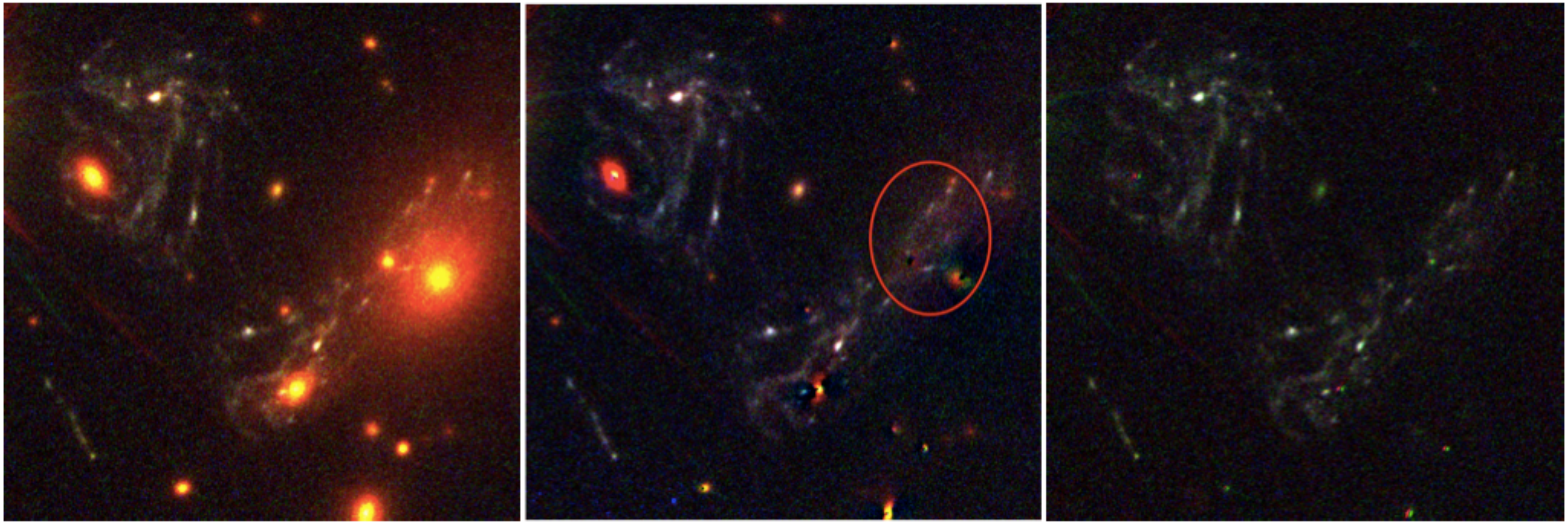
Multichannel data

galaxy cluster MACS~J1149+2223



Multichannel data

MACS~J1149+2223 cluster

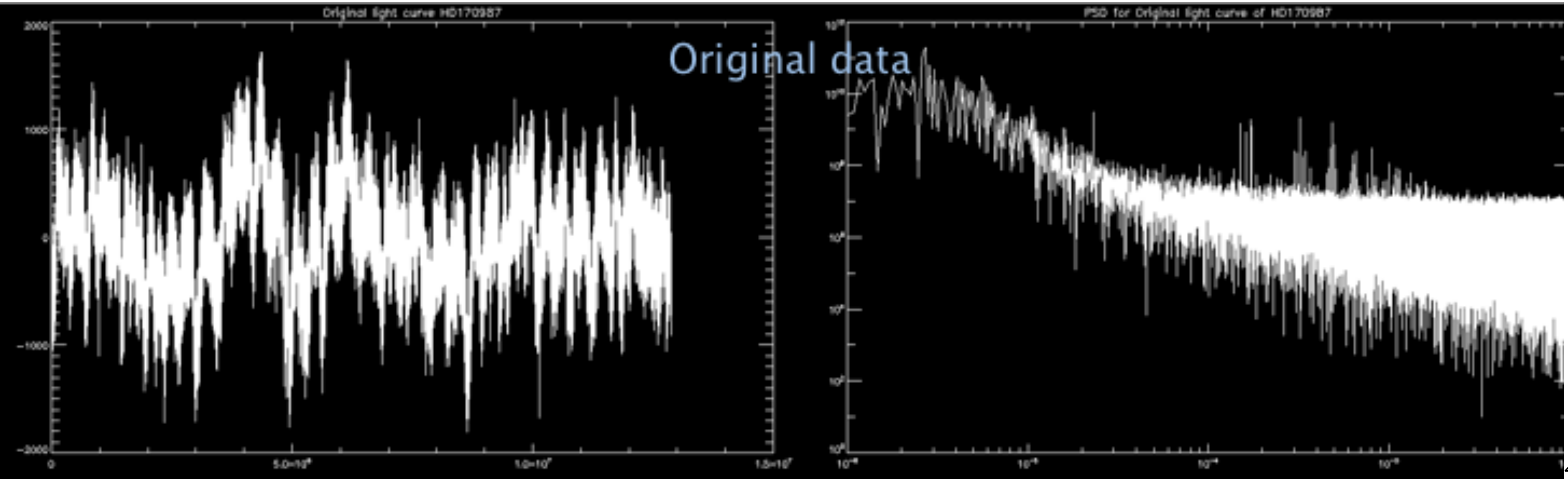


galfit subtraction of the galaxy members

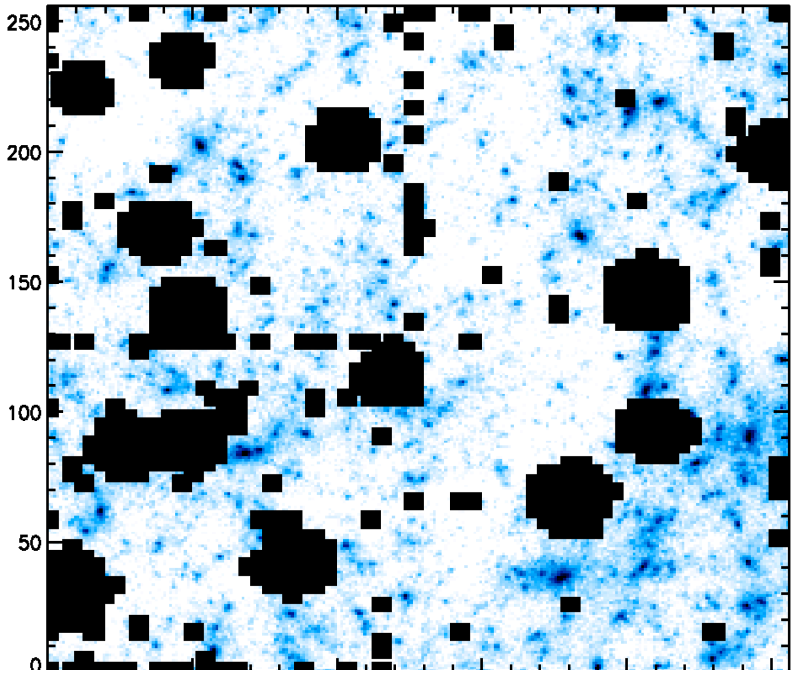
Missing Data

- Period detection in temporal series

COROT: HD170987



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



Interpolation of Missing Data

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.*

$$\mathbf{\Theta}_\Lambda = \mathbf{Id}_\Lambda$$

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

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

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

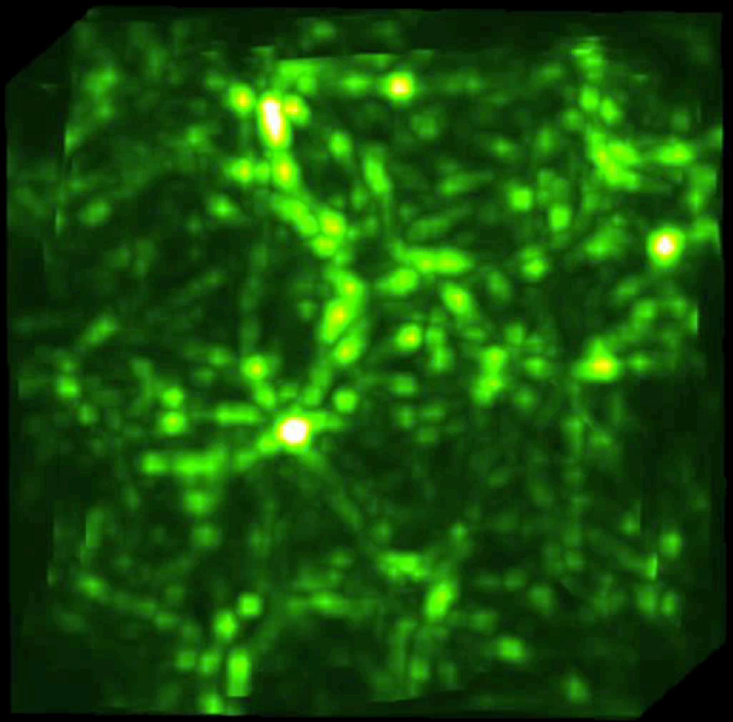
Iterative Hard Thresholding with a decreasing threshold.

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

20%

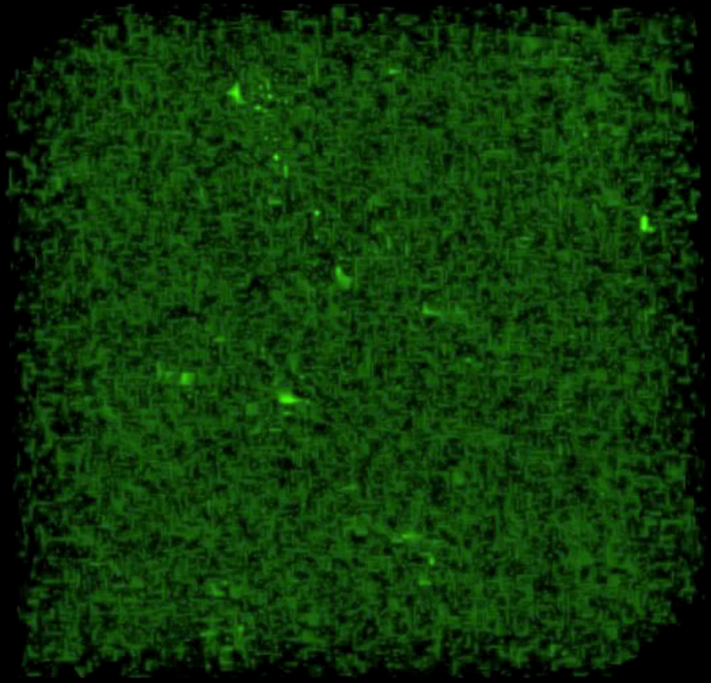
50%

80%

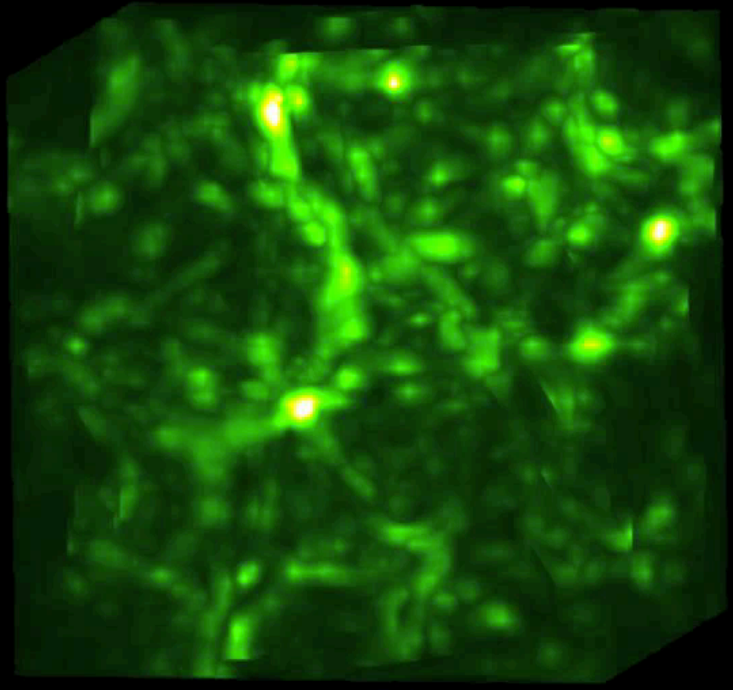


Original

Dictionary
BeamCurvelets



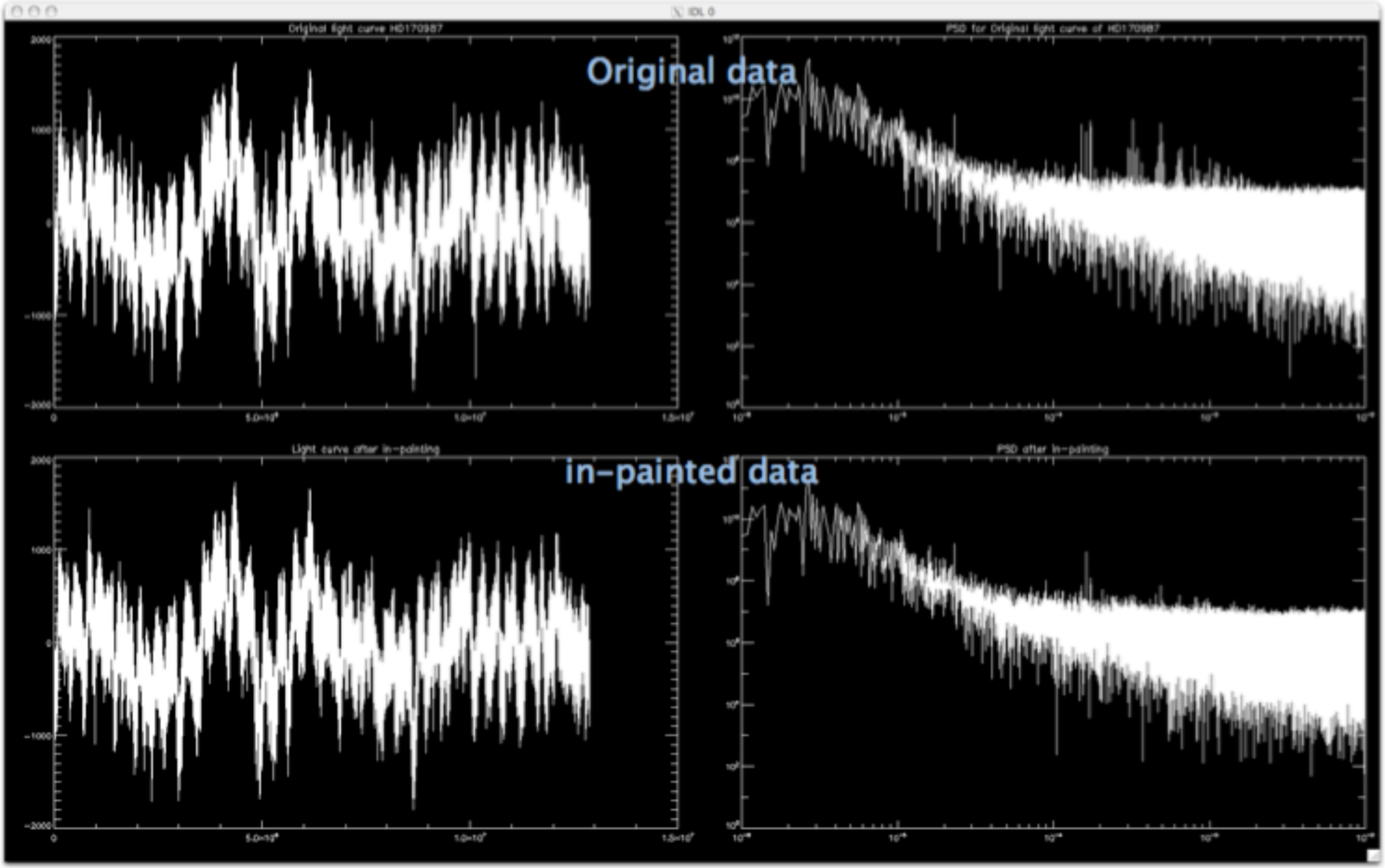
Mask



Inpainted

COROT: HD170987 with in-painting

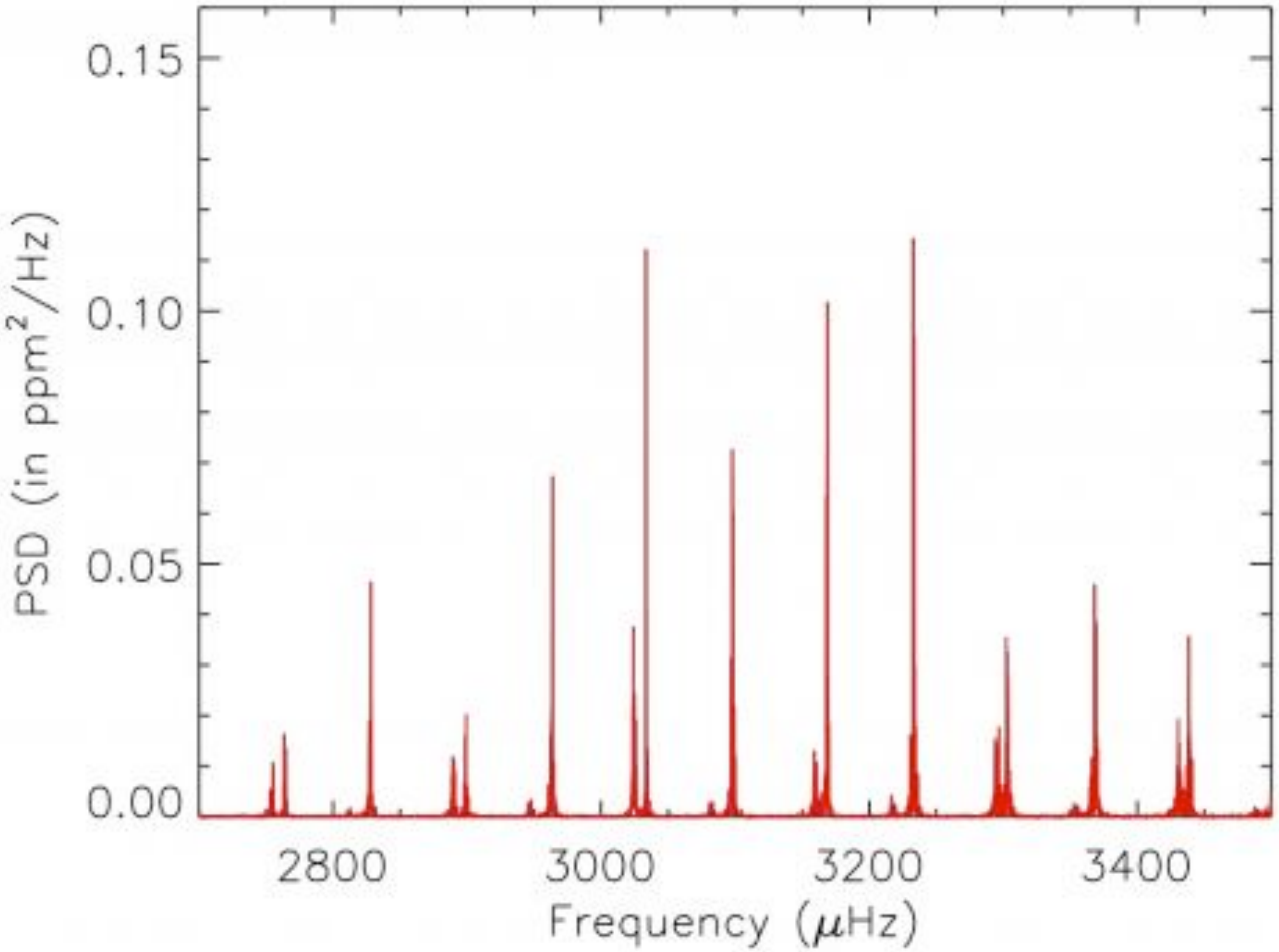
arXiv:1003.5178



Original data

in-painted data

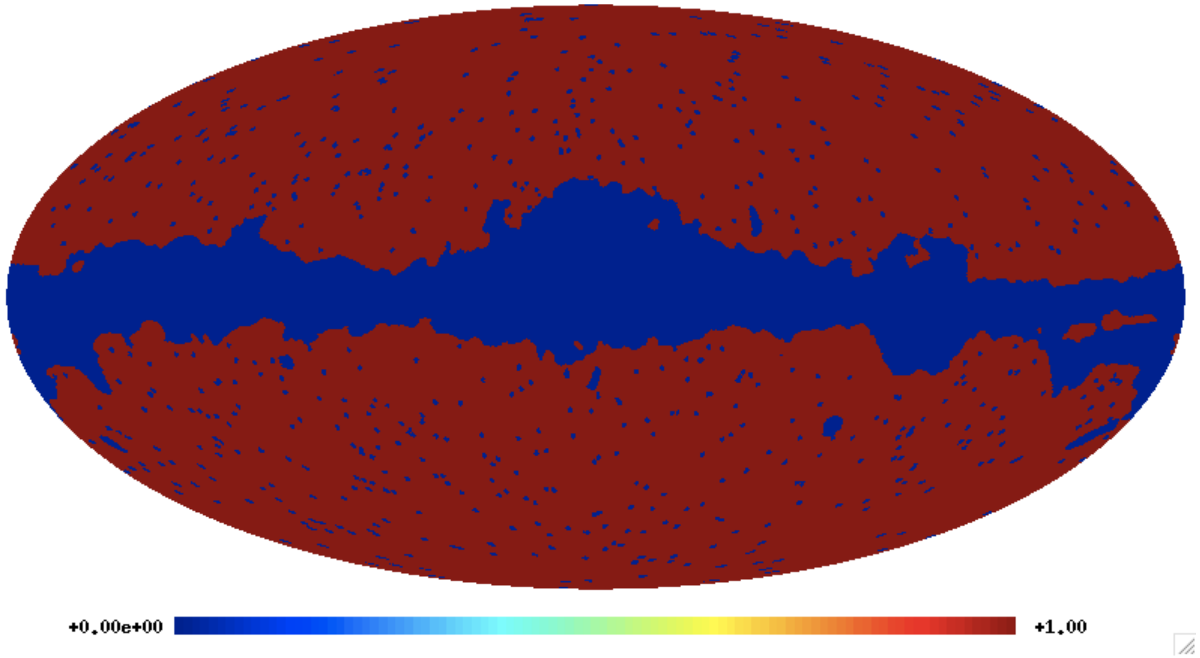
Sparse inpainting & asteroseismology

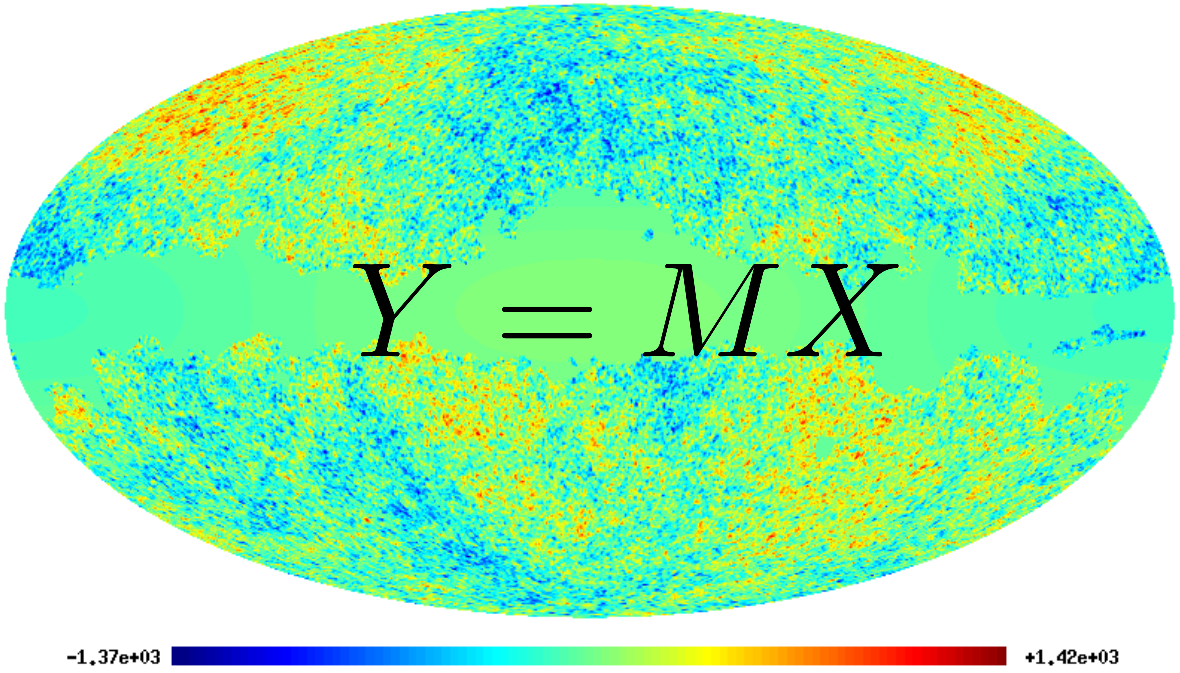


CoRoT: sparse inpainting is in the official pipeline.
Kepler: 18.000 stars have been processed.

Interpolation of Missing Data: Sparse Inpainting

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





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

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

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

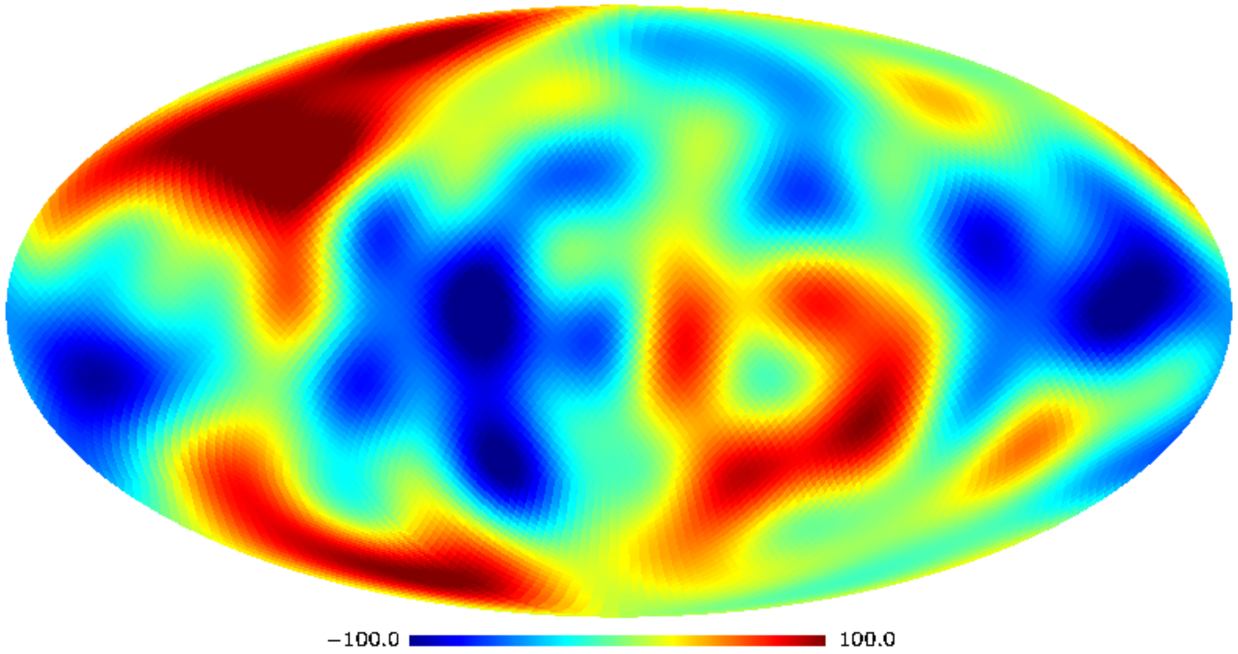
J.-L. Starck, A. Rassat, and M.J. Fadili, "Low-l CMB Analysis and Inpainting", **Astronomy and Astrophysics** , 550, A15, 2013.

J.-L. Starck, D.L. Donoho, M.J. Fadili and A. Rassat, "Sparsity and the Bayesian Perspective", **Astronomy and Astrophysics** , 552, A133, 2013.

Large CMB Scale Analysis

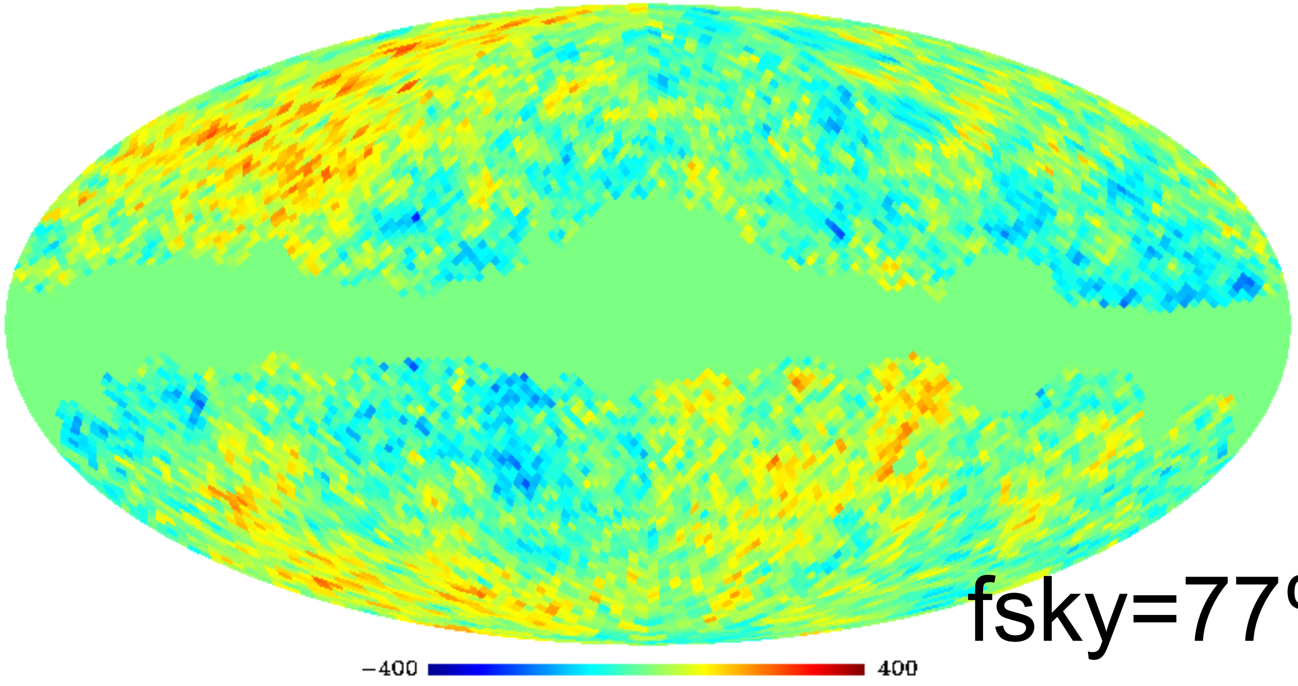
Simulated CMB (largest scale)

Simulated CMB (lmax=10)



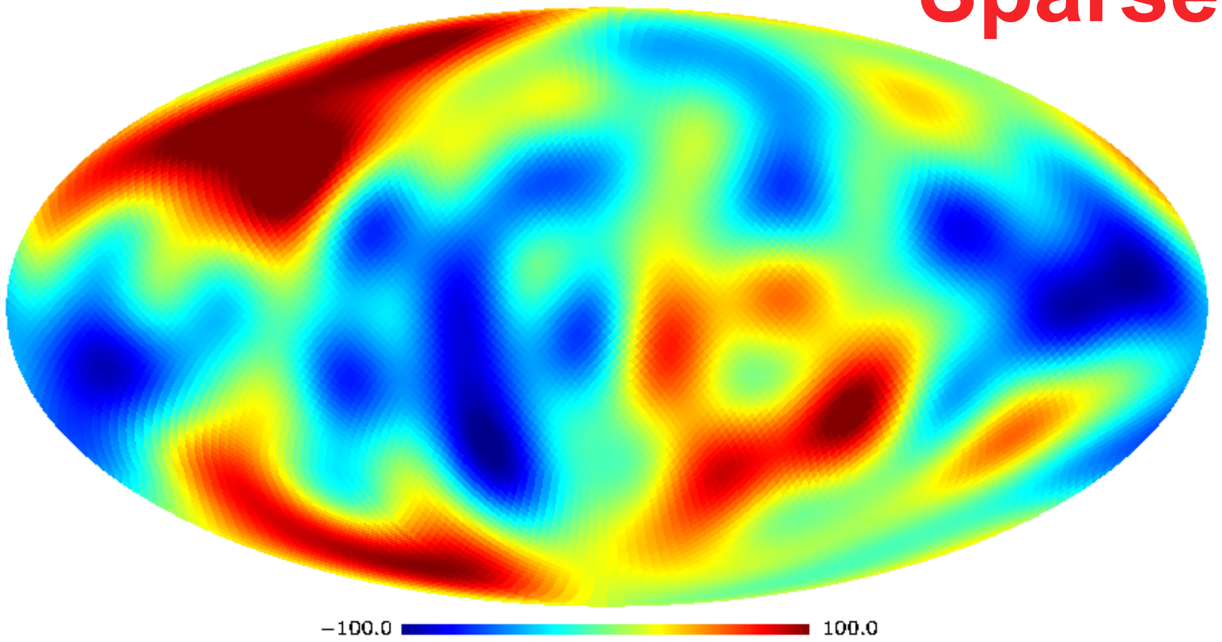
Masked Simulated Data (Fsky=77%)

Input Data



fsky=77%

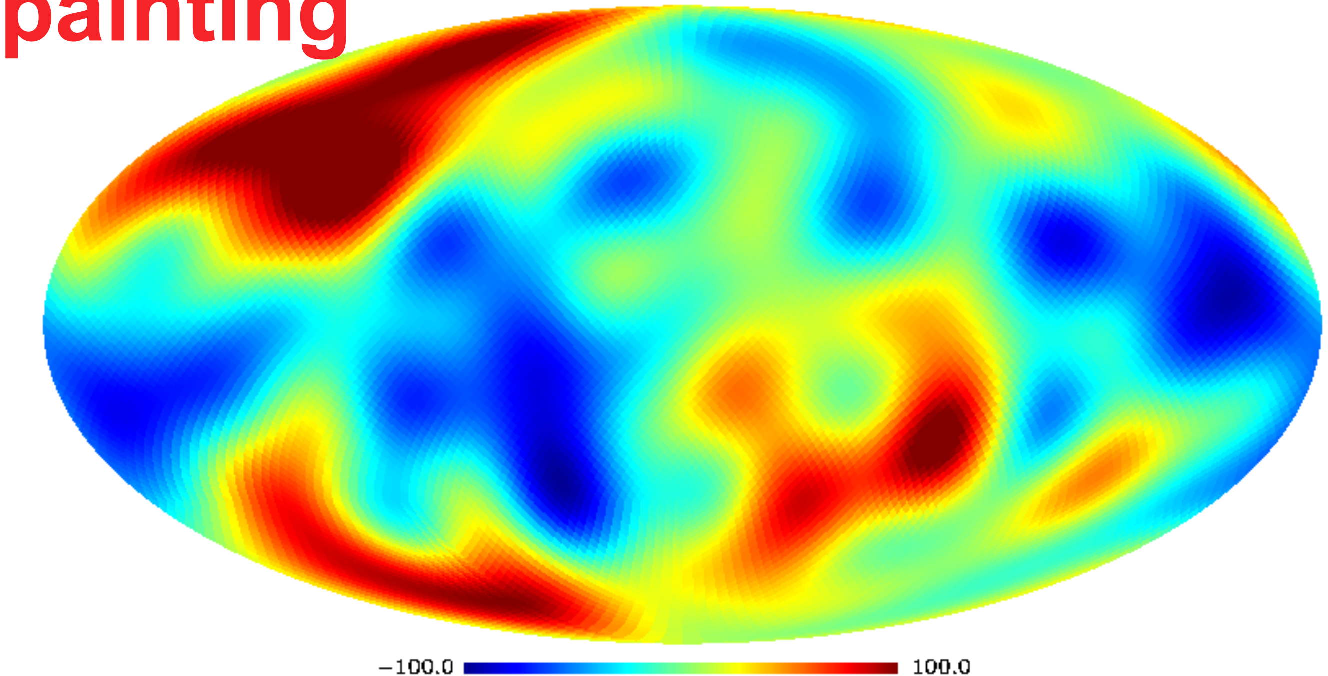
DR Sparse Constraint Inpainting: Mask Fsky = 87%



fsky=87%

Sparse Inpainting

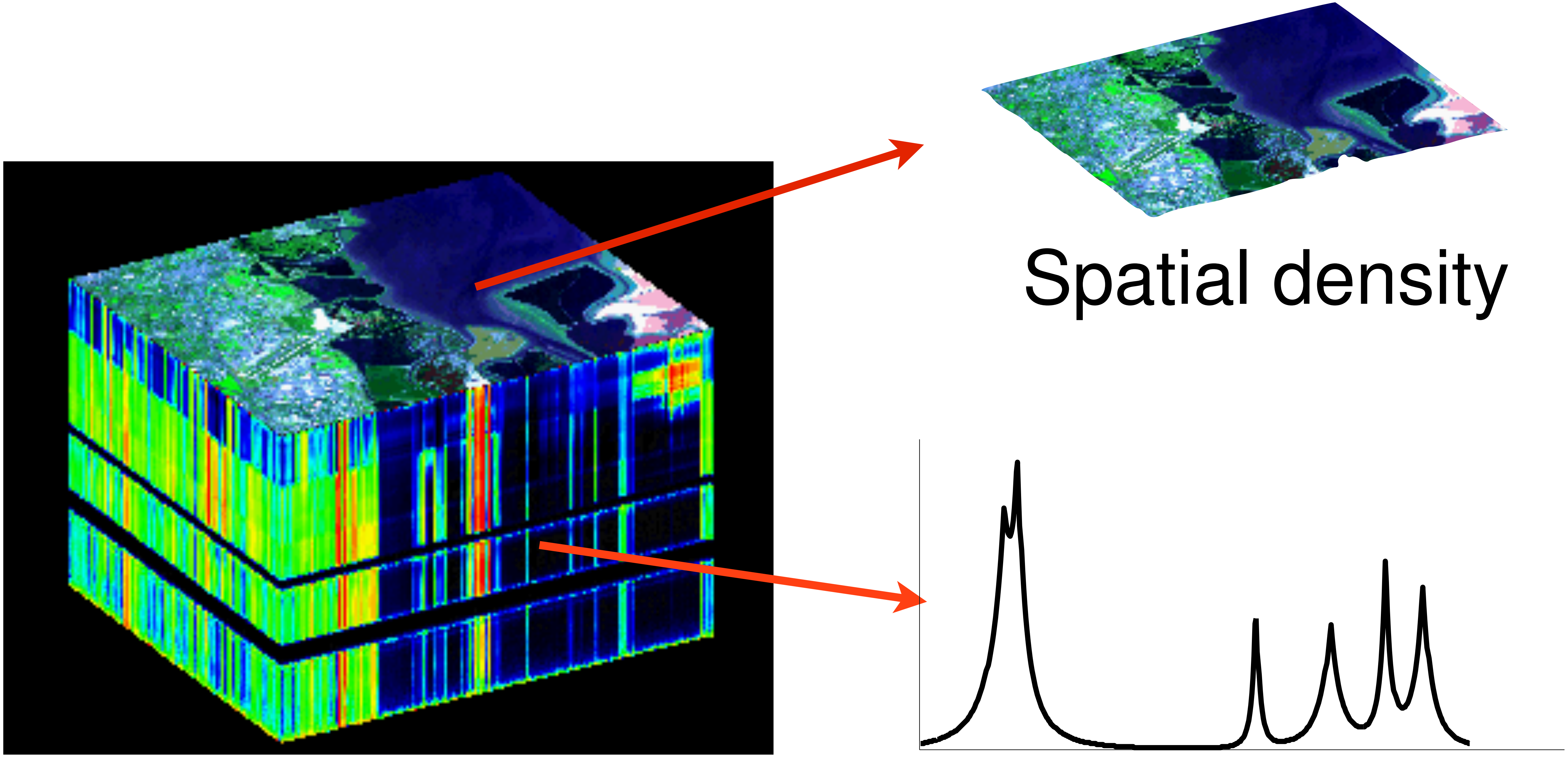
DR Sparse Constraint Inpainting: Mask Fsky = 77%



fsky=77%

J.-L. Starck, A. Rassat, and M.J. Fadili, "Low-l CMB Analysis and Inpainting", **Astronomy and Astrophysics**, 550, A15, 2013.

