Weak Shear Mass Reconstruction Using Wavelets

J.-L. Starck, S. Pires and A. Refregier

Dapnia/SEDI-SAP,

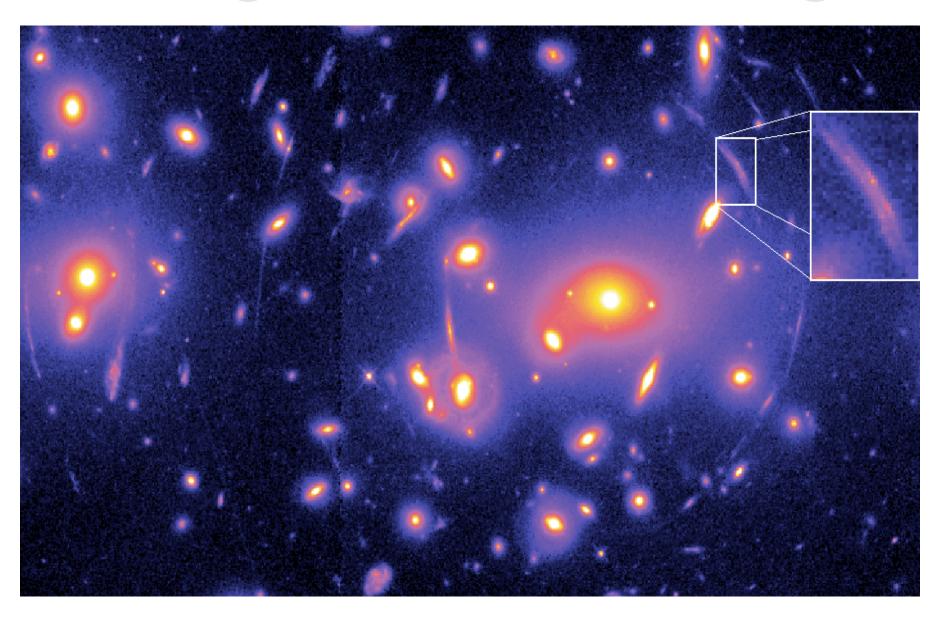
Service d'Astrophysique

CEA-Saclay, France.

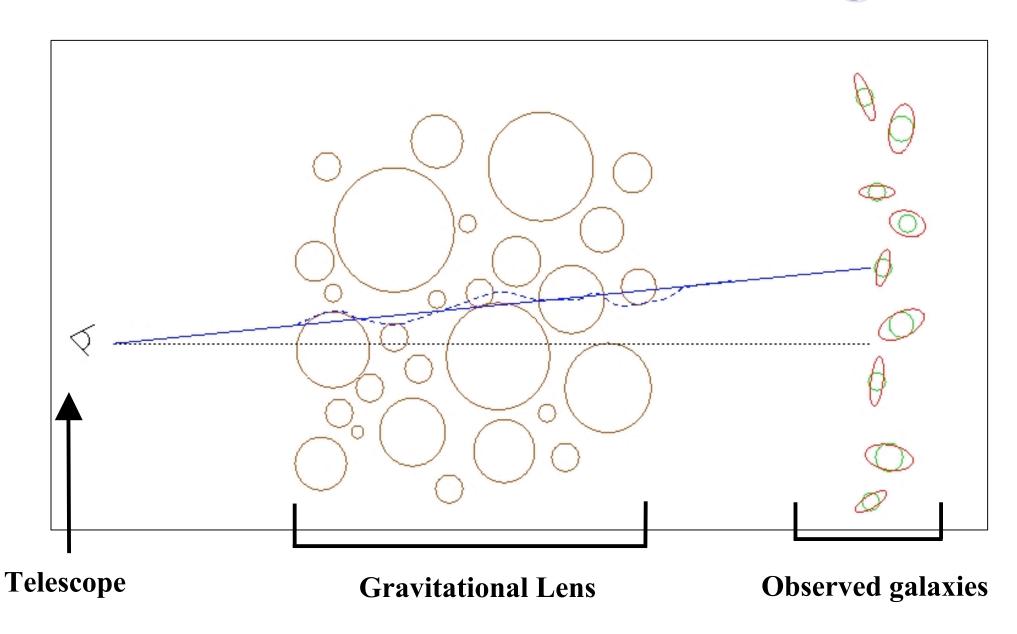
jstarck@cea.fr http://jstarck.free.fr

- 1) The Weak Lensing Mass Map Reconstruction Problem.
- 2) The Wavelet Transform
 - Relation between the "isotropic a trous wavelet transform" and the Undecimated bi-orthogonal wavelet transform
- 3) Mass map reconstruction using the Multiscale Entropy.
- 4) Experiments

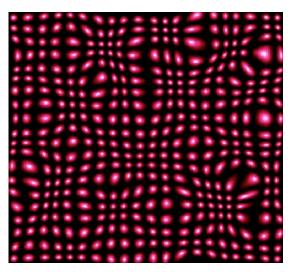
Strong Gravitational Lensing



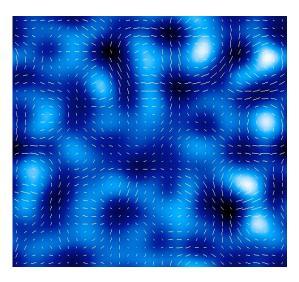
Weak Gravitational Lensing



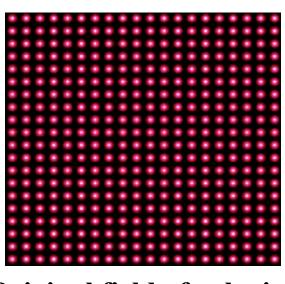
Weak Gravitational Lens



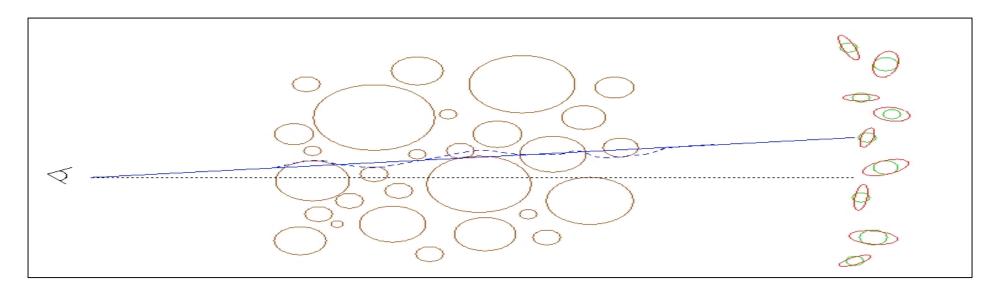
Observation



Dark Mass Map



Original field of galaxies



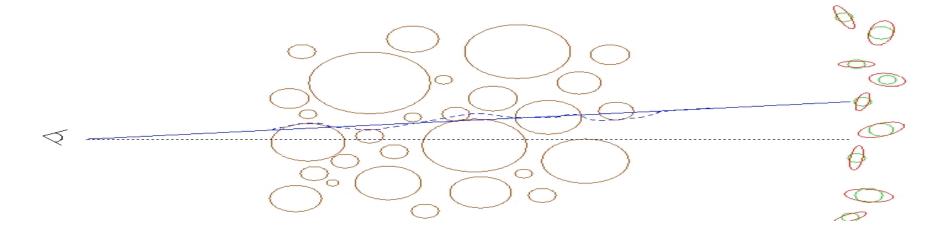
Weak Lensing by Large-Scale Structure

Distortion Matrix Ψ:

$$\Psi_{ij} = \frac{\partial \delta \theta_i}{\partial \theta_j} = \int dz \, g(z) \frac{\partial^2 \Phi}{\partial \theta_i \partial \theta_j} = \begin{pmatrix} \kappa + \gamma_1 & \gamma_2 \\ \gamma_2 & \kappa - \gamma_1 \end{pmatrix}$$

 $\delta\theta_i$ is the deflexion vector produced by lensing on the sky.

 Φ is the Newtonian potential, **z** is the distance, the weight function **g** reflects the fact that a lens is more effective when placed approximately half-way between the source and the observer, κ is proportial to the projected mass along the line of sight. The shear (γ_1, γ_2) describes stretches and compression.

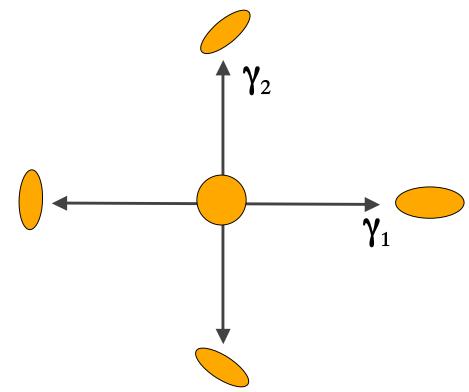


The shear map (γ_1, γ_2)

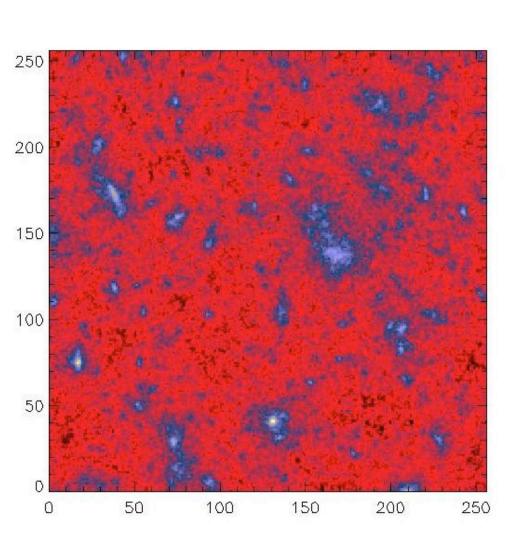
 γ_1 = deformation along the x-axis, and γ_2 at 45 degrees from it.

