Compressed Sensing in Astronomy

J.-L. Starck
jstarck@cea.fr
http://jstarck.free.fr

**Collaborators: J. Bobin, Stanford.
M. Sauvage, N. Barbay, CEA**


Compressed Sensing Impact in astronomy

$$y = \Theta x$$

x sparse in the dictionary $\mathbf{\Phi}$

$$\mu_{\Theta,\Phi} = \sqrt{N} \max_{i,k} \left| \left\langle \phi_i, \theta_k \right\rangle \right|$$

Typical Astronomical Data related to CS

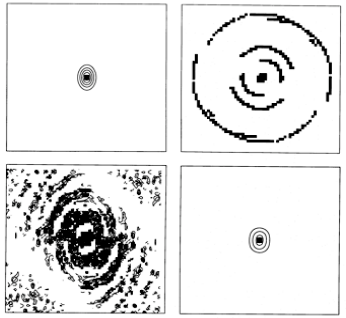
- (radio-) Interferometry
- Gamma Ray Instruments (Integral)
- Period detection in temporal series - Inpainting
- Herschel Data Compression Story: Part 1 and 2.

Radio-Interferometry

$$y = \Theta x$$

Θ = Fourier transform
Φ = Id (or Wavelet transform)

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



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

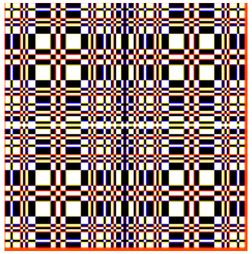
Wavelet - CLEAN minimizes well the l_0 norm

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

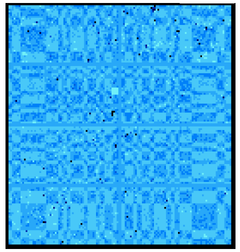
==> (Wiaux et al, 2009), (Cornwell et al, 2009), (Suskimo, 2009).

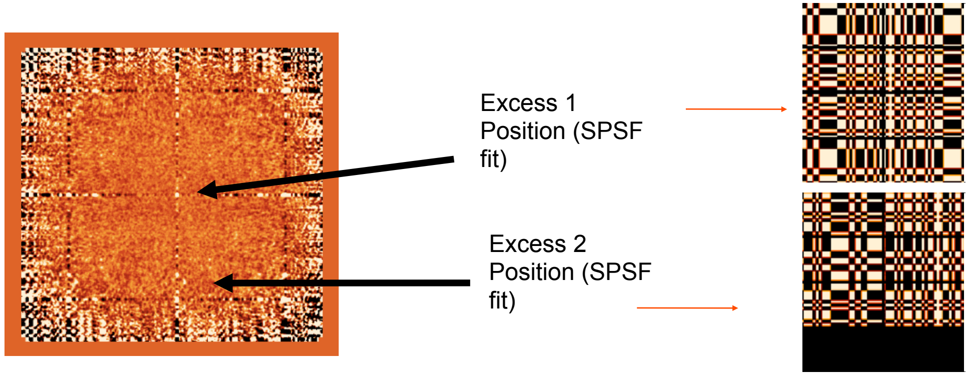
The futur Square Kilometre Array (SKA) radio-interferometer will certainly use such techniques.

Gamma Ray Instruments (Integral) - Acquisition with coded masks



INTEGRAL/IBIS Coded Mask

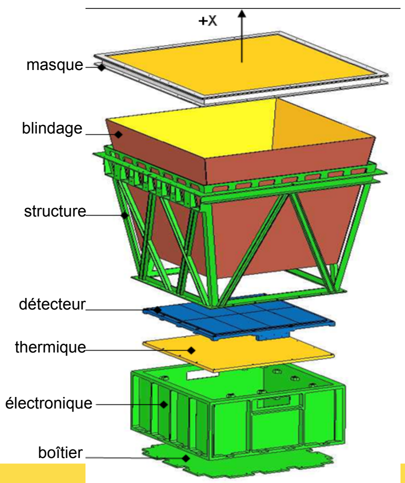




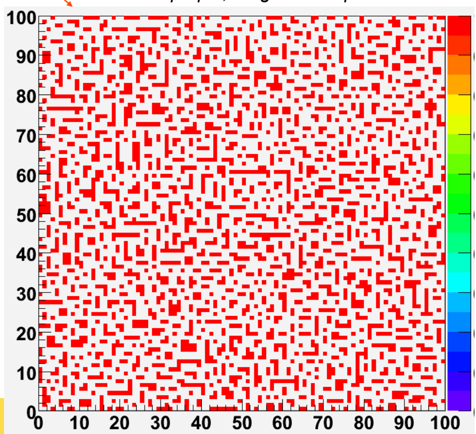
CS gives another point of view on some existing reconstruction methods

ECLAIRs

- ***ECLAIRs french-chinese satellite 'SVOM' (launch in 2014)***
  Gamma-ray detection in energy range 4 - 120 keV
  Coded mask imaging (at 460 mm of the detector plane)
- *<u>detector</u> 1024 cm^2 of Cd Te (80 x 80 pixels)*
- *<u>mask</u>   (100 x 100 pixels)*



blanc = opaque, rouge = transparent



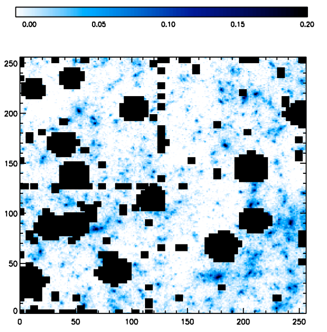
Compressed Sensing Impact in astronomy

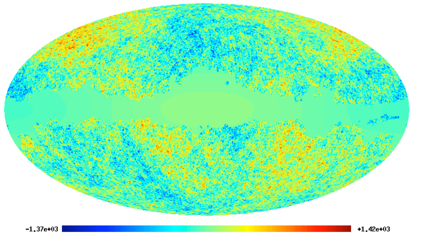
$$\min_{\alpha} \|\alpha\|_1 \quad \text{s.t.} \quad y = \Theta_\Lambda \Phi \alpha$$

- Interpolation of Missing Data: Inpainting

$$\boldsymbol{\Theta}_\Lambda = \mathbf{Id}_\Lambda \qquad \min_{\alpha} \parallel \alpha \parallel_0 \qquad y = M\Phi\alpha = Mx$$

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

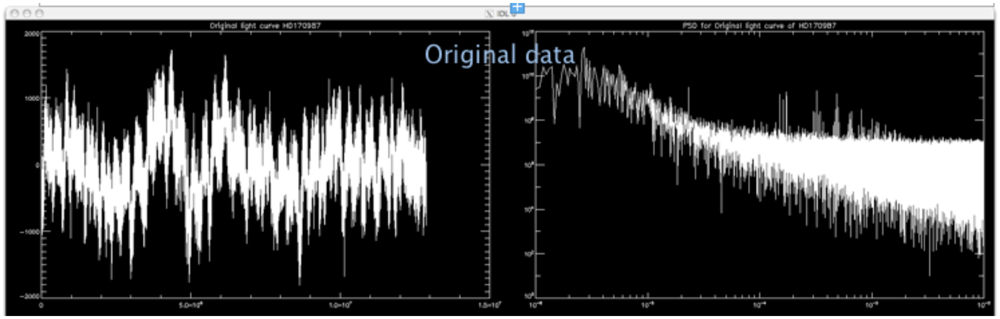




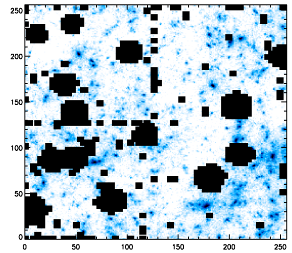
Missing Data

- **Period detection in temporal series**

COROT: HD170987



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



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.*

$$\boldsymbol{\Theta}_\Lambda = \mathbf{Id}_\Lambda \qquad \min_{\alpha} \|\alpha\|_{\ell_0} \text{ s.t. } y = Mx$$

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

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

Iterative Hard Thresholding with a decreasing threshold.