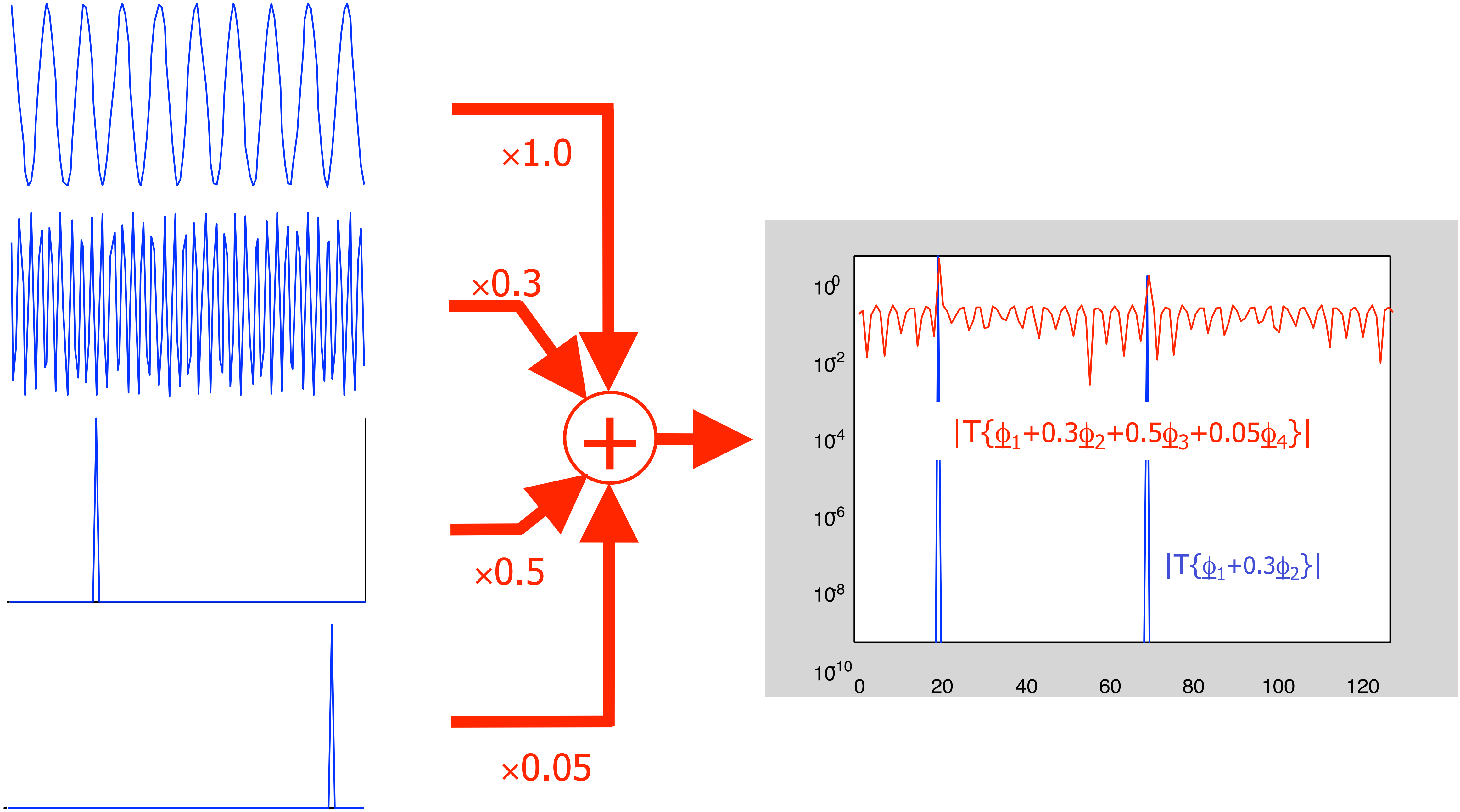
Blind Source Separation



Spatial density

Spectrum

Example of 1D mixture



×1.0

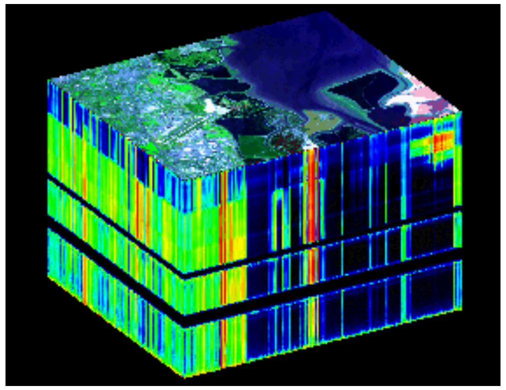
×0.3

×0.5

×0.05

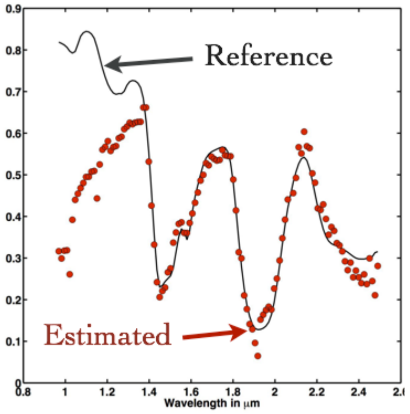
Data Modelization

- Hyperspectral data

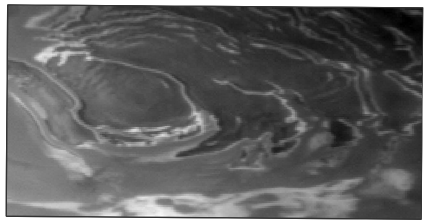


Hyperspectral data cube of Mars Express

=

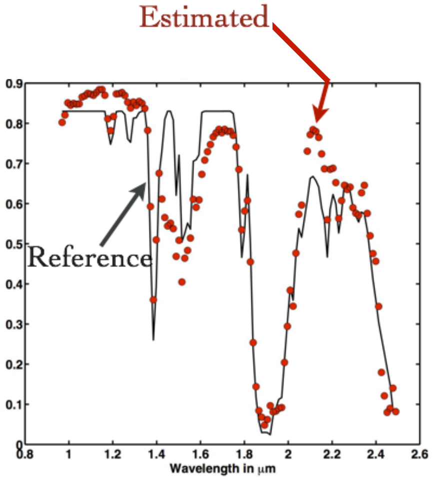


H₂O ice spectrum

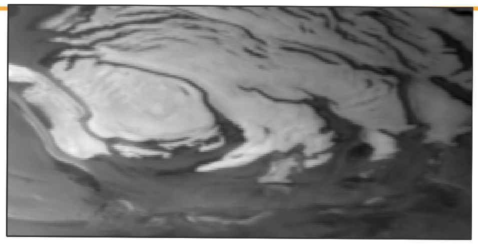


H_2O ice concentration

+



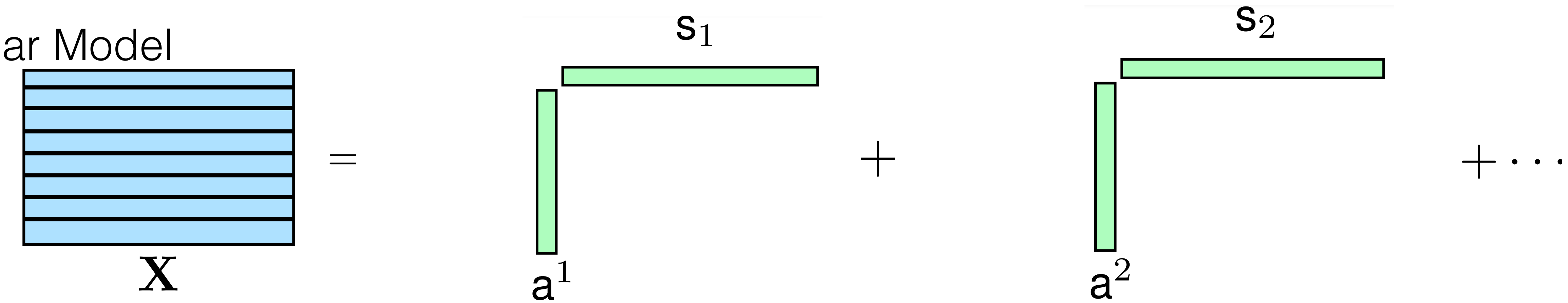
CO_2 ice spectrum



CO_2 ice concentration

+ …

- Linear Model



- Matrix formulation

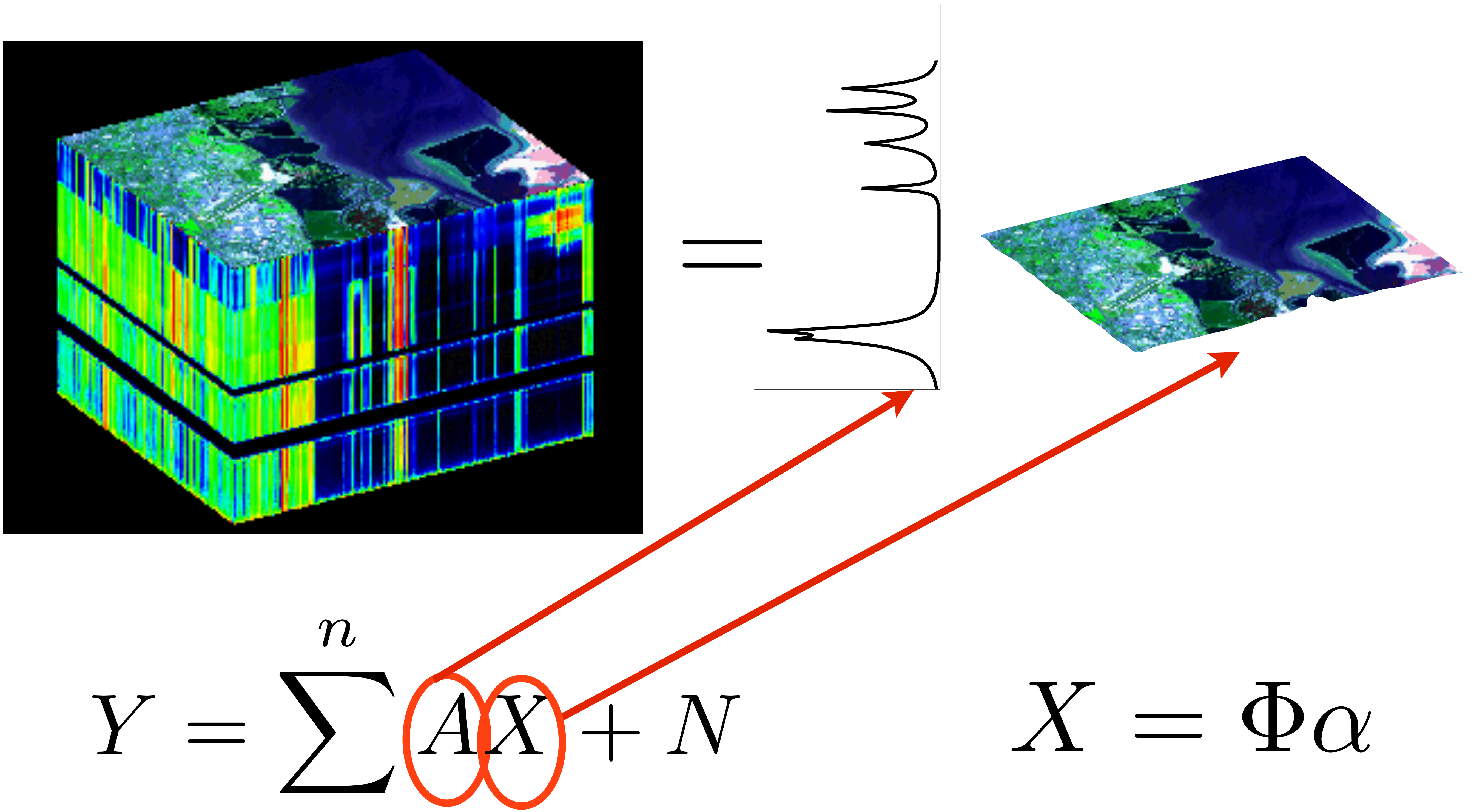
assumed to be known

sources

$$\mathbf{X} = \sum_{k=1}^{n} a^k s_k = \mathbf{AS}$$

mixing matrix

Data Modelization

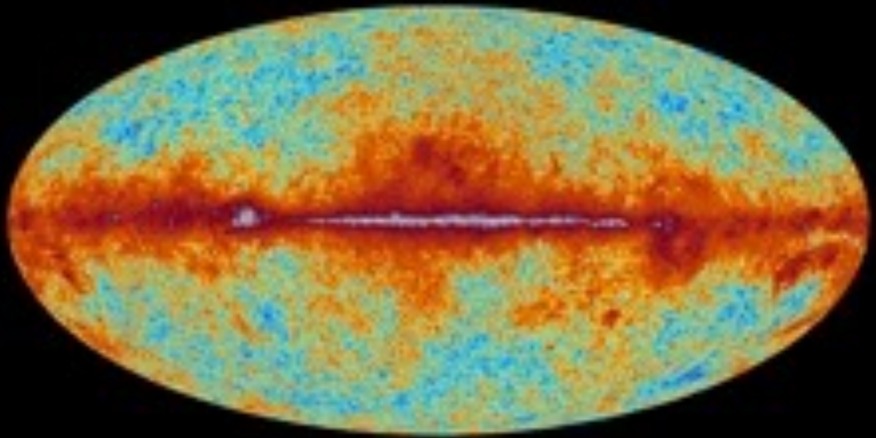


$$Y = \sum_{k=1}^{n} AX + N$$

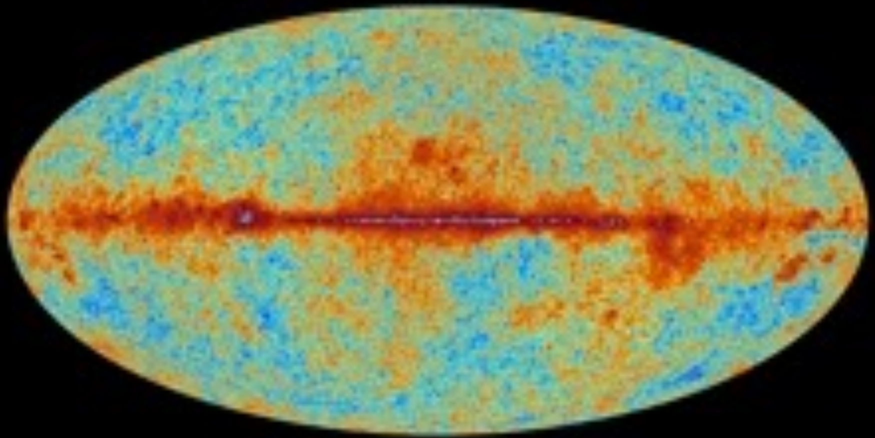
$$X = \Phi\alpha$$

planck

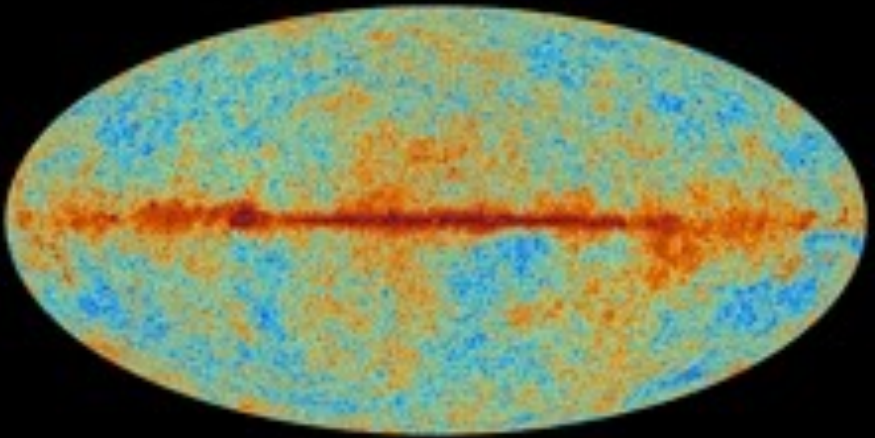
The sky as seen by Planck



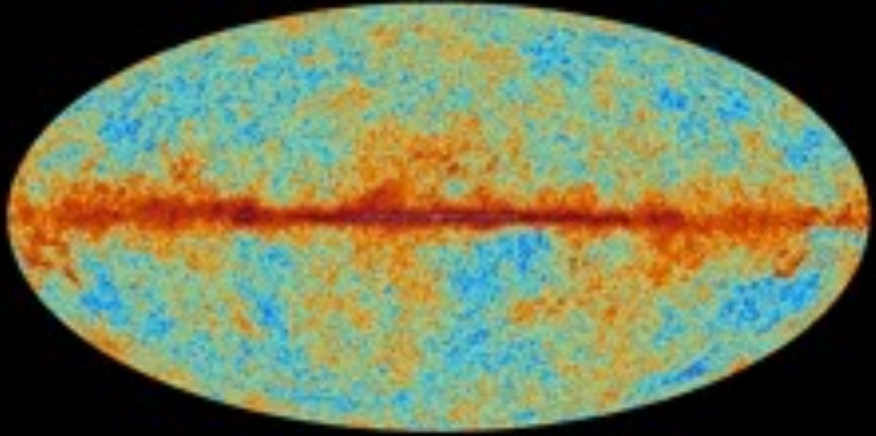
30 GHz



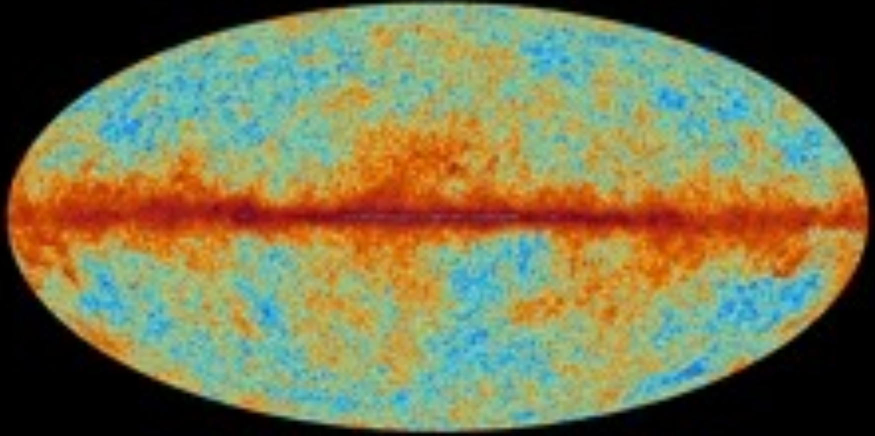
44 GHz



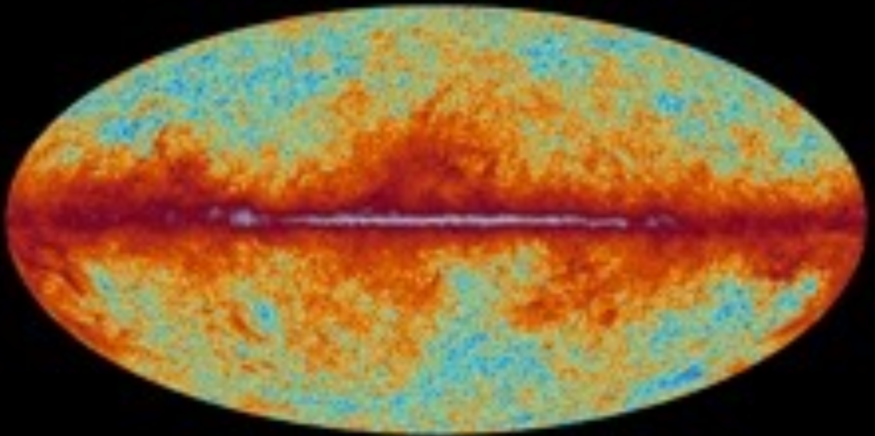
70 GHz



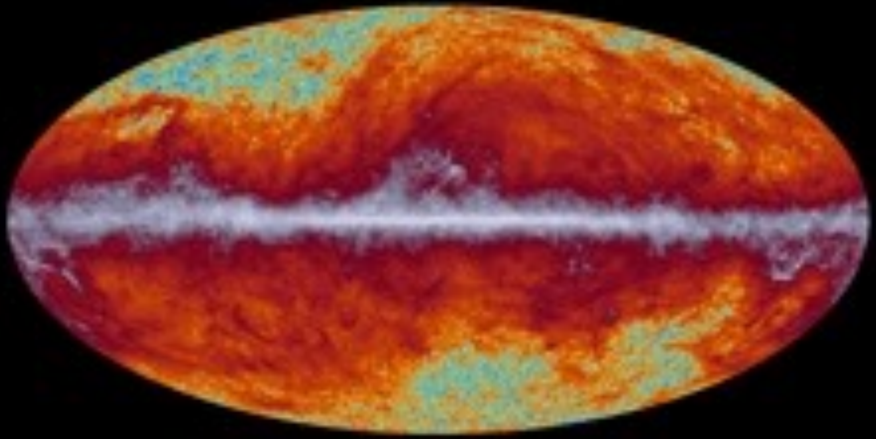
100 GHz



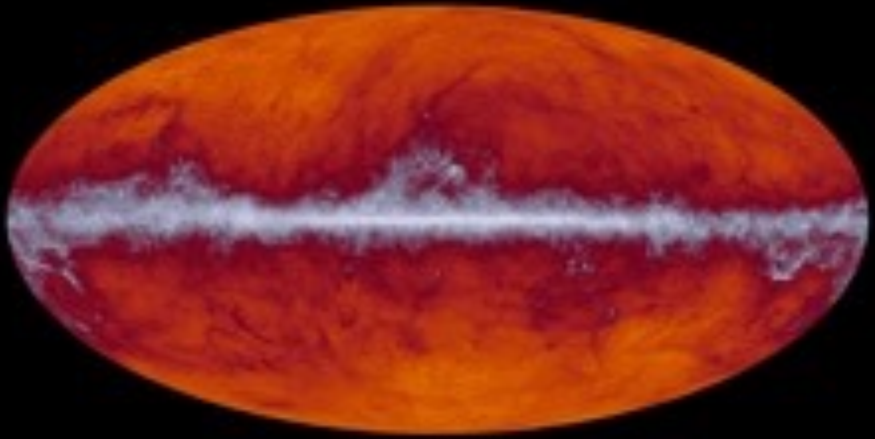
143 GHz



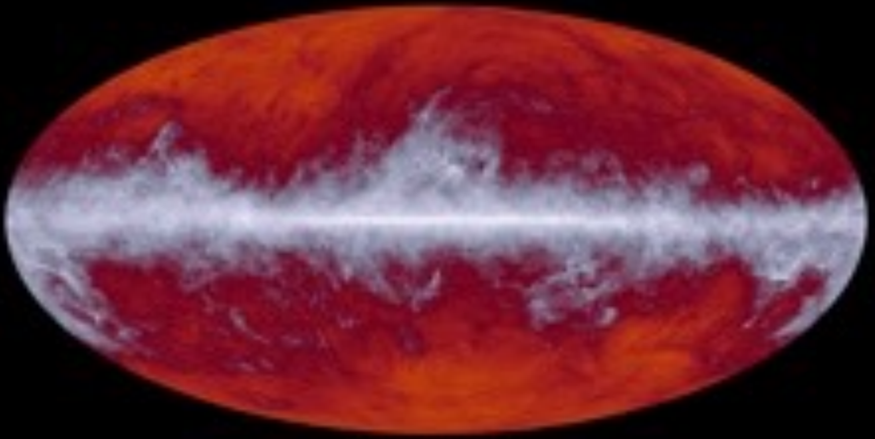
217 GHz



353 GHz

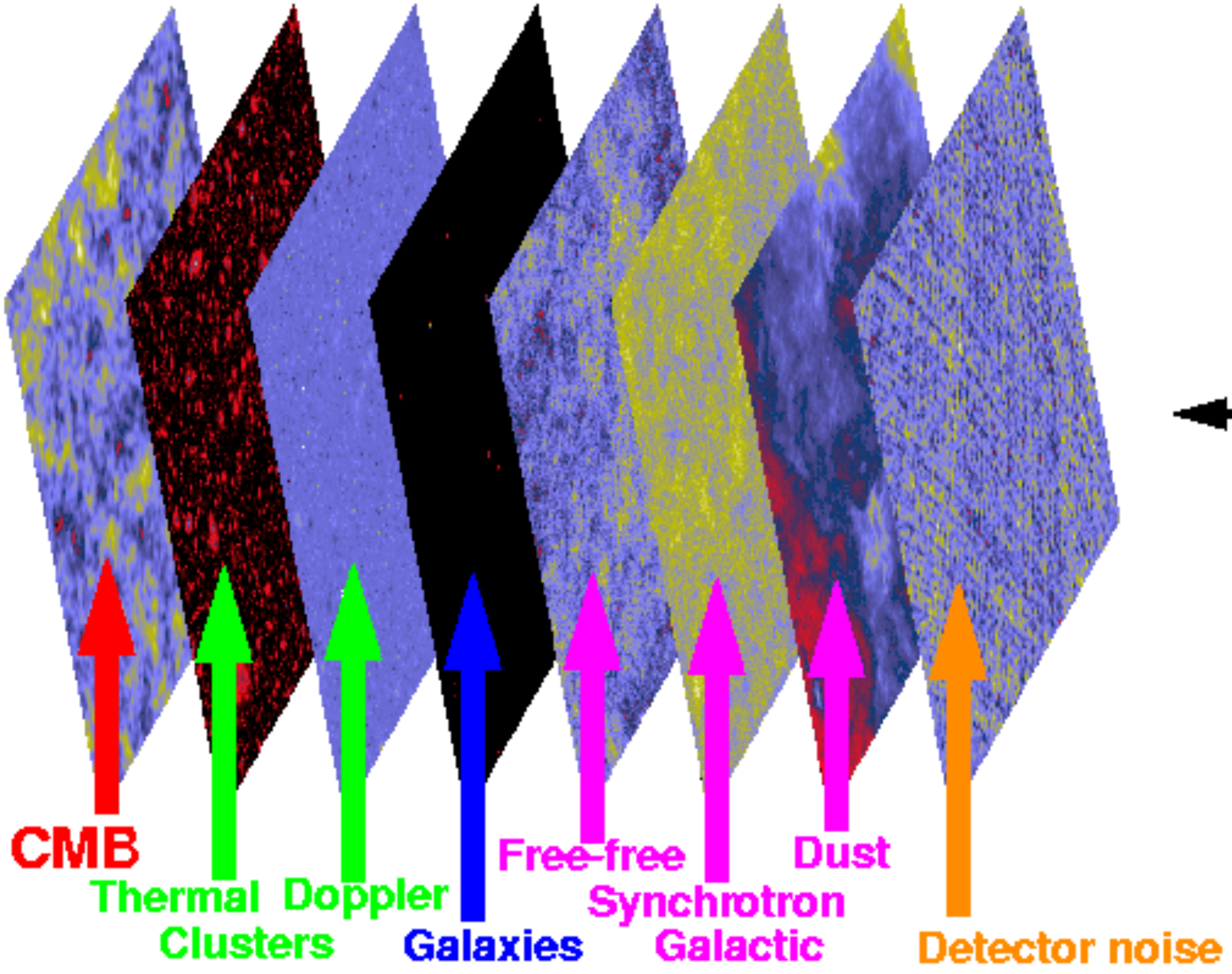


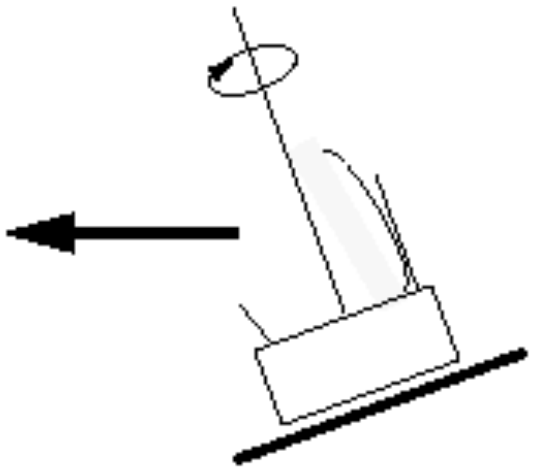
545 GHz

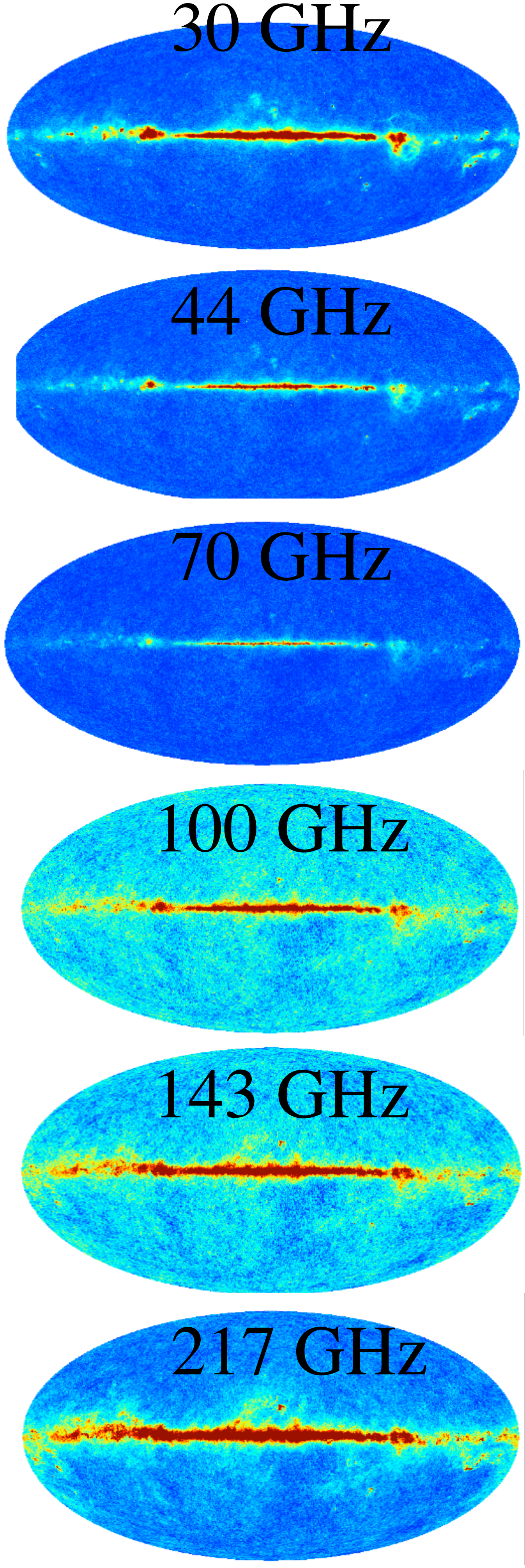


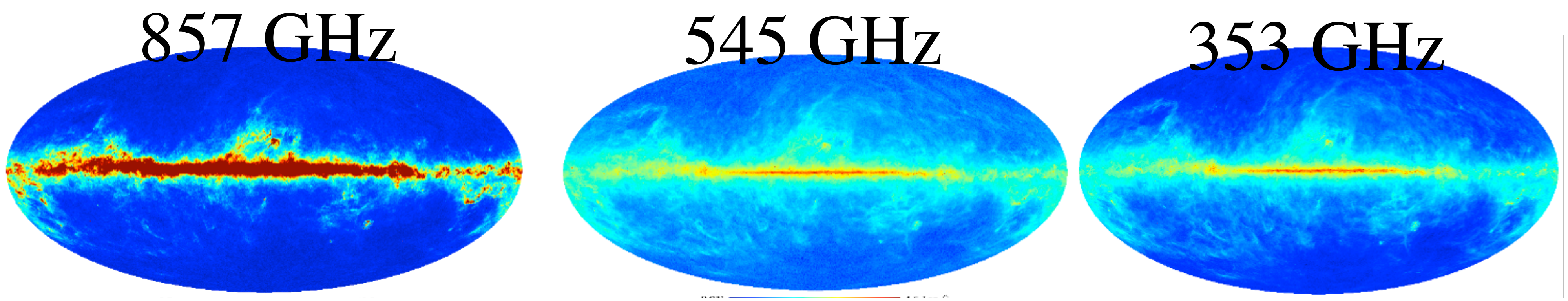
857 GHz

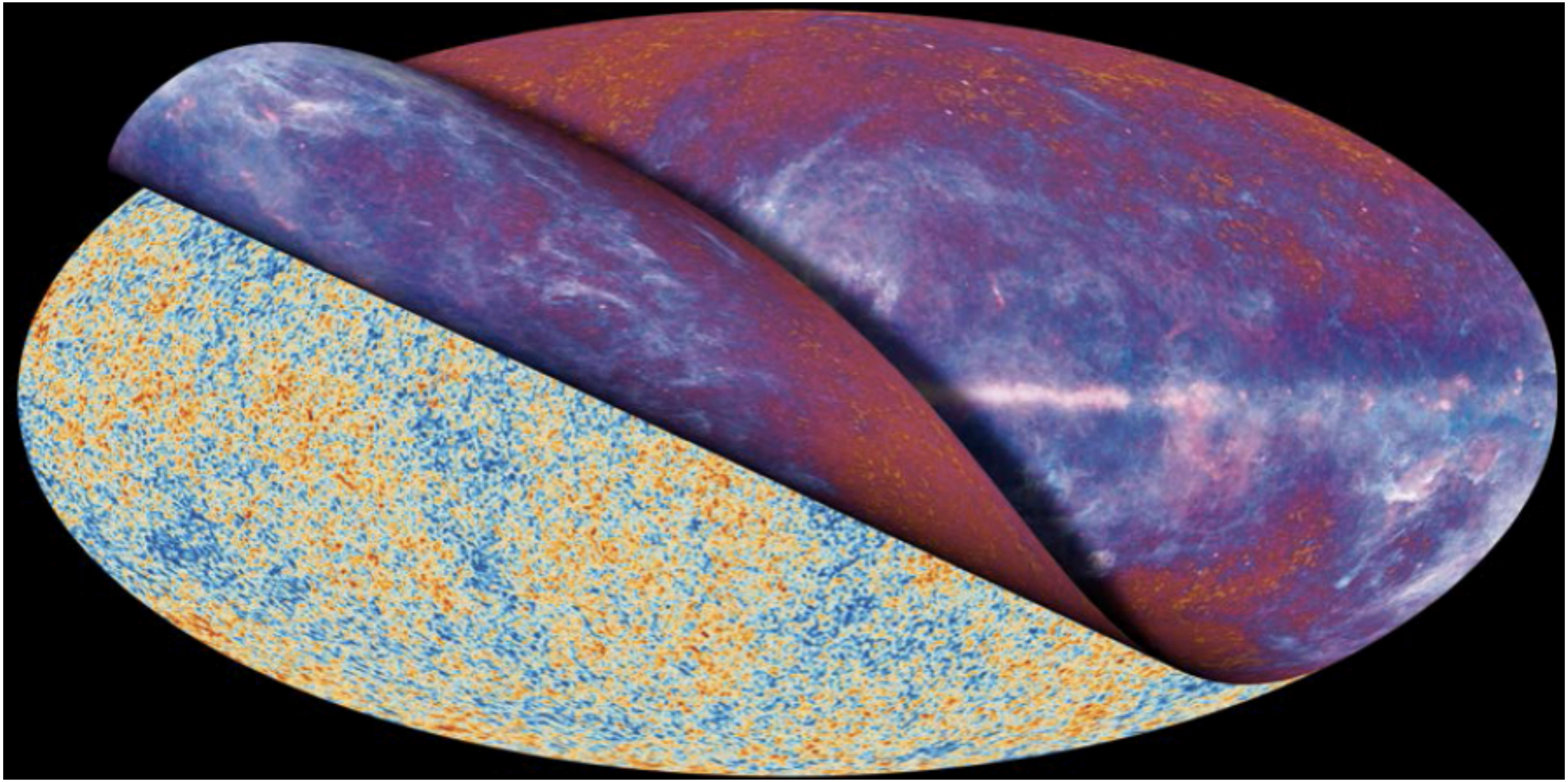
Planck Component Separation











Sparse Component Separation: the GMCA Method

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

•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.