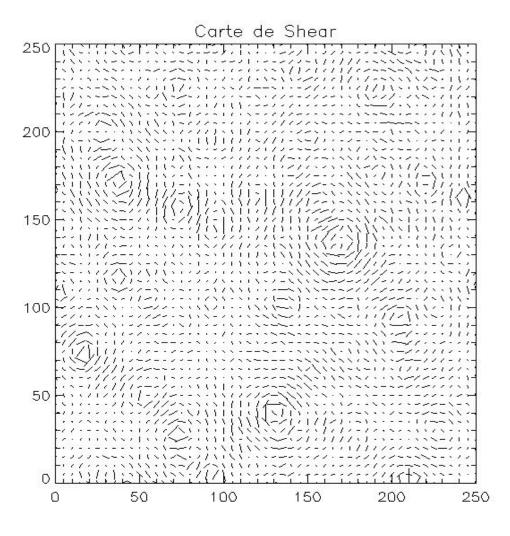
$$\gamma = \gamma_1 + i\gamma_2 = |\gamma| e^{2i\theta}$$

Where the modulus represents the amount of shear and the phase represents its direction.



Simulated Mass Map & Related Shear Mass Map



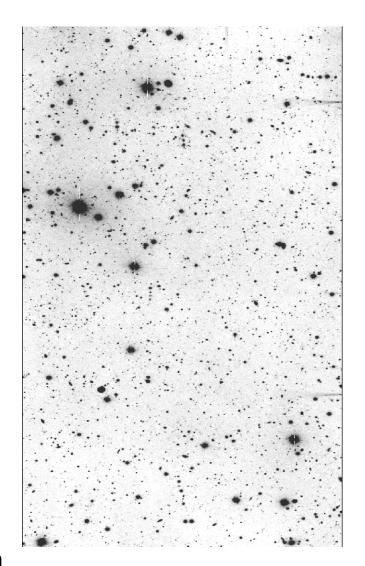


Deep Optical Images



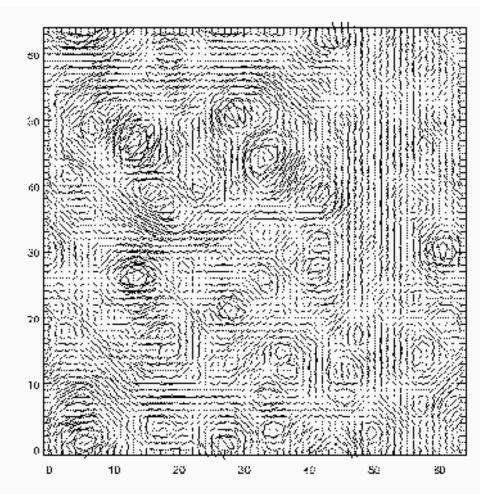
William Herschel Telescope La Palma, Canaries

16'x8' R<25.5 30 (15) gals/sq. arcmin



→Direct measure of the distribution of mass in the universe, as opposed to the distribution of light, as in other methods (eg. Galaxy surveys)

From the statistics of the shear field, weak lensing provides:



- Mapping of the distribution of Dark Matter on various scales
- Measurement of cosmological parameters.
- Measurement of the evolution of structures
- a mass-selected cluster catalog

Relation between the shear maps and the mass map κ

The relation between the weak shear maps γ_1 , γ_2 and the mass map κ are:

$$egin{array}{lll} \gamma_1 &=& (\partial_1 - \partial_2) \psi \ \gamma_2 &=& 2 \partial_1 \partial_2 \psi \end{array}$$

where ψ is the potential function defined by $\kappa = \nabla^2 \psi$.

In Fourier Space

$$egin{array}{lll} \hat{\kappa}(k_1,k_2) &=& k^2\hat{\psi}(k_1,k_2) \ \hat{\gamma_1}(k_1,k_2) &=& rac{k_1^2-k_2^2}{k^2}\hat{\kappa}(k_1,k_2) \ \hat{\gamma_2}(k_1,k_2) &=& rac{2k_1k_2}{k^2}\hat{\kappa}(k_1,k_2) \end{array}$$

with $k^2 = k_1^2 + k_2^2$.

An inverse problem

Noting $\hat{P}_1(k_1,k_2) = \frac{k_1^2 - k_2^2}{k^2}$ (with $\hat{P}_1(k_1,k_2) = 0$ when $k_1^2 = k_2^2$) and $\hat{P}_2(k_1,k_2) = \frac{2k_1k_2}{k^2}$ (with $\hat{P}_2(k_1,k_2) = 0$ when $k_1 = 0$ or $k_2 = 0$), the mass reconstruction consists in **searching** κ **such that it verifies both** $\gamma_1 = P_1 * \kappa$ **and** $\gamma_2 = P_2 * \kappa$. In practice, γ_1 and γ_2 are obtained through observations and are contaminated by noise. Then the relations between the observed data γ_{1b} , γ_{2b} and the true mass map κ are given by:

$$\left\{egin{array}{lll} \gamma_{1b} &=& P_1*\kappa+N_1 \ \gamma_{2b} &=& P_2*\kappa+N_2 \end{array}
ight.$$

Noise

 N_1 and N_2 can be considered as Gaussian noise with a standard deviation $\sigma_n=\frac{\sigma_\epsilon}{N_g}$, with $\sigma_\epsilon=0.3$, and $N_g=n_gA$. A is the surface of the pixel and n_g is the number of galaxies per arcmin². Typical n_g values are:

- $n_g = 100$ gal/arcmin² for spatial observations.
- $n_g = 20$ gal/arcmin² for ground observations.

The Inverse Filter: E and B mode

Noticing that $\hat{P_1}^2 + \hat{P_2}^2 = 1$, the least square estimation $\hat{\tilde{\kappa}}_l^{(E)}$ is:

$$\hat{ ilde{\kappa}}_l^{(E)} = \hat{P}_1 \hat{\gamma}_{1b} + \hat{P}_2 \hat{\gamma}_{2b}$$

The relation between this estimation and the true mass map is $\hat{\kappa}_l = \hat{\kappa} + \hat{N}$, where $\hat{N} = \hat{P}_1 \hat{N}_1 + \hat{P}_2 \hat{N}_2$. Another interesting feature is the term $\hat{\tilde{\kappa}}_l^{(B)} = P_2 * \gamma_{1b} - P_1 * \gamma_{2b}$. Indeed it should be free of any contamination from κ and can be used as a test for data quality estimation.

Therefore, the so called E and B mode are obtained by:

$$\hat{ ilde{\kappa}}_{l}^{(E)} = P_{1} * \gamma_{1b} + P_{2} * \gamma_{2b} \\ \hat{ ilde{\kappa}}_{l}^{(B)} = P_{2} * \gamma_{1b} - P_{1} * \gamma_{2b}$$

Both of them are noisy, and must be filtered before being analysed.

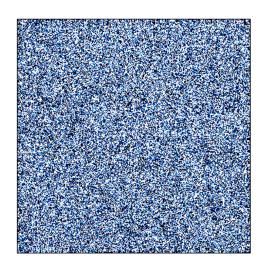
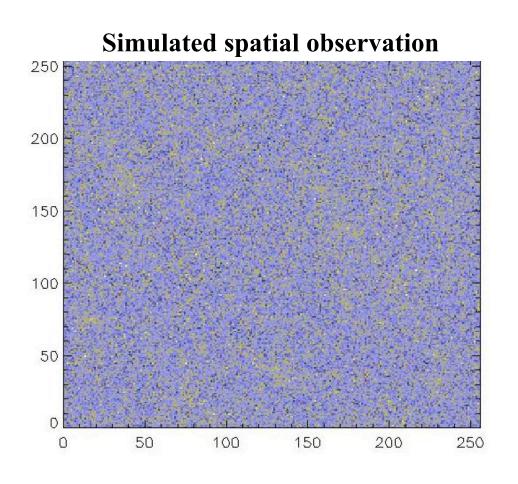


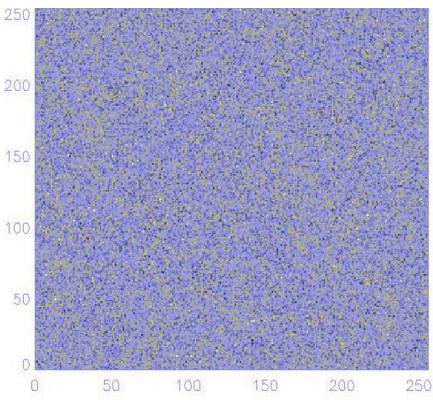
Figure 1: Noisy mass map $\kappa_l^{(E)}$ with $n_g=100$

The noise $N^{(E)}$ and $N^{(B)}$ in $\tilde{\kappa}_l^{(E)}$ and $\tilde{\kappa}_l^{(B)}$ is still Gaussian and uncorrelated. The inverse filtering does not amplify the noise (since $\hat{P}_1^2 + \hat{P}_2^2 = 1$), but $\tilde{\kappa}_l^{(E)}$ and $\tilde{\kappa}_l^{(B)}$ are dominated by the noise. This has motivated the development of different methods isuch the Gaussian filtering (kaiser, 1992) or the Maximun Entropy Method (Bridle et al, 1998; Marshall et al, 2002).

Problem: maps are very noisy



Simulated on ground observation



Linear Filtering

The standard method (kaiser, 1992) consists in convolving the noisy mass map $\tilde{\kappa}_l^{(E)}$ with a Gaussian window G with standard deviation σ_G :

$$ilde{\kappa}_G^{(E)} = G * ilde{\kappa}_l^{(E)} = G * P_1 * \gamma_{1b} + G * P_2 * \gamma_{2b}$$

The new estimation depends strongly on σ_G .

An alternative to the Gaussian filtering could be the Wiener filtering:

$$\hat{W}(u,v) = rac{|\hat{S}(u,v)|^2}{|\hat{S}(u,v)|^2 + |\hat{N}(u,v)|^2}$$

Where $\hat{S}(u,v)$ is a model of the true mass map and is in practice derived from the data. The Wiener filtering is known to be optimal when both the signal and the noise are a realization of a Gaussian Random Field. For weak lensing mass reconstruction, this assumption is obviously not true (cluster presents non Gaussianities), but experiments show it leads to reasonable results, generally better than the simple Gaussian filtering.