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

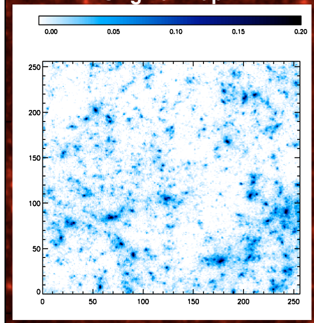
50%

80%

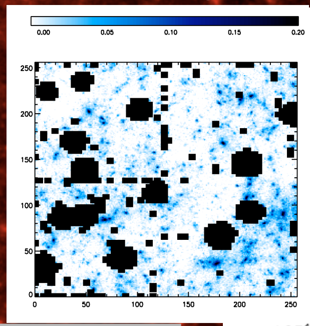
Inpainting :

S. Pires, J.-L. Starck, A. Amara, R. Teyssier, A. Refregier and J. Fadili, "FASTLens (FAst STatistics for weak Lensing) : Fast method for Weak Lensing Statistics and map making", MNRAS, 395, 3, pp. 1265-1279, 2009.

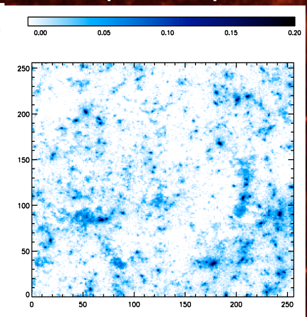
Original map

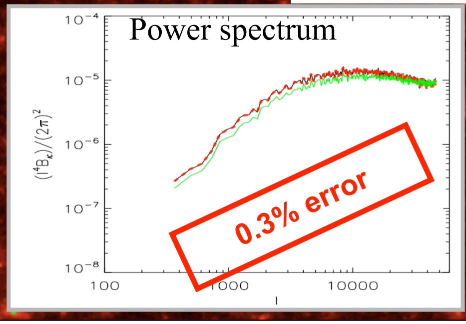


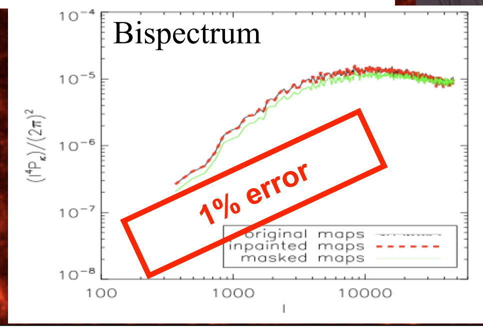
Masked map



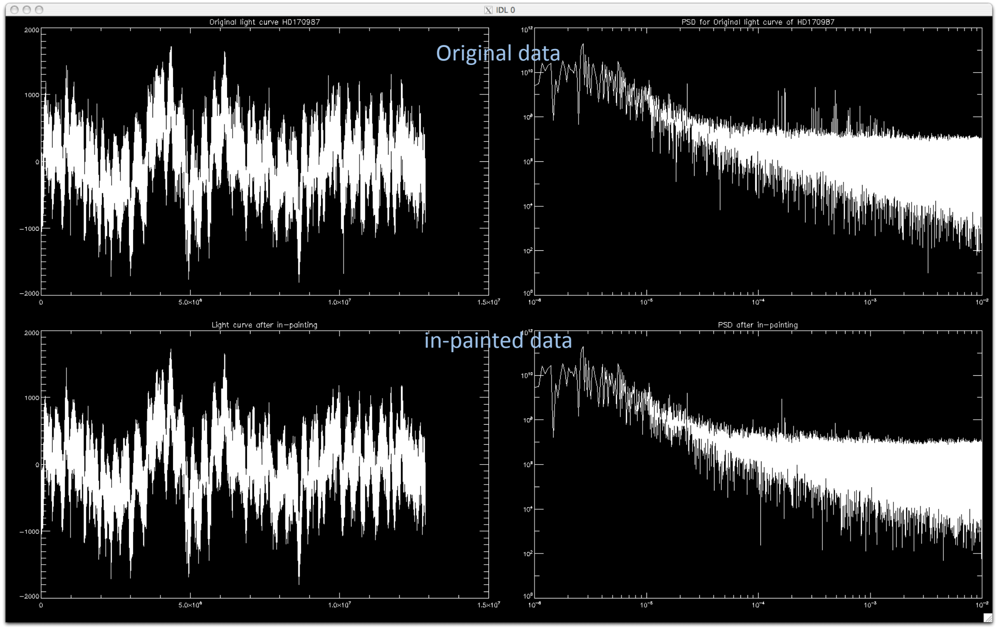
Inpainted map



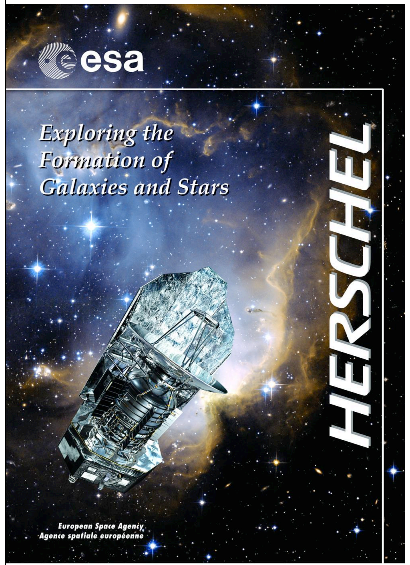




COROT: HD170987 with in-painting (Rafael A. Garcia, SAP)



The Herschel Compressed Sensing Story



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

Goals :
- Understand the beginning of stars formation (molecular clouds).
- Observe galaxies at their formation epoch.

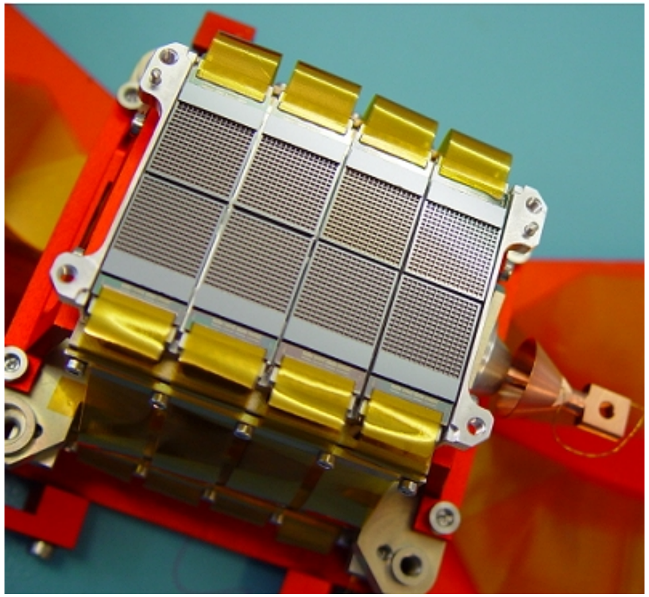
On bord: three instruments:

HIFI : spectromètre hétérodyne (SRON-NL)
PACS : Spectromètre et Photomètre 57-205 μm (MPE-D)
SPIRE : Spectromètre et Photomètre - 200-607 μm

The shortest wavelength band, 57-210 microns, is covered by PACS (Photodetector Array Camera and Spectrometer).