$$\{\alpha, A\} = \text{Argmin}_{\alpha, A} \|\mathbf{Y} - \mathbf{A}\alpha\mathbf{\Phi}\|^2_{F,\mathbf{\Sigma}} + \sum_j \lambda_j \|\alpha_j\|_1$$

1) Estimate X assuming A is fixed :

$$\{\alpha\} = \text{Argmin}_{\alpha} \ \|\mathbf{Y} - \mathbf{A}\alpha\mathbf{\Phi}\|^2_{F,\mathbf{\Sigma}} + \sum_j \lambda_j \|x_j \mathbf{W}\|_1$$

=> Sparse coding (proximal theory, etc)

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

$$\{A\} = \text{Argmin}_A \|\mathbf{Y} - \mathbf{A}\alpha\mathbf{\Phi}\|^2_{F,\mathbf{\Sigma}}$$

=> Least square estimator

BSS experiment : Noiseless case

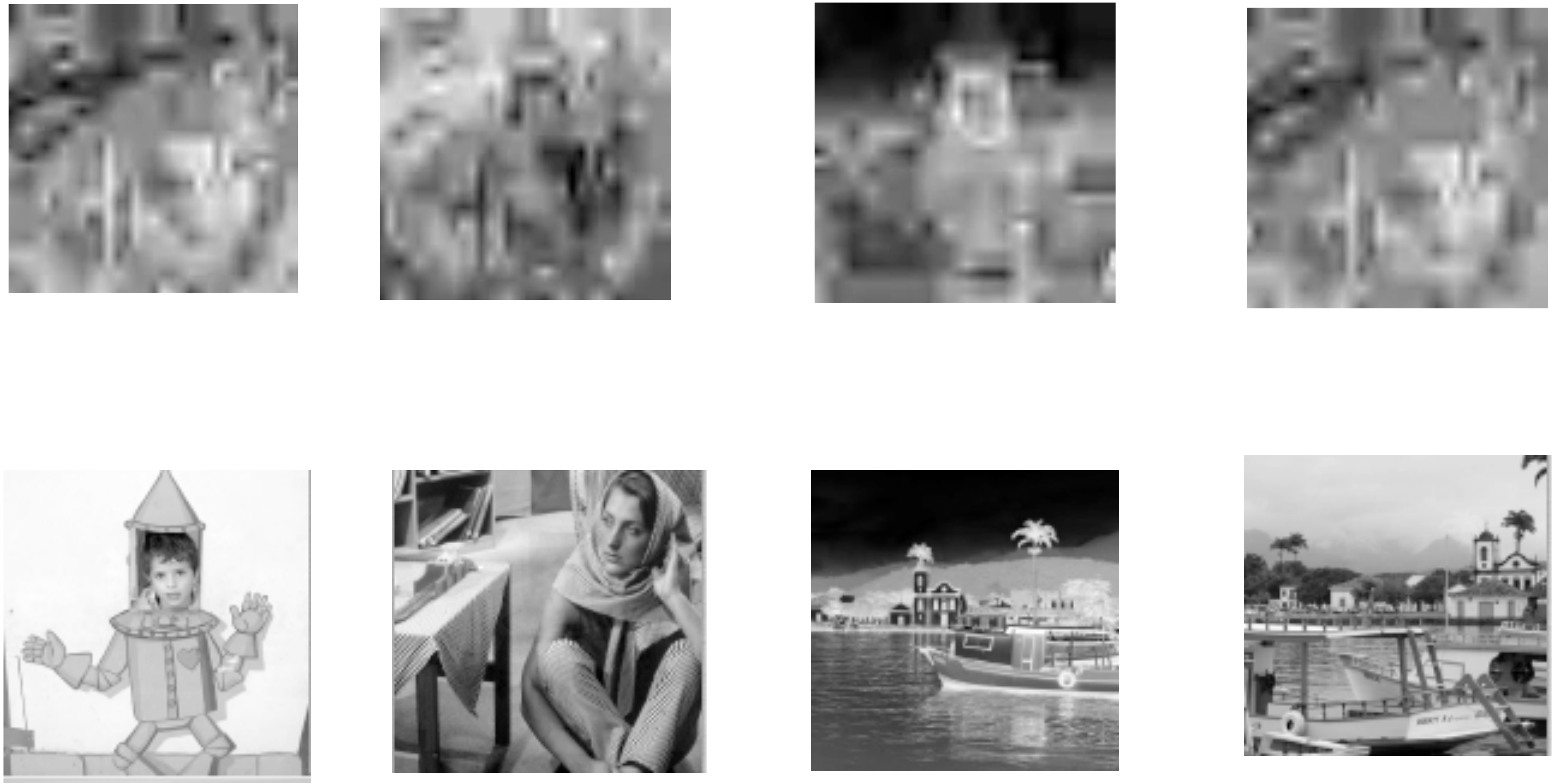
Original
Sources

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.



GMCA Limitations for PLANCK

GMCA Model: $Y = A X + N$

But three main problems:
i) A is spatially variant.
ii) This model does not take into account the beam.
iii) Noise is not homogeneous.

- Limitation of GMCA:

- One matrix to describe the whole sky (i.e. the simplest model !)
- PSF were not taken into account properly

Component Separation: more problems

1) The beam: $\forall i;\ y_i = \boxed{b_i} \star \left(\sum_j a_{ij} x_j \right) + n_i$

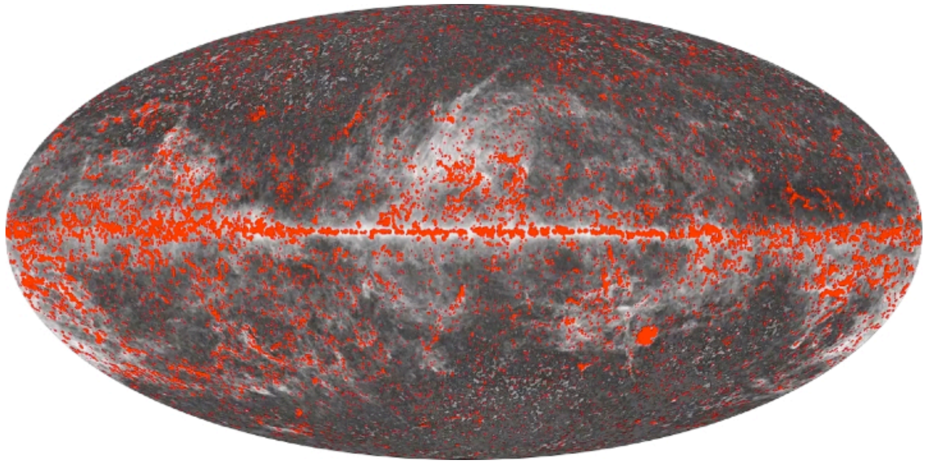
Globally: $\mathbf{Y} = \mathcal{H}(\mathbf{AX}) + \mathbf{N}$ $\mathcal{H}$ **is singular !**

where $\mathcal{H}$ is the multichannel convolution operator

2) Spectral behavior **varies spatially** for some components (dust, synchroton).

$$\mathbf{Y}[k] = \mathcal{H}(\mathbf{A_k X})[k] + \mathbf{N}[k]$$

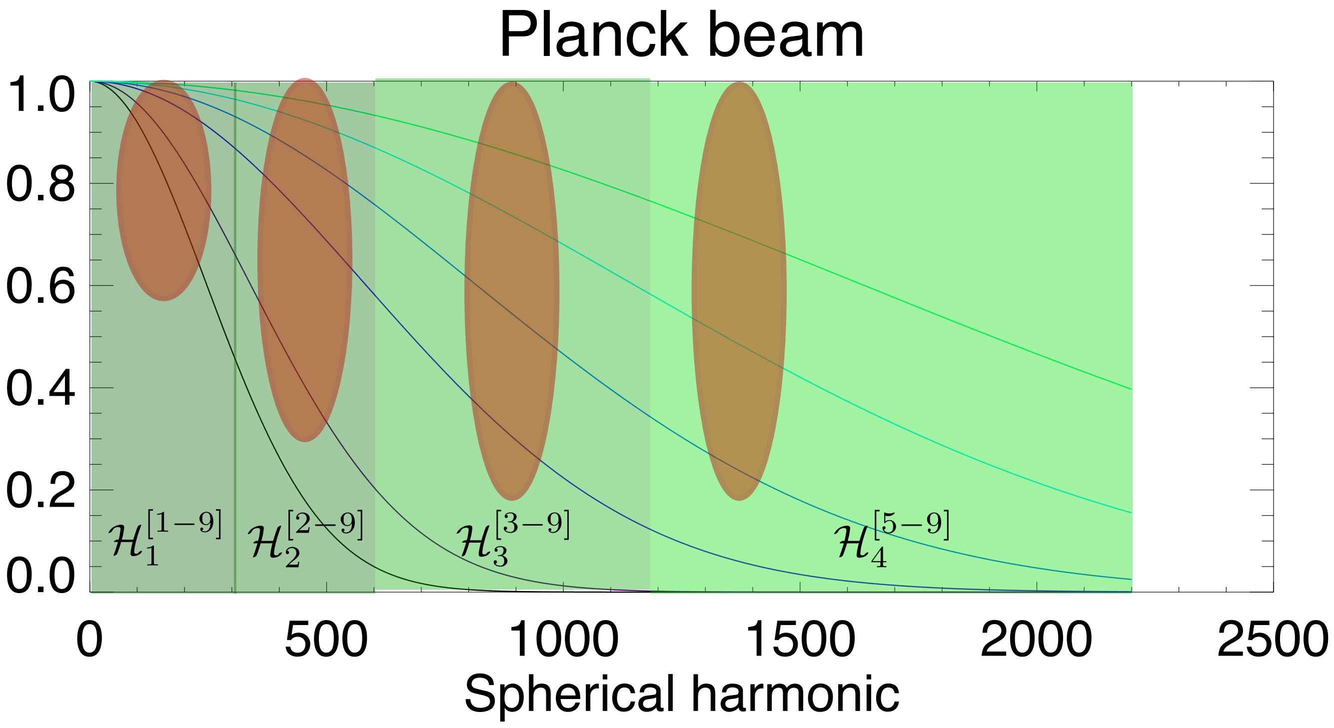
3) Point sources:

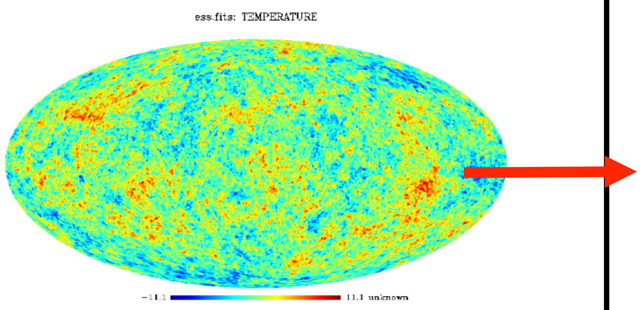


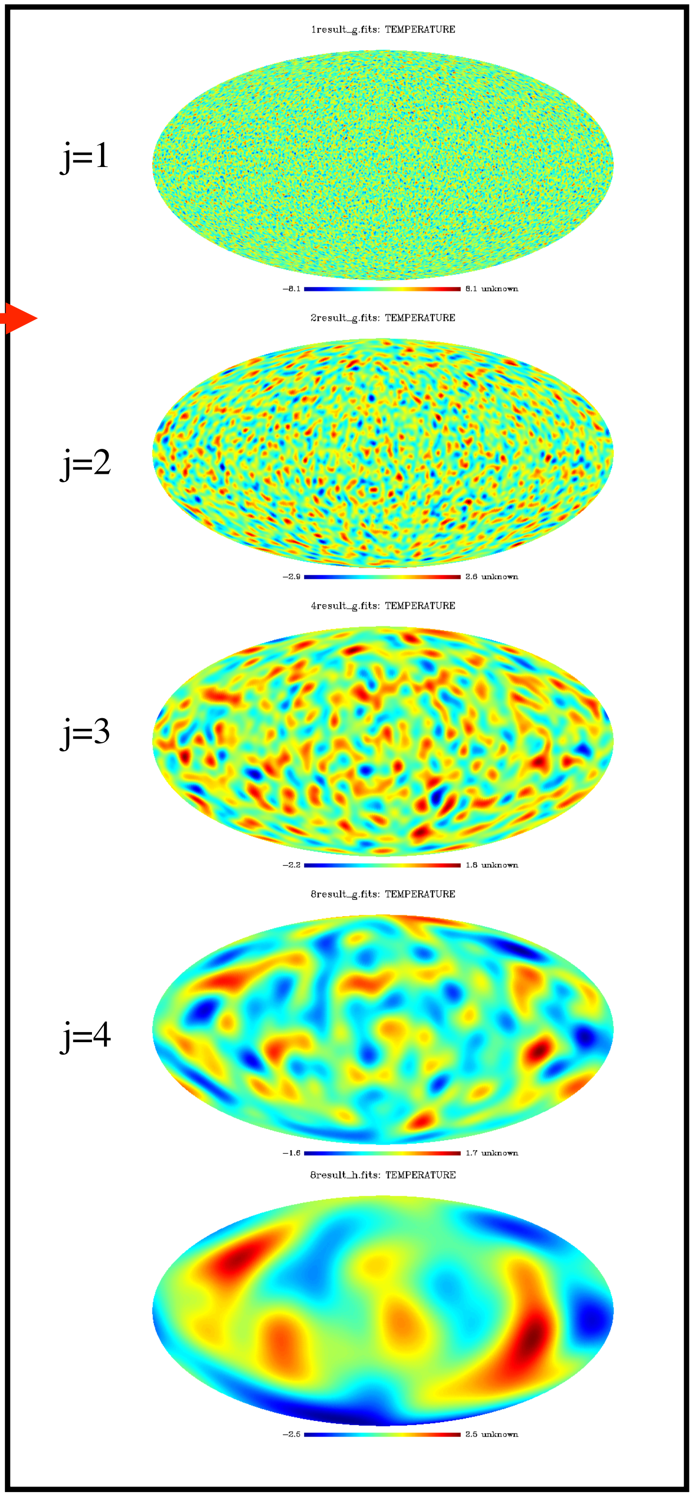
Component Separation

=> Use Wavelets to work at different resolutions:

Undecimated Wavelet Transform

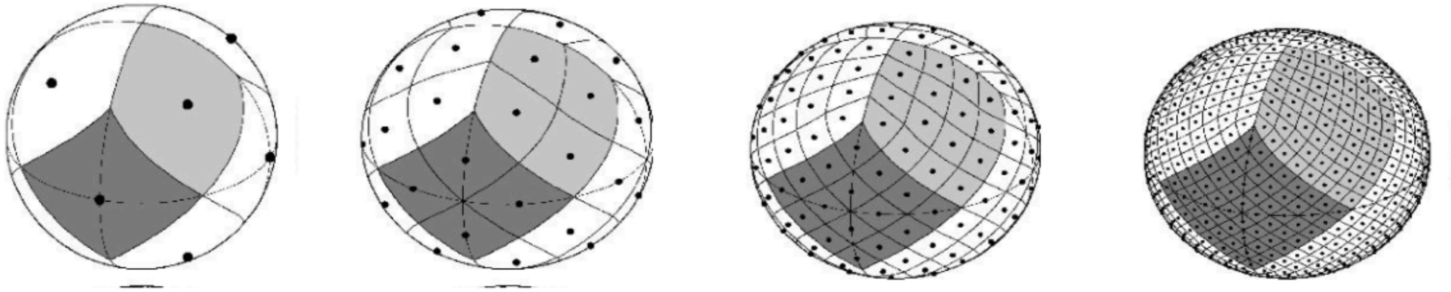






=> Assume the mixing matrix varies smoothly

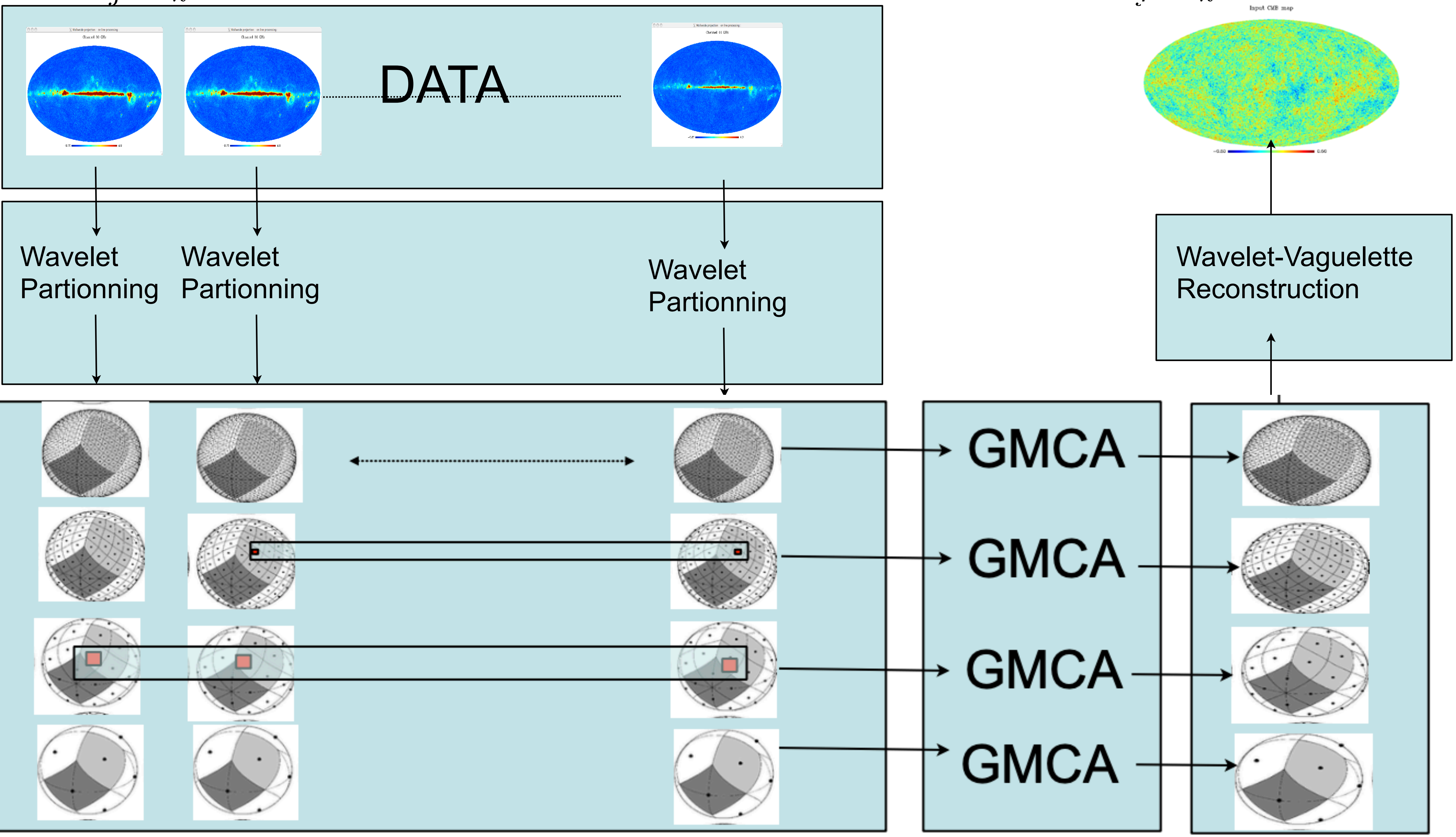
Partitionning of the Wavelet Scales



Local GMCA (LGMCA)

$$f = \sum_j \sum_k \langle Kf, \Psi_{j,k} \rangle \, \psi_j, k \qquad \text{with } K^*\Psi_{j,k} = \psi_{j,k}$$

$$\tilde{f} = \sum_j \sum_k \Delta(\langle y, \Psi_{j,k} \rangle)\psi_j, k$$



DATA

Wavelet Partionning

Wavelet Partionning

Wavelet Partionning

GMCA

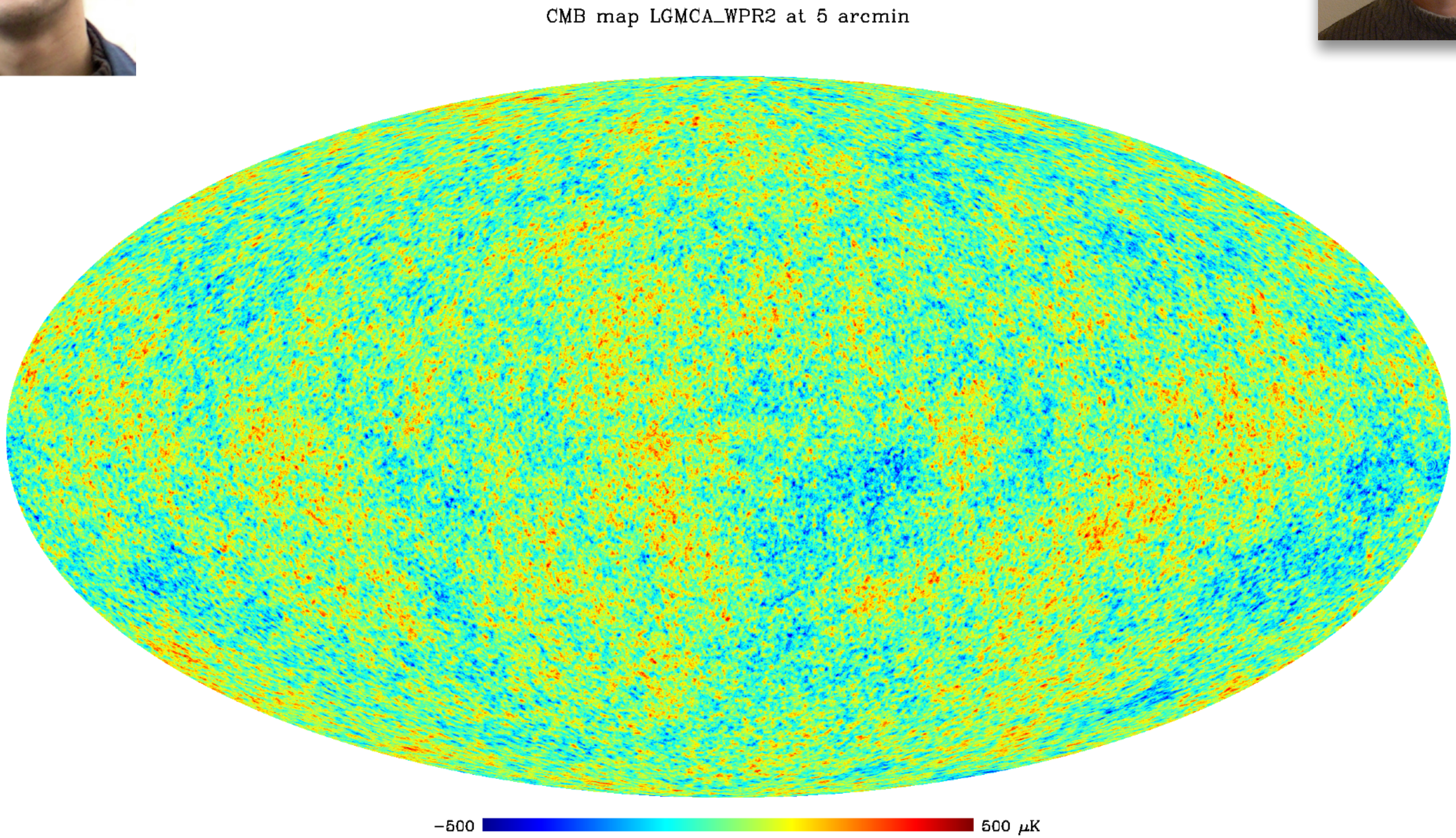
GMCA

GMCA

GMCA

Wavelet-Vaguelette Reconstruction

Full Sky Sparse WMAP + Planck-PR2 Map

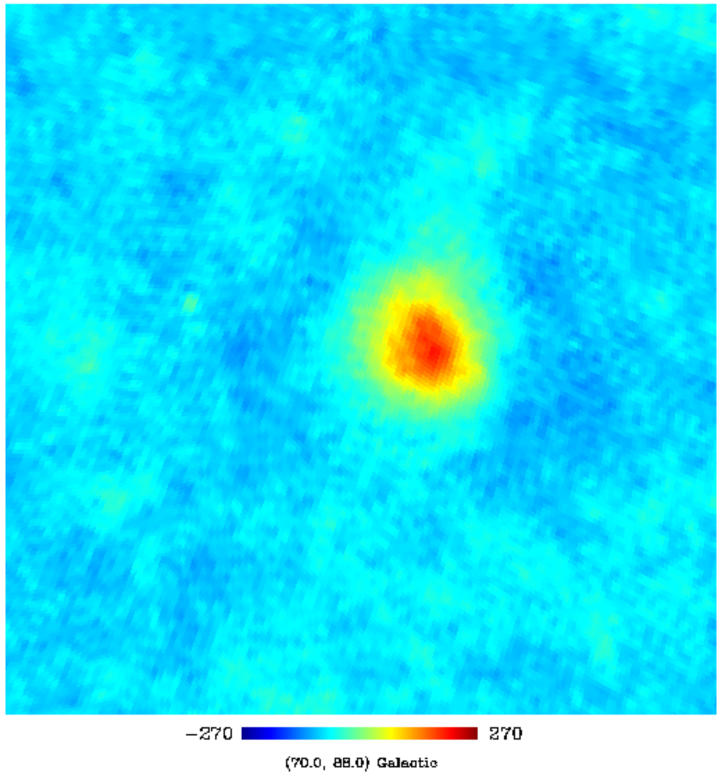


Bobin J., Sureau F., Starck J-L, Rassat A. and Paykari P., Joint Planck and WMAP CMB map reconstruction, A&A, 563, 2014

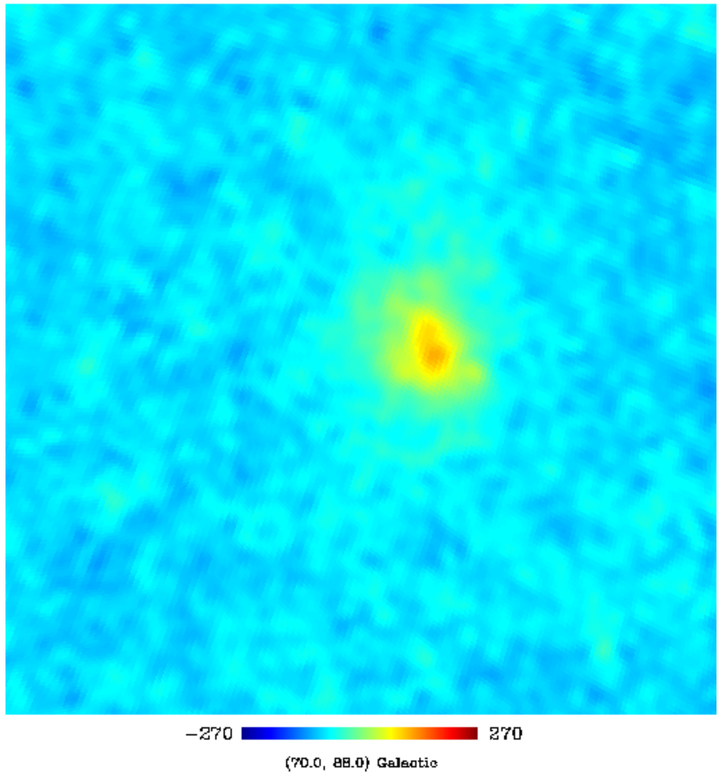
Bobin J., Sureau F., Starck, CMB reconstruction from the WMAP and Planck PR2 data, submitted to A&A, 2015

Traces of tSZ effect

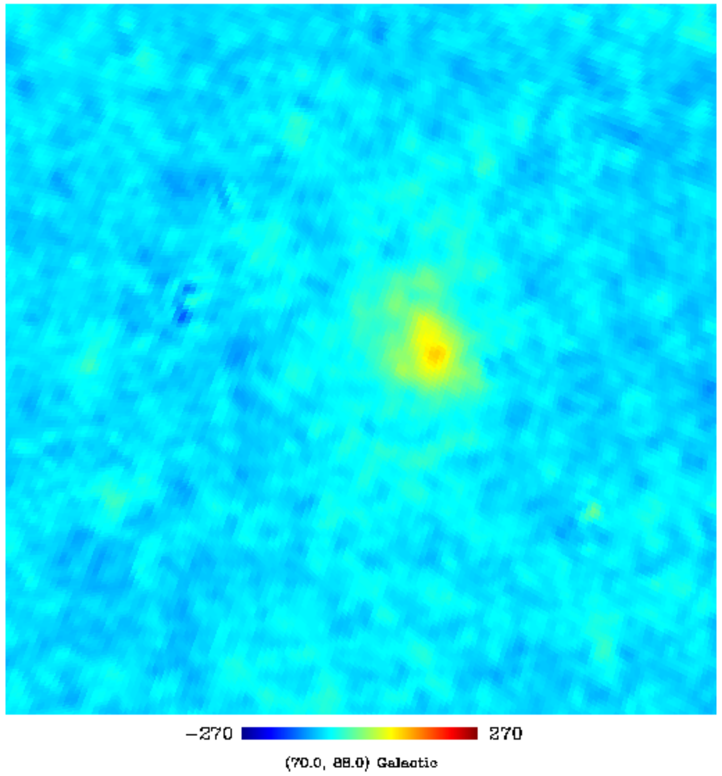
Coma: 217GHz PR2–HFI – NILC



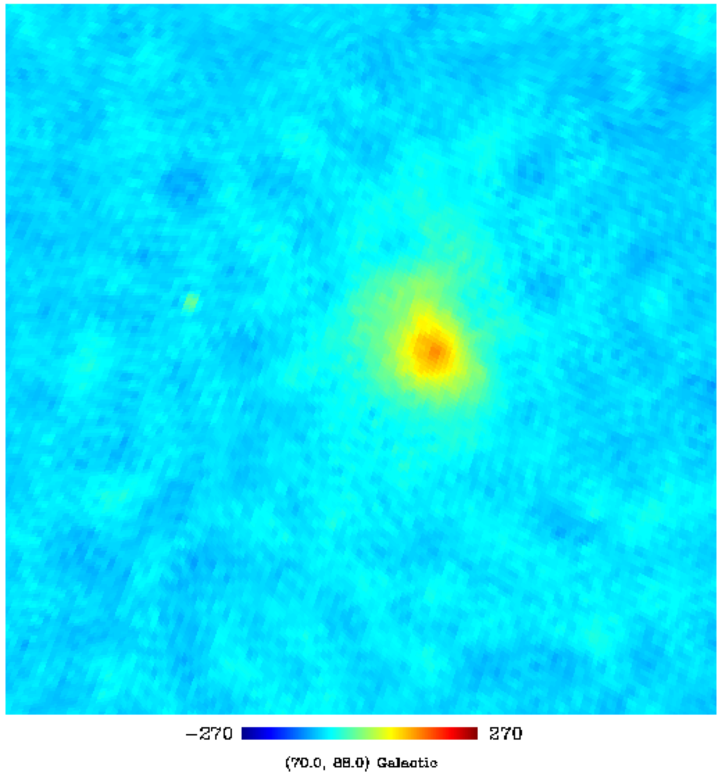
Coma: 217GHz PR2–HFI – SEVEM



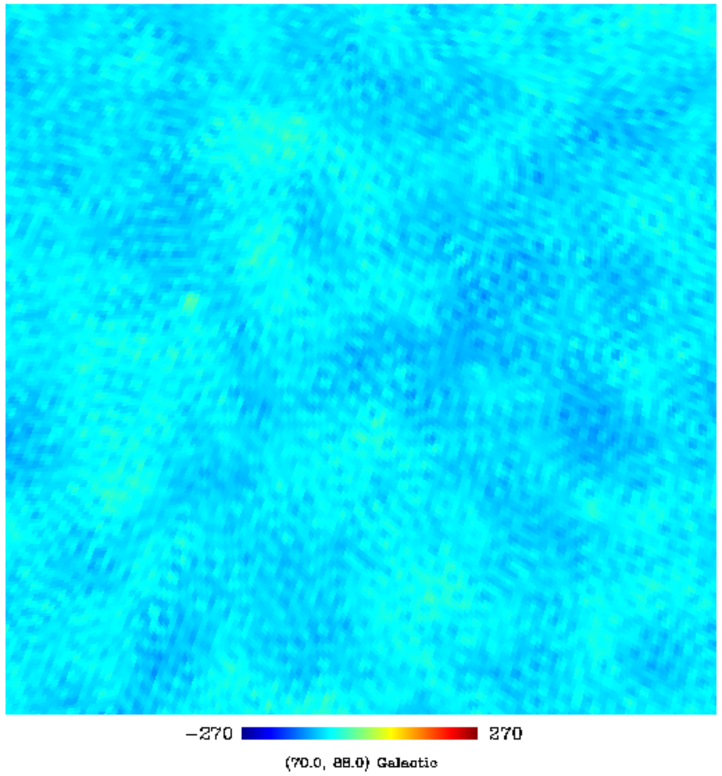
Coma: 217GHz PR2–HFI – SMICA



Coma: 217GHz PR2–HFI – CR



Coma: 217GHz PR2–HFI – GMCA_WPR2



Inverse Problems in Astrophysics

- ➡Part 1: Introduction to Inverse Problems
- ➡Part 2: From Fourier to Wavelets
- ➡Part 3: Wavelet and Beyond
- ➡Part 4: Sparse Regularization
- ➡Part 5: Application to Unmixing and Inpainting
- ➡**Part 6: Compressed Sensing**
- ➡Part 7: Deep Learning

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.