Maximun Entropy Method

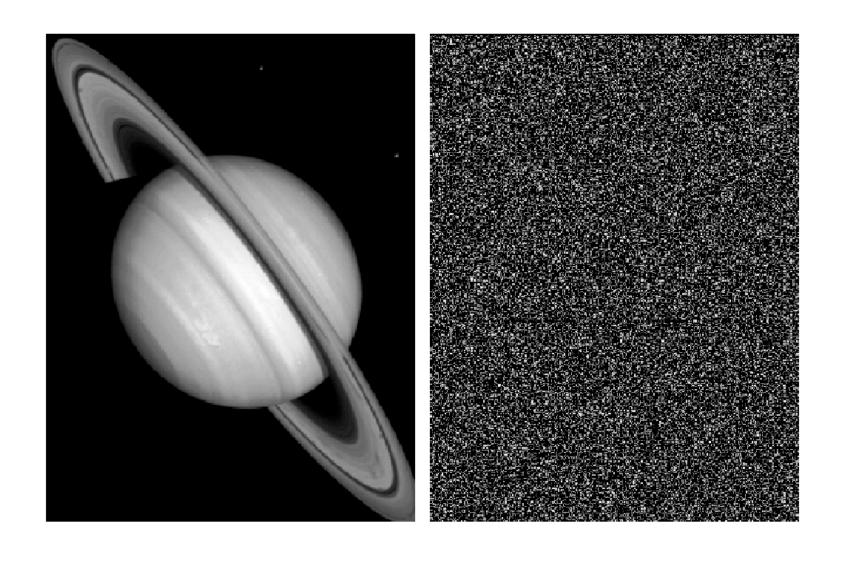
The Maximum Entropy Method (MEM) is well-known and widely used in image analysis in astronomy: The solution is found by minimizing

$$J(ilde{\kappa}) = \sum_{pixels} rac{{(ilde{\kappa}_l^{(E)} - ilde{\kappa})}^2}{2\sigma^2} + lpha H(ilde{\kappa}) = rac{\chi^2}{2} + lpha H(ilde{\kappa})$$

Several definitions of entropy exists. The most common is the Gull and Skilling definition:

$$H_g(\kappa) = \sum_x \sum_y \kappa(x,y) - m(x,y) - \kappa(x,y) \ln \left(\frac{\kappa(x,y)}{m(x,y)} \right)$$

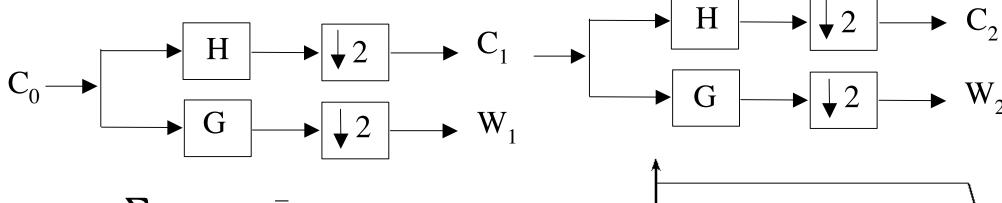
where m is a model. It has recently proposed by Bridle et al (1998) and Marshall et al (2002) for weak lensing mass reconstruction.



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



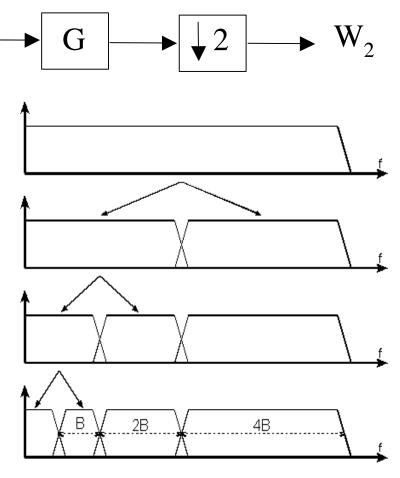
$$c_{j+1,l} = \sum_{h} h_{k-2l} c_{j,k} = (\overline{h} * c_j)_{2l}$$

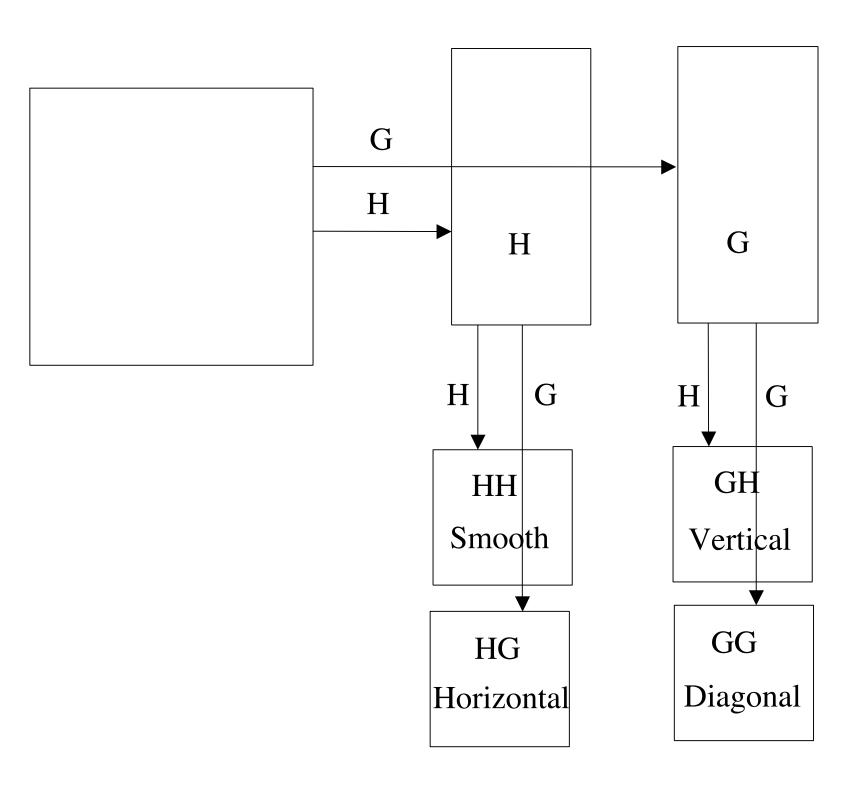
$$w_{j+1,l} = \sum_{h} g_{k-2l} c_{j,k} = (\overline{g} * c_j)_{2l}$$

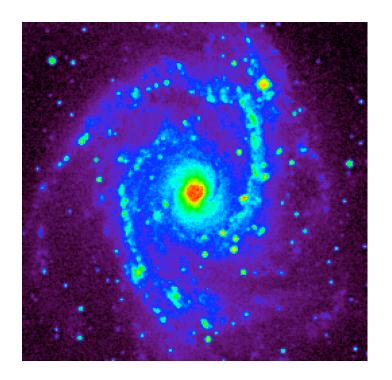
Reconstruction:

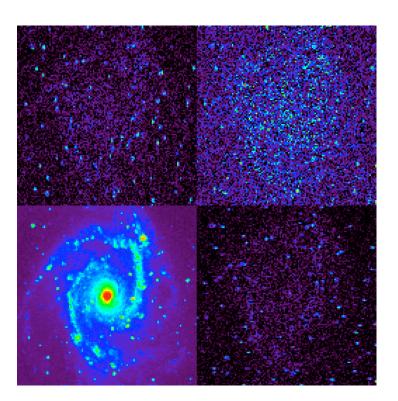
$$c_{j,l} = \sum_{k} \tilde{h}_{k+2l} c_{j+1,k} + \tilde{g}_{k+2l} w_{j+1,k} = \tilde{h} * \check{c}_{j+1} + \tilde{g} * \check{w}_{j+1}$$

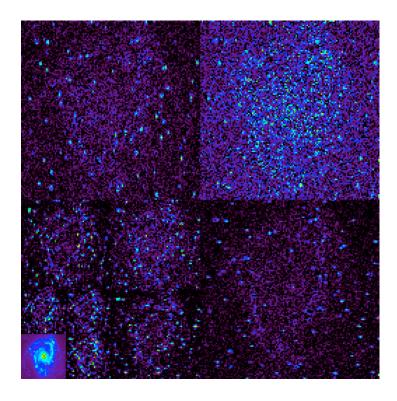
$$\check{x} = (x_1, 0, x_2, 0, x_3, \dots, 0, x_j, 0, \dots, x_{n-1}, 0, x_n)$$



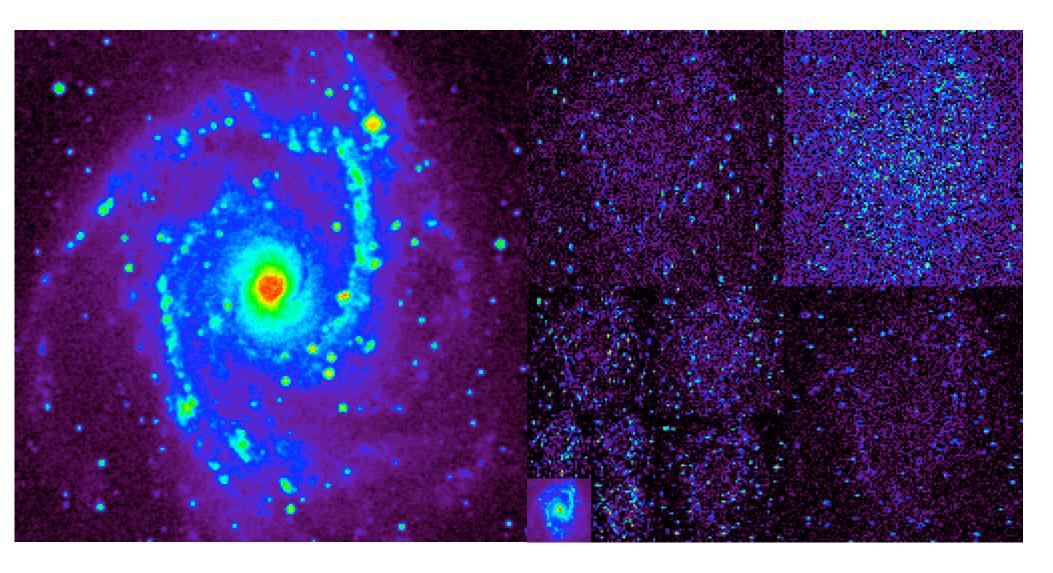








NGC2997 WT



The Filter Bank

In order to get an exact reconstruction, two conditions are required for the filters:

- Dealiasing condition: $\hat{h}(\nu+\frac{1}{2})\hat{\tilde{h}}(\nu)+\hat{g}(\nu+\frac{1}{2})\hat{\tilde{g}}(\nu)=0$
- Exact restoration: $\hat{h}(
 u)\hat{ ilde{h}}(
 u)+\hat{g}(
 u)\hat{ ilde{g}}(
 u)=1$

The Isotropic Undecimated Wavelet Transform

• Filters do not need to verify the dealiasing condition. We need only the exact restoration condition:

$$\hat{h}(
u)\hat{ ilde{h}}(
u)+\hat{g}(
u)\hat{ ilde{g}}(
u)=1$$

- Filters do not need to be (bi) orthogonal.
- Filters must be symmetric.
- In 2D, we want h(x, y) = h(x)h(y) for fast calculation and more important, h(x, y) must nearly isotropic.