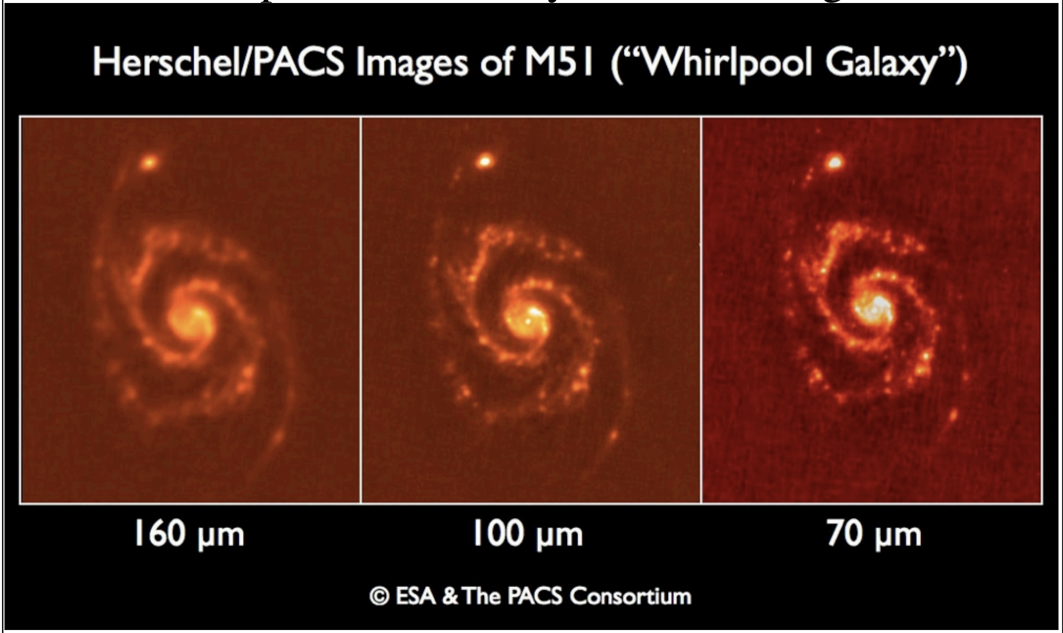
Launch: **May 14, 2009.**

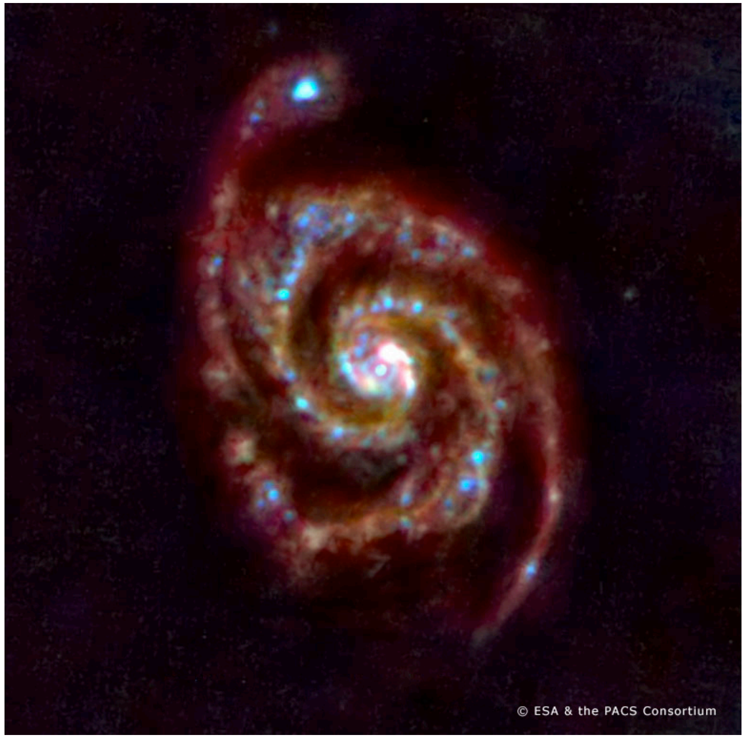
PACS: 8 matrices of 16x16 pixels, cooled down to 300 mK.



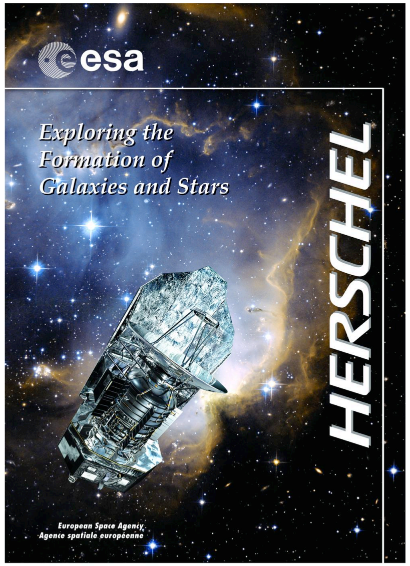
Herschel Status

- Herschel produces already beautiful images:




© ESA & the PACS Consortium

HERSCHEL DATA COMPRESSION PROBLEM



Herschel data transfer problem:

-no time to do sophisticated data compression on board.
-a compression ratio of 8 must be achieved.

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

CS may offer another alternative.

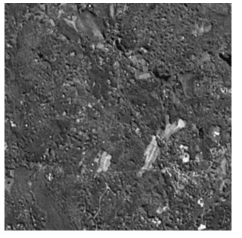
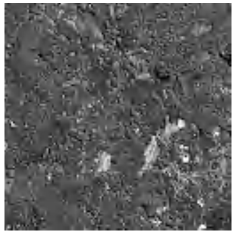
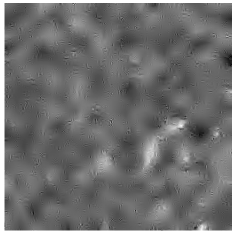
Bobin, J.-L. Starck, and R. Ottensamer, "Compressed Sensing in Astronomy", IEEE Journal of Selected Topics in Signal Processing, Vol 2, no 5, pp 718--726, 2008.