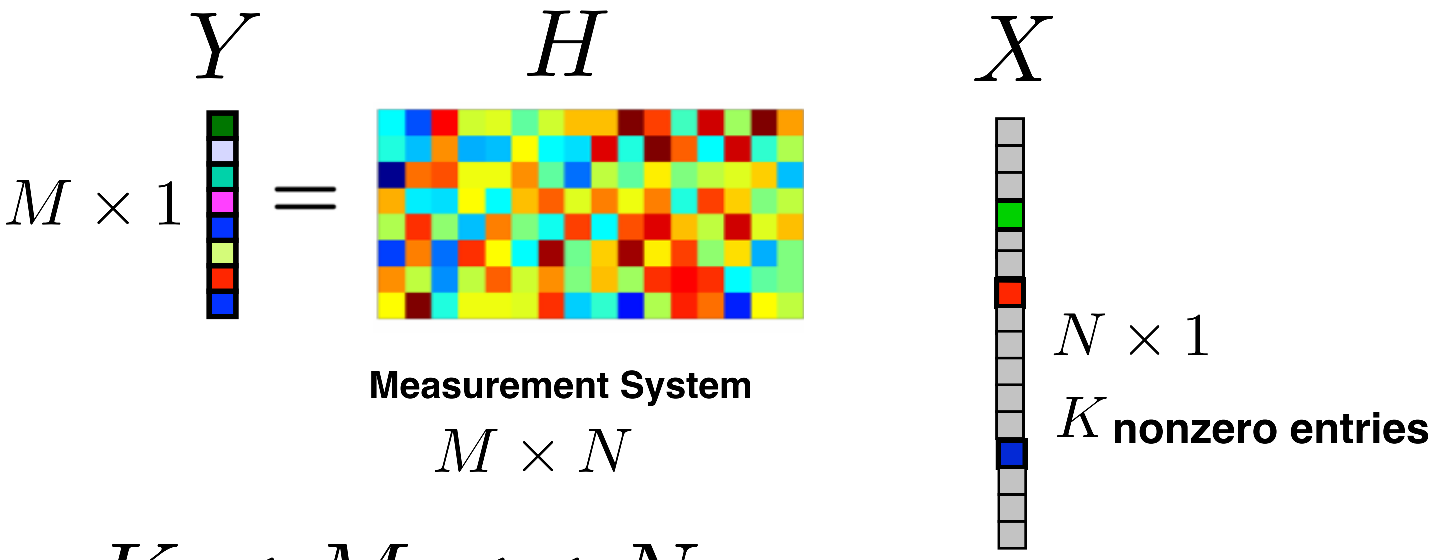
A non linear sampling theorem

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

Replace samples with *few linear projections* $Y = HX$

Incoherent Measurements

X = K sparse signal



Y H X

$M \times 1$ $=$

Measurement System

$M \times N$

$N \times 1$

K **nonzero entries**

$$K < M << N$$

$$\min_X \| X \|_1 \quad \text{s.t.} \quad Y = HX$$

Reconstruction via non linear processing:

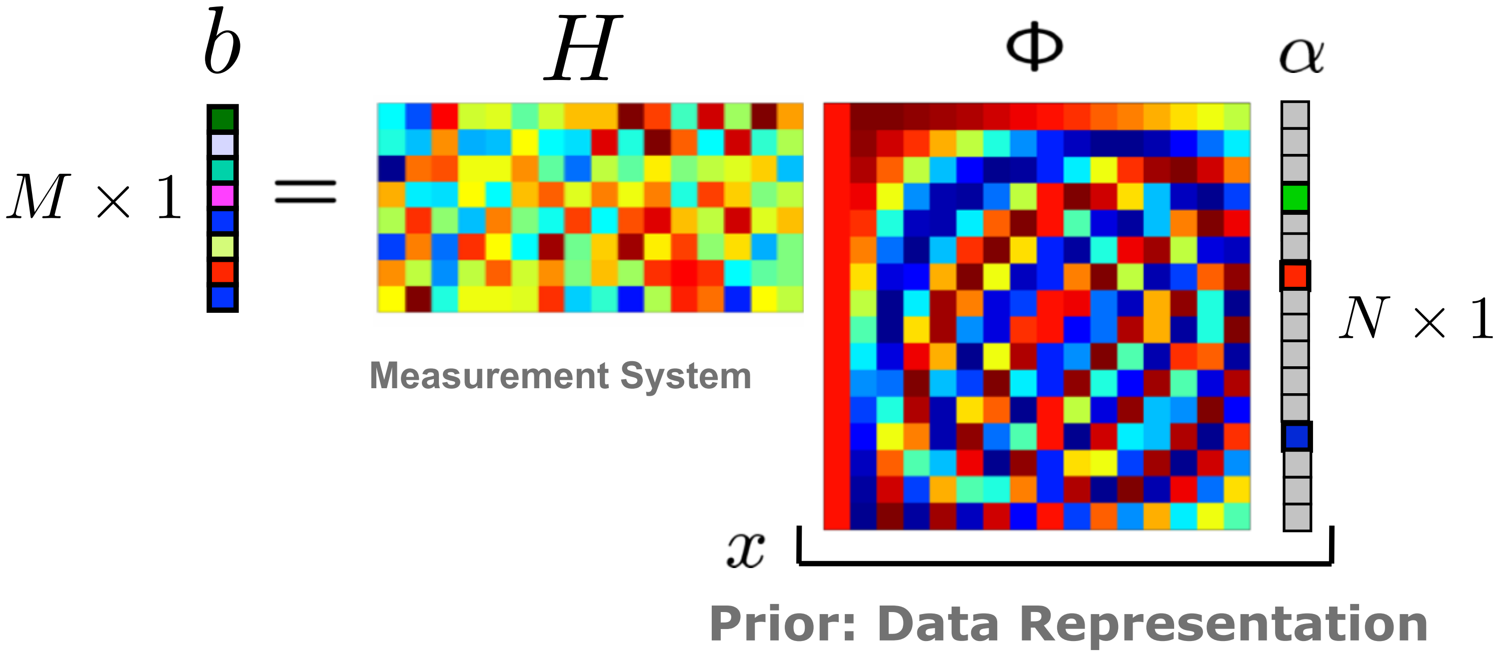
Mutual Incoherence

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

optimally incoherent

ex: Fourier/Dirac

perfectly coherent ensembles



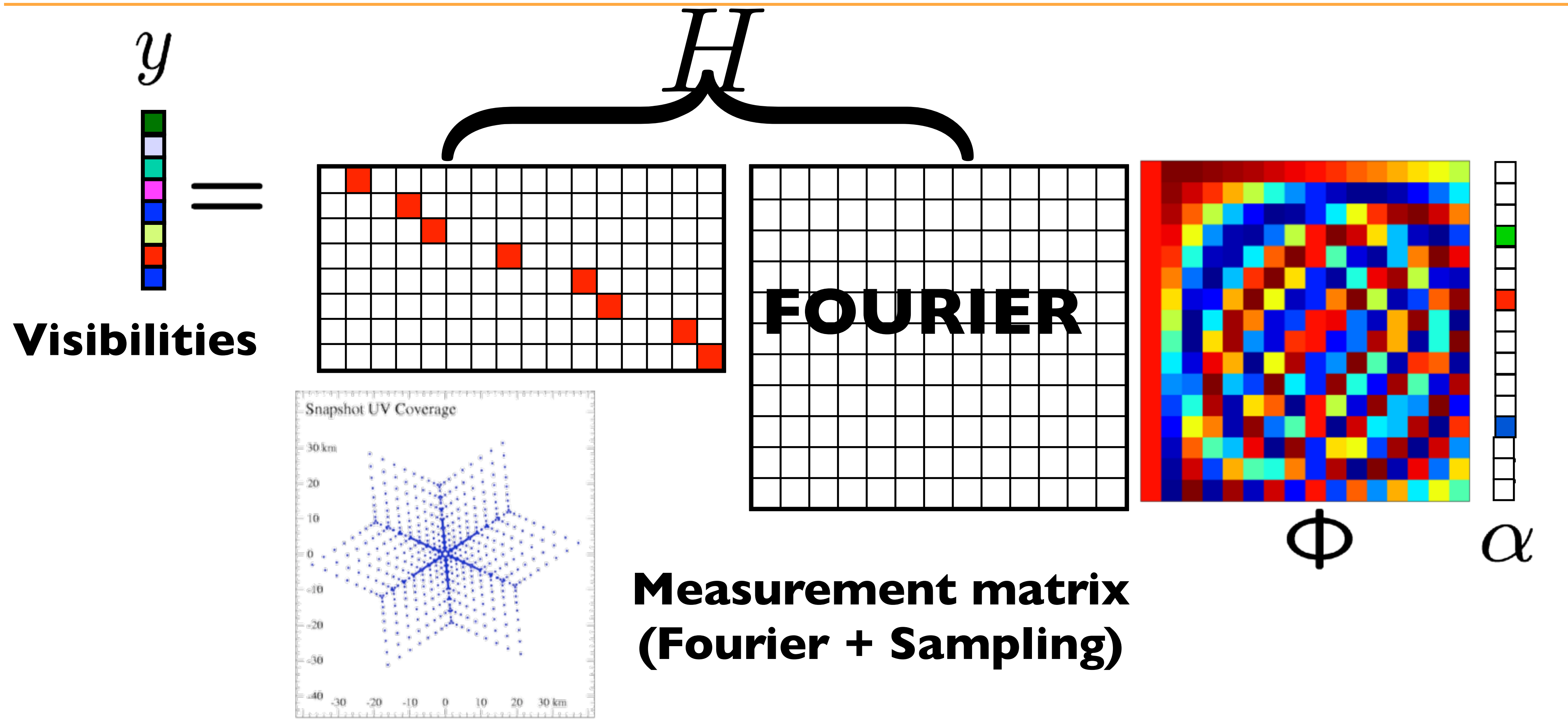
Incoherence between a sparse representation and a measurement matrix (i.e. mutual incoherence):

$$\mu(H, \mathbf{\Phi}) = \max_{i,j} |\langle H_i, \phi_j \rangle|$$

measures how an atom of the sparse representation spreads in the measurement ensemble.

$$\frac{1}{\sqrt{n}} \leq \mu(H, \mathbf{\Phi}) \leq 1$$

Radio-Interferometry Sparse Recovery



Measurement matrix (Fourier + Sampling)

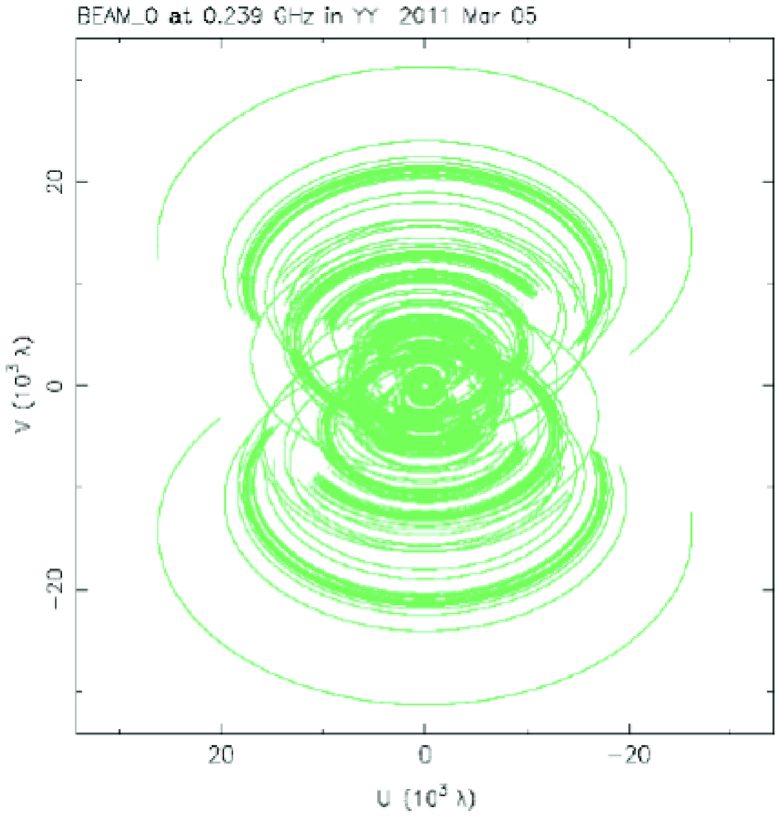
$$Y = HX + N$$

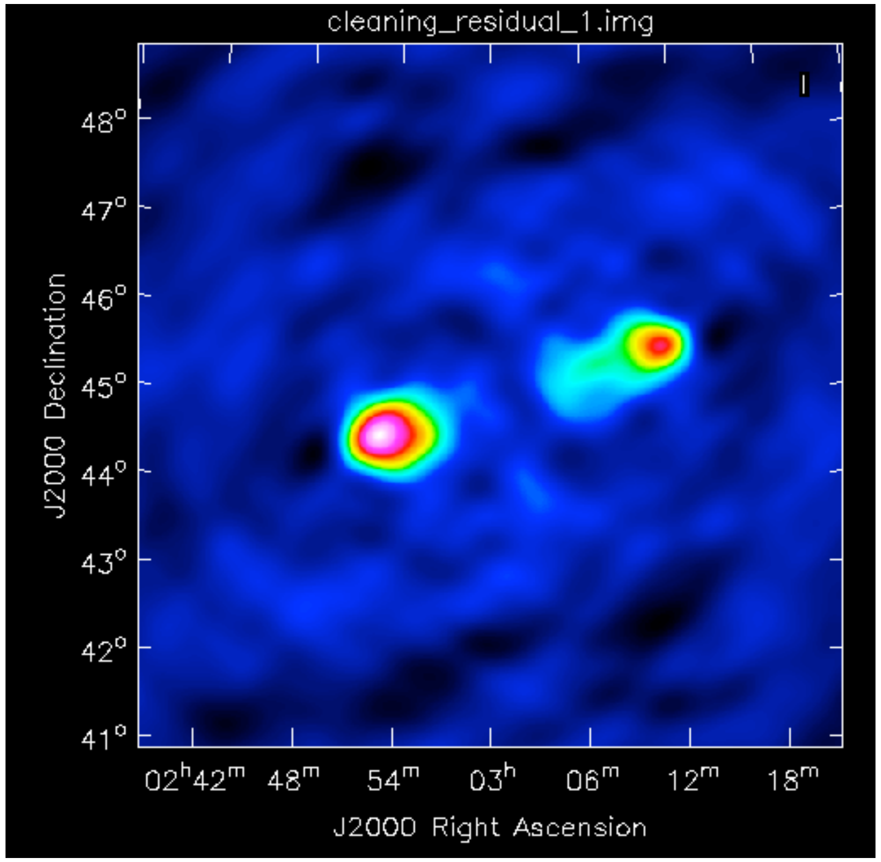
$$X = \Phi\alpha$$

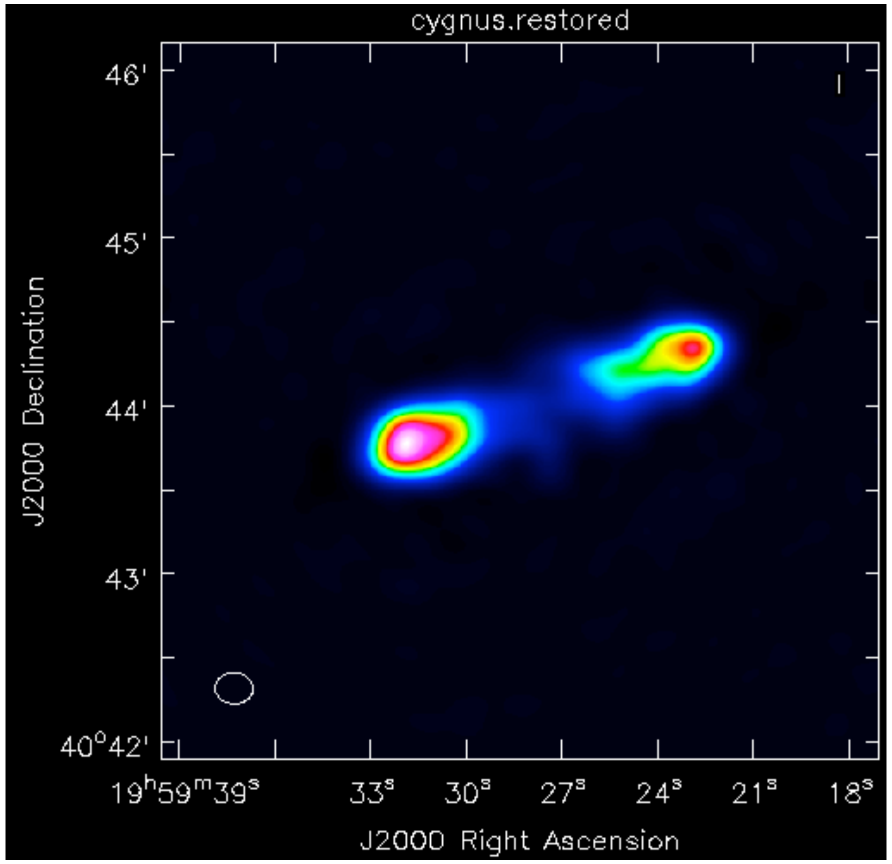
$$\min_{\alpha} \|\alpha\|_p^p \quad \text{subject to} \quad \|Y - H\Phi\alpha\|^2 \leq \epsilon$$

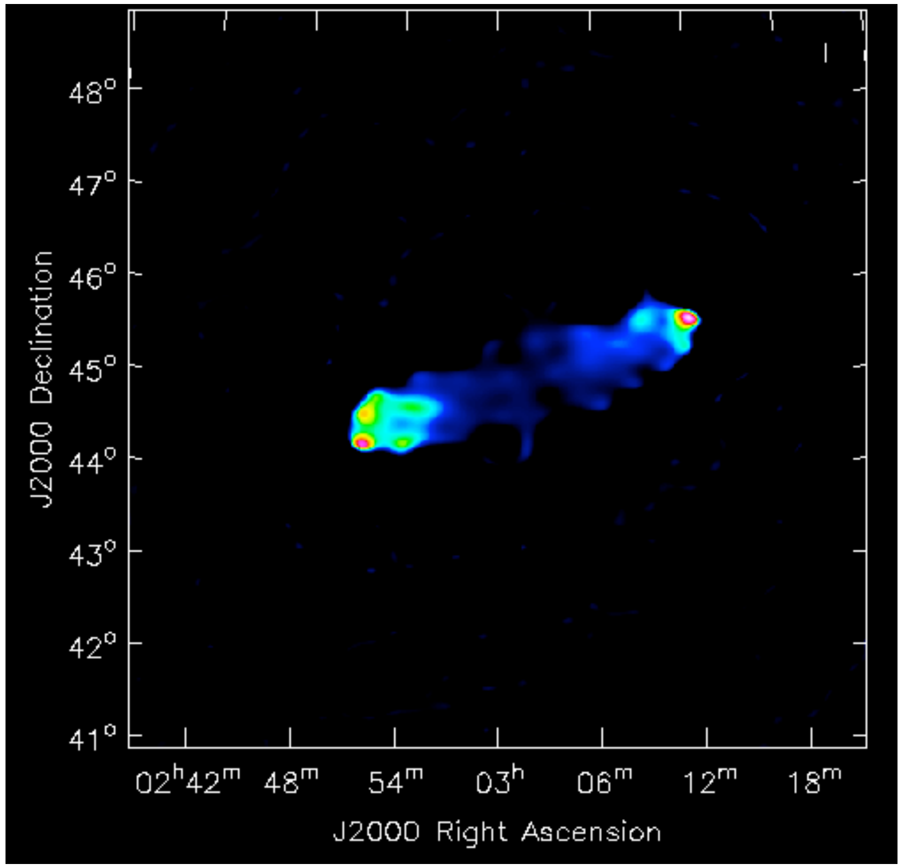
- Photometry: similar to CLEAN on point sources.
- Resolution: improved by a factor larger than 2 for SNR > 10.
- Extended objects reconstruction much better than CLEAN and Multiscale CLEAN.
- Improved image quality (RMS better by a factor of 10 compared to CLEAN)

Compressed Sensing & LOFAR Cygnus A Data









Garsden et al, "**LOFAR Image Sparse Reconstruction**", A&A, 575, A90, 2015.

http://arxiv.org/abs/1406.7242

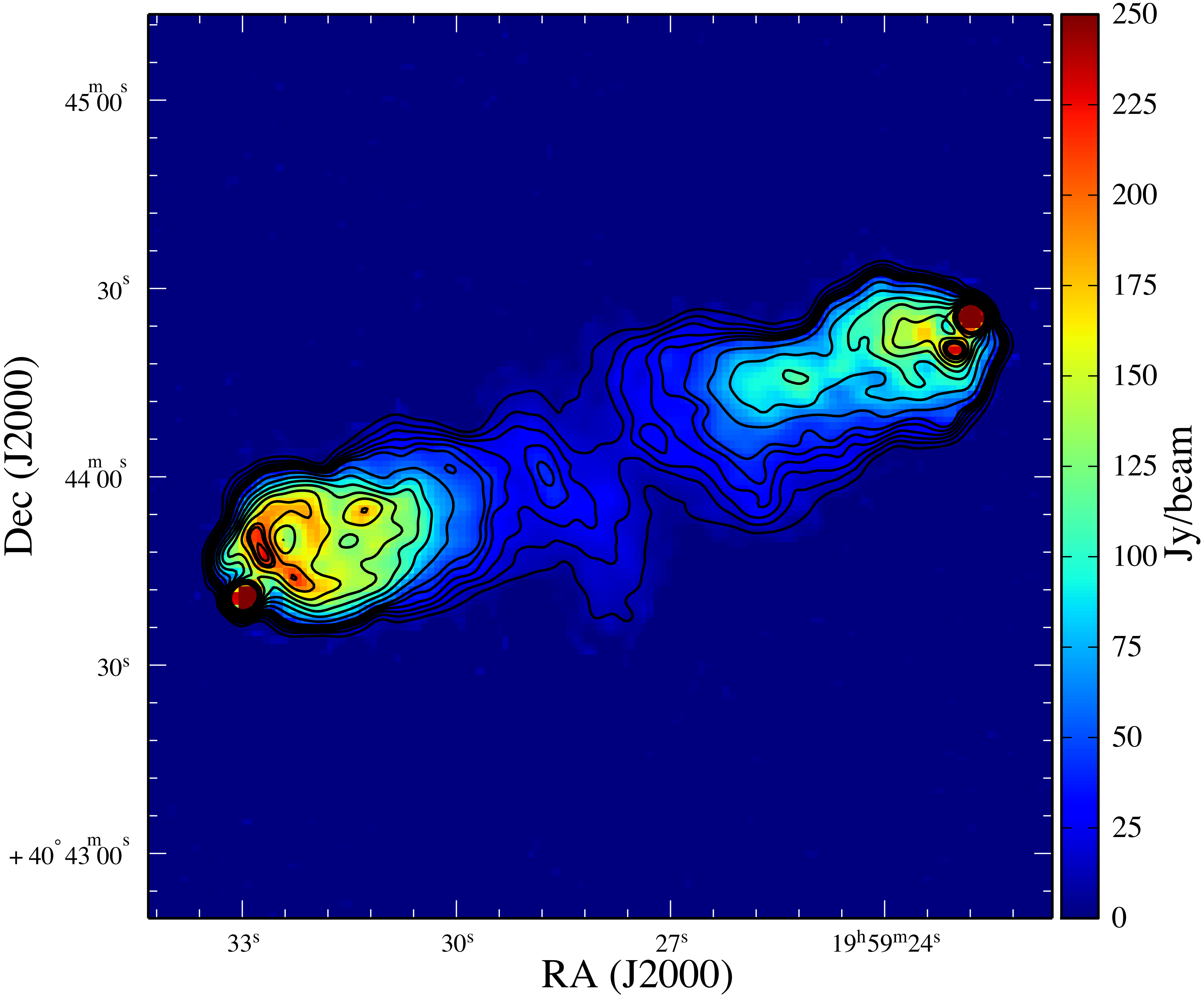
Garsden et al, “LOFAR Image Sparse Reconstruction”, A&A, 575, A90, 2015, ArXiv:1406.7242.

J. Girard

H. Garsden

S. Corbel

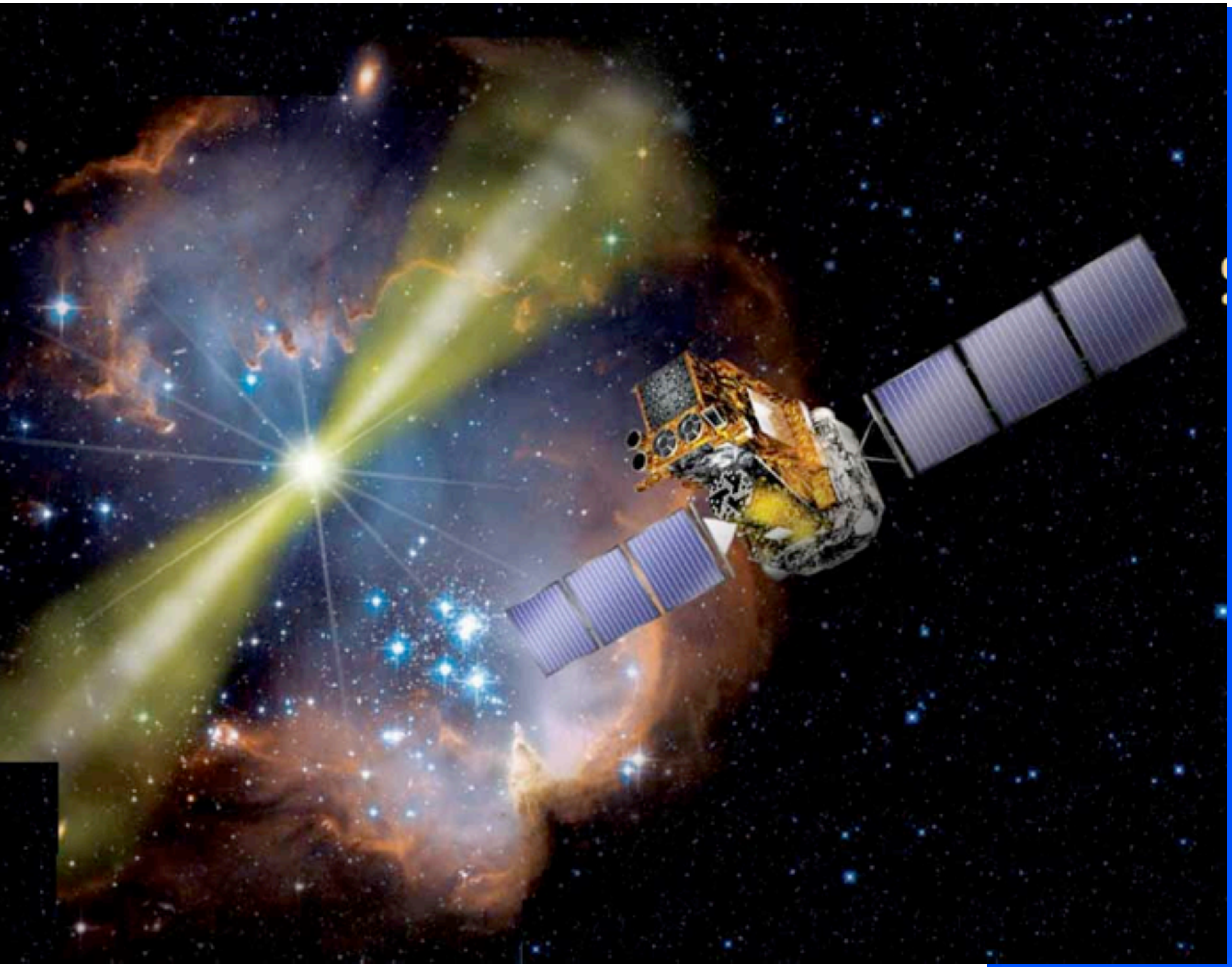
C. Tasse

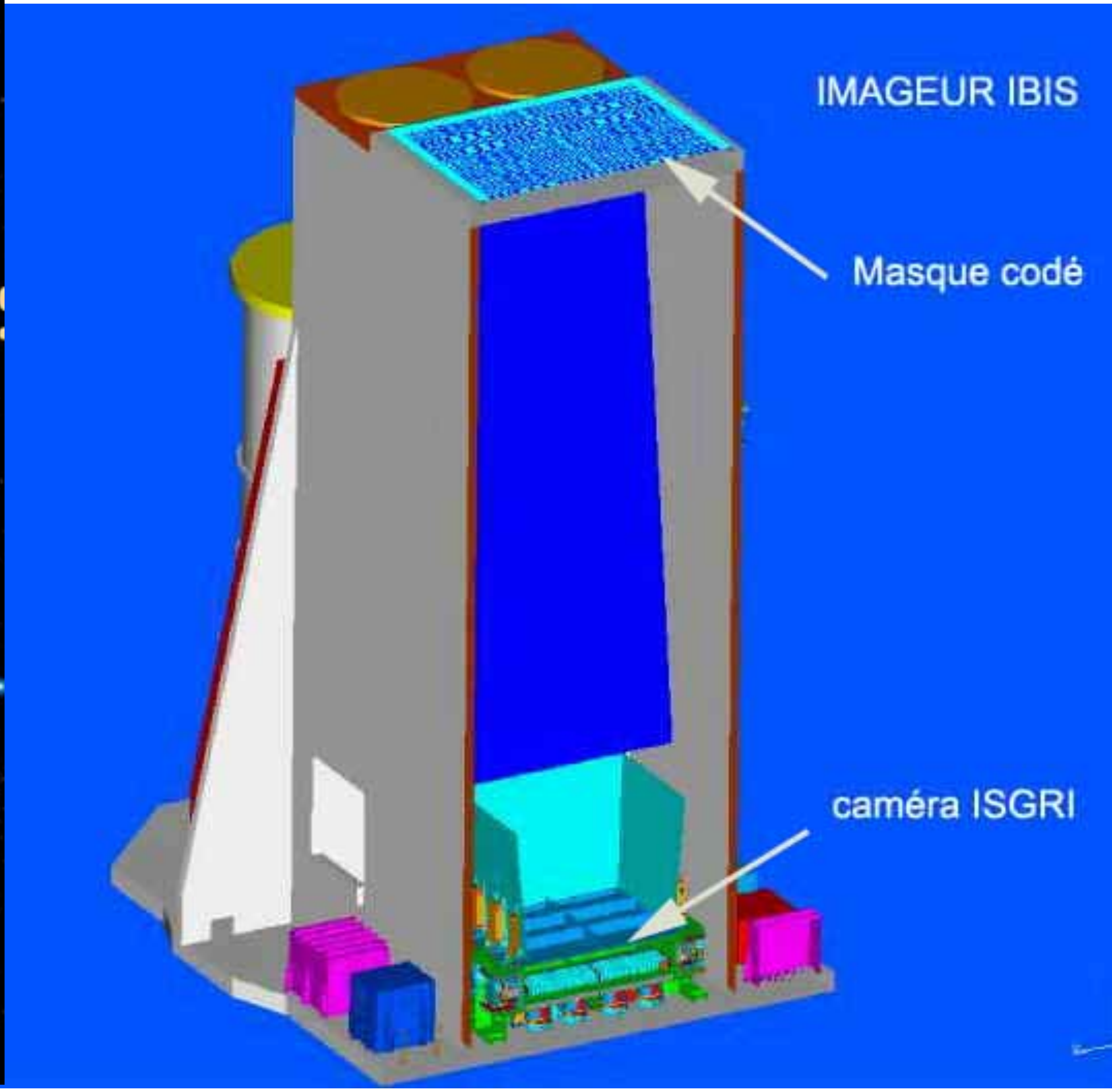


Colorscale: reconstructed 512x512 image of Cygnus A at 151 MHz (with resolution 2.8” and a pixel size of 1”). Contours levels are [1,2,3,4,5,6,9,13,17,21,25,30,35,37,40] Jy/Beam from a 327.5 MHz Cyg A VLA image (Project AK570) at 2.5” angular resolution and a pixel size of 0.5”.

Recovered features in the CS image correspond to real structures observed at higher frequencies.