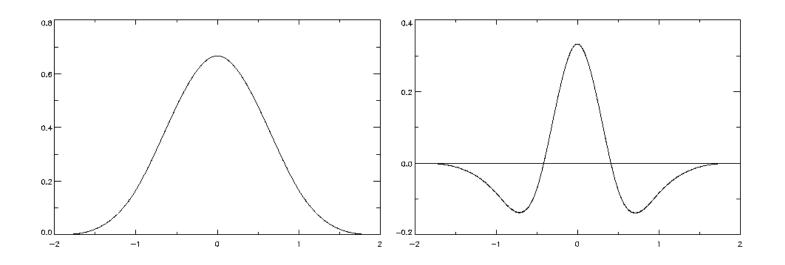
h is derived from a B_3 spline: $h_{1D}(k) = [1, 4, 6, 4, 1]/16$, and in 2D $h_{2D} = h_{1D}h_{1D} =$

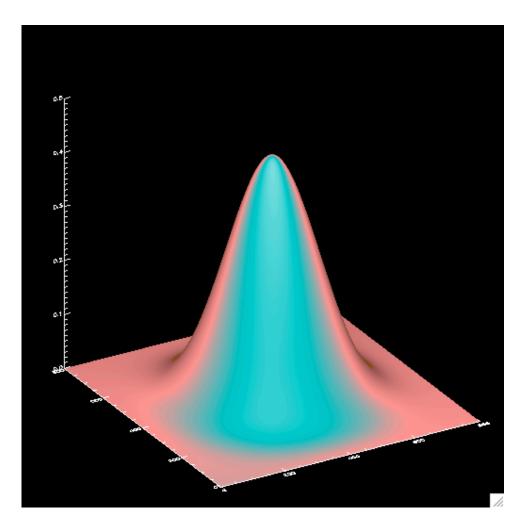
The Isotropic Wavelet and Scaling Functions

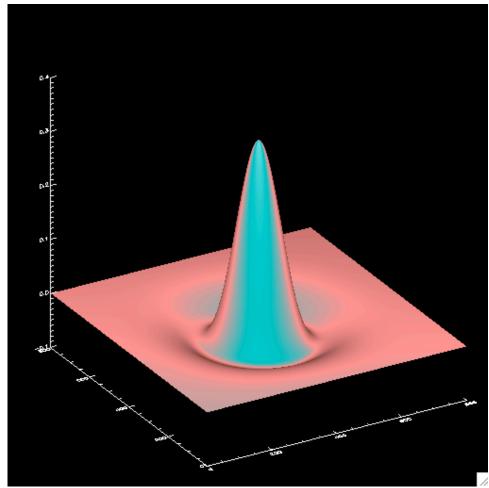
$$B_3(x) = \frac{1}{12} (|x-2|^3 - 4|x-1|^3 + 6|x|^3 - 4|x+1|^3 + |x+2|^3)$$

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

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

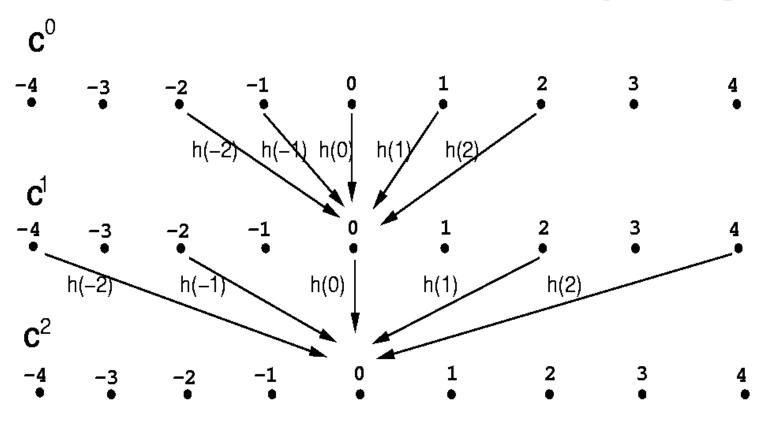




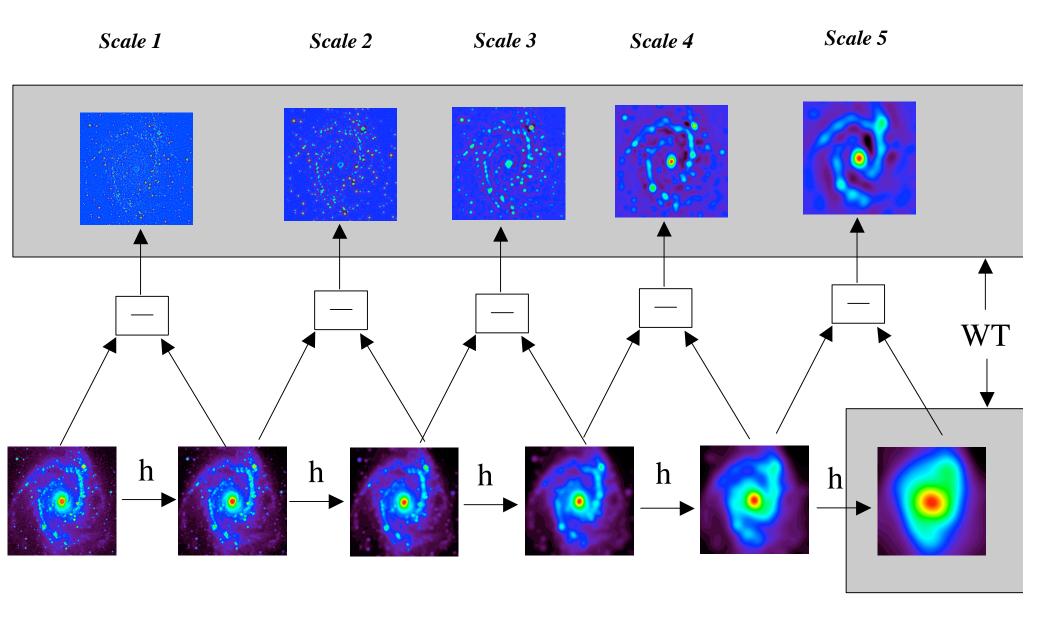


Passage from c_0 to c_1 , and from c_1 to c_2

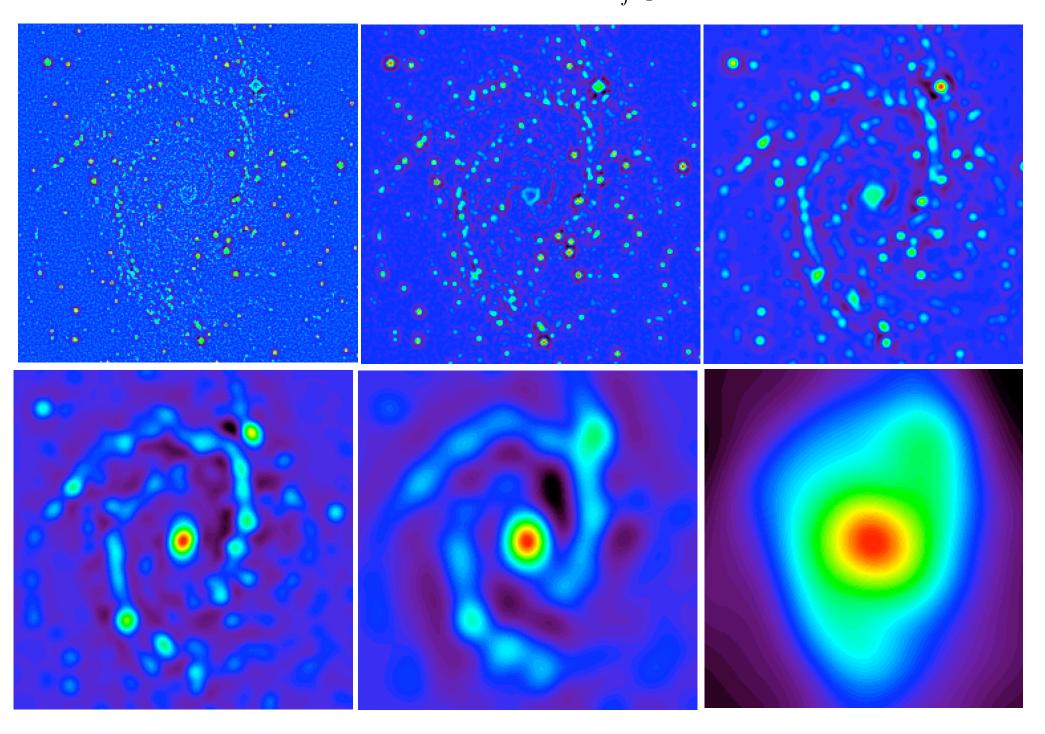
h=[1,4,6,4,1]/16

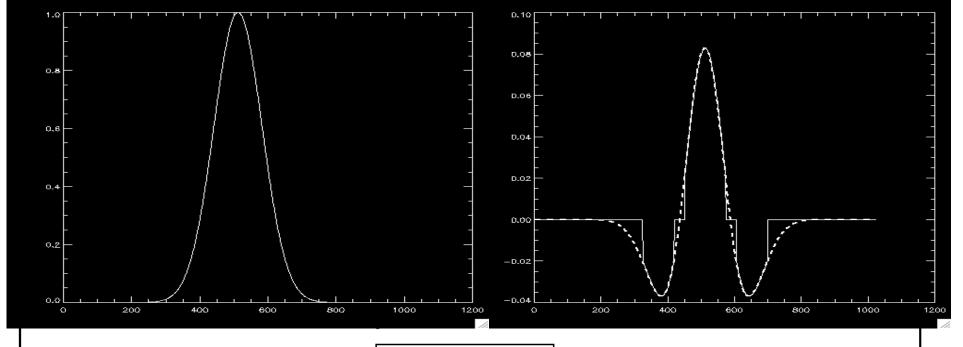


ISOTROPIC UNDECIMATED WAVELET TRANSFORM



Isotropic Undec. WT: $I(k,l) = c_{J,k,l} + \sum_{j=1}^{J} w_{j,k,l}$





Drawbacks

- Data restoration: ringing artefacts may appear around strongest sources, which is due to the negative part of the wavelet function.
- Large scale structure thresholding may create artefacts.
- For some data, we may prefer an anisotropic decomposition.