http://fr.arxiv.org/abs/0802.0131

Compressed Sensing For Data Compression

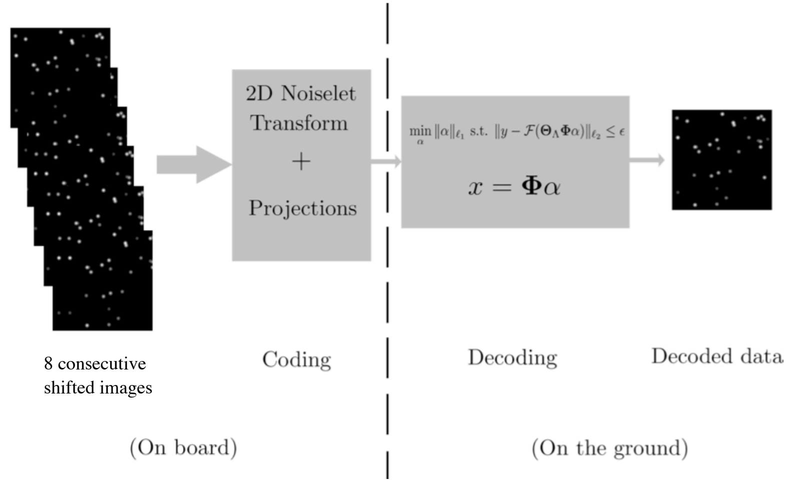
Compressed Sensing presents several interesting properties for data compress:

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

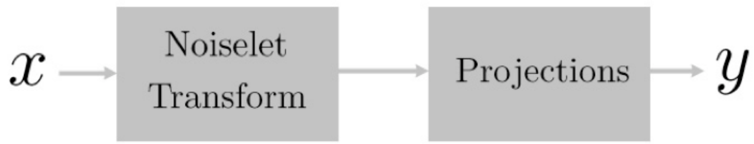
  

But clearly not as competitive to JPEG or JPEG2000 to compress an image.

The proposed Herschel compression scheme



The coding scheme



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

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

Advantages:
Low computational cost (O(n))
Most astronomical data are sparsely represented in a wide range of wavelet bases

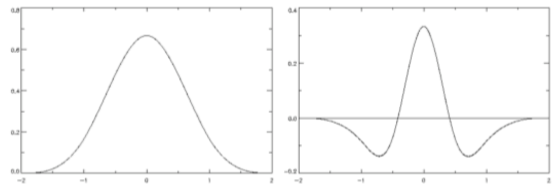
The decoding scheme

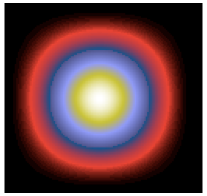
$$\begin{tikzcd} y \arrow[r] & \boxed{\begin{array}{c} \min\limits_{\alpha} \|\alpha\|_{\ell_1} \text{ s.t. } \|y - \mathbf{\Theta}_\Lambda \mathbf{\Phi}\alpha\|_{\ell_2} \leq \epsilon \\ x = \mathbf{\Phi}\alpha \end{array}} \arrow[r] & x \\ & \text{Physical priors} \arrow[u] & \end{tikzcd}$$

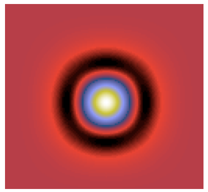
Wavelet Transform in Astronomy

The Isotropic Wavelet and Scaling Functions

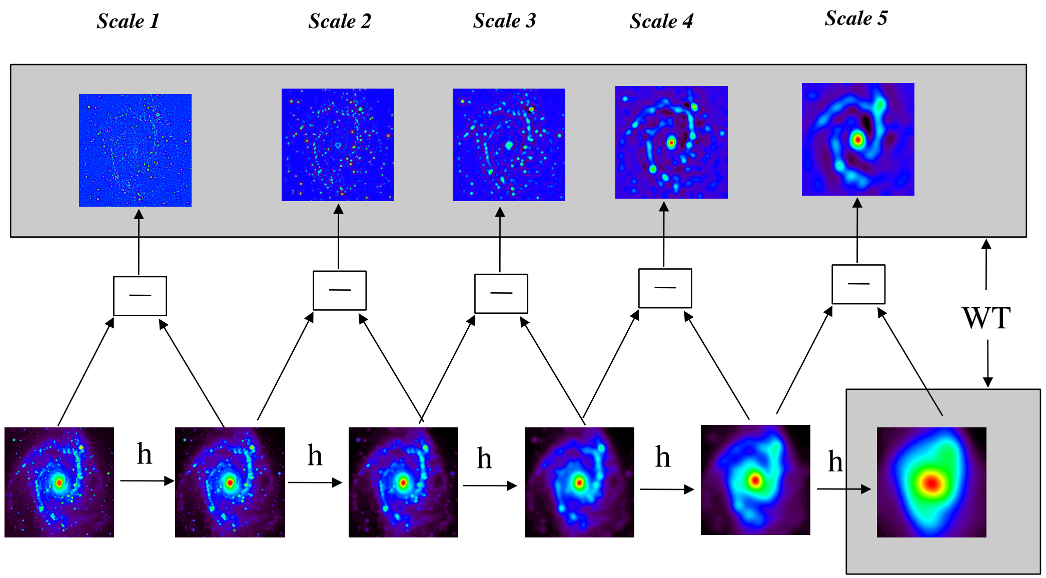
$$
\begin{aligned}
B_3(x) &= \frac{1}{12}(\mid x-2\mid^3 - 4\mid x-1\mid^3 + 6\mid x\mid^3 - 4\mid x+1\mid^3 + \mid x+2\mid^3) \\
\psi(x,y) &= B_3(x)B_3(y) \\
\frac{1}{4}\psi(\frac{x}{2},\frac{y}{2}) &= \phi(x,y) - \frac{1}{4}\phi(\frac{x}{2},\frac{y}{2})
\end{aligned}
$$







ISOTROPIC UNDECIMATED WAVELET TRANSFORM

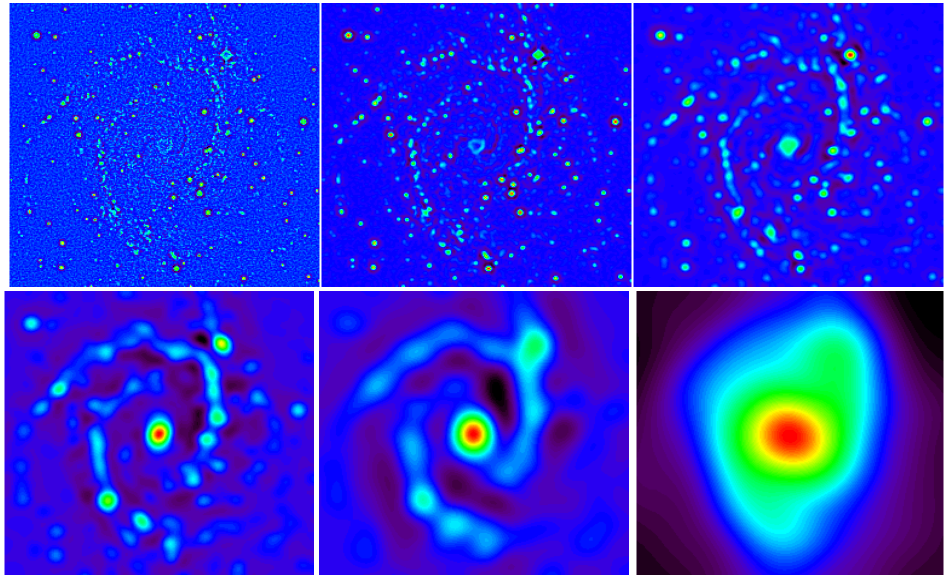


Isotropic Undecimated Wavelet Transform (a trous algorithm)