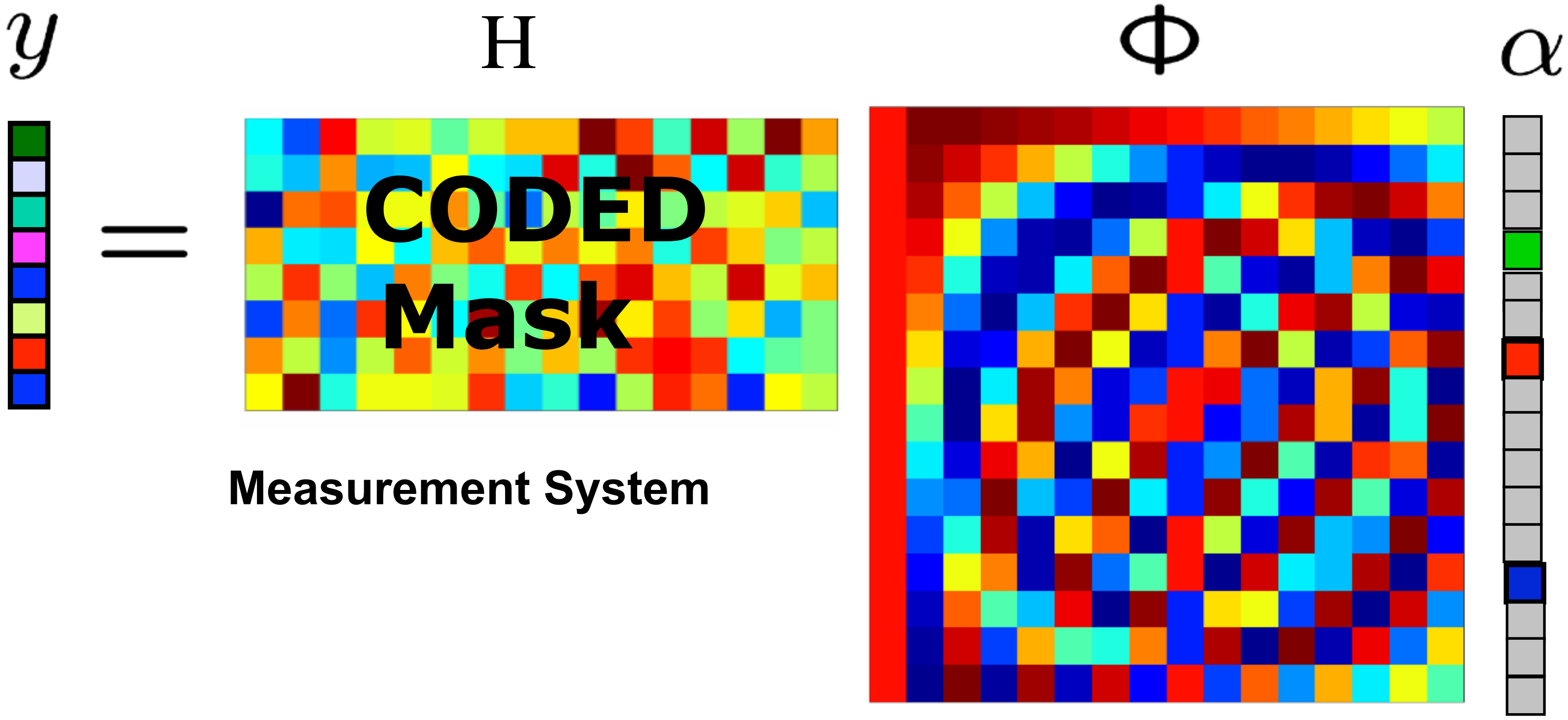
IBIS Camera on board of the Integral satellite

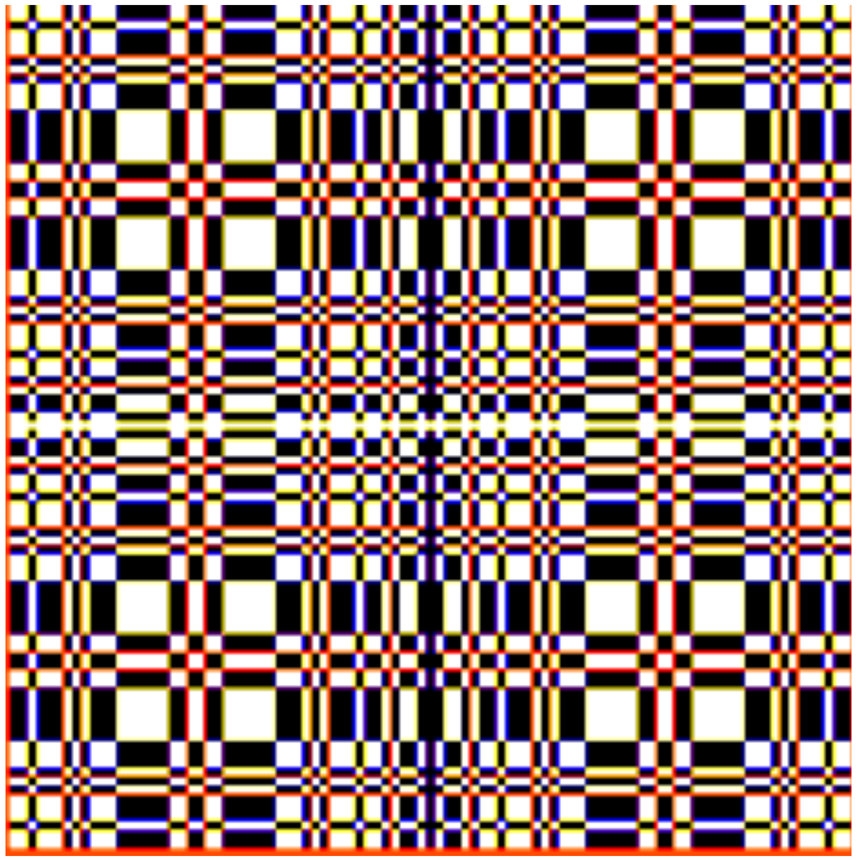




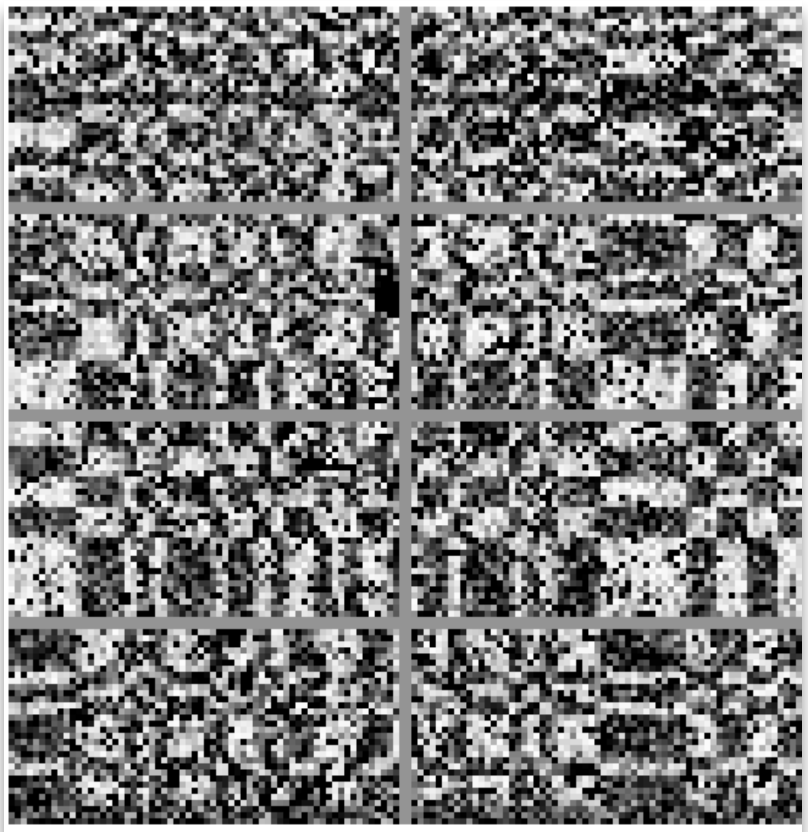
Launch, October 4, 2002.

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'***
  Gamma-ray detection in energy range 4 - 120 keV
  Coded mask imaging (at 460 mm of the detector plane)

Physical mask pattern

Figure 3 : Le motif sélectionné dans le cadre de la mission SVOM, C3A2S, est formé d'un maillage de 47×47 éléments constitué de trous et de pavés opaques. Le choix final du dessin du masque codé d'ECLAIRs est le fruit de nombreuses années d'études. Crédit APC-CEA.

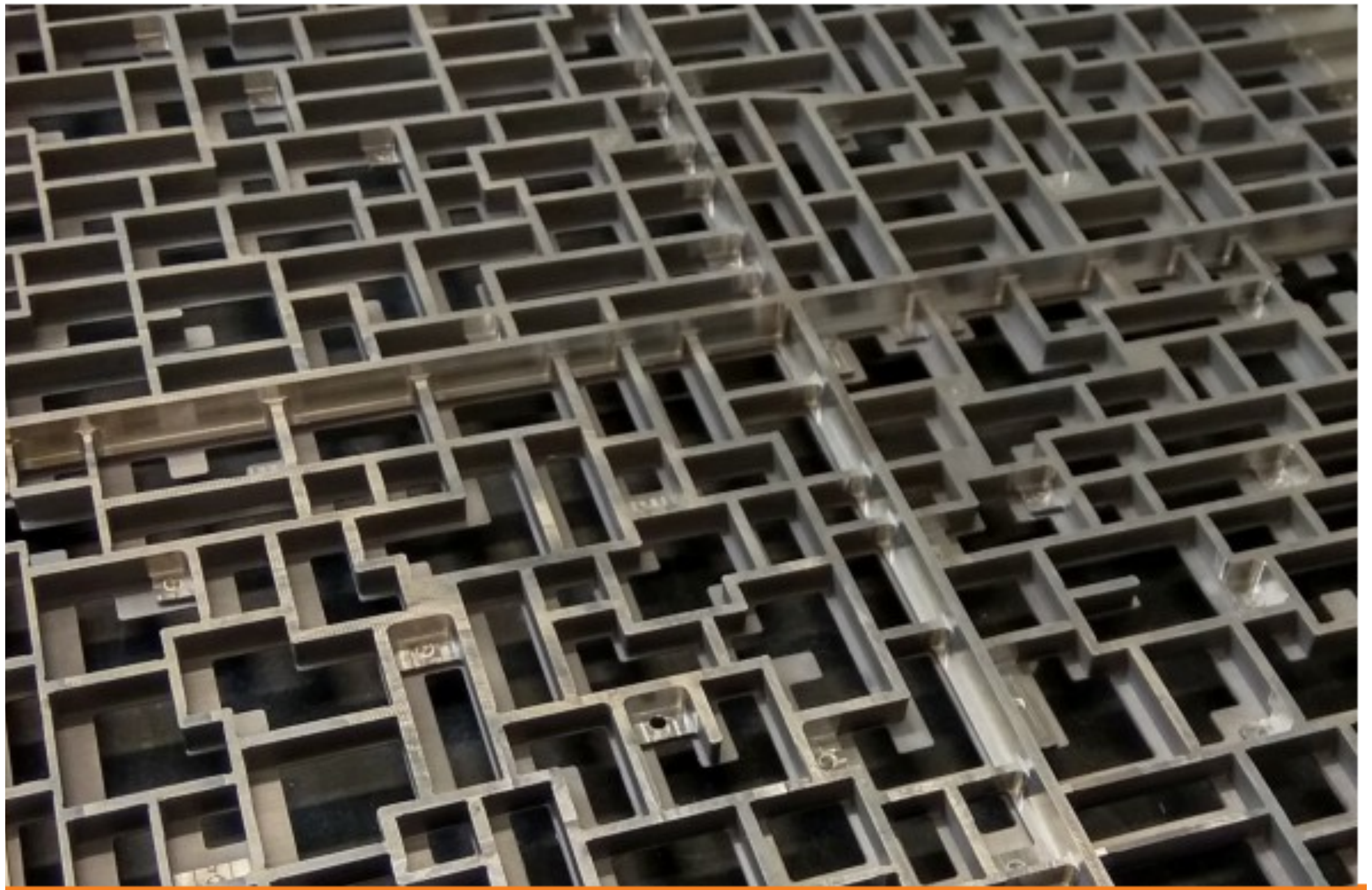
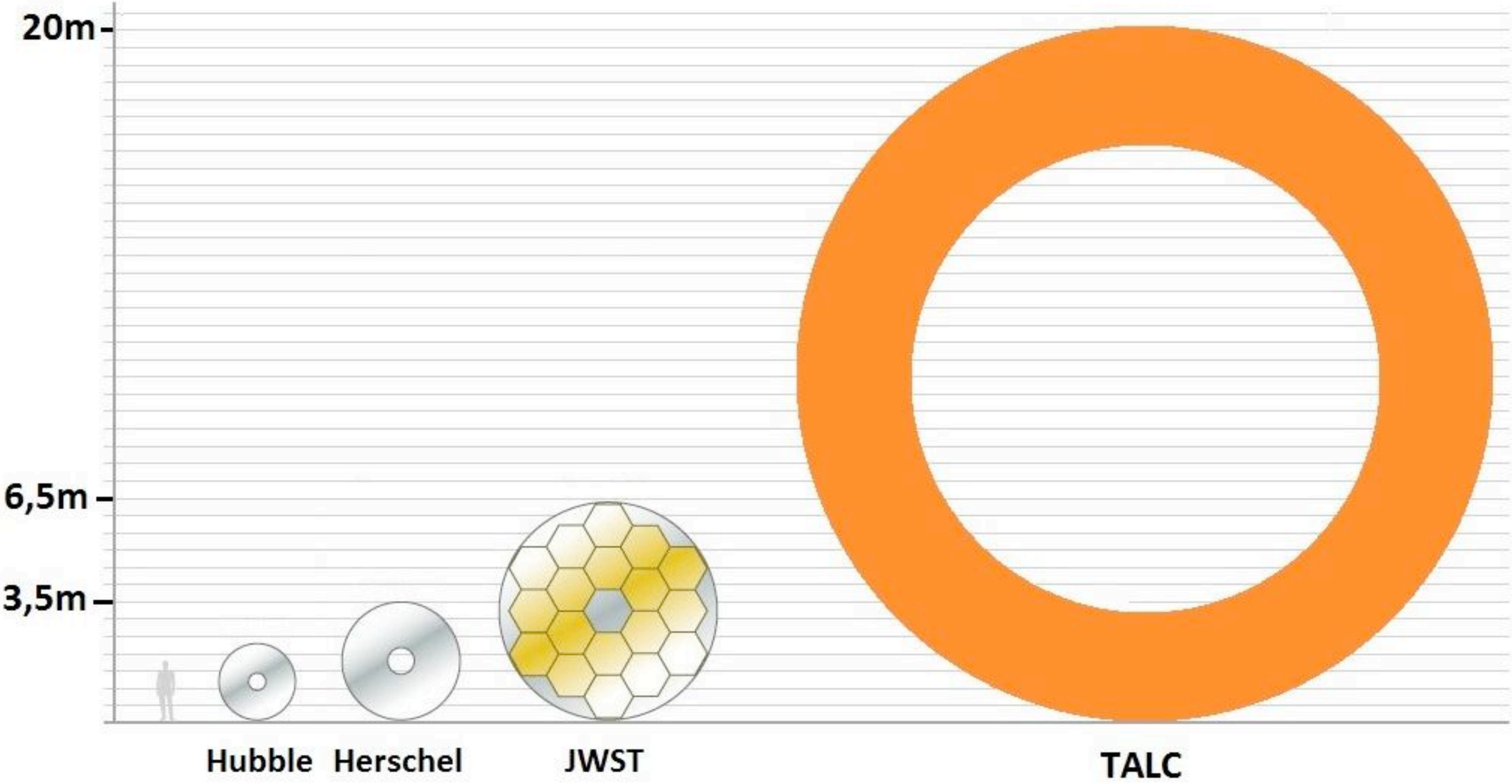


Figure 6 : Photo du modèle prototype STM. Crédit : APC.

Spatial Anular Telescope: TALC

Sauvage et al, “a Thinned Aperture Light Collector for space far-infrared studies”, submitted to CNES Call for Ideas, 2013.



Paradigm Shift in Statistics/Signal Processing

	Modeling	Sampling	Inverse Problems
20th century	band limited signals	Shannon Nyquist sampling	linear l2 norm regularization
↓	↓	↓	↓
21st century	sparse/compressible signals	Compressed Sensing	non-linear l0-l1 regularization

Inverse Problems in Astrophysics

➡Part 1: Introduction to Inverse Problems

➡Part 2: From Fourier to Wavelets

➡Part 3: Wavelet and Beyond

➡Part 4: Sparse Regularization

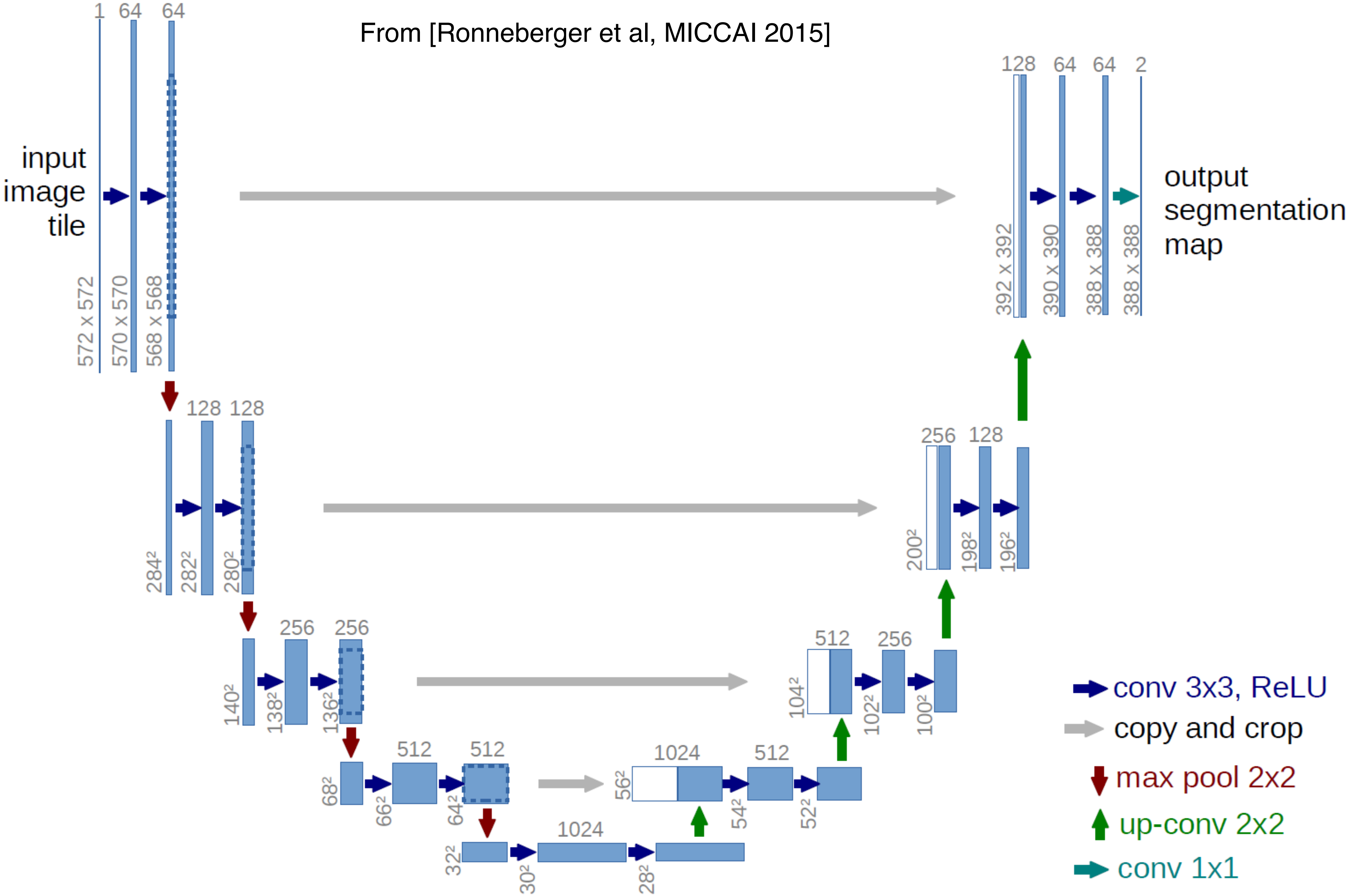
➡Part 5: Application to Unmixing and Inpainting

➡Part 6: Compressed Sensing

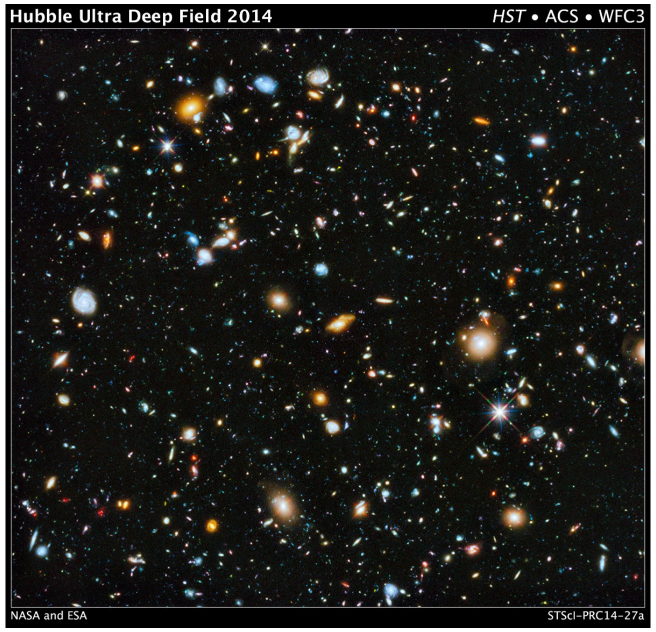
➡**Part 7: Deep Learning**

U-NET

From [Ronneberger et al, MICCAI 2015]



Galaxies Survey Image Deconvolution



Sandard deconvolution framework:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

Noise

True Image

PSF Convolution

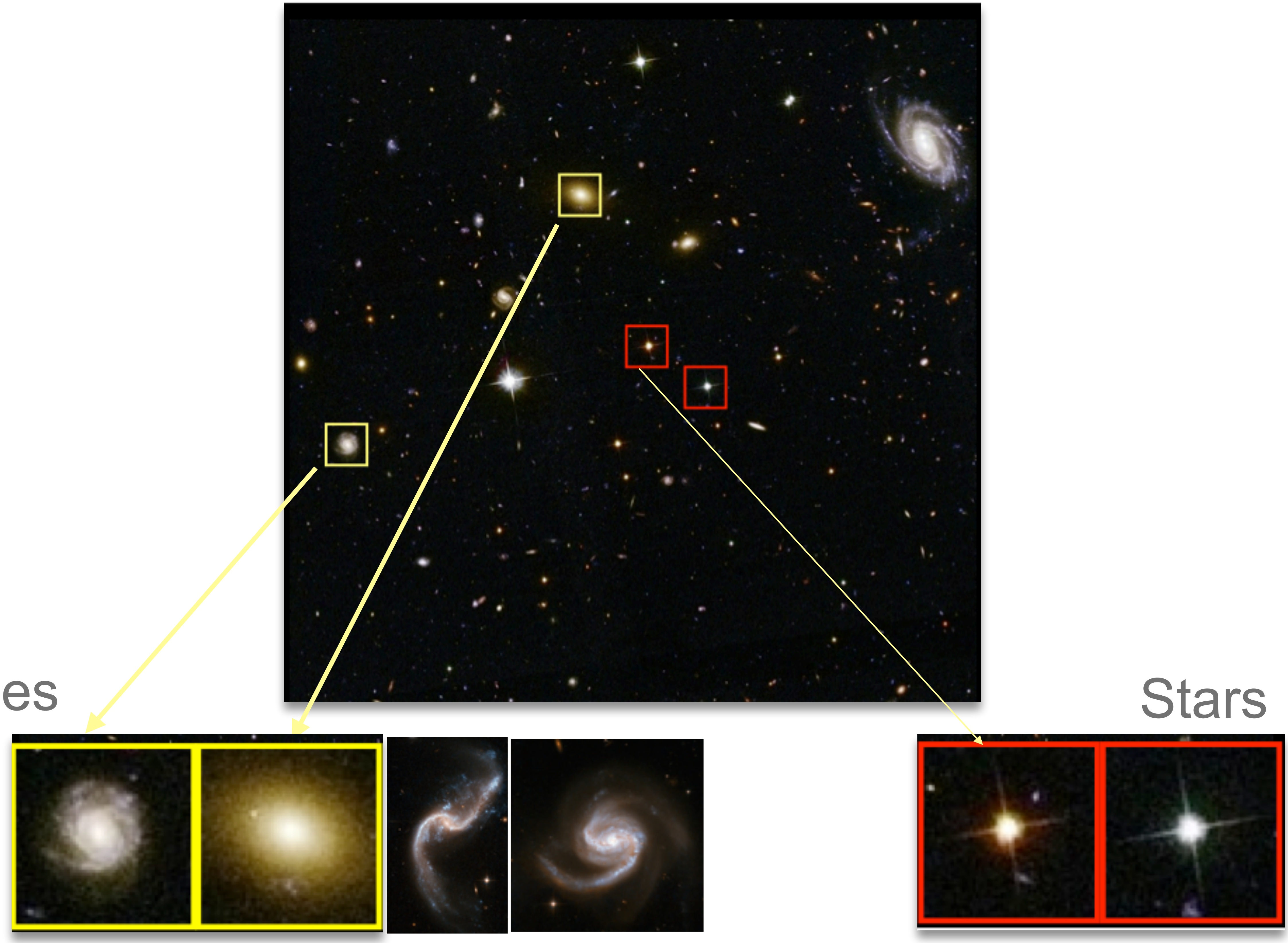
Observed Image

Sandard deconvolution framework:

$$\underset{\mathbf{X}}{\operatorname{argmin}} \quad \frac{1}{2}\|\mathbf{Y} - \boxed{\mathbf{HX}}\|_2^2 + \|\Phi^t\mathbf{X}\|_p \quad \text{s.t.} \quad \mathbf{X} \geq 0$$

H is huge !!!

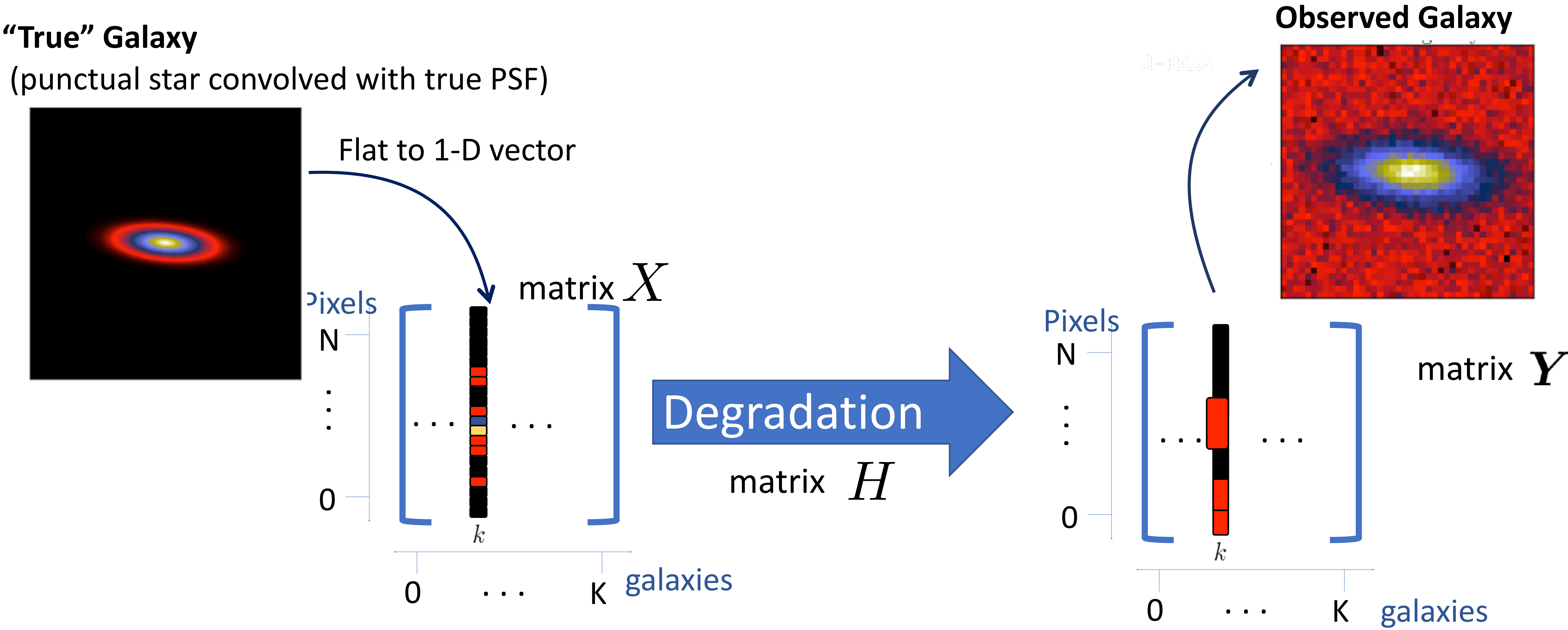
Detection + Classification stars/galaxies



Galaxies

Stars

Object Oriented Deconvolution



“True” Galaxy

(punctual star convolved with true PSF)

Flat to 1-D vector

matrix X

Pixels

N

⋮

0

. . . k . . .

0 . . . K galaxies

Degradation

matrix H

Observed Galaxy

Pixels

N

⋮

0

. . . k . . .

0 . . . K galaxies

matrix Y

$$\mathbf{Y} = \mathcal{H}(\mathbf{X}) + \mathbf{N}$$

$[n^0, n^1, \dots, n^n]$

$[H^0x^0, H^1x^1, \dots, H^nx^n]$

$[y^0, y^1, \dots, y^n]$

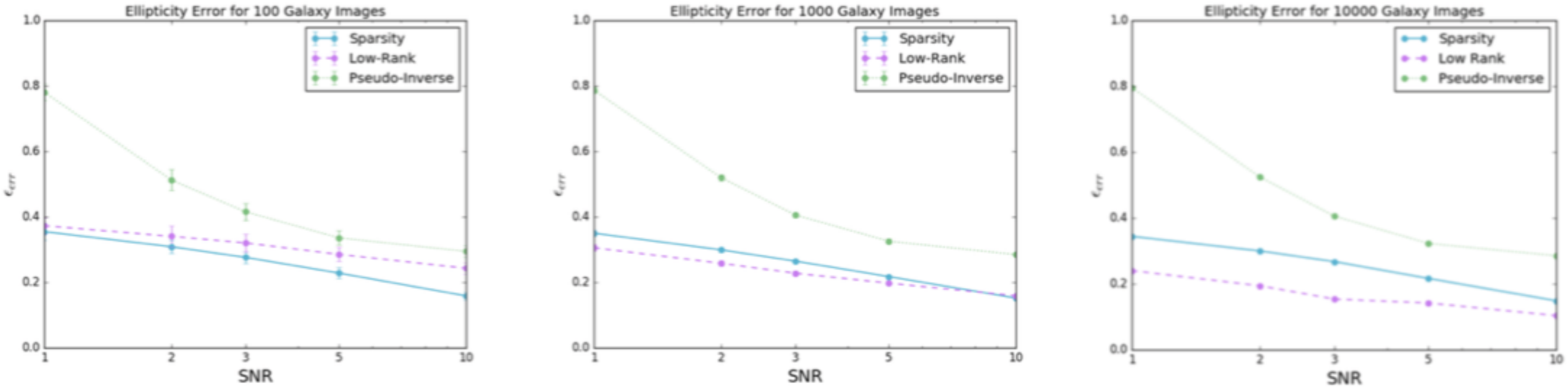
Big Astronomical Image Deconvolution

Object Oriented Deconvolution

For each galaxy, we use the PSF related to its center pixel:

Sparsity
$$\underset{\mathbf{X}}{\operatorname{argmin}} \quad \frac{1}{2}\|\mathbf{Y} - \mathcal{H}(\mathbf{X})\|_2^2 + \lambda\|\Phi^t\mathbf{X}\|_p \quad \text{s.t.} \quad \mathbf{X} \geq 0$$

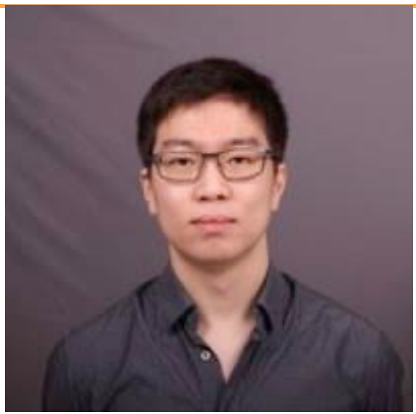
Low-Rank
$$\underset{\mathbf{X}}{\operatorname{argmin}} \quad \frac{1}{2}\|\mathbf{Y} - \mathcal{H}(\mathbf{X})\|_2^2 + \lambda\|\mathbf{X}\|_* \quad \text{s.t.} \quad \mathbf{X} \geq 0$$



Farrens et al, Space Variant Deconvolution of Galaxy Survey Images, A&A 2017.

Deep Learning Space Variant Deconvolution

F. Sureau, A. Lechat and J.-L. Starck, "Deep learning for a space-variant deconvolution in galaxy surveys", A&A 641, A67 (2020)

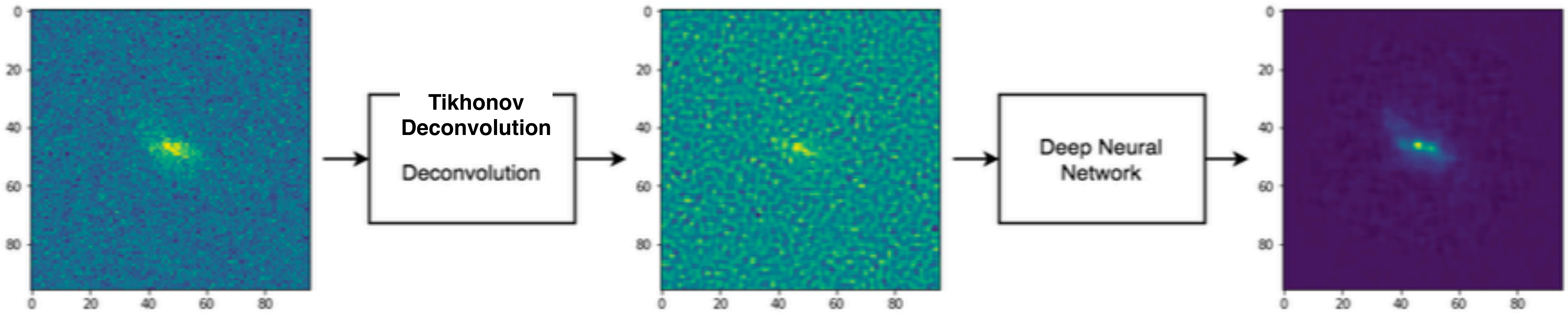


F. Sureau

A. Lechat

Idea : integrate physical modeling (i.e. Beam)+ galaxy representation with Deep Learning

- Tikhonet: Post processing approach after deconvolution with mere Tikhonov regularization



==> Space Variant Deconvolution with TikhoNet

- ADMMnet: learn a "regularizer" (denoiser) to use in ADMM algorithm derived from convex optimization

Space Variant Tikhonov Deconvolution

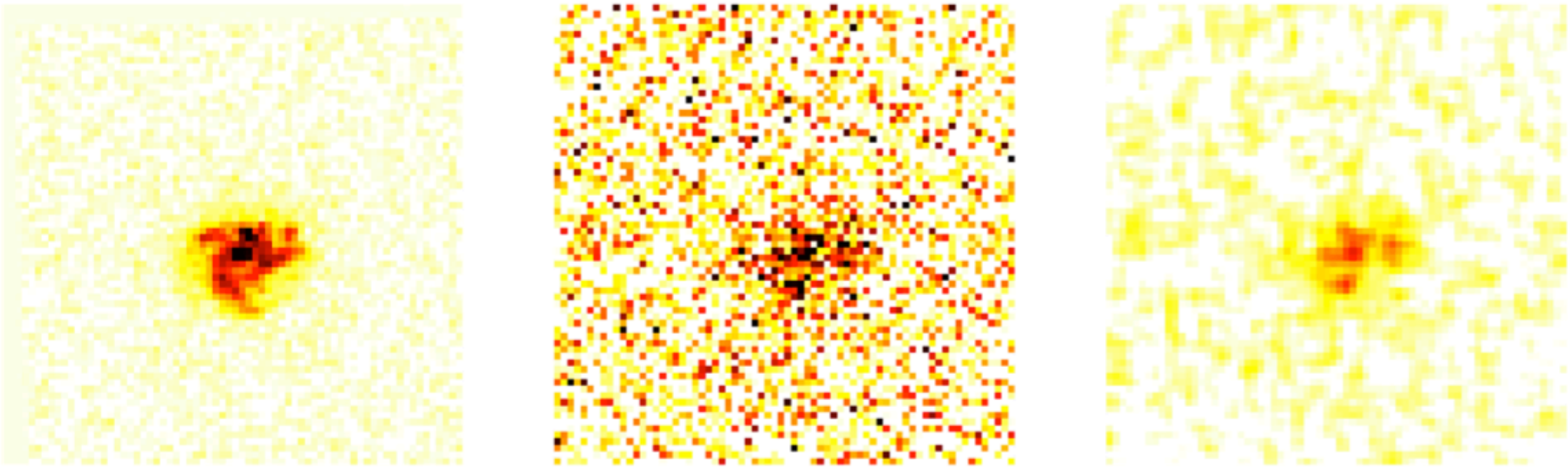
The Tikhonov solution for the space variance variant PSF deconvolution is:

$$\underset{\mathbf{X}}{\arg\min} \frac{1}{2}\|\mathbf{Y} - \mathcal{H}(\mathbf{X})\|_F^2 + \|\mathcal{L}(\mathbf{X})\|_F^2$$

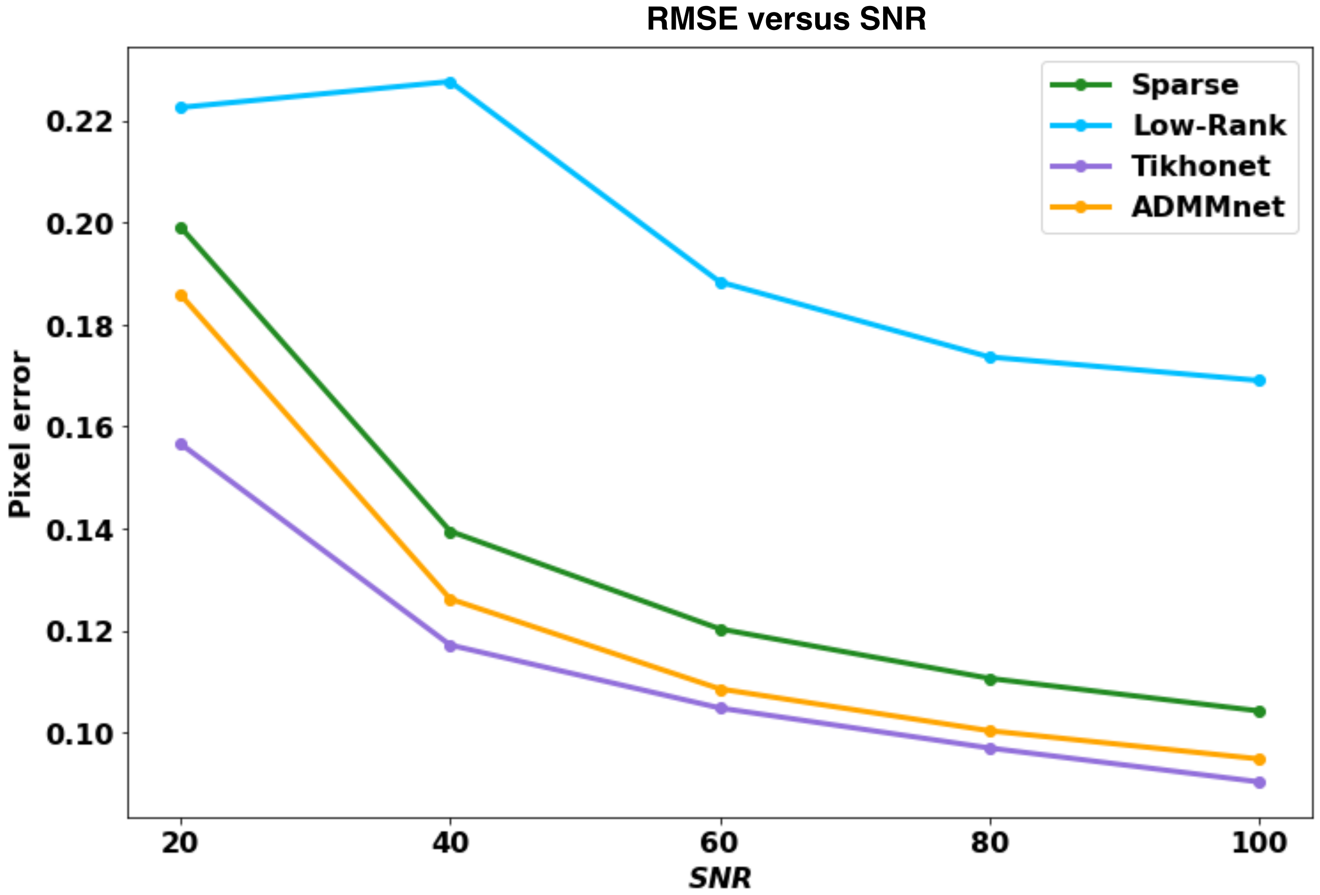
where $\mathcal{L}$ is similarly built as $\mathcal{H}$.