2D Undecimated Wavelet Transform

The à trous algorithm can be extended to 2D:

$$c_{j+1,k,l} = (\bar{h}^{(j)}\bar{h}^{(j)}*c_j)_{k,l}$$

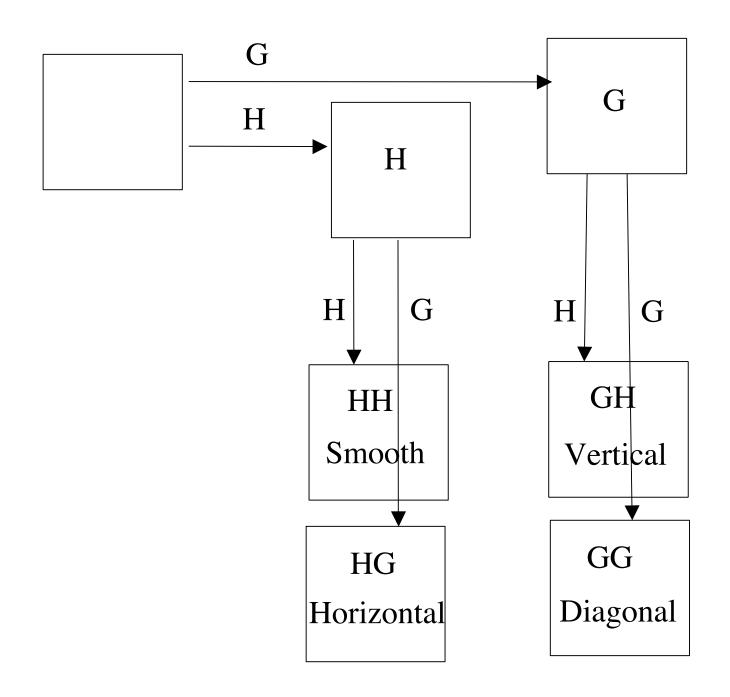
$$w_{j+1,1,k,l} = (\bar{g}^{(j)}\bar{h}^{(j)}*c_j)_{k,l}$$

$$w_{j+1,2,k,l} = (\bar{h}^{(j)}\bar{g}^{(j)}*c_j)_{k,l}$$

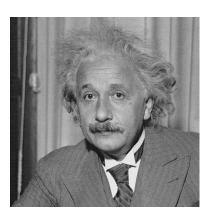
$$w_{j+1,3,k,l} = (\bar{g}^{(j)}\bar{g}^{(j)}*c_j)_{k,l}$$

where hg * c is the convolution of c by the separable filter hg (i.e convolution first along the columns per h and then convolution along the lines per g).

Undecimated bi-orthogonal Wavelet Transform



Undecimated Wavelet Transform



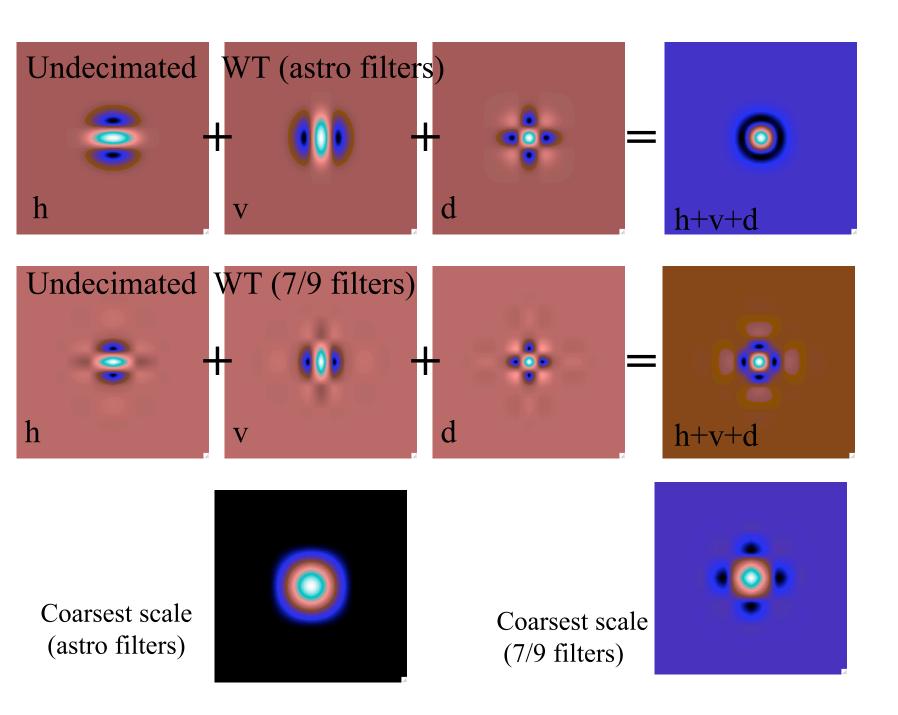
Non (bi-) Orthogonal Directional Undecimated WT using the "astro" filter bank

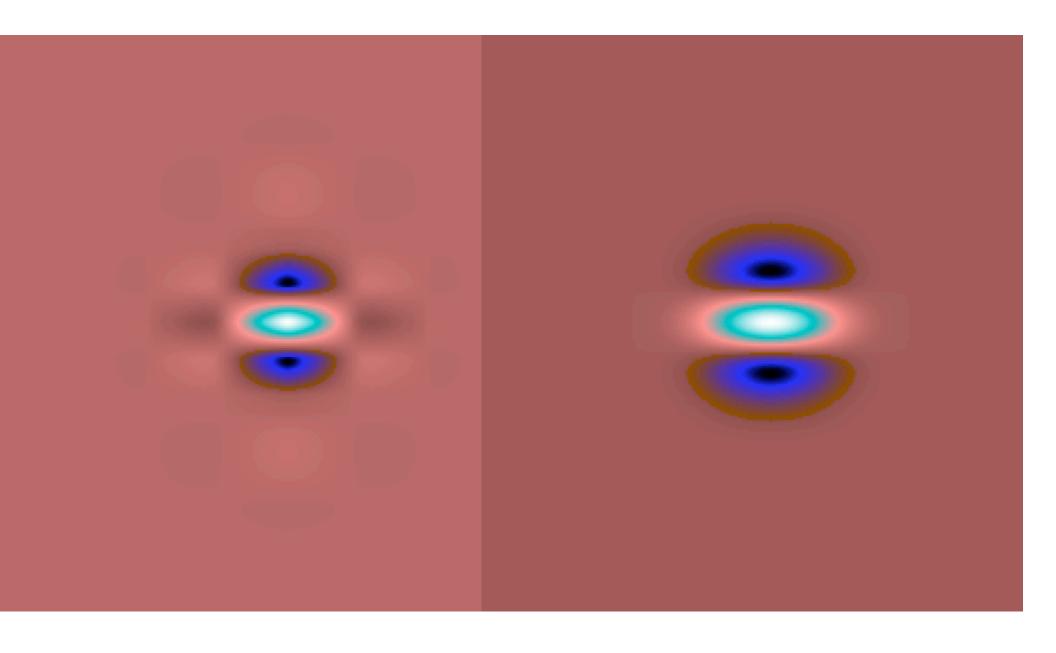
$$egin{array}{lcl} h & = & [1,4,6,4,1]/16 \ g & = & Id-h = [-1,-4,10,-4,-1]/16 \ \tilde{h} & = & \tilde{g} = Id \end{array}$$

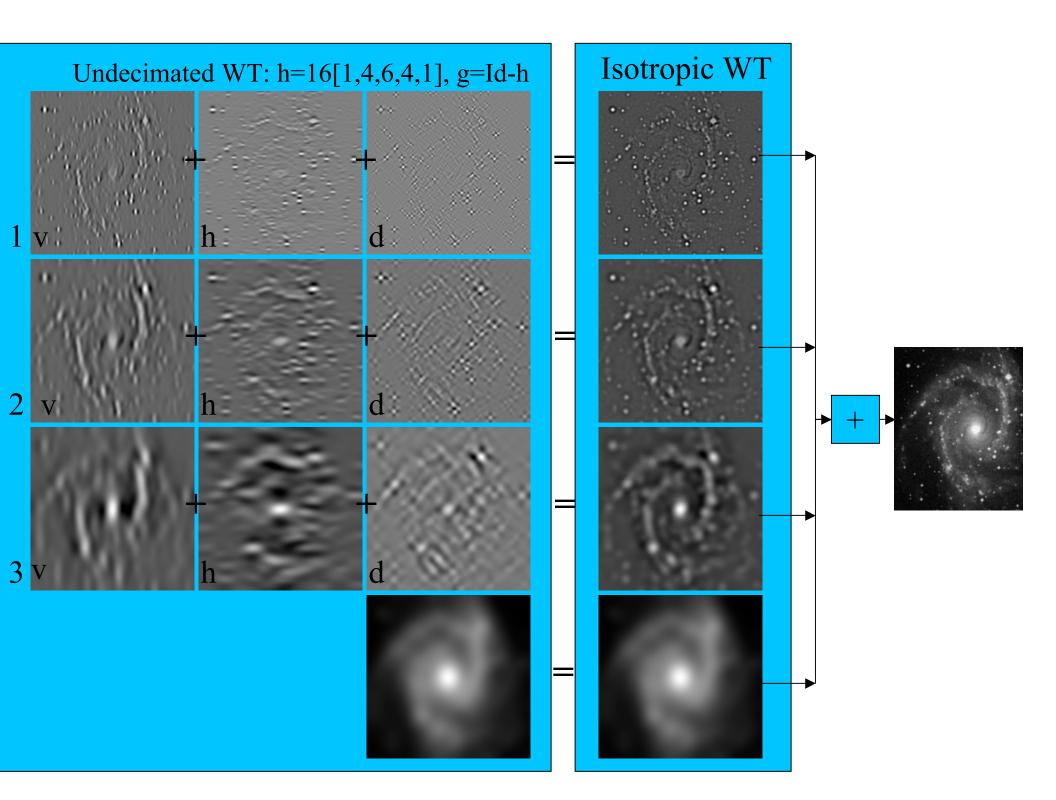
In two dimensions, the detail signal is contained in three sub-images