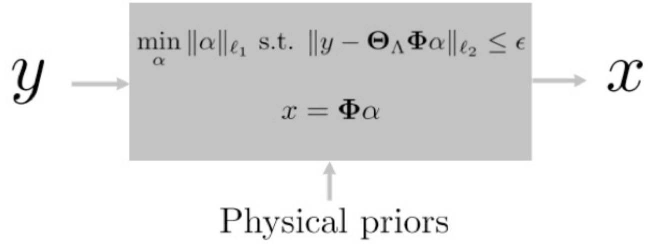
$\varphi = B_3 - \text{spline}, \quad \frac{1}{2}\psi(\frac{x}{2}) = \frac{1}{2}\varphi(\frac{x}{2}) - \varphi(x)$

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

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



Herschel image packets decompression



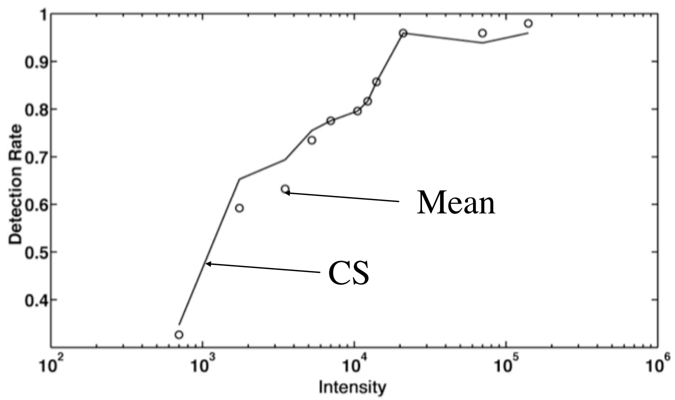
Height consecutive observations of the same field can be decompressed together (forward-backward splitting algorithm)

$$\alpha^{(t+1)} = \mathrm{SoftThresh}_{\mu_t\lambda^{(t)}}\left(\alpha^{(t)} + \mu_t \frac{1}{P}\sum_{i=1}^{P} \boldsymbol{\Phi}^* \mathcal{S}_{-d_i}\left(\Theta^*_{\Lambda_i}\left(y_i - \Theta_{\Lambda_i}\mathcal{S}_{d_i}\left(\boldsymbol{\Phi}\alpha^{(t)}\right)\right)\right)\right)$$

where $\mu_t \in (0, 2P/\sum_i \Theta^2_{\Lambda_i}\boldsymbol{\Phi}^2)$.

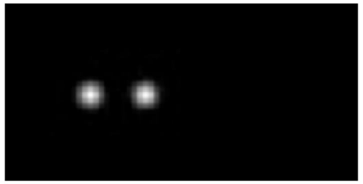
At each iteration, the sought after image is reconstructed from the coefficients $\alpha^{(t)}$ as $x^{(t)} = \boldsymbol{\Phi}\alpha^{(t)}$.

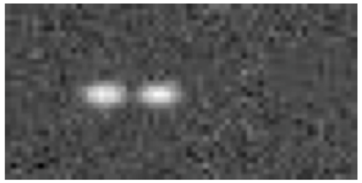
Sensitivity: CS versus mean of 8 images



Resolution: CS versus Mean

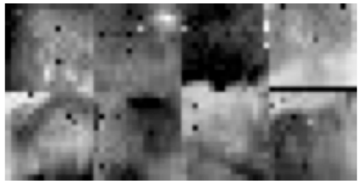
Simulated image

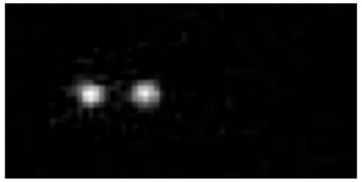




Mean of six images

Simulated noisy image with flat and dark





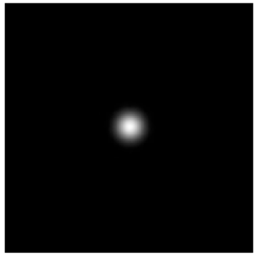
Compressed sensing reconstructed images

Resolution limit versus SNR

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

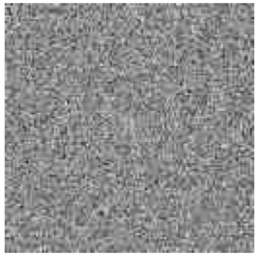
THE CS-BASED COMPRESSION ENTAILS A RESOLUTION GAIN EQUAL TO A 30% OF THE SPATIAL RESOLUTION PROVIDED BY MO6.

Data Fusion: JPEG versus Compressed Sensing

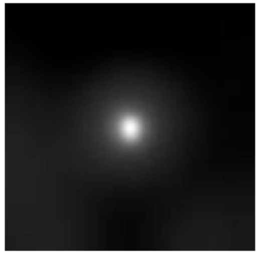


Simulated source

One of the 10 observations



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

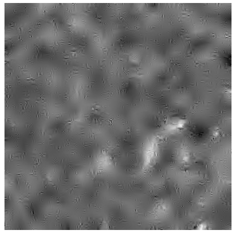


Reconstruction from the 10 compressed sensing images (CR=4)