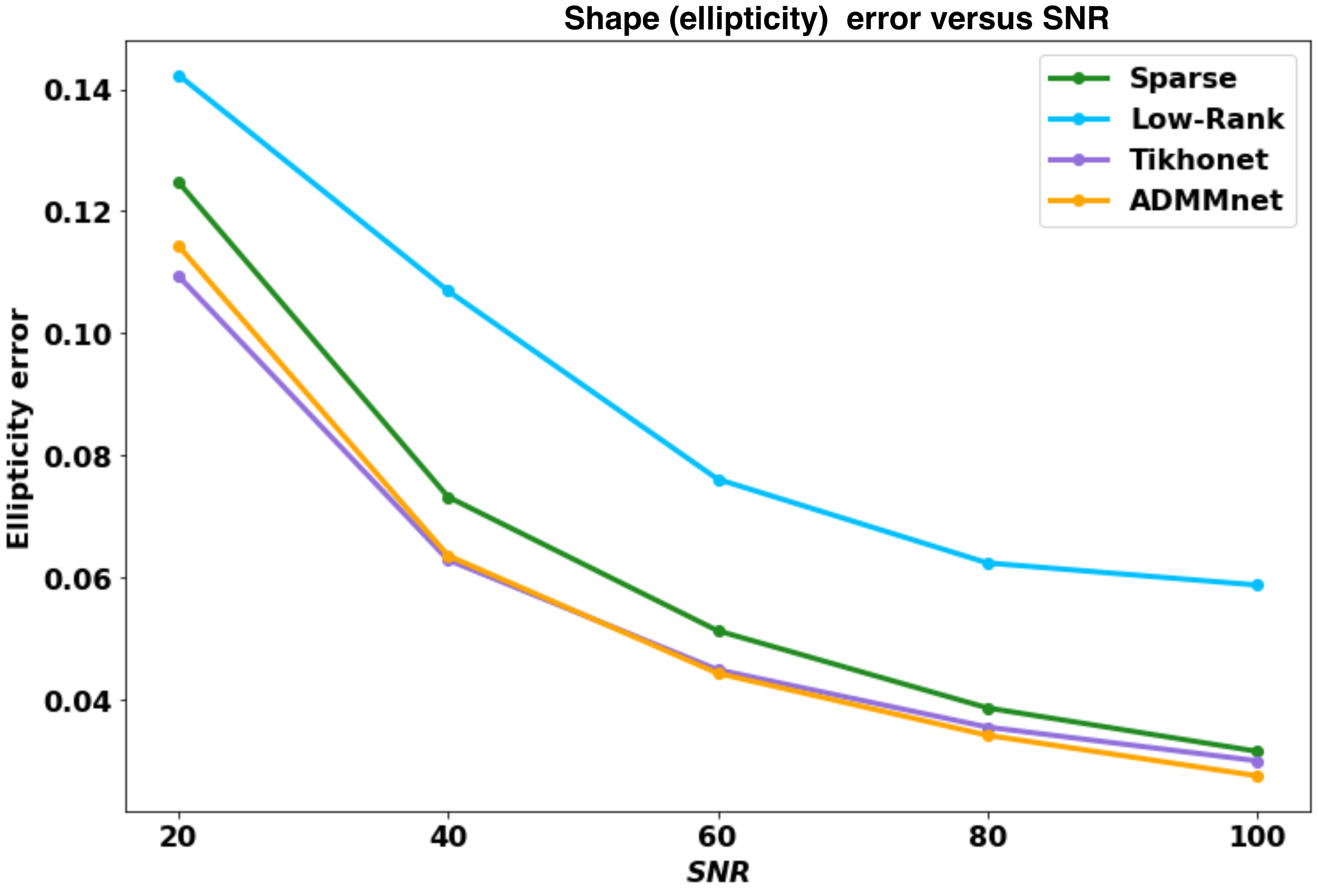
The closed-form solution of this linear inverse problem is given by:

$$\left\{\tilde{\mathbf{x}}_\mathbf{i} = \left(\mathbf{H}_\mathbf{i}^T\mathbf{H}_\mathbf{i} + \lambda_i\mathbf{L}_\mathbf{i}^T\mathbf{L}_\mathbf{i}\right)^{-1}\mathbf{H}_\mathbf{i}^T\mathbf{y}_\mathbf{i}\right\}_{i=1..n_g}$$



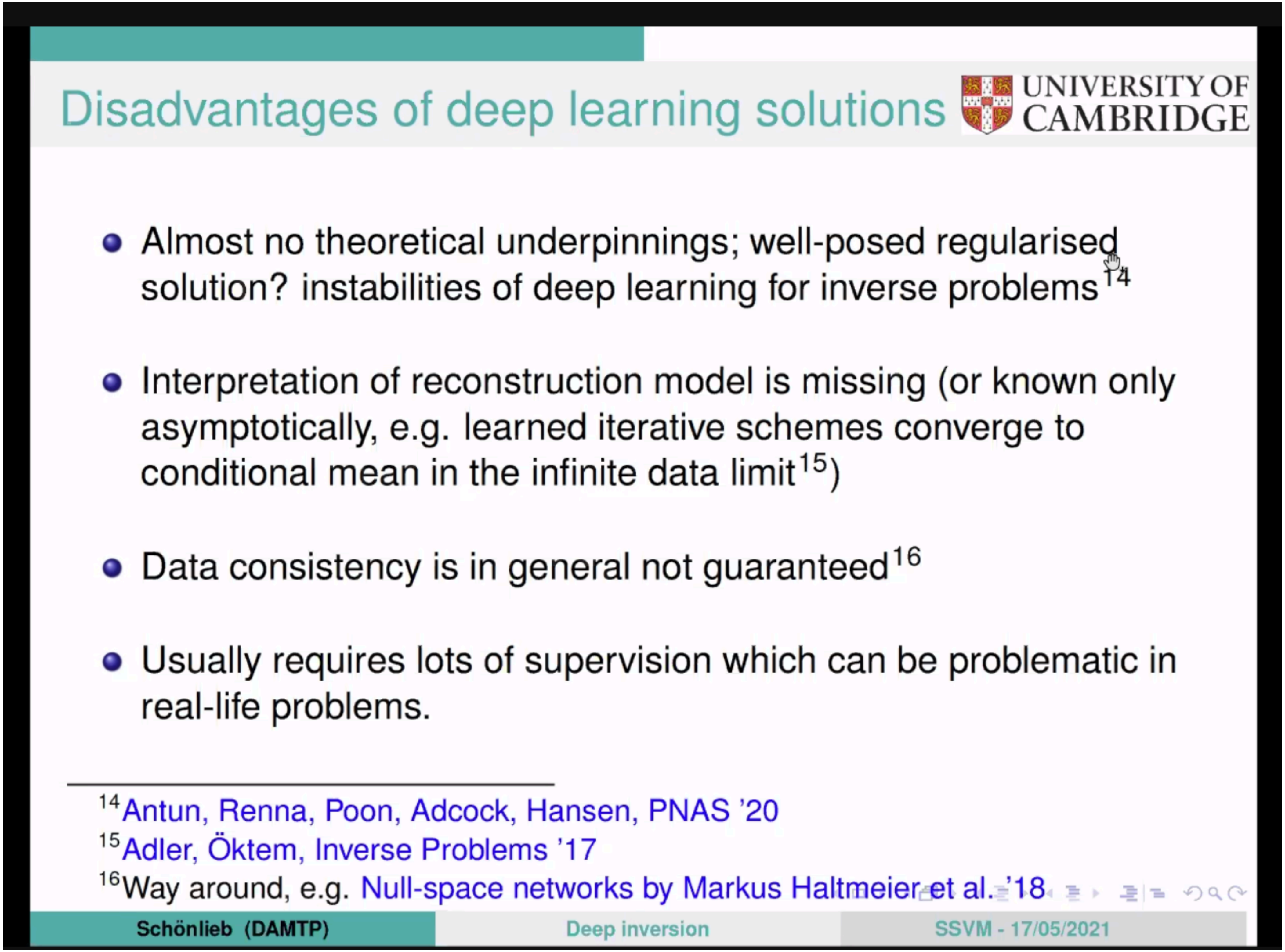
Results per SNR





Is Deep Learning Perfect ?

From Carola Bibiane Schönlieb, *Machine Learned Regularisation for Inverse Imaging Problems, 17 May 2021, SSVM Conference*



Disadvantages of deep learning solutions

UNIVERSITY OF CAMBRIDGE

- Almost no theoretical underpinnings; well-posed regularised solution? instabilities of deep learning for inverse problems[14]
- Interpretation of reconstruction model is missing (or known only asymptotically, e.g. learned iterative schemes converge to conditional mean in the infinite data limit[15])
- Data consistency is in general not guaranteed[16]
- Usually requires lots of supervision which can be problematic in real-life problems.

---

[14] Antun, Renna, Poon, Adcock, Hansen, PNAS ’20

[15] Adler, Öktem, Inverse Problems ’17

[16] Way around, e.g. Null-space networks by Markus Haltmeier et al. ’18

Schönlieb (DAMTP) Deep inversion SSVM - 17/05/2021

N. Gottschling et al, The troublesome kernel: why deep learning for inverse problems is typically unstable, ArXiv 2001.01258

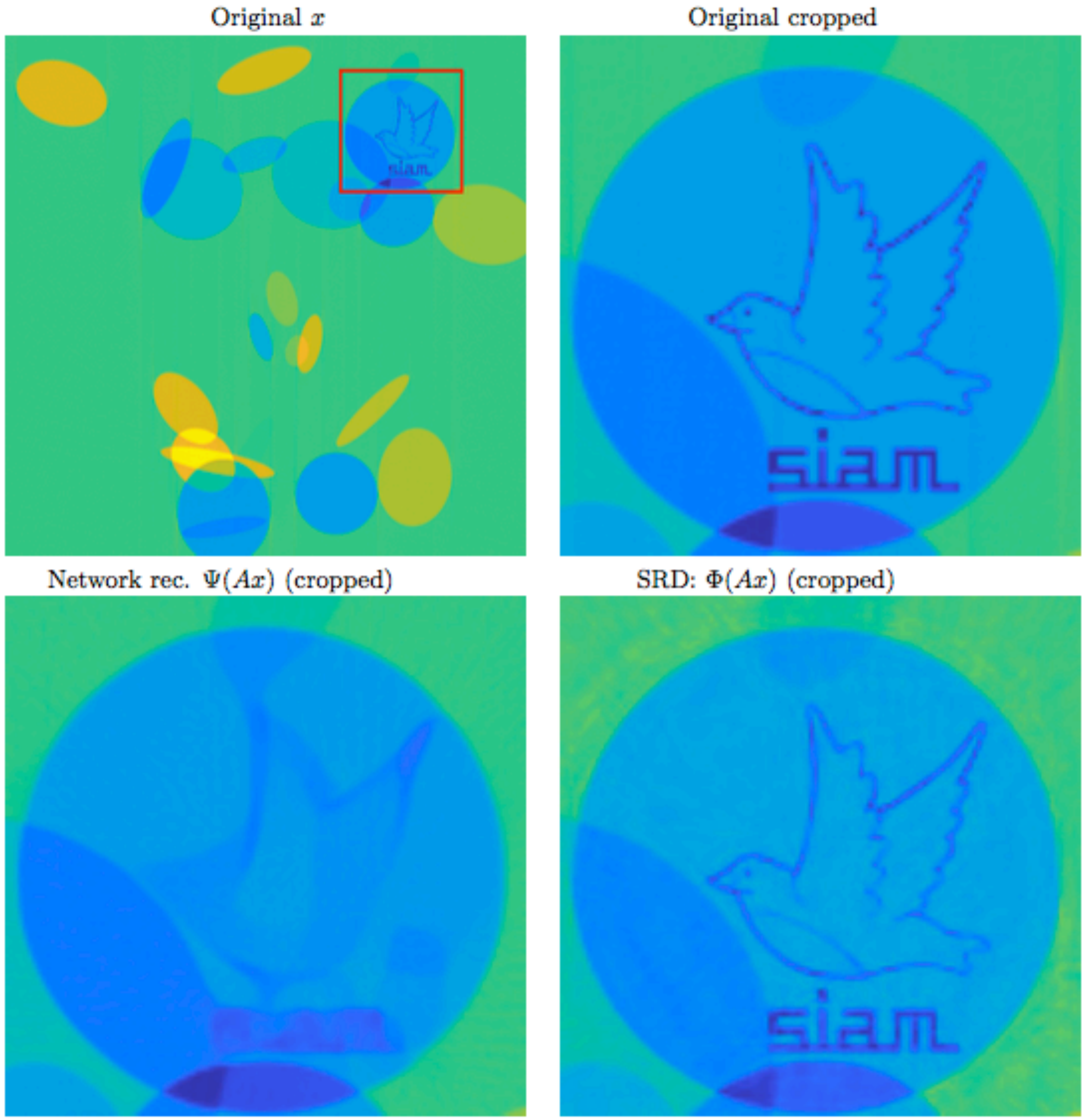
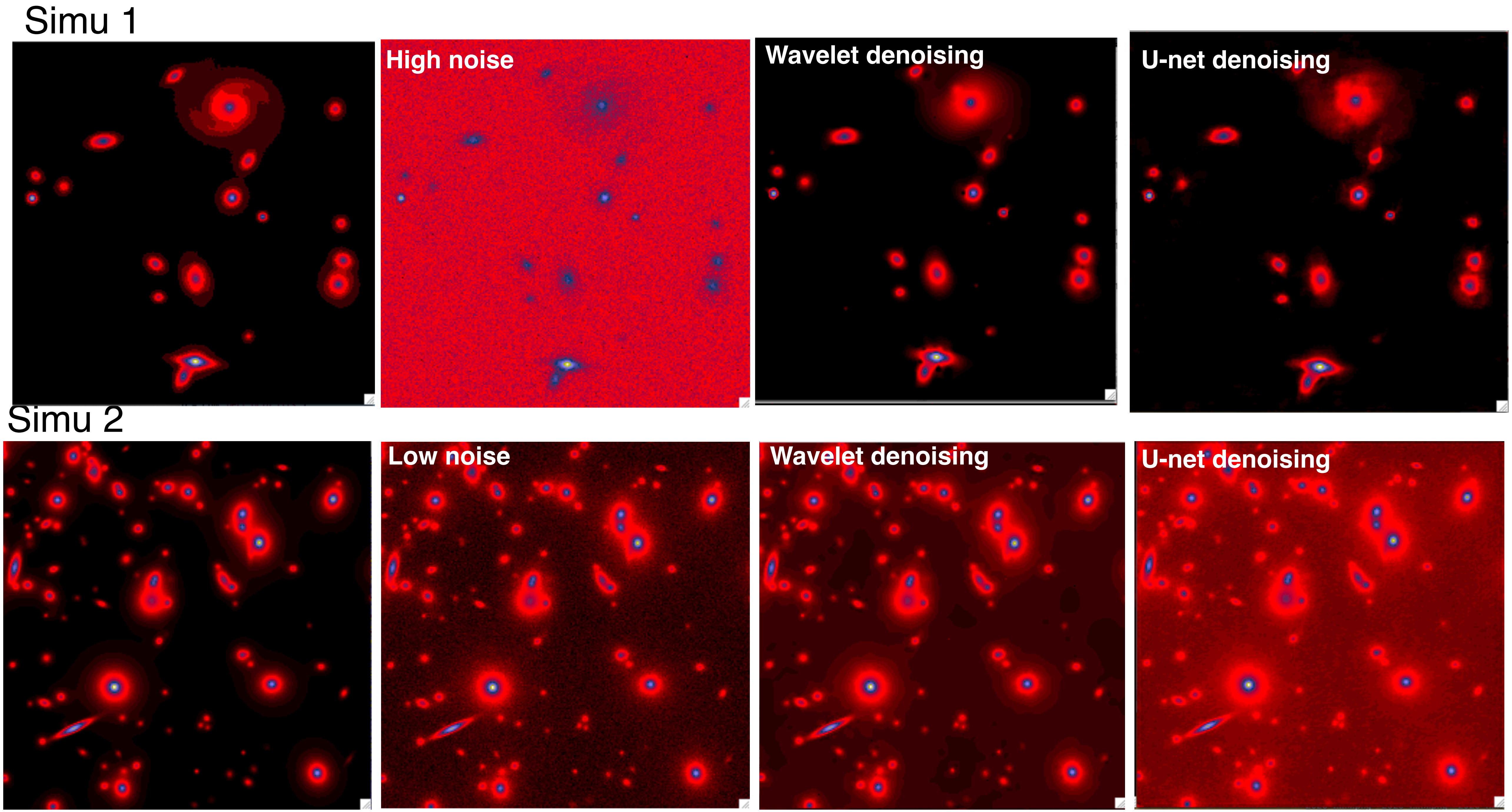
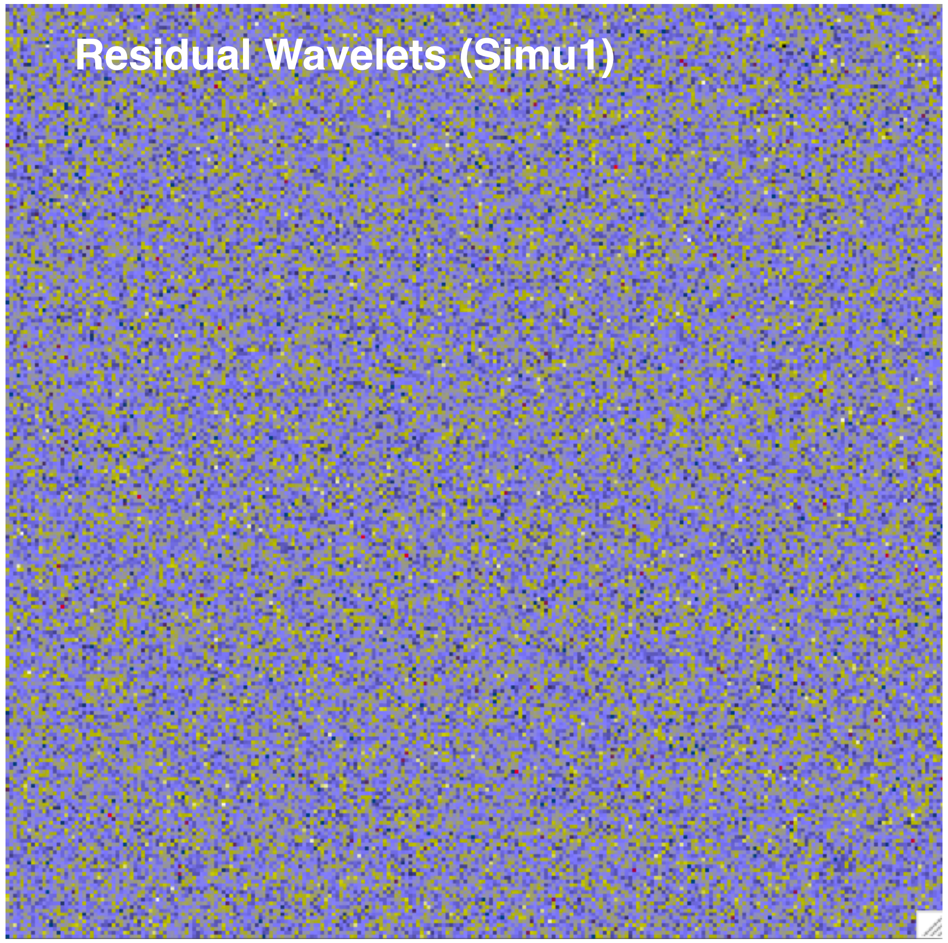


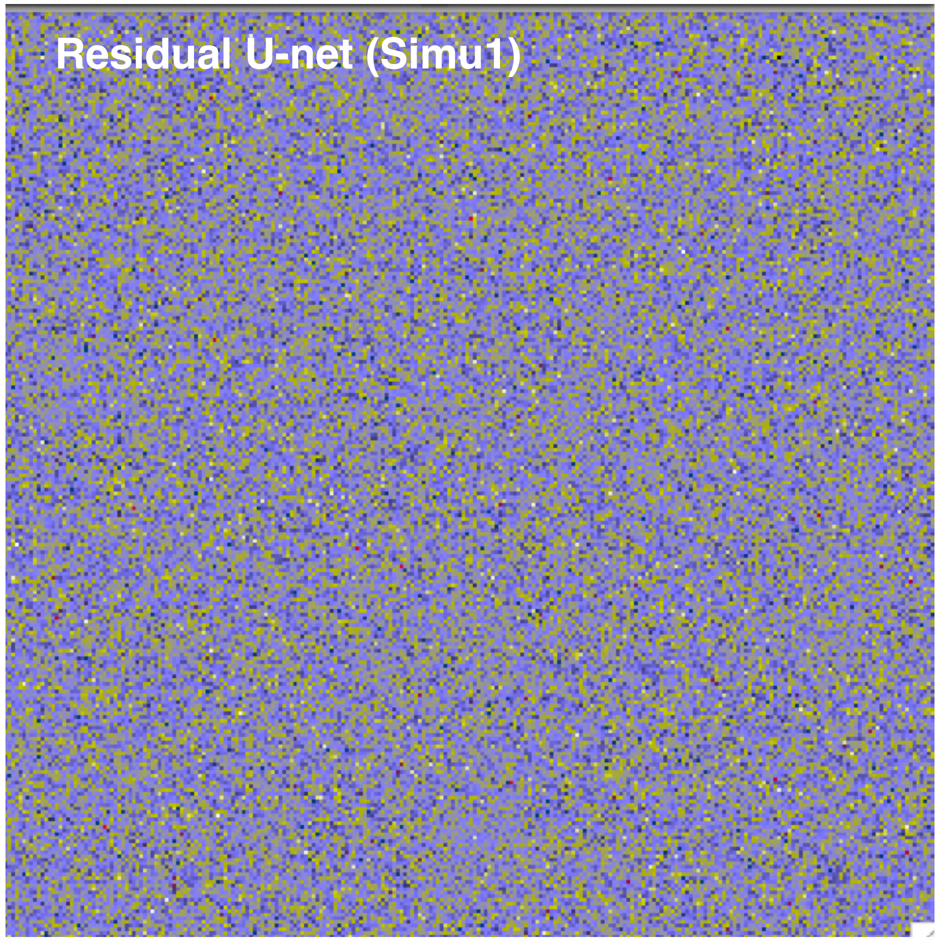
Figure 3: **(False negatives).** The FBPConvNet map $\Psi : \mathbb{R}^m \to \mathbb{R}^N$ [28] is trained to recover images comprised of ellipses from a Radon sampling operator $A \in \mathbb{R}^{m\times N}$. Top left: The image x containing a bird and the SIAM logo. This is a feature the network has not seen. Top right: Cropped original image. Lower left: The cropped FBPConvNet reconstruction from measurements Ax. Lower right: The cropped reconstruction of x from measurements Ax using a sparse regularization decoder (SRD) $\Phi : \mathbb{R}^m \to \mathbb{R}^N$. See Section 9 for further details.

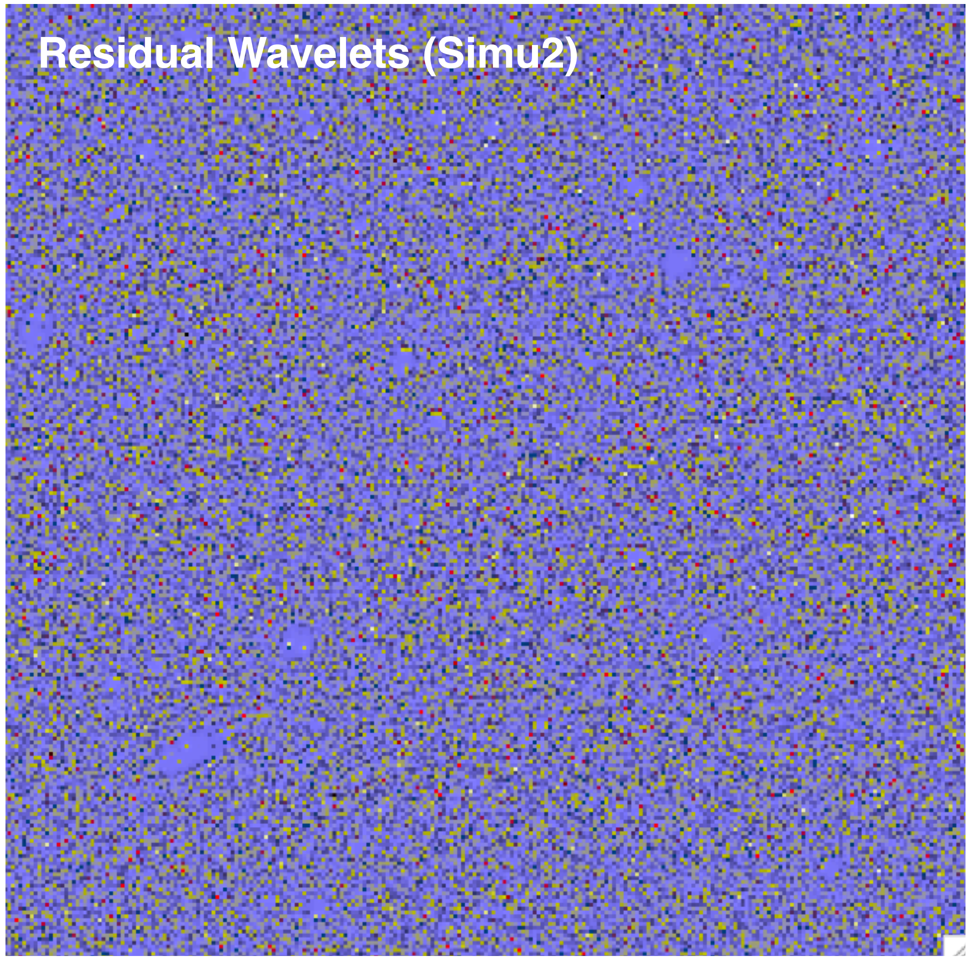
. V. Antun, F. Renna, C. Poon, B. Adcock, and A. C. Hansen. On instabilities of deep learning in image reconstruction – does AI come at a cost? arXiv preprint arXiv:1902.05300, 2019.

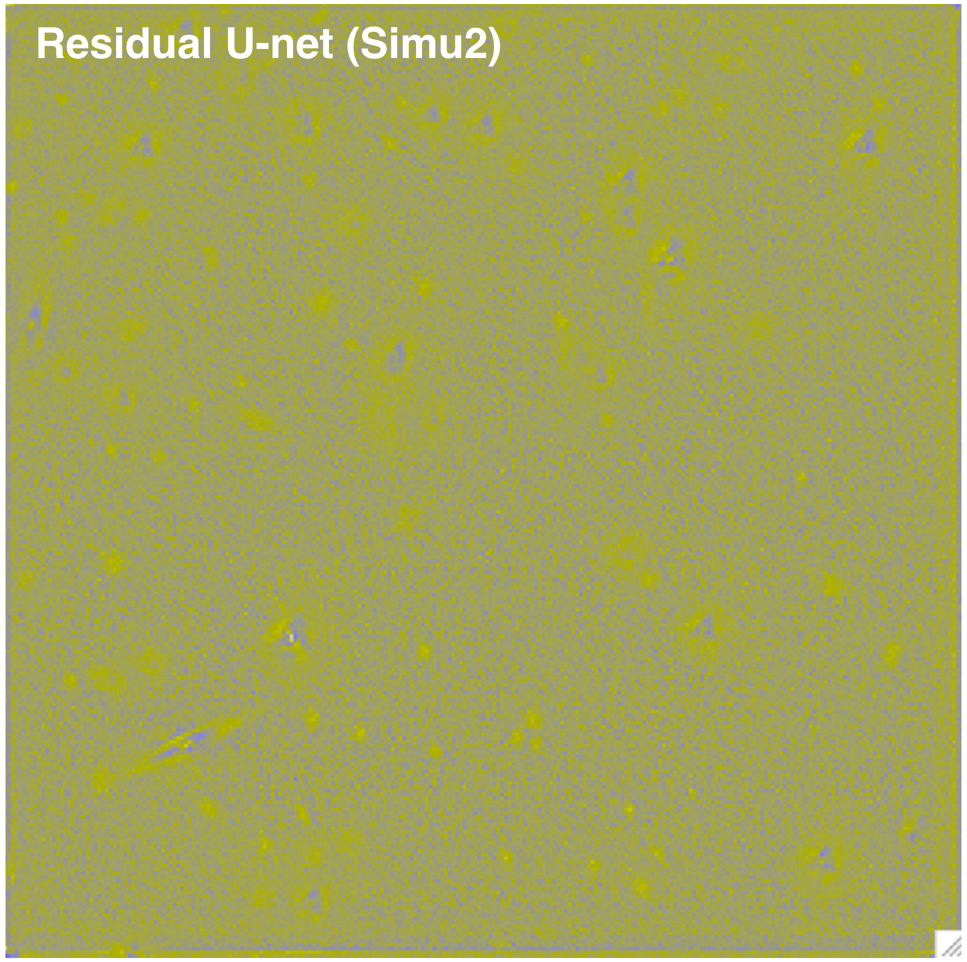


	Wavelets	U-Net
Simu 1:RMSE	**1,1**	1,18
Simu 2: RMSE	**17,35**	29,49

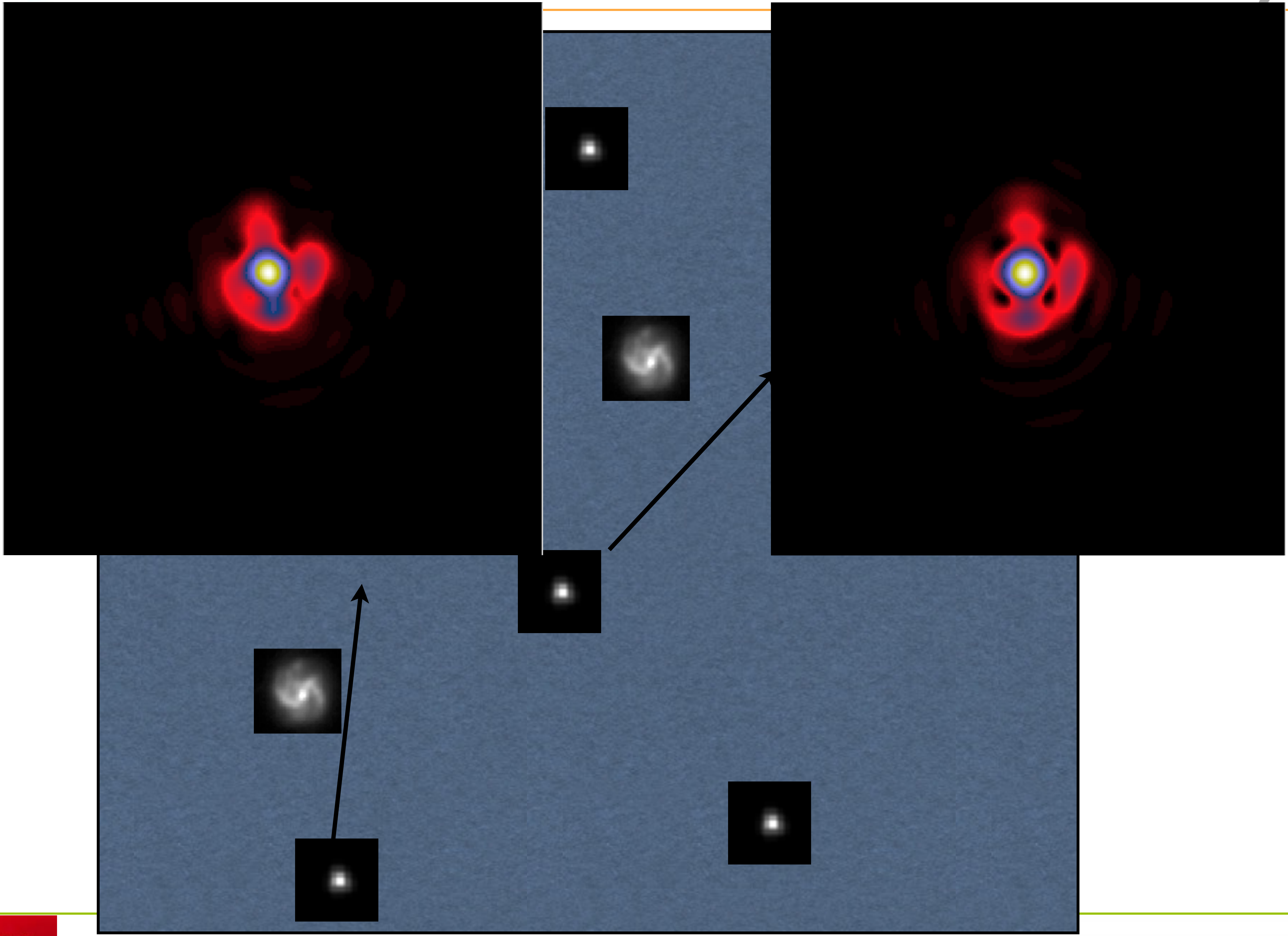

Residual Wavelets (Simu1)


Residual U-net (Simu1)