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

| Undecimated Wavelet Transform: h=1/16[1,4,6,4,1] g= Id-h | | - | * |
|--|---|---|---|
| | • | - | * |
| | | | |
| | | | |







The Surprise

Because the decomposition is redundant, there are many way to reconstruct the original image from its wavelet transform. For a given (h,g) filter bank, any filter bank (\tilde{h},\tilde{g}) which verifies the equation $\hat{h}(\nu)\hat{\tilde{h}}(\nu)+\hat{g}(\nu)\hat{\tilde{g}}(\nu)=1$ leads to an exact reconstruction. For instance, if we choose $\tilde{h}=h$ (the synthesis scaling function $\tilde{\phi}=\phi$) we obtain a filter \tilde{g} defined by:

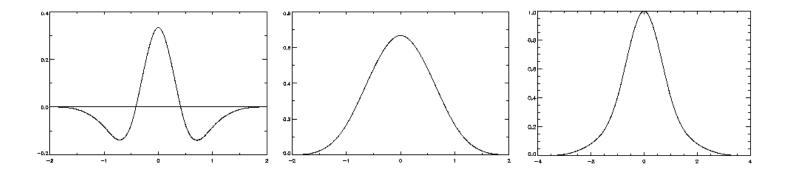
$$ilde{g}=h+Id$$

if h = [1, 4, 6, 4, 1]/16, then g = [1, 4, 22, 4, 1]/16. g is positive. This means that g is not related anymore to a wavelet function. The synthesis scaling function related to \tilde{g} is defined by:

$$\frac{1}{2}\tilde{\phi}(\frac{x}{2}) = \phi(x) + \frac{1}{2}\phi(\frac{x}{2})$$

Reconstruction Using the Scaling Function

$$egin{array}{lll} s_l &=& \displaystyle\sum_k c_{J,k} \phi_{J,l}(k) + \displaystyle\sum_k \displaystyle\sum_{j=1}^J ilde{\phi}_{j,l}(k) w_{j,k} \end{array}$$



| vertical 1 • | horizontal • | diagonal • |
|--------------|-----------------|---------------|
| 2 | 0 | * |
| 3 | 0 | |
| 4 | 0 | |

Another Interesting Filter Bank

Deriving h from a spline scaling function, for instance $B_2 = [1, 2, 1]/2$ or $B_3 = [1, 4, 6, 4, 1]/16$ (note that $B_3 = B_2 * B_2$, we define the following the filter bank:

$$egin{array}{lll} h & = & ilde{h} = B_l \ g & = & Id - h * h \ ilde{g} & = & Id \end{array}$$

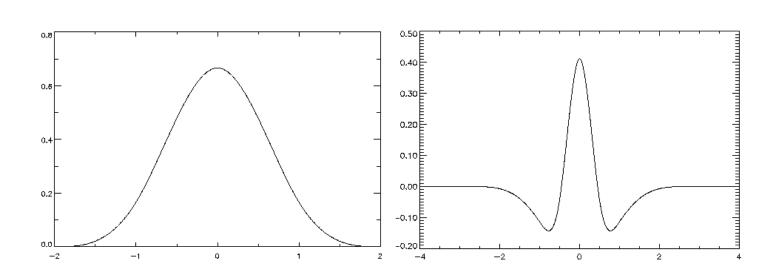
which leads to an analysis/synthesis with the following functions:

$$\phi(x) = \tilde{\phi}(x) = B_l(x)$$

$$\hat{\psi}(\nu) = \frac{\hat{\phi}^2(\nu) - \hat{\phi}^2(2\nu)}{\hat{\phi}(\nu)}$$

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

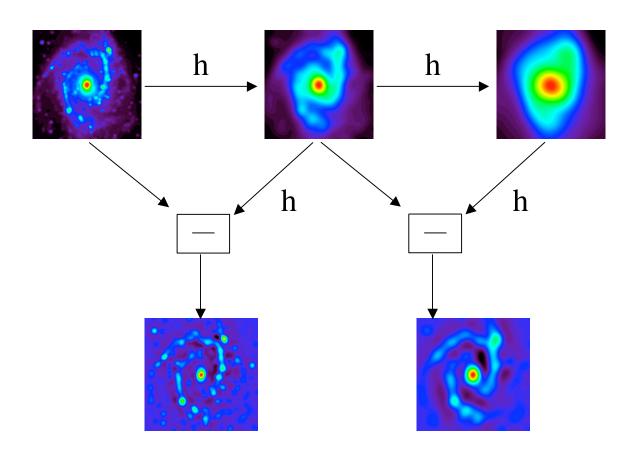
CEA-Saclay, DAPNIA/SEDI-SAP



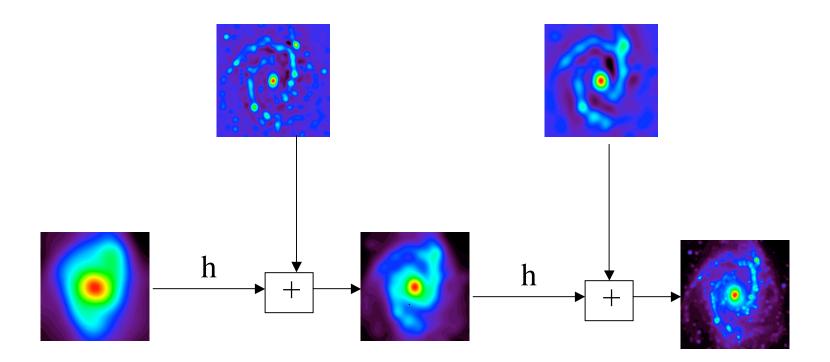
$$\phi(x) = \tilde{\phi}(x) = B_3(x)
\hat{\psi}(\nu) = \frac{\hat{\phi}^2(\nu) - \hat{\phi}^2(2\nu)}{\hat{\phi}(\nu)}
\frac{1}{2}\tilde{\psi}(\frac{x}{2}) = \phi(x)
h = \tilde{h} = [1, 4, 6, 4, 1]/16
g = Id - h * h = [1, 8, 28, 56, 70, 56, 28, 8, 1]/256
\tilde{g} = Id$$

MODIFIED ISOTROPIC UNDECIMATED WAVELET TRANSFORM

$$h = h_{1d} # h_{1d}, g = Id - h * h$$



RECONSTRUCTION



Multiscale Enrtopy

The Multiscale Entropy concept (Pantin et al, 1996) consists in replacing the standard MEM prior (i.e. Gull and Skilling entropy) by a wavelet based prior. The entropy is now defined by $H(\kappa) = \sum_{j=1}^{J-1} \sum_{k,l} h(w_{j,k,l})$. where w_j are the wavelet coefficients of κ at scale j. The information content of an image can be seen as sum of information at different scales. The function k defines the amount of information relative to a given wavelet coefficients. We minimize:

$$J(ilde{\kappa}) = rac{\parallel ilde{\kappa}_l^{(E)} - ilde{\kappa} \parallel^2}{2\sigma_D^2} + lpha \sum_{j=1}^J \sum_{k,l} h_n((\mathcal{W} ilde{\kappa})_{j,k,l})$$

where σ_D the noise standard deviation in $\tilde{\kappa}_l^{(E)}$ and J is the number of scales.

Constraint in the Wavelet Domain

The multiscale entropy is derived from the noise modeling:

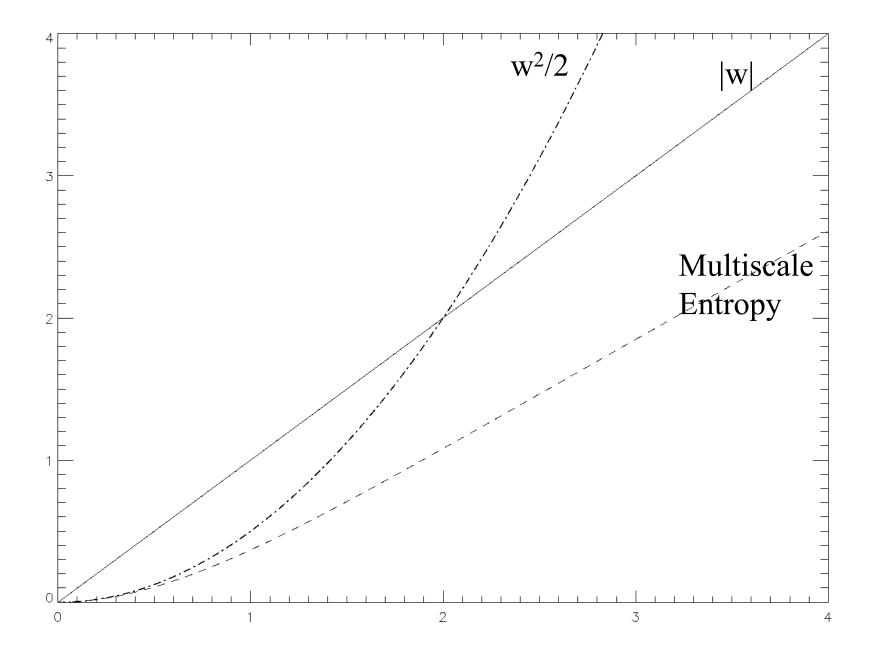
$$h(w_{j,k,l}) = \int_0^{|w_{j,k,l}|} P_n(|w_{j,k,l}| - u) (\frac{\partial h(x)}{\partial x})_{x=u} du$$

where $P_n(w_{j,k,l})$ is the probability that the coefficient $w_{j,k,l}$ can be due to the noise: $P_n(w_{j,k,l}) = \text{Prob}(W > |w_{j,k,l}|)$. For Gaussian noise, we have:

$$P_n(w_{j,k,l}) = \frac{2}{\sqrt{2\pi}\sigma_j} \int_{|w_{j,k,l}|}^{+\infty} \exp(-W^2/2\sigma_j^2) dW = \operatorname{erfc}(\frac{|w_{j,k,l}|}{\sqrt{2}\sigma_j})$$

and

$$h(w_{j,k,l}) = rac{1}{\sigma_j^2} \int_0^{|w_{j,k,l}|} u \operatorname{erfc}(rac{\mid w_{j,k,l} \mid -u}{\sqrt{2}\sigma_j}) du$$



In (Pantin et Starck, 1996), it has been suggested to not apply the regularization on wavelet coefficients which are clearly detected (i.e. significant wavelet coefficients). The new Multiscale Entropy is:

$$h_n(w_{j,k,l}) = ar{M}(j,k,l) h(w_{j,k,l})$$

where $\bar{M}(j, k, l) = 1 - M(j, k, l)$, and M is the multiresolution:

$$M(j,k,l) = \left\{egin{array}{ll} 1 & ext{if } w_{j,k,l} ext{ is significant} \ 0 & ext{if } w_{j,k,l} ext{ is not significant} \end{array}
ight.$$

 $w_{j,k,l}$ is significant if $|w_{j,k,l}| > k\sigma_j$, where σ_j is the noise standard deviation at scale j, and k is a constant, generally taken between 3 and 5. The number of false detections depends on both the ϵ value and the image size.