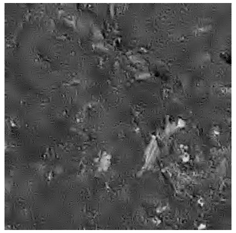
JPEG-2000 Versus Compressed Sensing

Compression Rate: 25

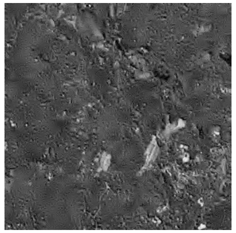
One observation



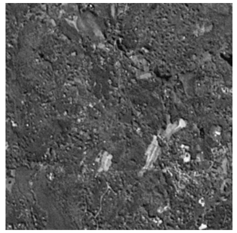
10 observations

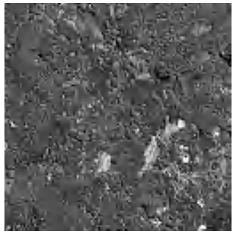


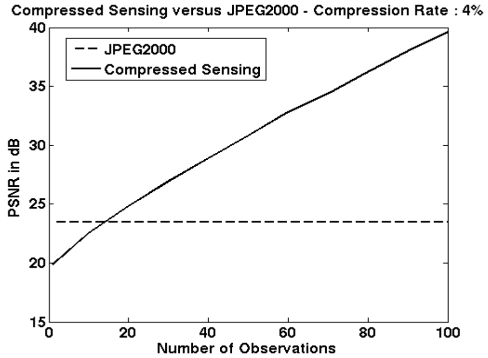
20 observations

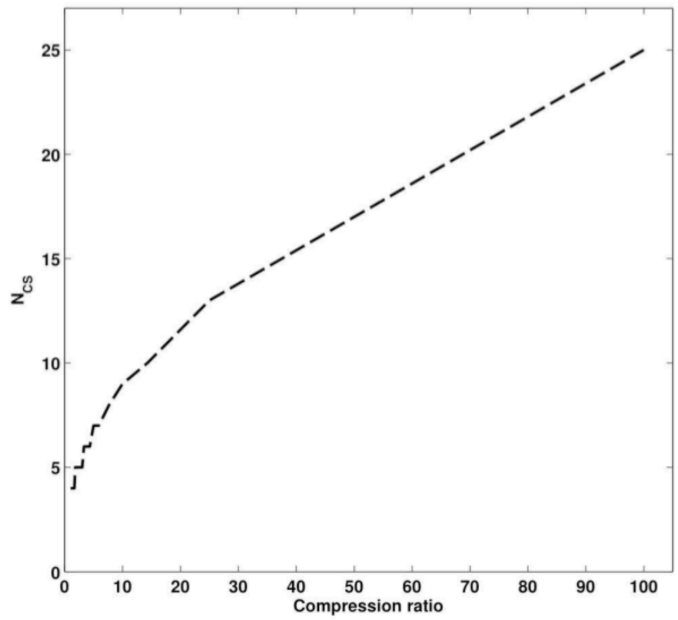


100 observations








N_CS
Compression ratio
0
5
10
15
20
25
0
10
20
30
40
50
60
70
80
90
100

ESA wants to test CS

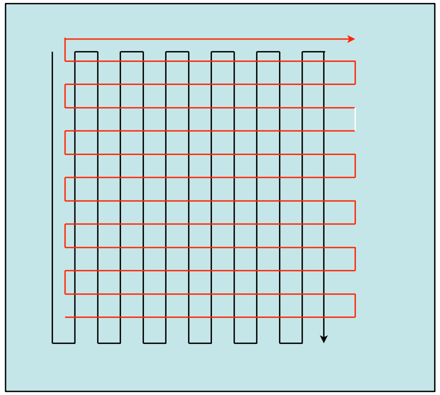
- CS compression **is** implemented in the Herschel on-board software (as an option).
- Astronet Grant: 1 postdoc for 3 years (Nicolas Barbay), from 2009 to 2011.
- CS Tests in flight **started** in November.
- The CS decompression is fully integrated in the data processing pipeline.
- Software developments required for an efficient decompression (taking into account dark, flat-field, PSF, etc) not yet finished.
  Main problem: the background drift:

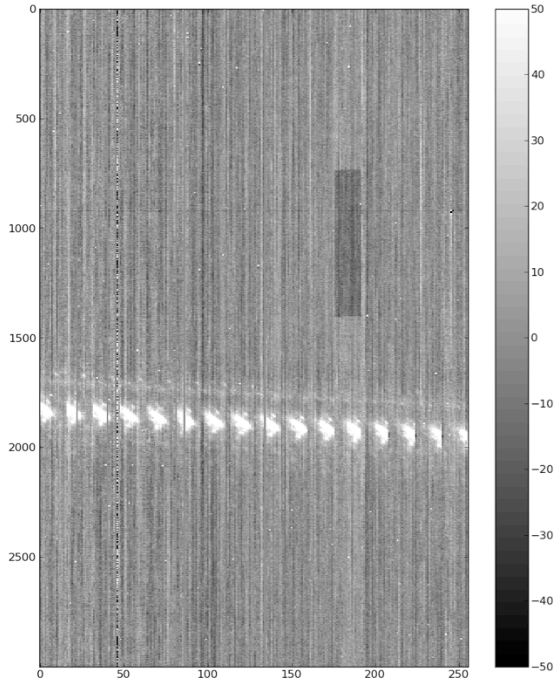
$$Y(i,j,t) = s(i,j) + B(i,j,t) + N(i,j)$$

We need to recover s and B

Observed Data During the Calibration Phase, November 2010, without any compression.

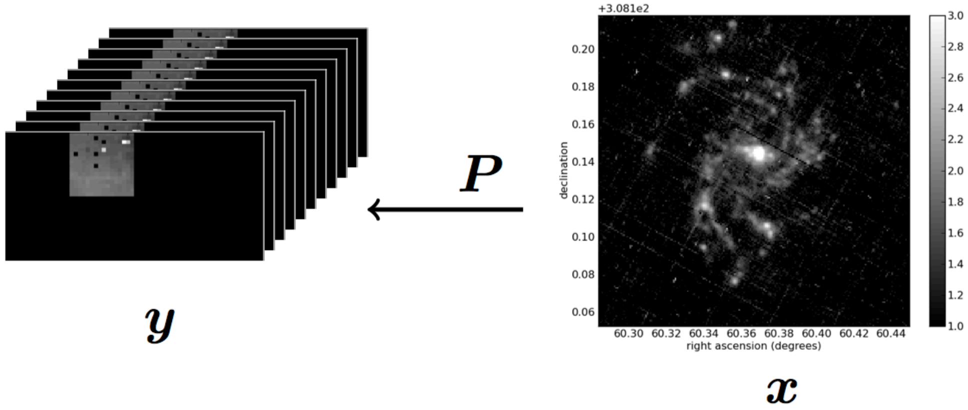
Two scans (16 x 16 pixels at 40 Hz during 25 min each, we obtained 60000 images for each scan, at 85-130 μm). The two scanning are at 90 degrees.





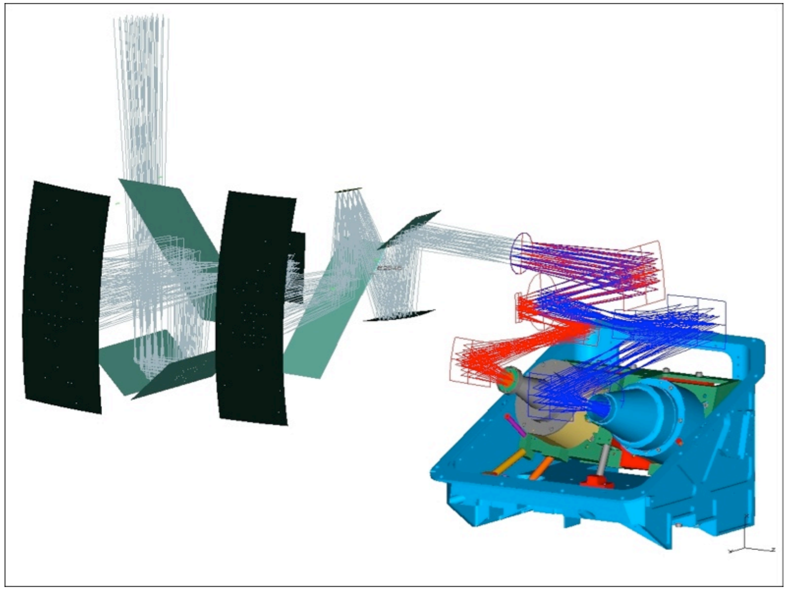
$$Y(i,j,t) = s(i,j) + B(i,j,t) + N(i,j)$$

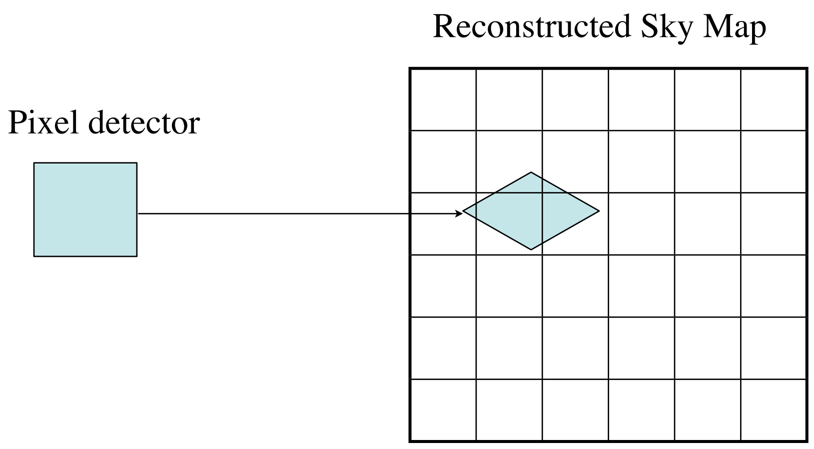
Reconstruction of one sky map from 60000 frames: redundancy of around 200, at each position of the sky image x[i,j].