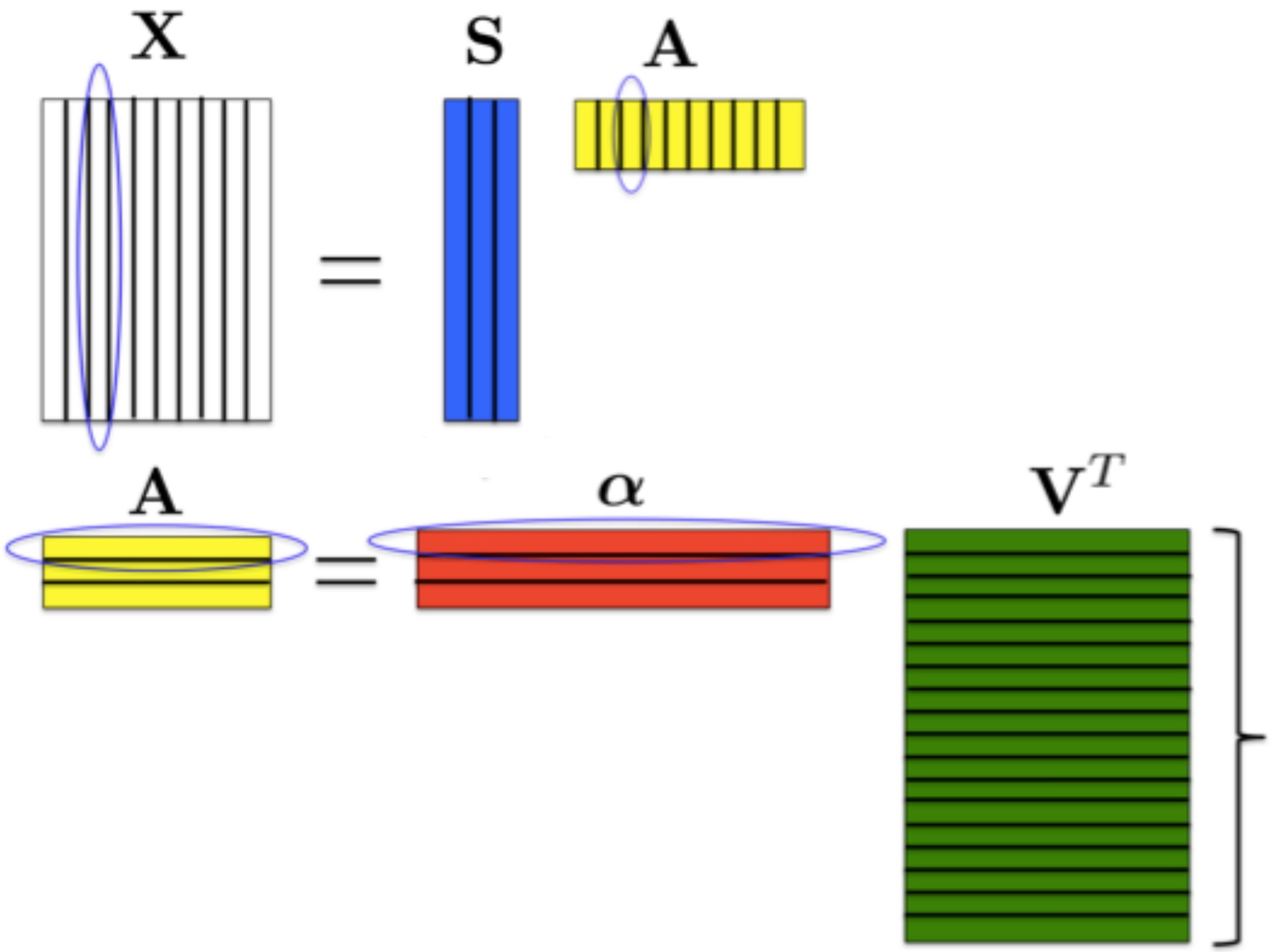

Residual Wavelets (Simu2)


Residual U-net (Simu2)

Space Variant PSF



PSF Field Recovery Equation



$$\min_{S,\alpha} \frac{1}{2}\|Y - M(S\alpha V^\top)\|_2^2 + \sum_{i=1}^{r} \|w_i \odot \Phi s_i\|_1 + \iota_+(S\alpha V^\top) + \iota_\Omega(\alpha)$$

Data fidelity term

Sparsity on the eigen PSF: the PSF should have a sparse representation in an appropriate basis

Positivity Constraint

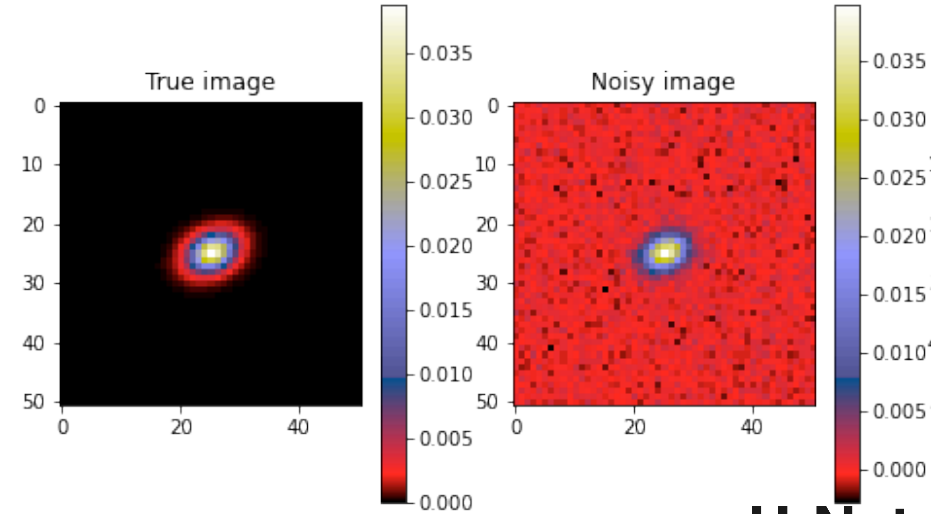
Smoothness of the PSF field constraint : the smaller the difference between two PSFs' positions u_i , u_j , the smaller the difference between their estimated representations

CFIS Simulated PSF Denoising

Aziz Ayed

Tobias Liaudat

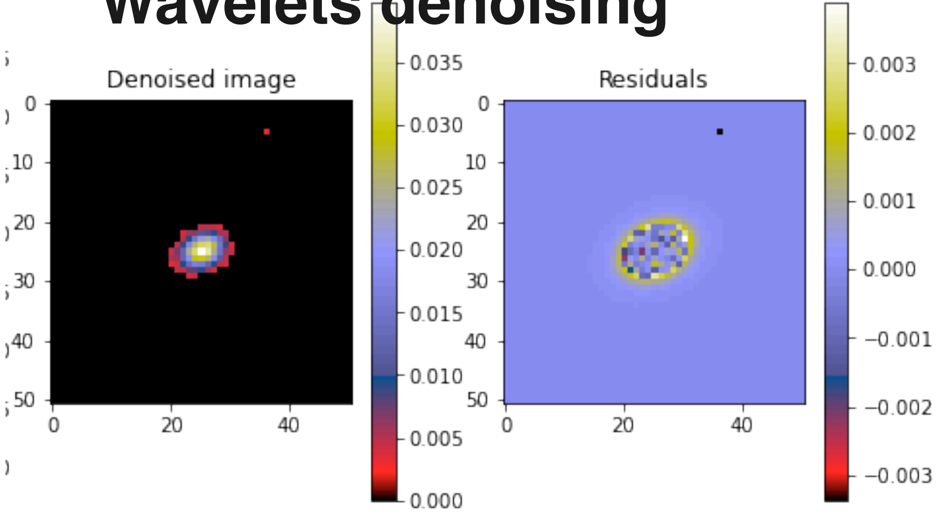
	Train size	Test size	e1 range	e2 range	R2 range	SNR range
Dataset	36 000	9 000	-0.15 - 0.15	-0.15 - 0.15	2.5 - 8	0 - 200



True image

Noisy image

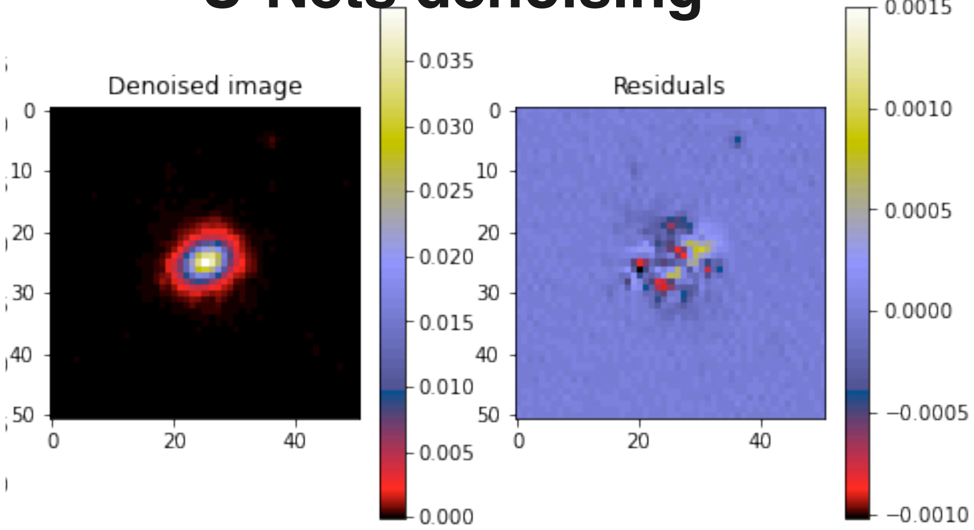
Wavelets denoising



Denoised image

Residuals

U-Nets denoising



Denoised image

Residuals

	Wavelets	U-Nets
Pixel RMSE	1.681e-04	**5.879e-05**

CFIS Simulated PSF Field Recovery


Aziz Ayed


Tobias Liaudat

Sparsity on the eigen PSF: the PSF should have a sparse representation in an appropriate basis

$$\min_{S,\alpha} \frac{1}{2}\|Y - M(S\alpha V^\top)\|_2^2 + \boxed{\sum_{i=1}^{r} \|w_i \odot \Phi s_i\|_1} + \iota_+(S\alpha V^\top) + \iota_\Omega(\alpha)$$

Data fidelity term

Positivity Constraint

Smoothness of the PSF field constraint : the smaller the difference between two PSFs' positions u_i , u_j , the smaller the difference between their estimated representations

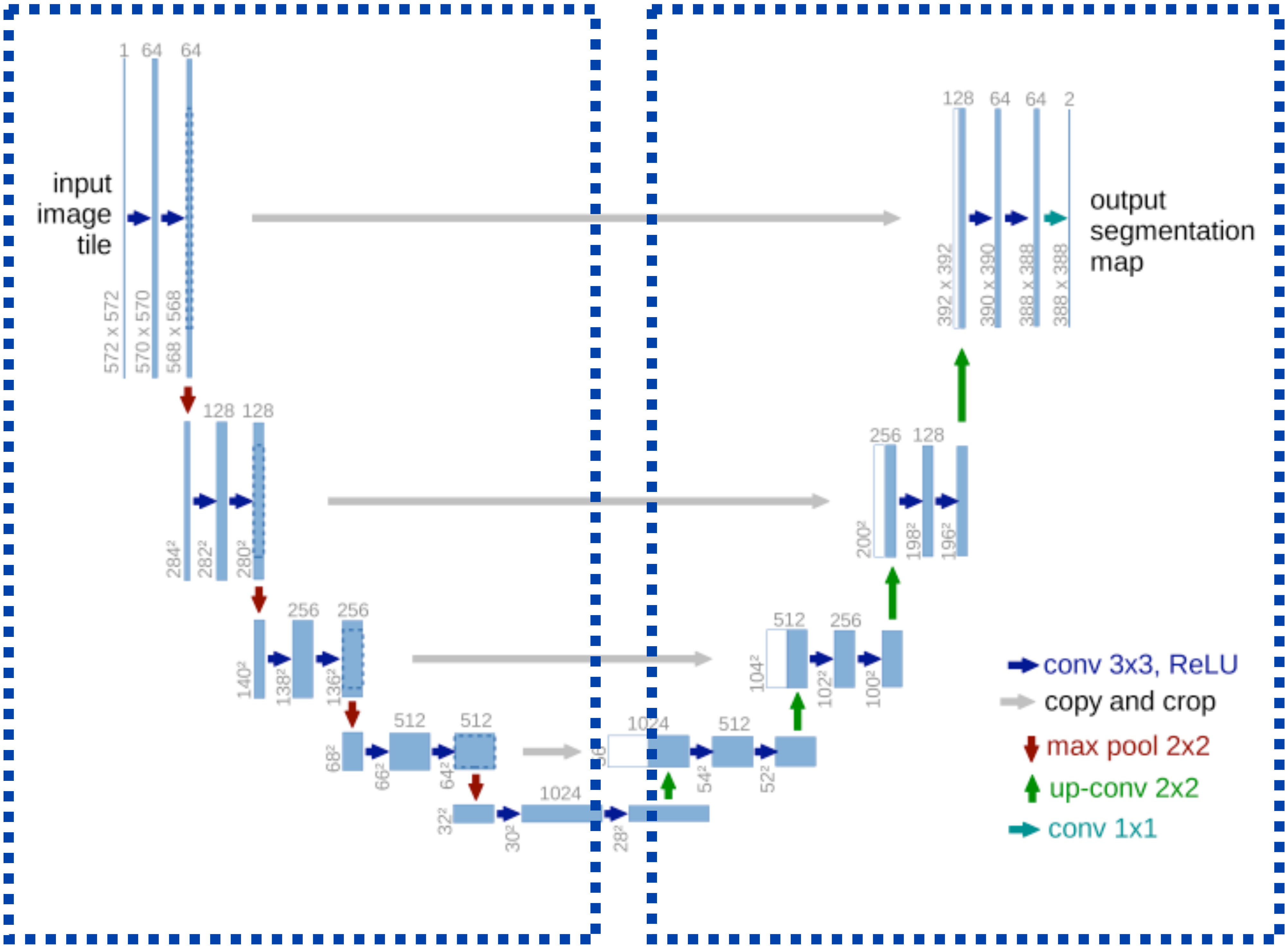
Replace the prox operator (i.e. wavelet thresholding) by the U-net PSF denoising

	MCCD - Wavelets	MCCD - U-Nets
Pixel RMSE	**8.34199e-05**	2.66538e-04

U-nets

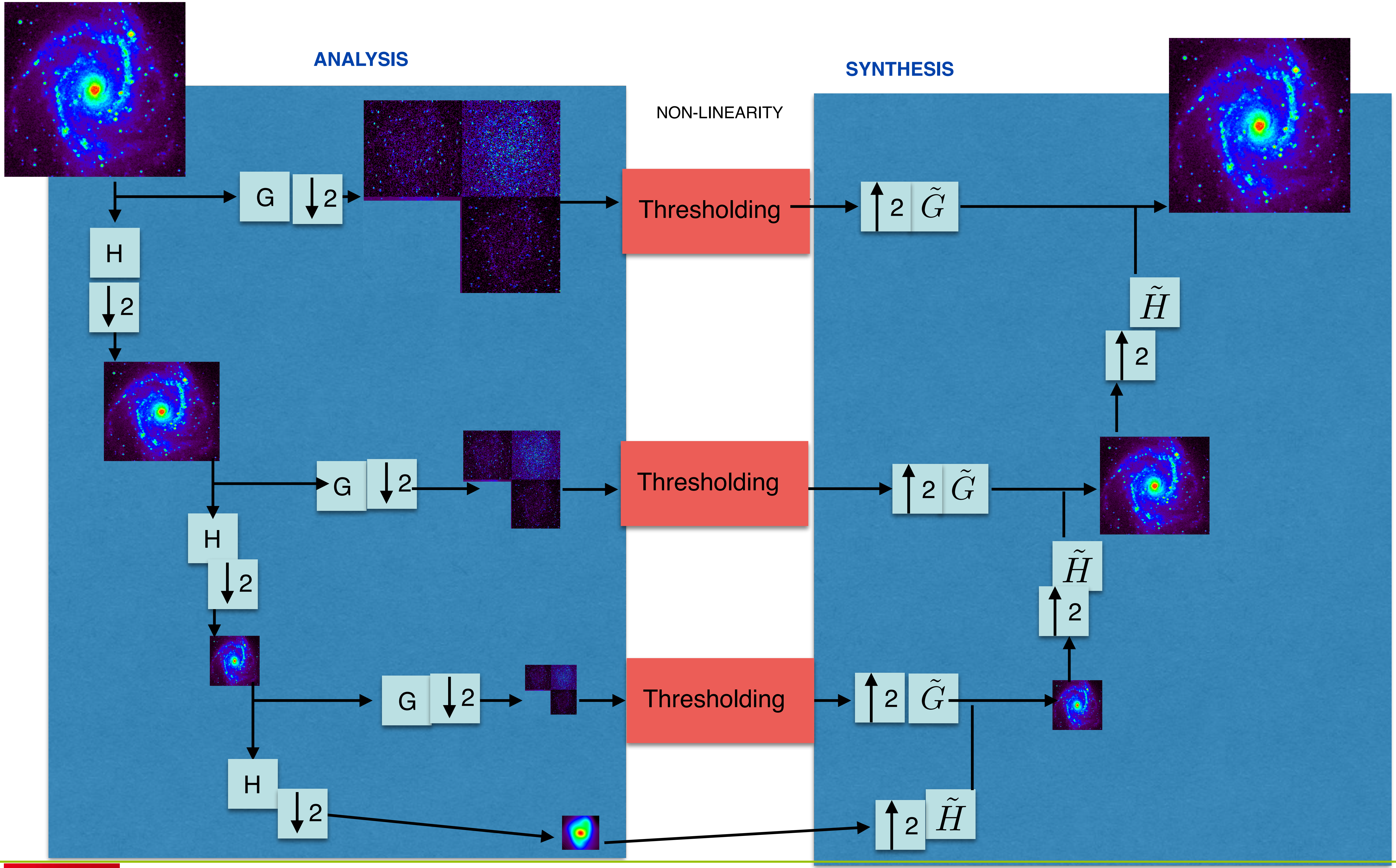
ANALYSIS

SYNTHESIS



From [Ronneberger et al, MICCAI 2015]

Orthogonal Wavelets



U-nets versus Wavelet

Many similarities:

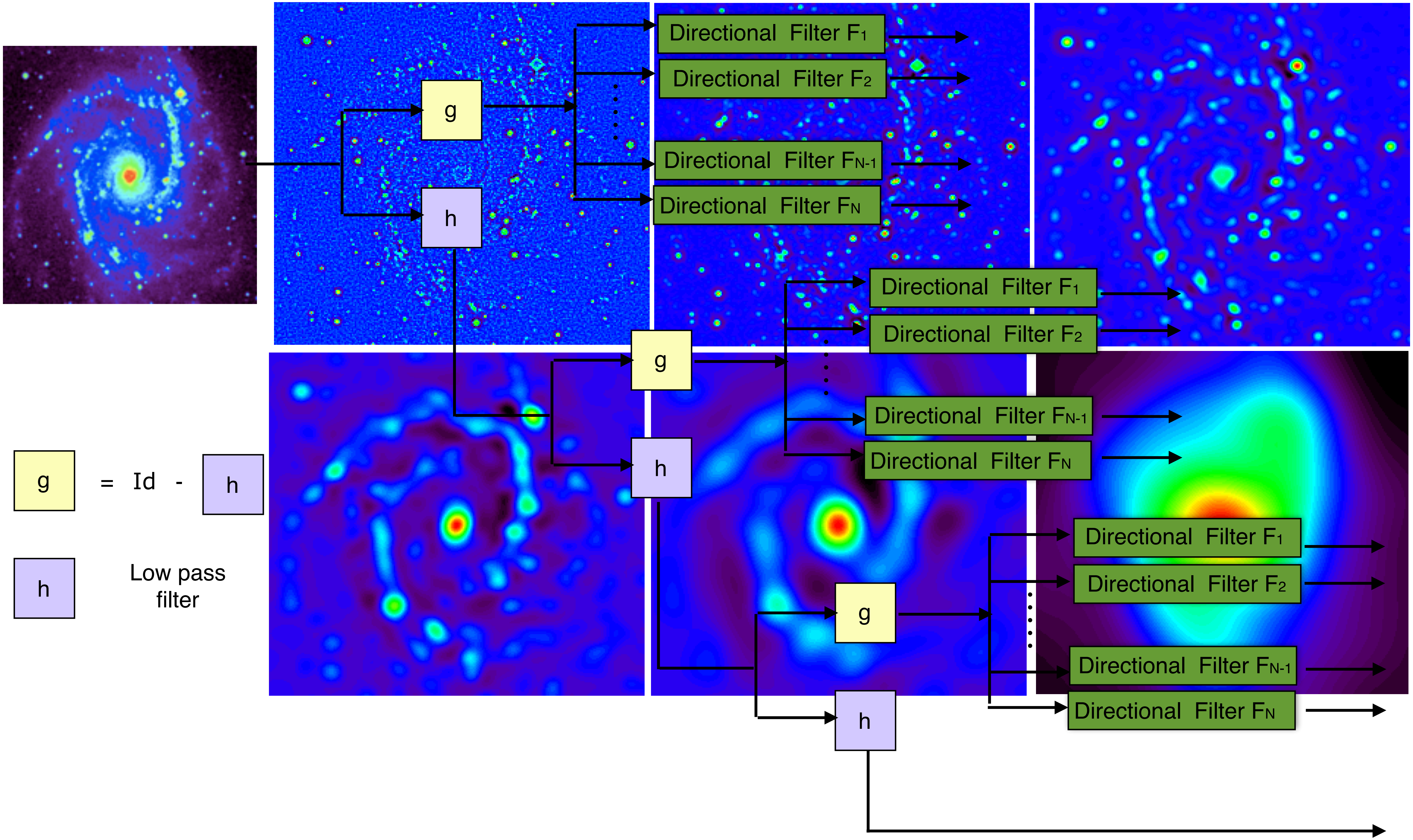
- **Expanding path and contracting path**: the U-nets two parts are very similar to ***synthesis and analysis*** concepts in sparse representations. This has motivated the use of wavelets to implement in the U-net average pooling and unpooling in the expanding (Ye et al. 2018b; Han & Ye 2018).
- **Multi-scale** : Similarly to wavelets, U-nets present a multi-scale approach, which **analyse the signal at different resolutions.**
- **Non-Linearities: Both use non-linearities** (thresholding for the wavelet and RELU for the U-nets).

And Significant Differences:

- **Massive Learning**: made possible by the flourishing of optimization algorithms with the power of computers (GPU) and the use of large available data sets for training.
- **Highly non-linear processing: wavelets** apply only one non-linearity, while U-nets are everywhere non-linear (analysis and synthesis).
- **Exact Reconstruction for Wavelets:** note however that some recent NN are reversible.

Can we add massive learning in a wavelet framework ?

Curvelets



g

h

Directional Filter F_1

Directional Filter F_2

Directional Filter F_{N-1}

Directional Filter F_N

g

h

Directional Filter F_1

Directional Filter F_2

Directional Filter F_{N-1}

Directional Filter F_N

g

h

Directional Filter F_1

Directional Filter F_2

Directional Filter F_{N-1}

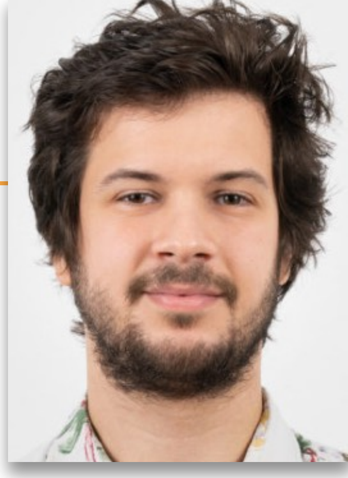
Directional Filter F_N

g = Id - h

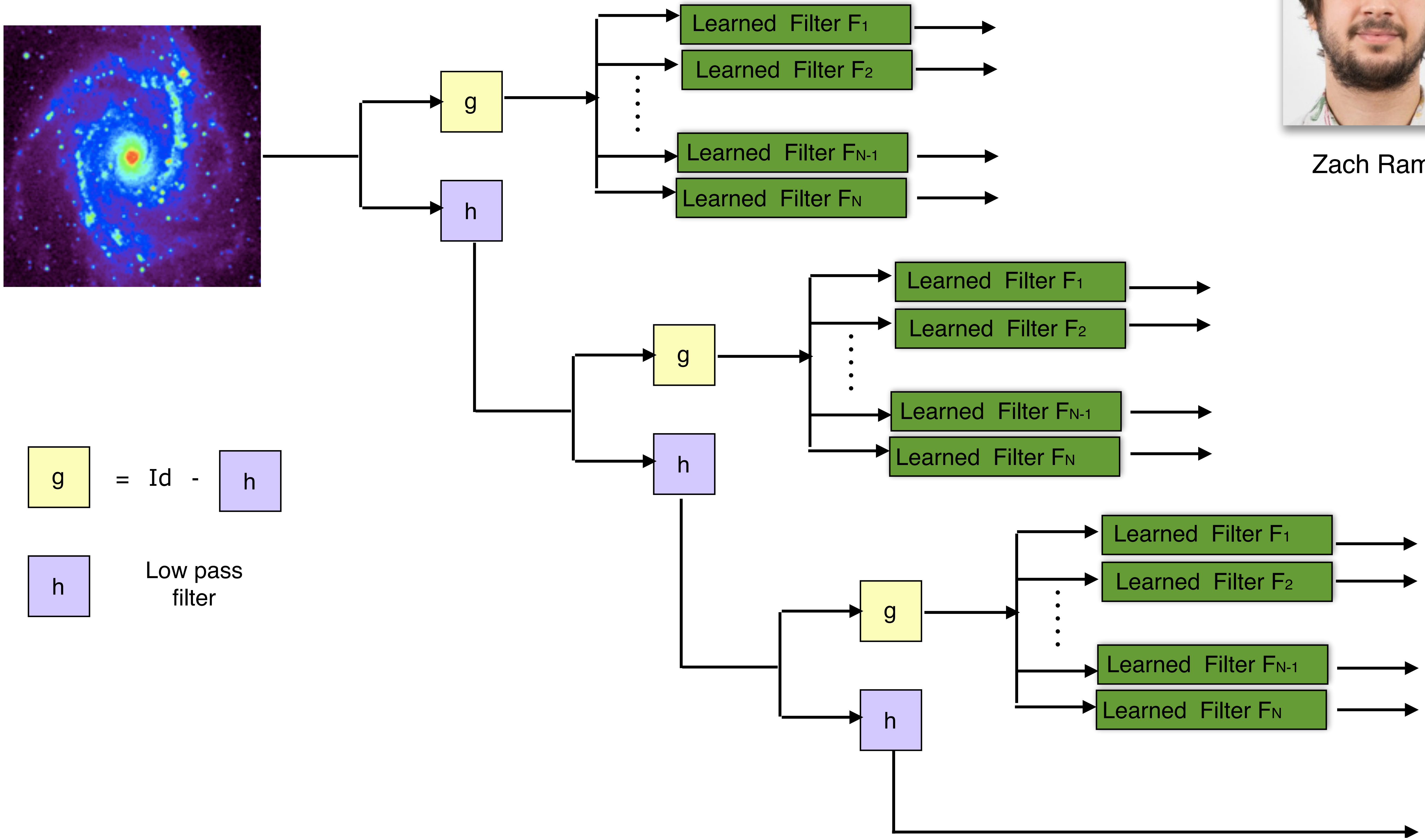
h Low pass filter

J.-L. Starck, E. J. Candès, and D. L. Donoho, “The Curvelet Transform for Image Denoising,” *IEEE Trans. IMAGE Process.*, vol. 11, no. 6, 2002.

Learnlets Analysis

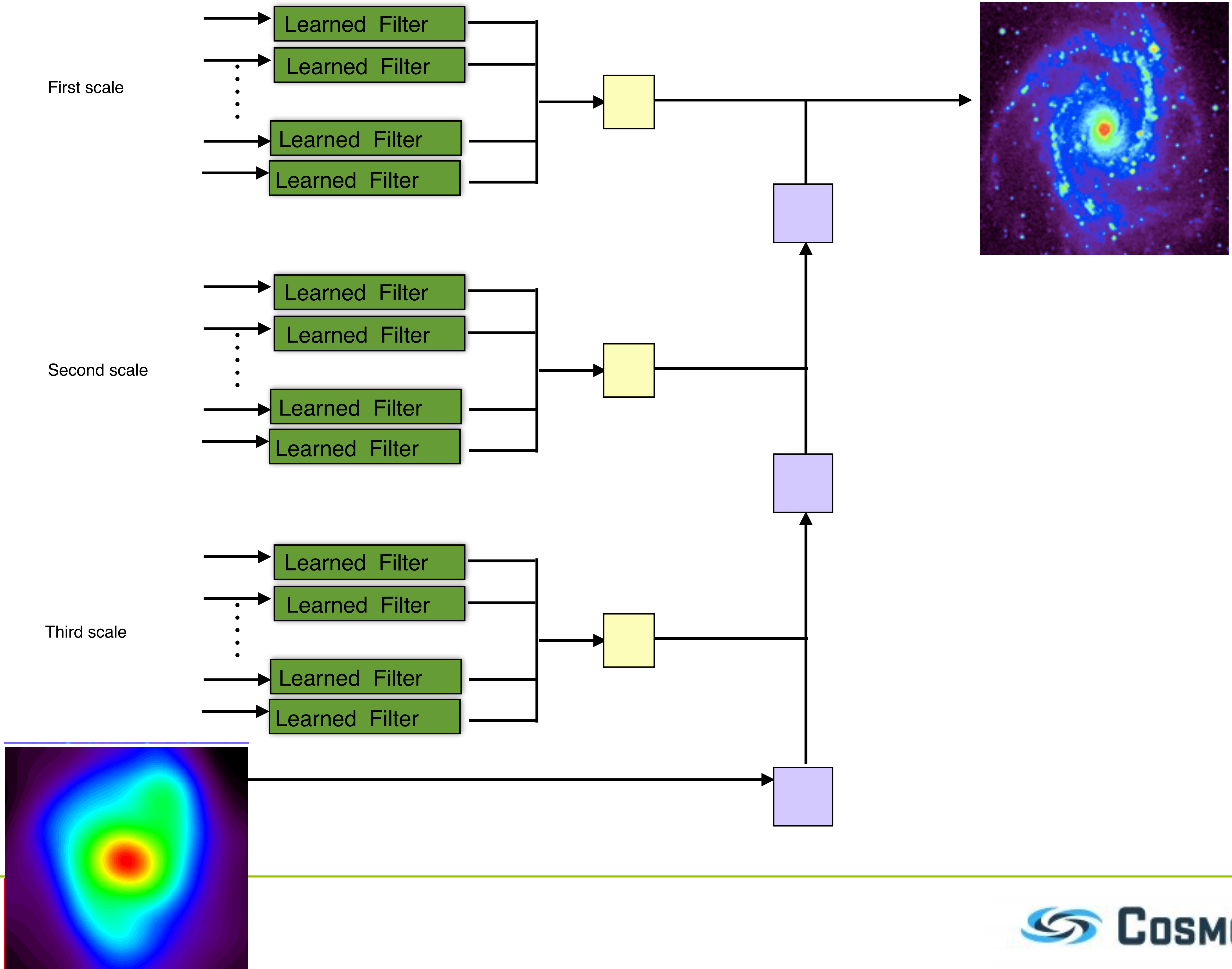


Zach Ramzi



Z. Ramzi, J. -L. Starck, T. Moreau and P. Ciuciu, "Wavelets in the Deep Learning Era," *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1417-1421, doi: 10.23919/Eusipco47968.2020.9287317.

Learnlets Synthesis



First scale

Learned Filter

Learned Filter

Learned Filter

Learned Filter

Second scale

Learned Filter

Learned Filter

Learned Filter

Learned Filter

Third scale

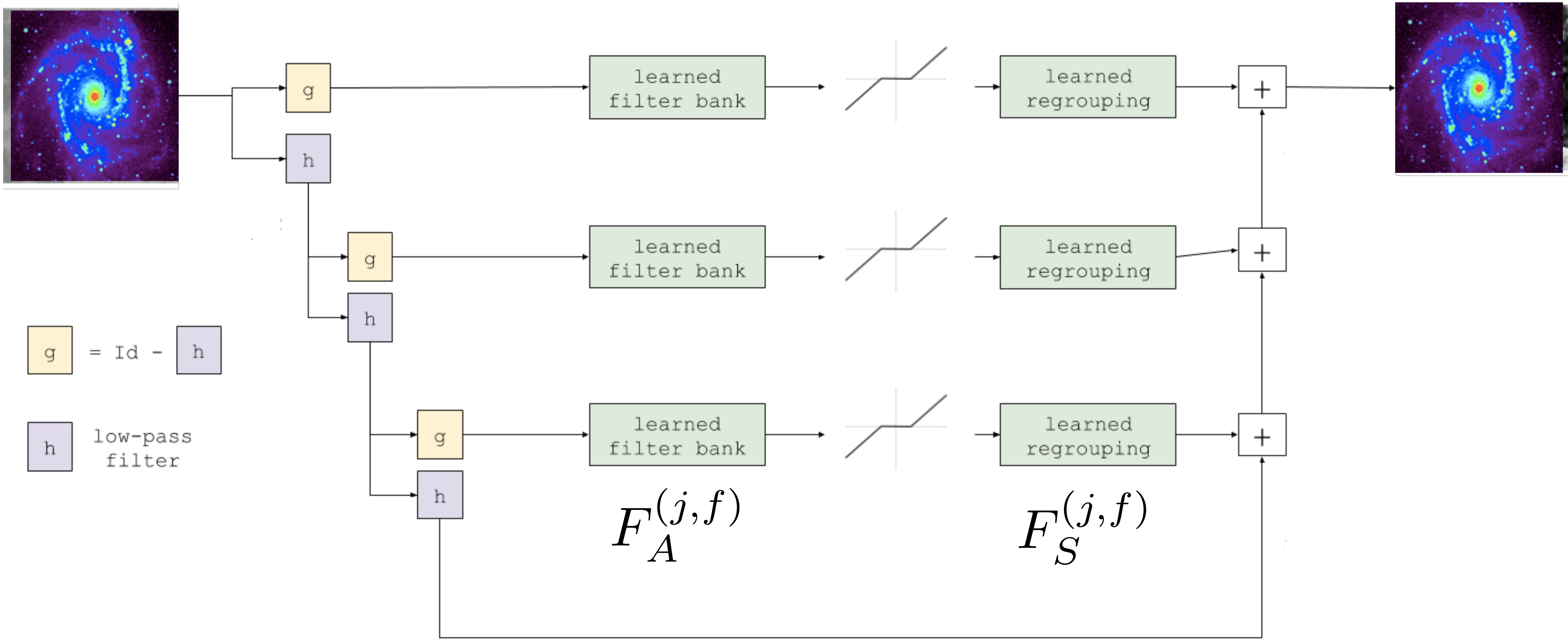
Learned Filter

Learned Filter

Learned Filter

Learned Filter

Learnlet Denoising



Learnlet Soft-Thresholding

The threshold level is computed as a linear function of the estimated propagated noise level at a given scale.

Soft Thresholding at scale j and for filter f:

$$\tilde{\alpha}_{j,f} = \mathrm{sign}(\alpha_{j,f})\ [\alpha_{j,f} - T_{j,f}]_{+}$$

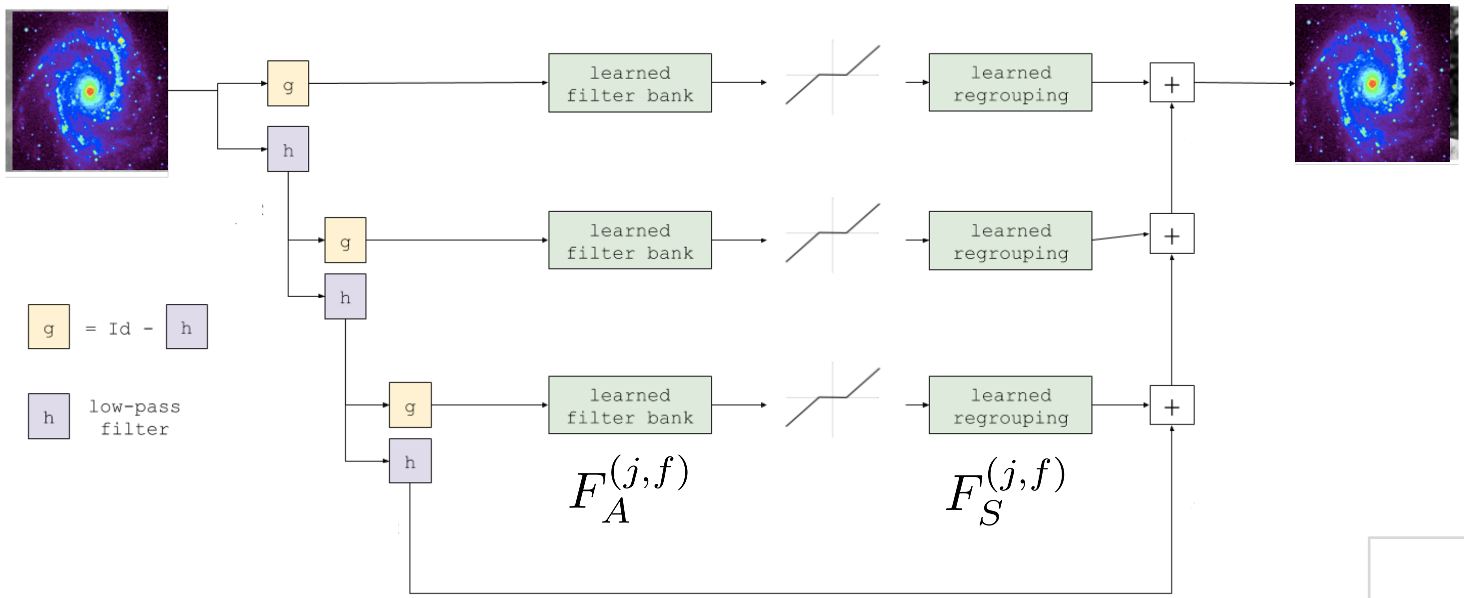
$$T_{j,f} = k_\theta\ \sigma_D \longrightarrow$$

Estimated noise level

Learnable thresholding level

Learnlet Exact Reconstruction

Exact reconstruction using **a compensation synthesis filter** computed as a function of the others, corresponding to an identity analysis filter.