Benjamini & Hochberg FDR

False Discovery Rate

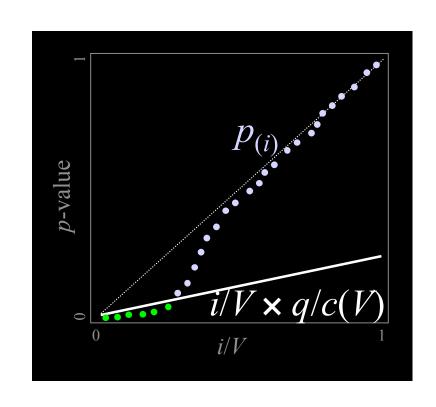
- FDR = expected (# false predictions/ # total predictions)
- Control the *proportion* of false positive in all positive at the desired significance level.

- Select desired limit q on FDR
- Order p-values, $p_{(1)} \le p_{(2)} \le ... \le p_{(V)}$
- Let r be largest i such that

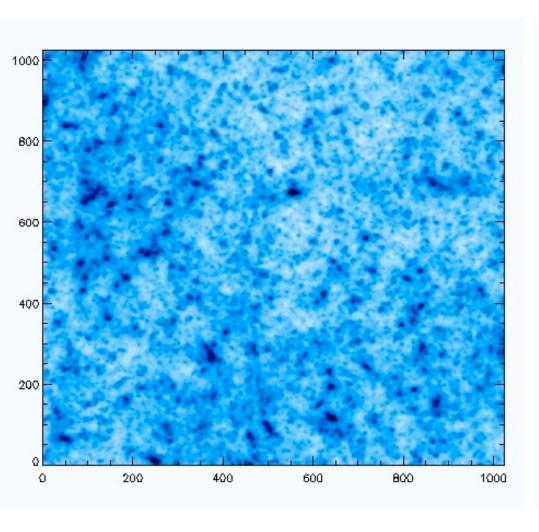
$$p_{(i)} \le i/V \times q/c(V)$$

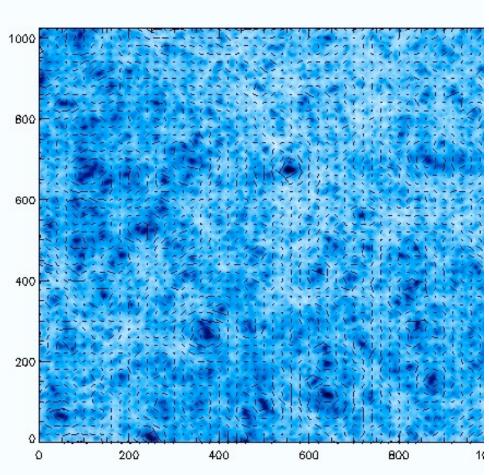
Reject all hypotheses corresponding to

$$p_{(1)}, \ldots, p_{(r)}$$
.