We remove from each pixel y[i,j,t] its median(y[i,j, t-75: t+75]), to suppress the drift.

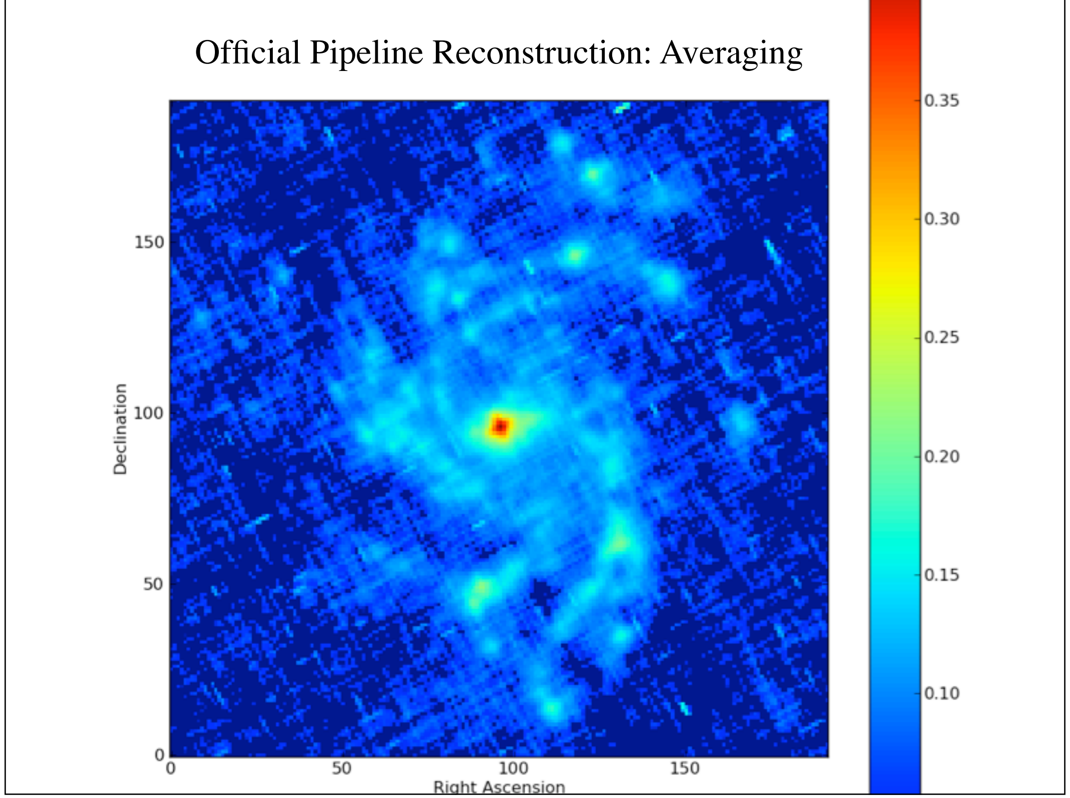
Optical Distortion: more than ten mirrors in the instrument



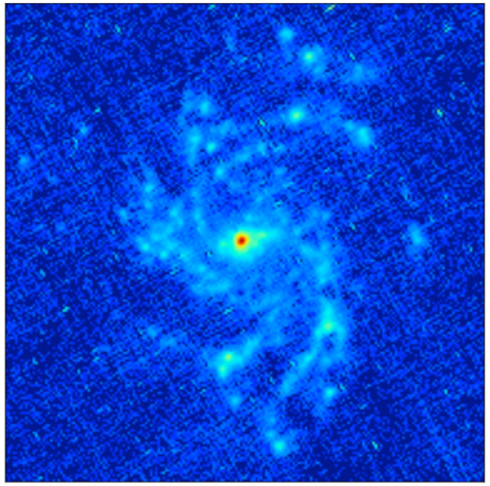


A = **Θ**P, where P is the projection to the camera pixels, and **Θ** is the CS random linear operator.

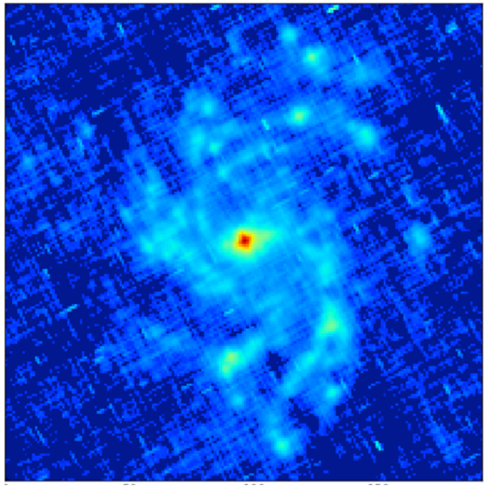
$$\min_{\alpha} \quad \|y - A\Phi\alpha\|^2 + \lambda\|\alpha\|_1 \ ,$$