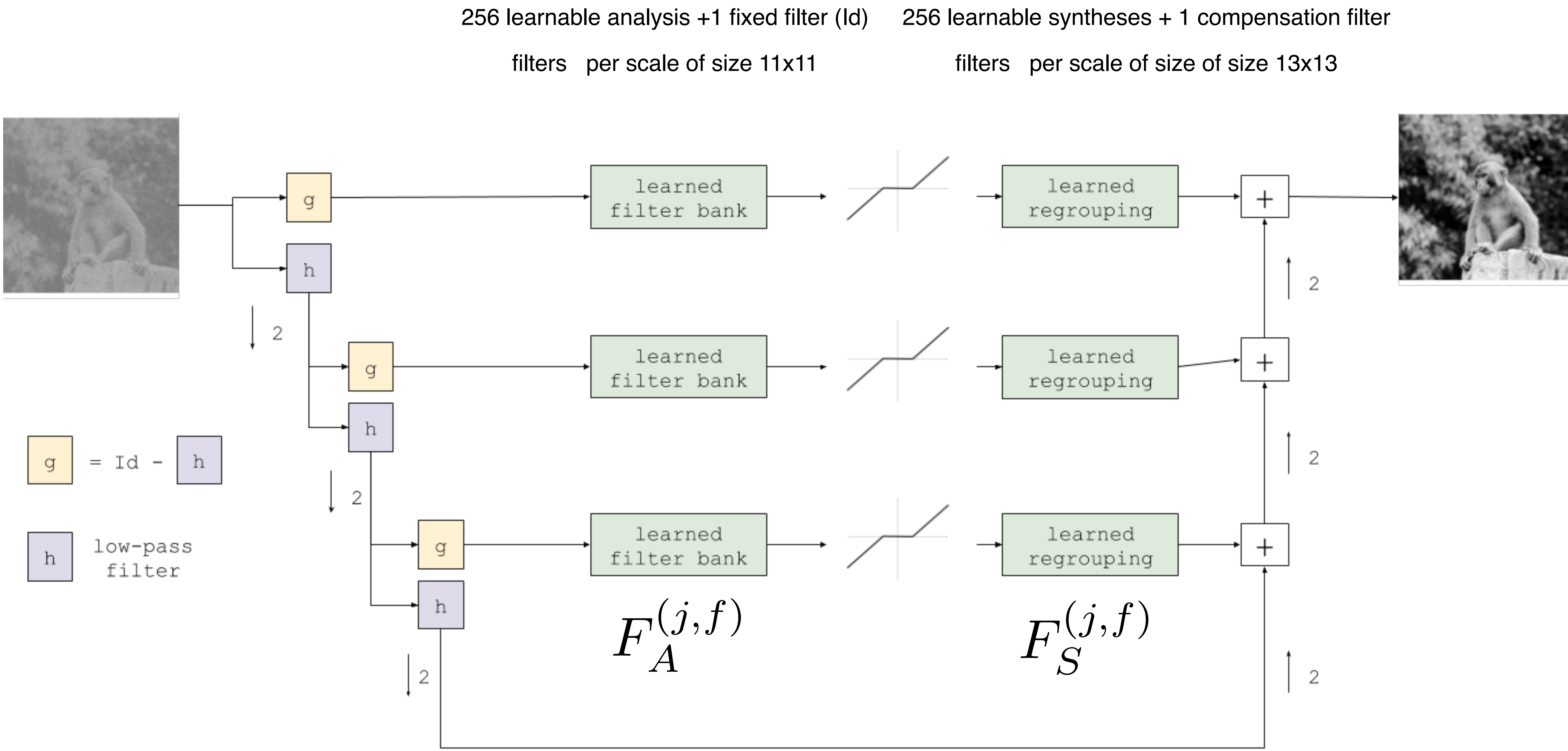


Synthesis filters

Analysis filters

$$F_S^{(j,1)} = Id - \sum_{f=2}^{N} F_S^{(j,f)} \otimes F_A^{(j,f)}$$

Learnlet Denoising



- This amounts to 372k trainable parameters, only **one hundredth** of the size of a U-net.
- Frame with **exact reconstruction**
- Filters with **zero mean average** (convenient for noise modeling)

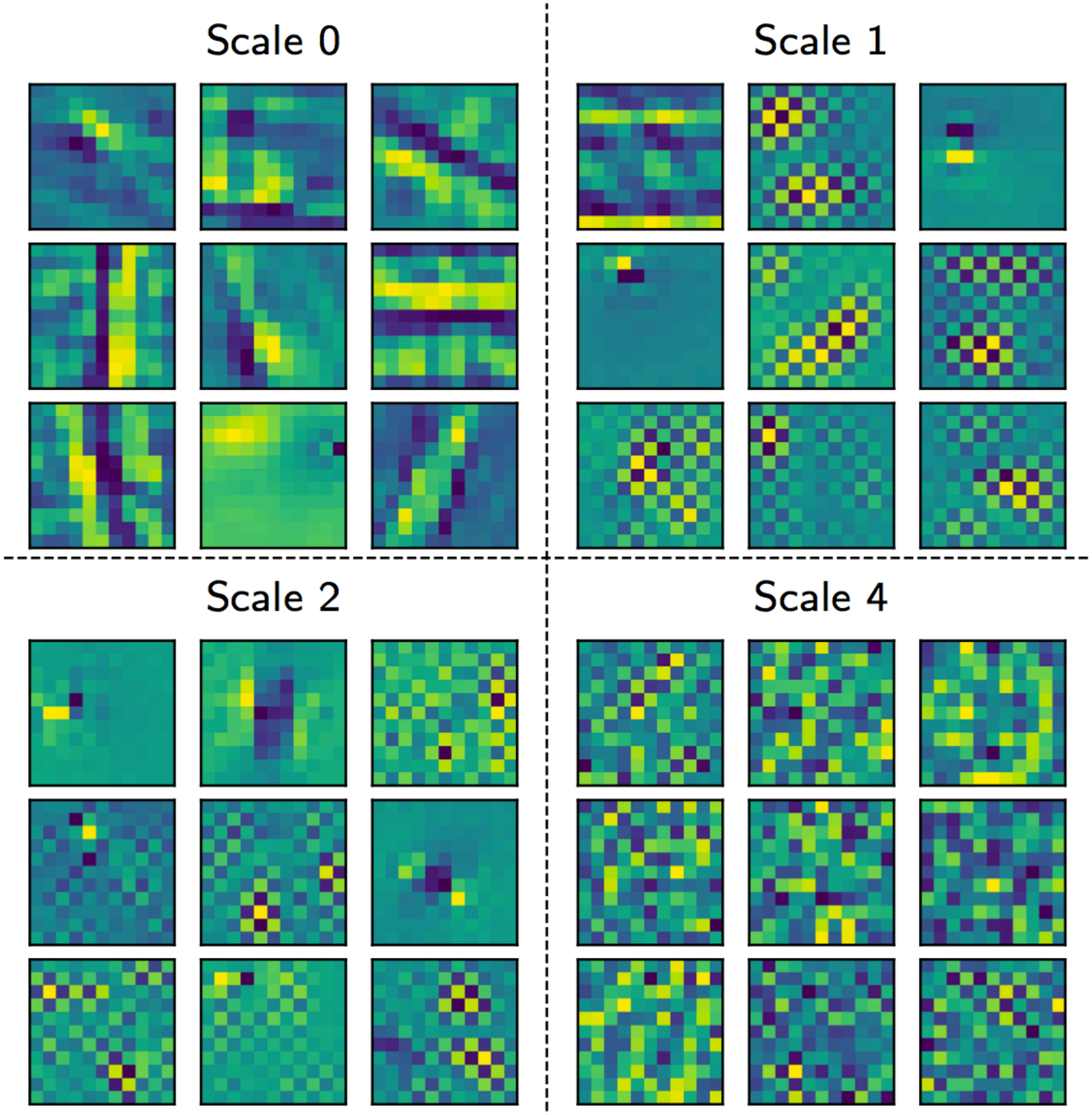
Training and Evaluating

Datasets: BSD500 for training and validation; BSD68 for testing.

BSD500 sample

BSD68
sample

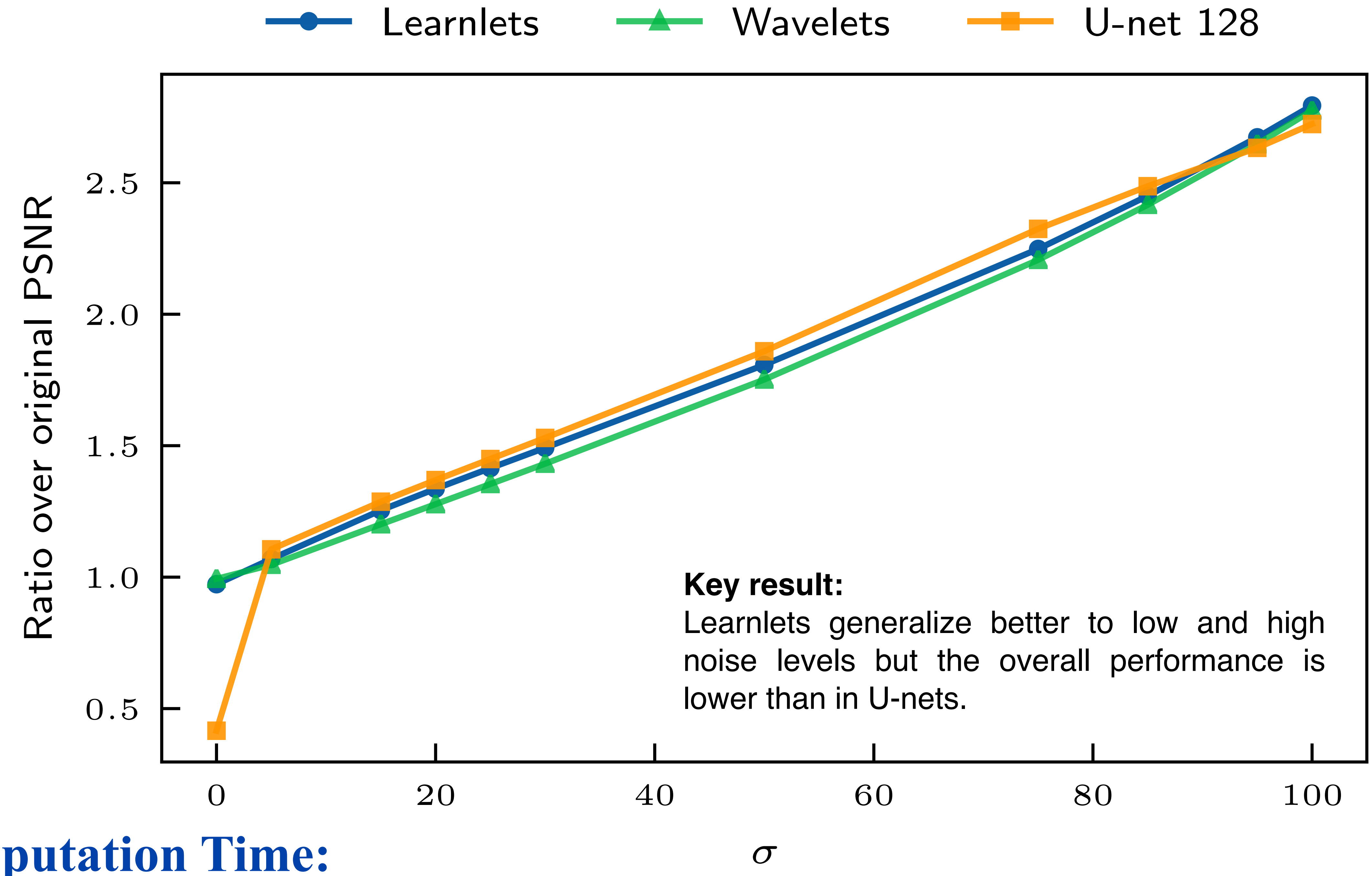
Learnlet Filters



Learnlets' analysis filters constitute meaningful designs (and they make sense!).

Quantitative Results

Generalisation:

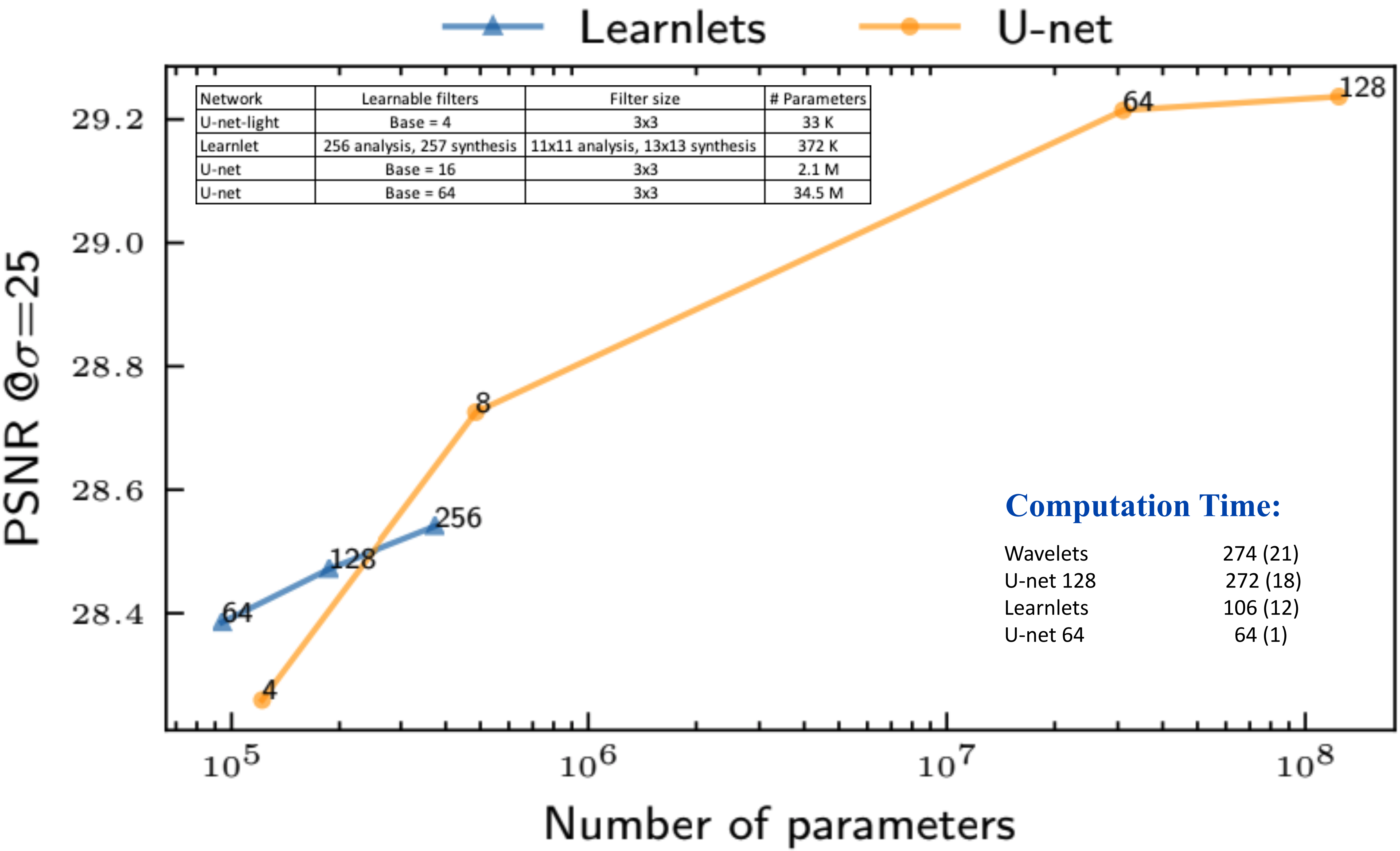


Key result:
Learnlets generalize better to low and high noise levels but the overall performance is lower than in U-nets.

Computation Time:

Model name	Wavelets	U-net 128	Learnlets	U-net 64
Denoising runtime in ms (std)	274 (21)	272 (18)	106 (12)	64 (1)

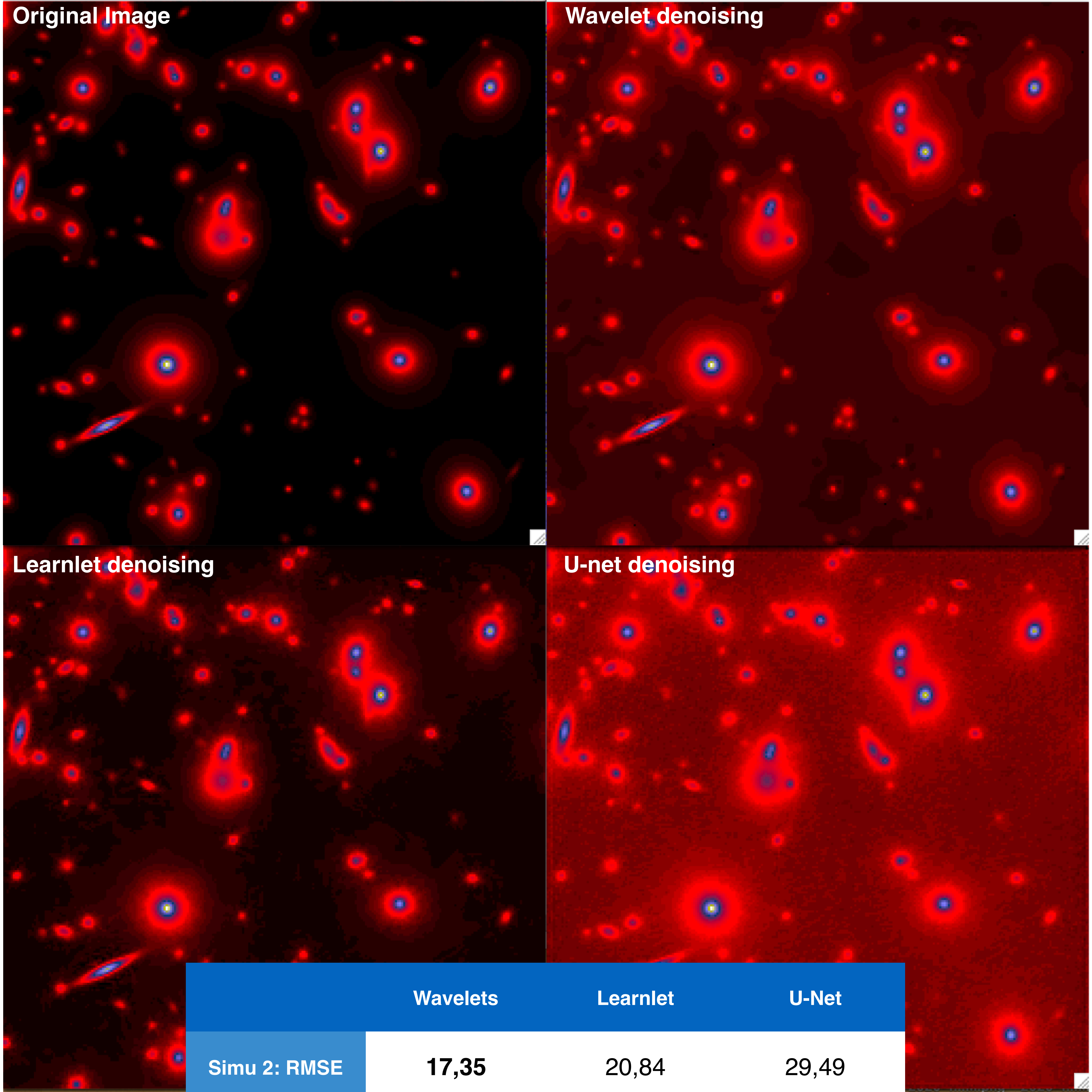
U-Net-light



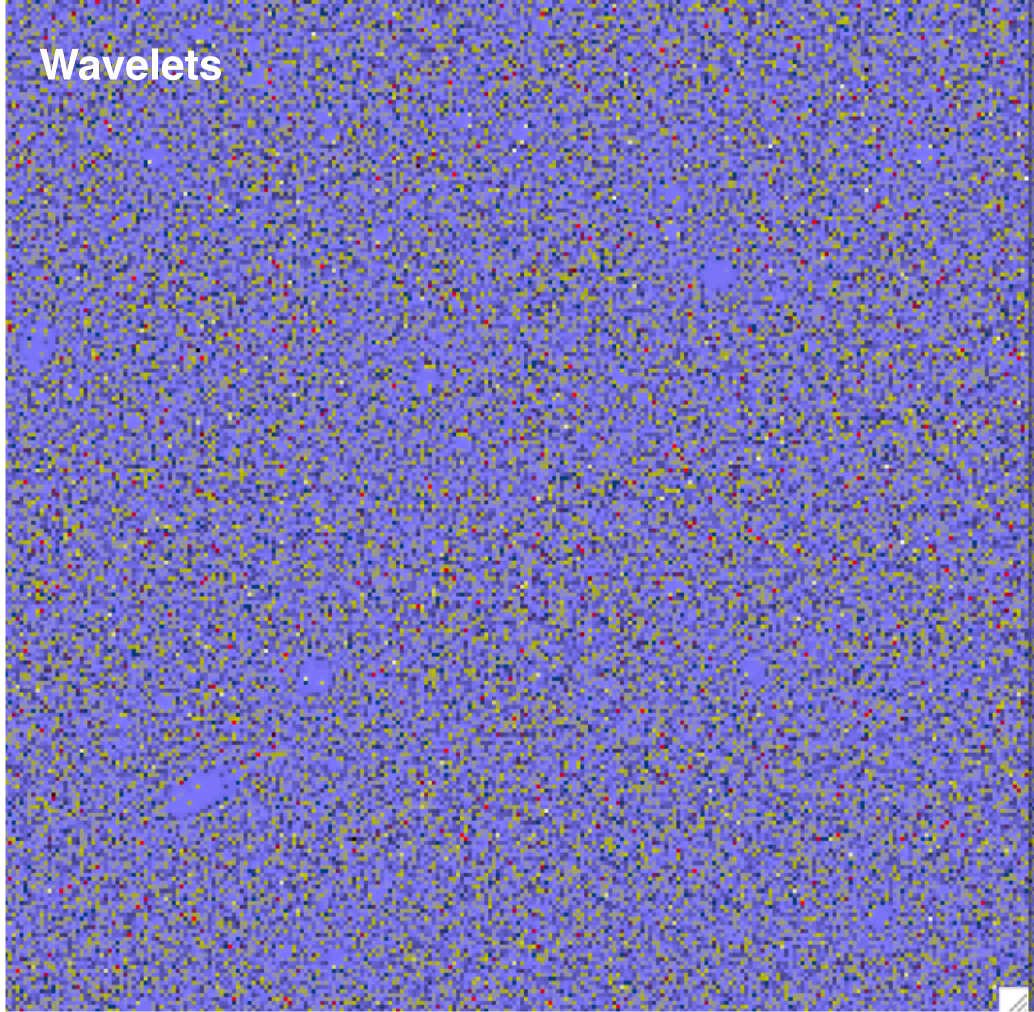
Network	Learnable filters	Filter size	# Parameters
U-net-light	Base = 4	3x3	33 K
Learnlet	256 analysis, 257 synthesis	11x11 analysis, 13x13 synthesis	372 K
U-net	Base = 16	3x3	2.1 M
U-net	Base = 64	3x3	34.5 M

Computation Time:

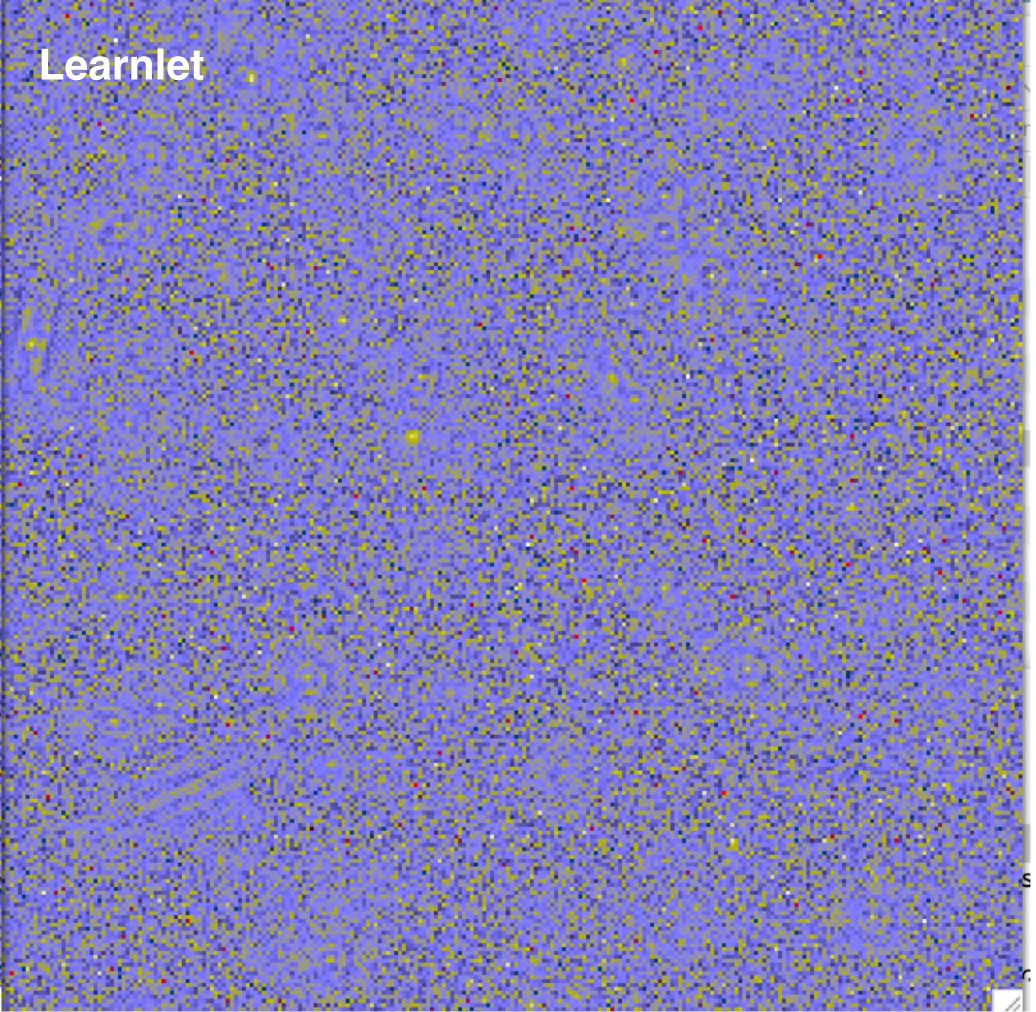
Wavelets	274 (21)
U-net 128	272 (18)
Learnlets	106 (12)
U-net 64	64 (1)



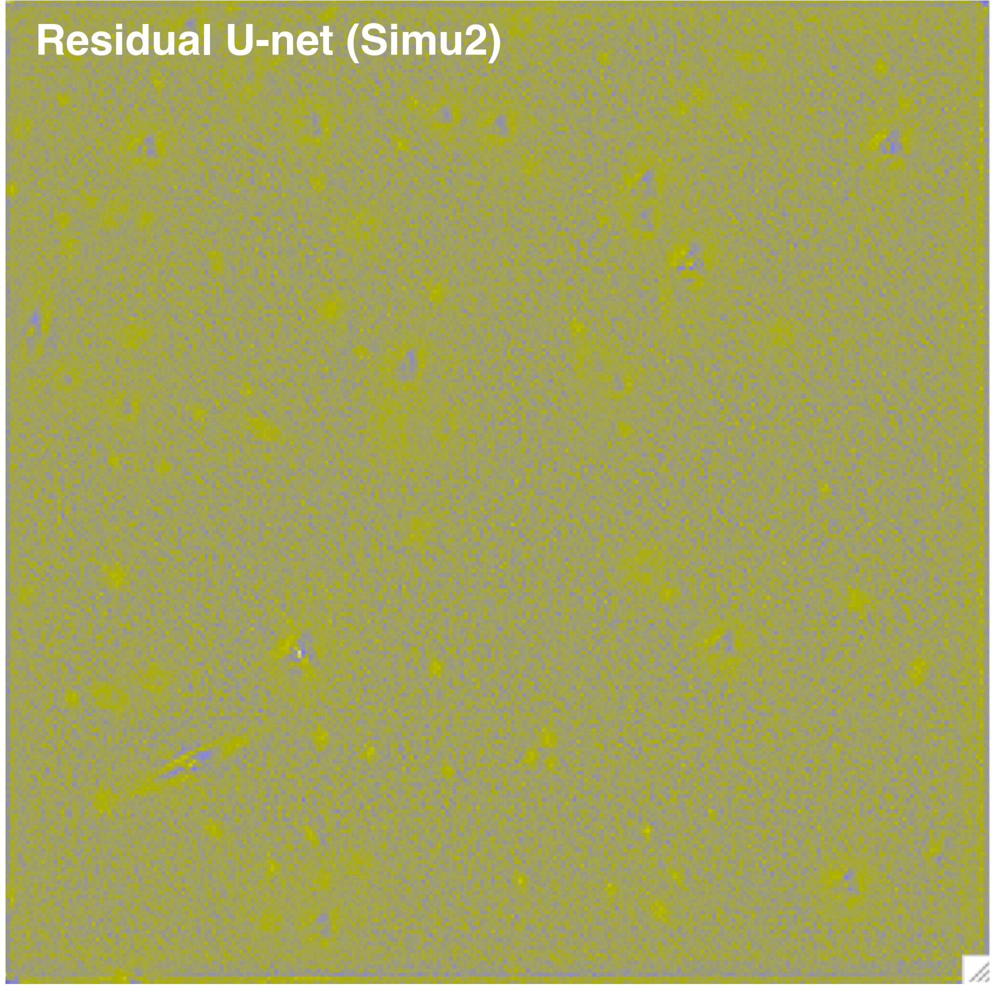
	Wavelets	Learnlet	U-Net
Simu 2: RMSE	**17,35**	20,84	29,49



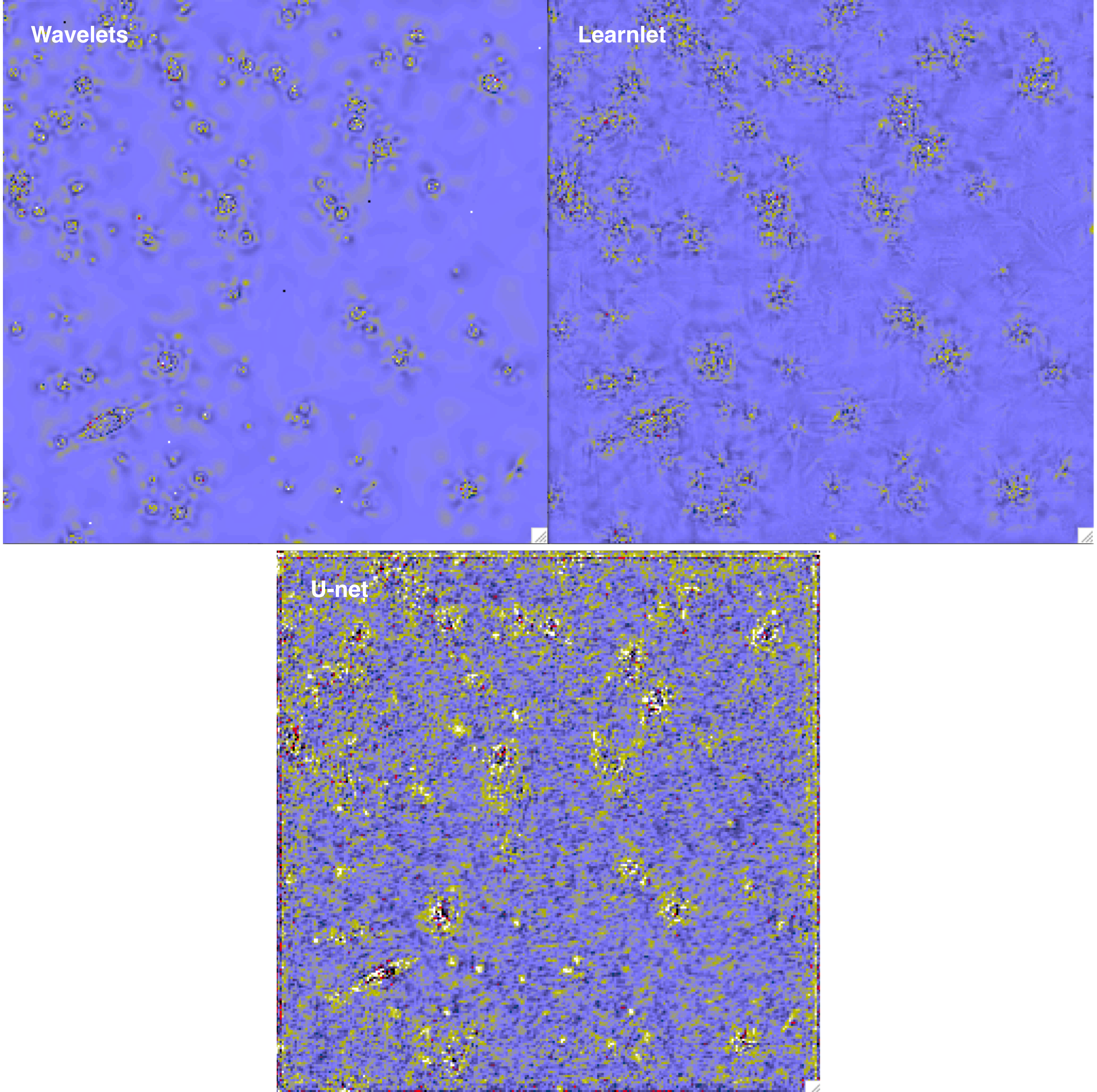
Wavelets



Learnlet



Residual U-net (Simu2)


Wavelets
Learnlet
U-net

CFIS Simulated PSF Field Recovery

Aziz Ayed

Tobias Liaudat

Sparsity on the eigen PSF: the PSF should have a sparse representation in an appropriate basis

$$\min_{S,\alpha} \frac{1}{2}\|Y - M(S\alpha V^\top)\|_2^2 + \boxed{\sum_{i=1}^{r} \|w_i \odot \Phi s_i\|_1} + \iota_+(S\alpha V^\top) + \iota_\Omega(\alpha)$$

Data fidelity term

Positivity Constraint

Smoothness of the PSF field constraint : the smaller the difference between two PSFs' positions u_i , u_j , the smaller the difference between their estimated representations

Replace the prox operator (i.e. wavelet thresholding) by the U-net PSF denoising

Replace the prox operator (i.e. wavelet thresholding) by the Learnlet PSF denoising

	MCCD - Wavelets	MCCD - U-Nets	MCCD - Learnlets
Pixel RMSE	8.34199e-05	2.66538e-04	**5.76903e-05**

Conclusions

- ✓ **Sparse recovery techniques we were using in astrophysics very efficient, but now clearly out performed by Deep Learning.**

- ✓ **Need to be very cautious with Deep Learning (problem of generalisation, data consistency, lack of theoretical guarantees, etc).**

- ✓ **Learnlets bridge the gap between sparsity and neural networks**
  - ➡ Massive learning as in neural networks, **but**
  - ➡ **Linearity**: adaptability to all noise models (non stationary, correlated, Poisson, etc...)
  - ➡ **Exact reconstruction** and **Very fast** (GPU implementation).
  - ➡ **a better understanding** of how results are obtained, with all the theoretical guarantees existing in the area of parsimony.

- ✓ **Learnlet Performances**
  - ➡ Better than Wavelets or Curvelets, but Inferior to U-nets.
  - ➡ Learn 370k parameters, while U-net need 34 millions parameters (100 times more).
  - ➡ U-net-lights do not improve over learnlets.
  - ➡ **Generalize better** than U-nets (learnlets can replace wavelets in any existing pipeline without any risk).
  - ➡ **BUT:** **it doesn't explain** all the gain from the Neural Networks (maybe explained by the additional non-linearities).

==> TRADE OFF BETWEEN PERFORMANCE AND GENERALISATION