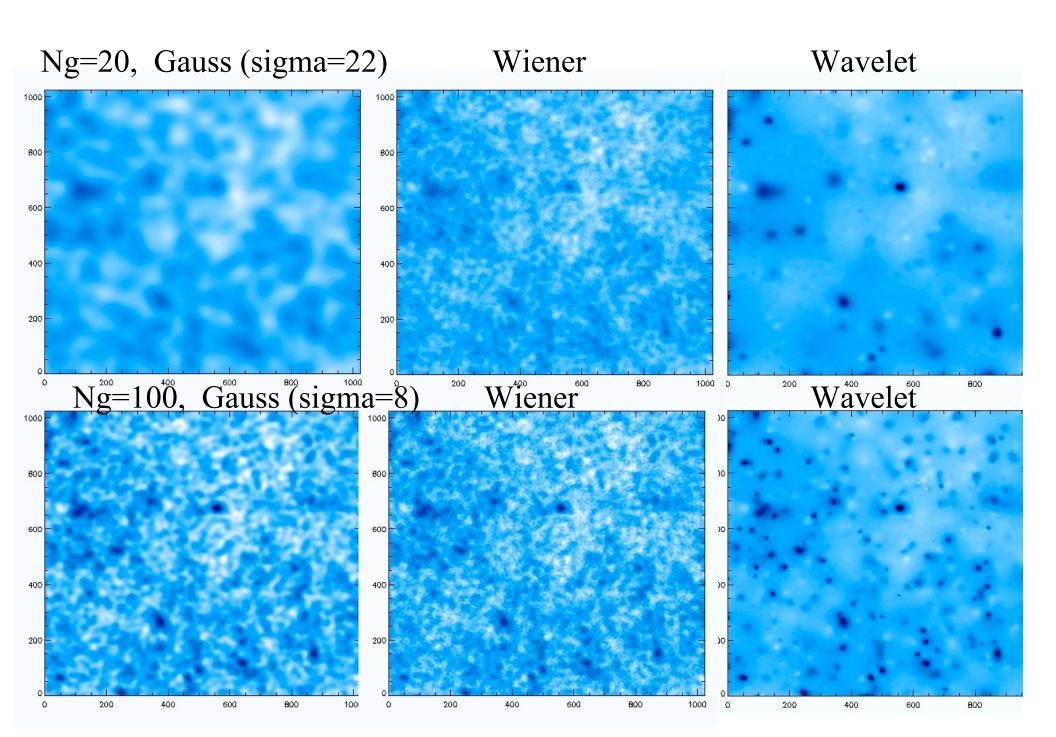


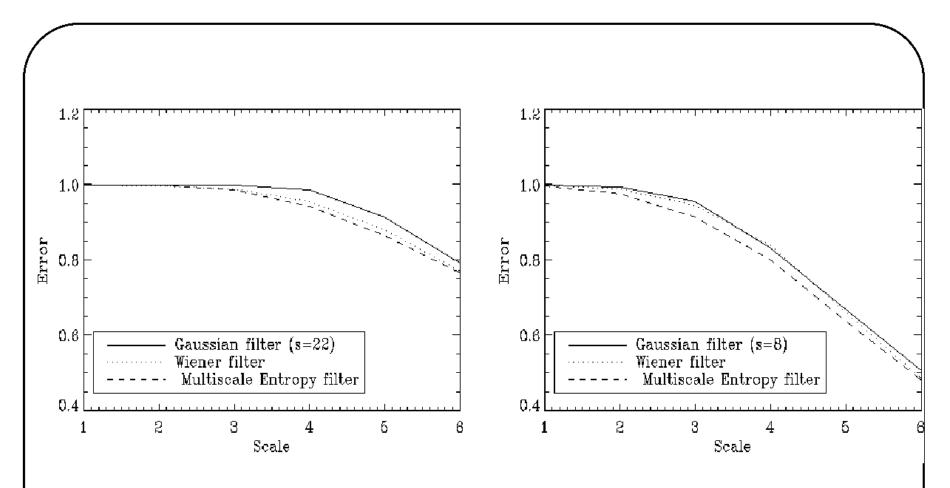
EXPERIMENTS: Simulated Data



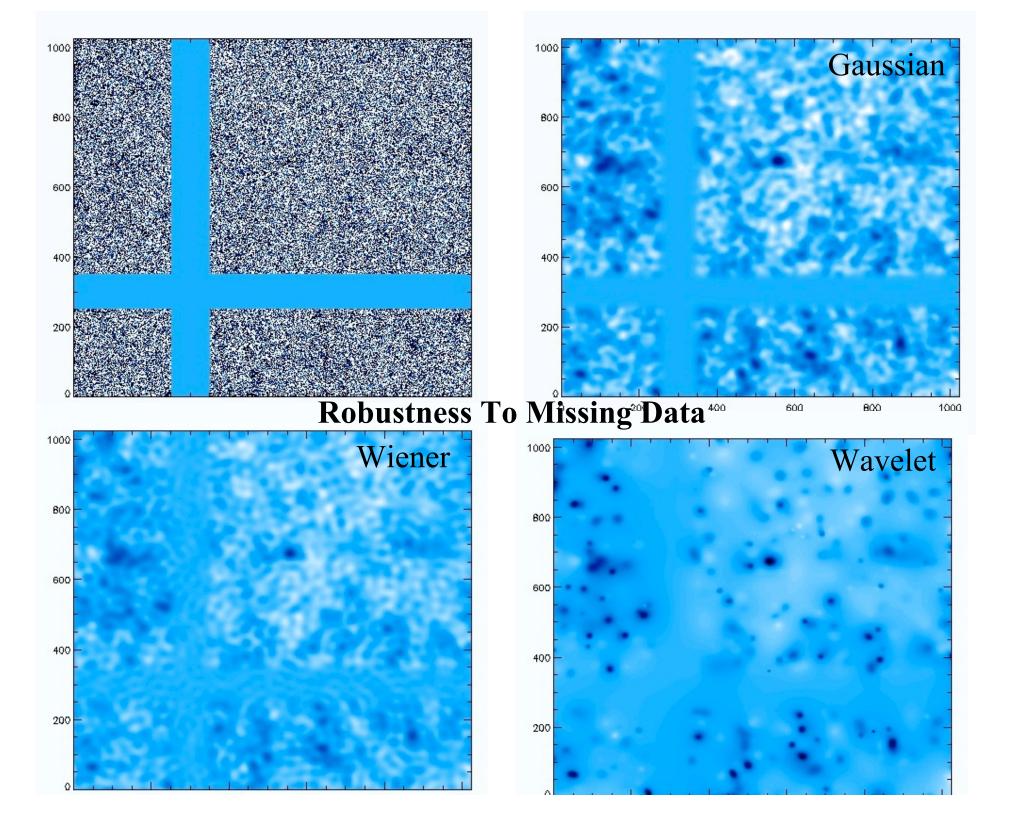


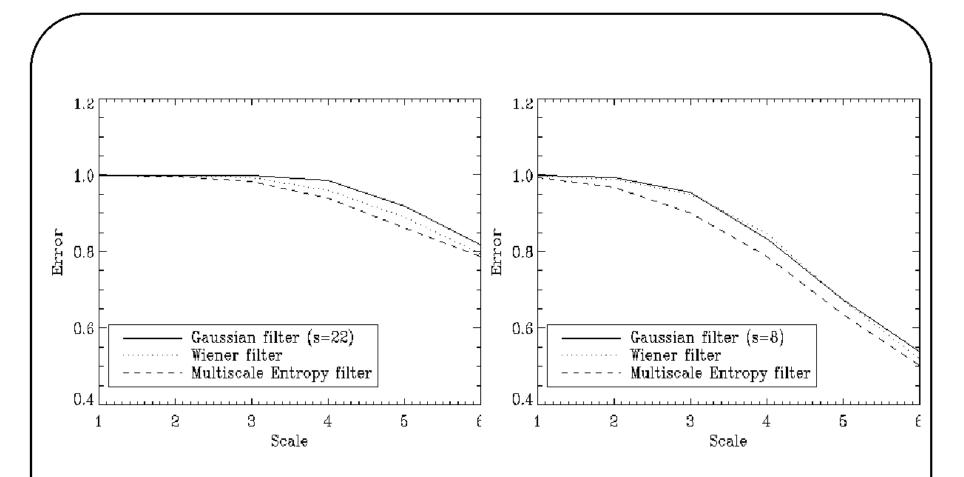
Denoising, for ng=20 and ng=100





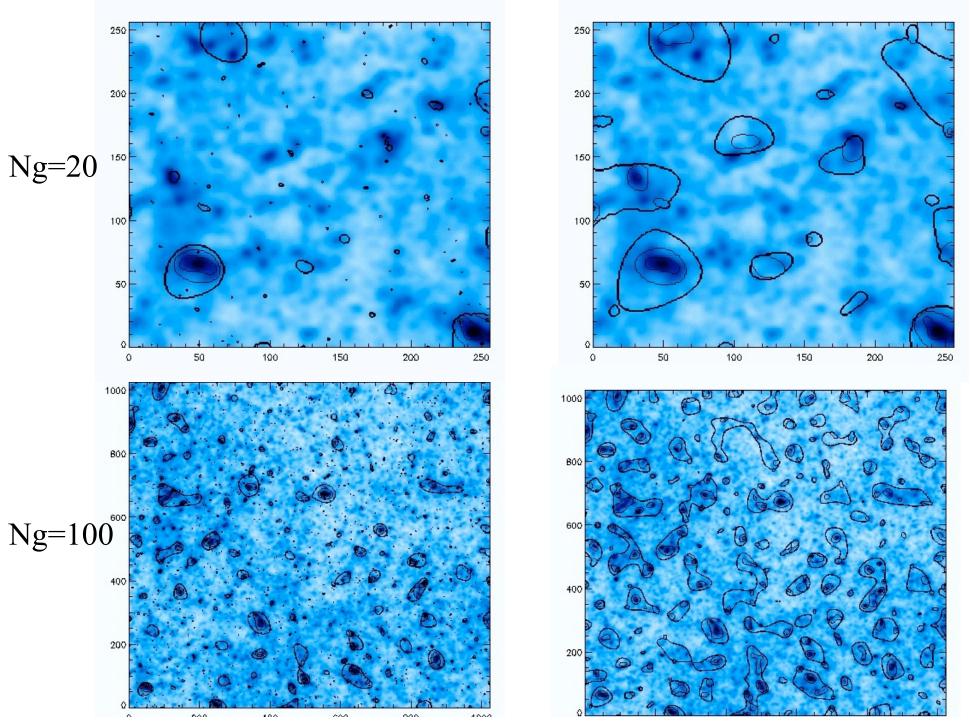
Standard deviation versus scale for the on ground simulation (left) and the spatial simulation (right).



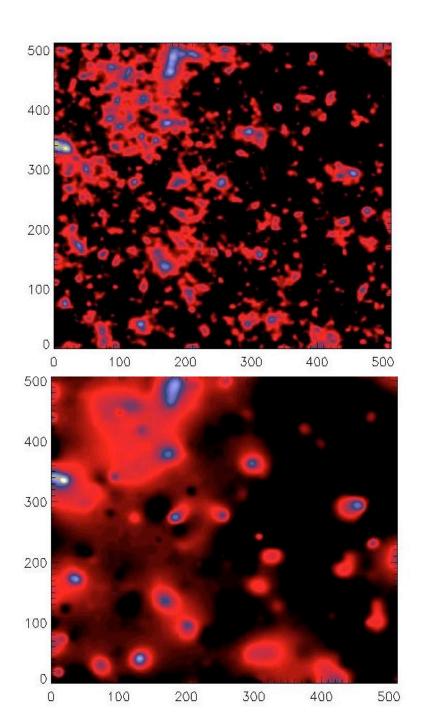


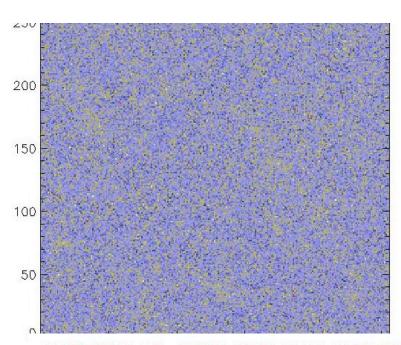
Standard deviation versus scale for the on ground simulation (left) and the spatial simulation (right) for the simulation with missing data

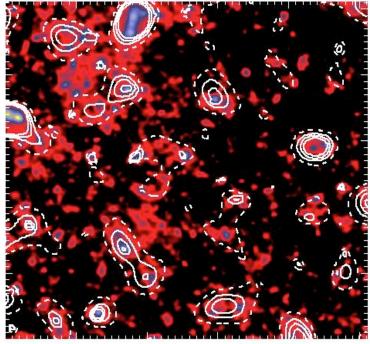
CLUSTER DETECTION
Ksigma Detectiom FDR



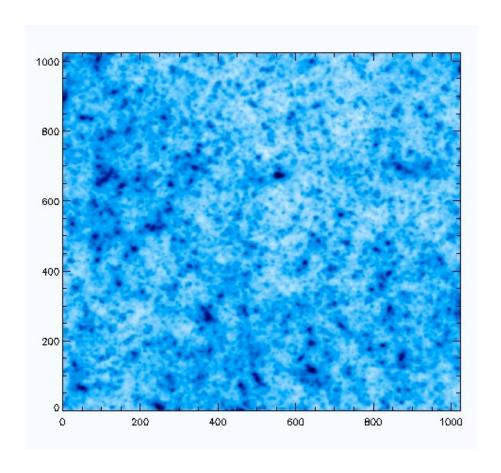
Reconstructed Dark Matter Map

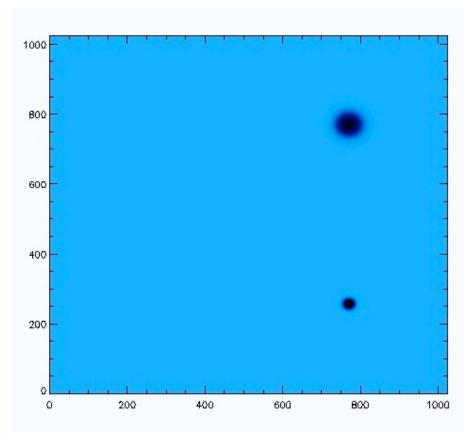






E and B Mode





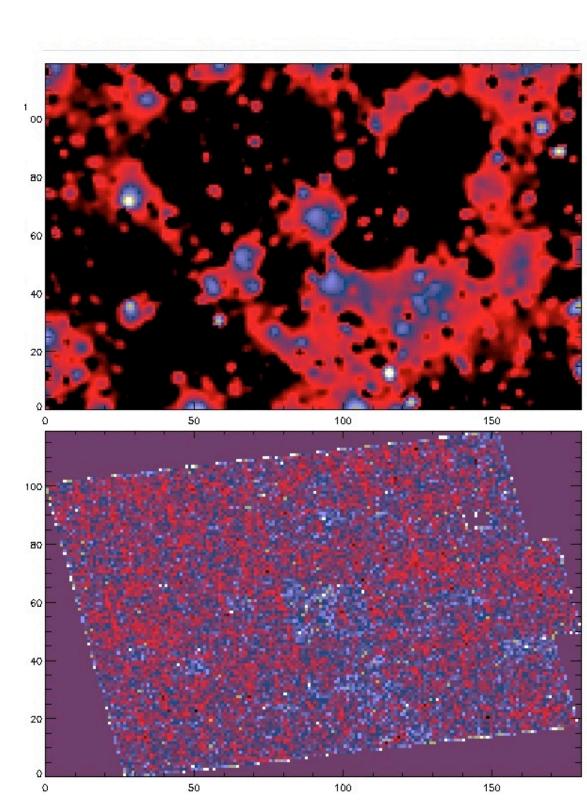
E MODE B MODE

HST Cosmos DATA

- Field: 1 degree x 1.5 degree

- Sampled with: 180*119 pixels

- pixel size : 0.5*0.5 arcmin2



Conclusions

We have presented a new way to reconstruct the weak shear E and B modes.

- We have proposed a new wavelet decomposition.
- We have modified the Multiscale Entropy method in order to take into account the FDR.
- This new method outperforms standard techniques currently used for the weak shear mass reconstruction (visual aspect, rms, rms per scale, ...).
- The method is robust to the missing data problem.
- We have also shown that the B-mode should be filtered as well by the same method, allowing us to better control the quality of the solution.

Our method allows us also to build a catalog of clusters and the use of FDR leads to a clear improvement in sensibility, compared to what has been done previously with wavelets.