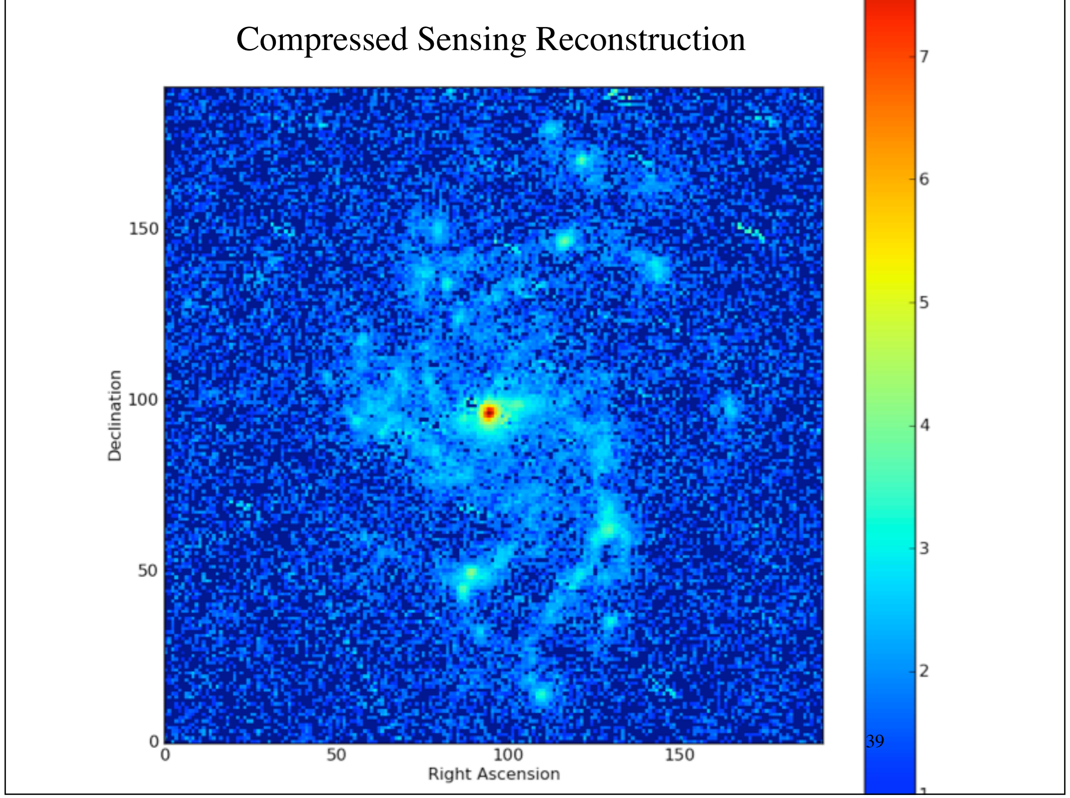

Official Pipeline Reconstruction: Averaging
Declination
Right Ascension
0
50
100
150
0.35
0.30
0.25
0.20
0.15
0.10

Map from Uncompressed Data

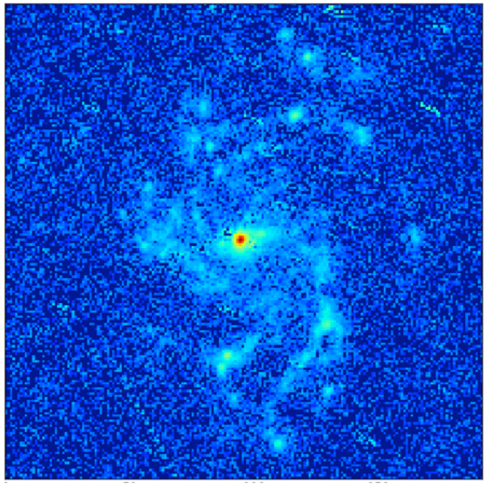


Official Pipeline Reconstruction: Averaging

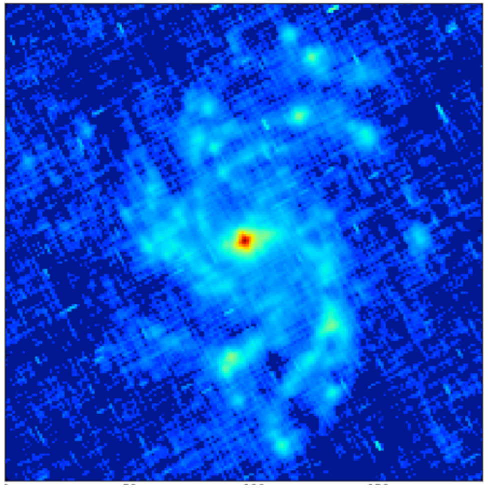



Compressed Sensing Reconstruction
Declination
Right Ascension
0
50
100
150
0
50
100
150
7
6
5
4
3
2
1

Compressed Sensing Reconstruction



Official Pipeline Reconstruction: Averaging

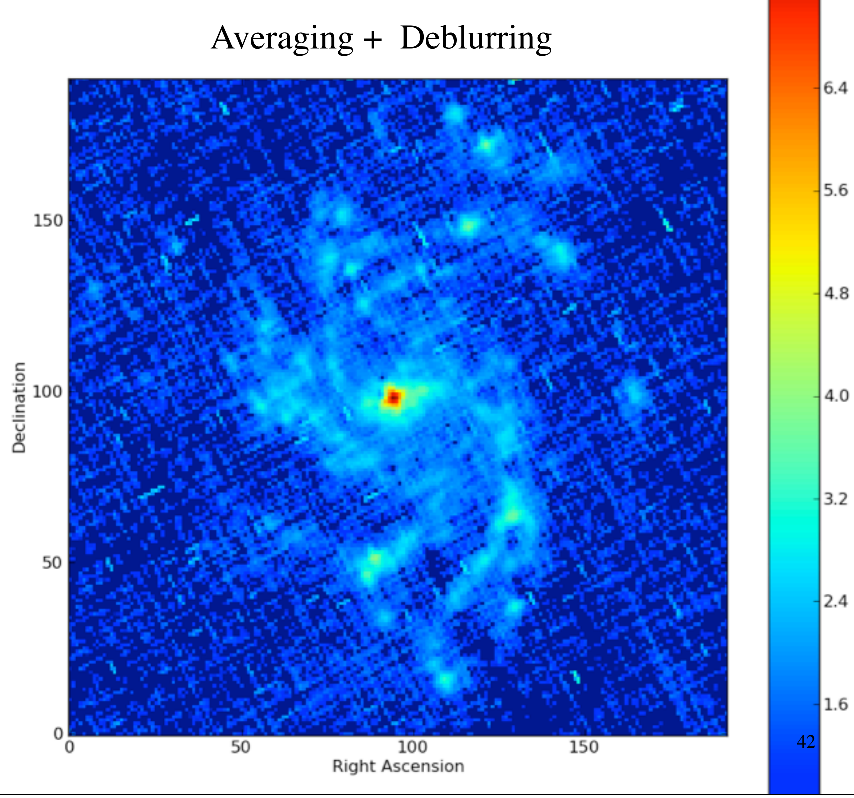


Averaging + Deblurring

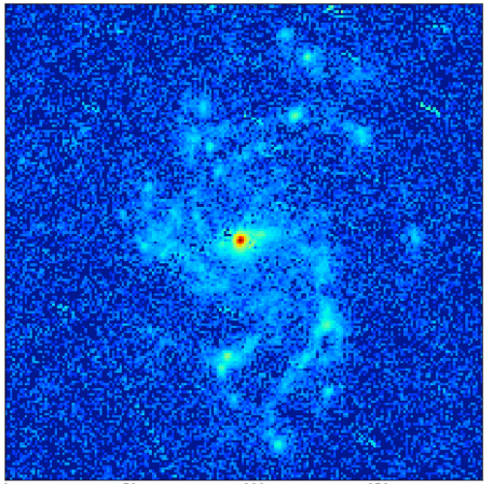
$$\min_{\alpha} \quad \|y - A\Phi\alpha\|^2 + \lambda\|\alpha\|_1 \quad ,$$

A = $\boldsymbol{\Theta}$P, where P is the projection to the camera pixels, and $\boldsymbol{\Theta}$ is the CS random linear operator.

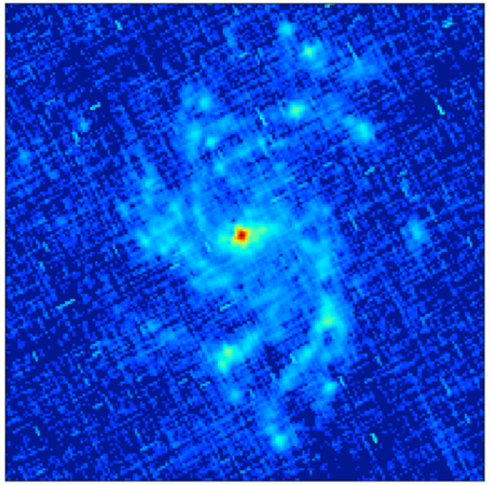
Averaging + Deblurring

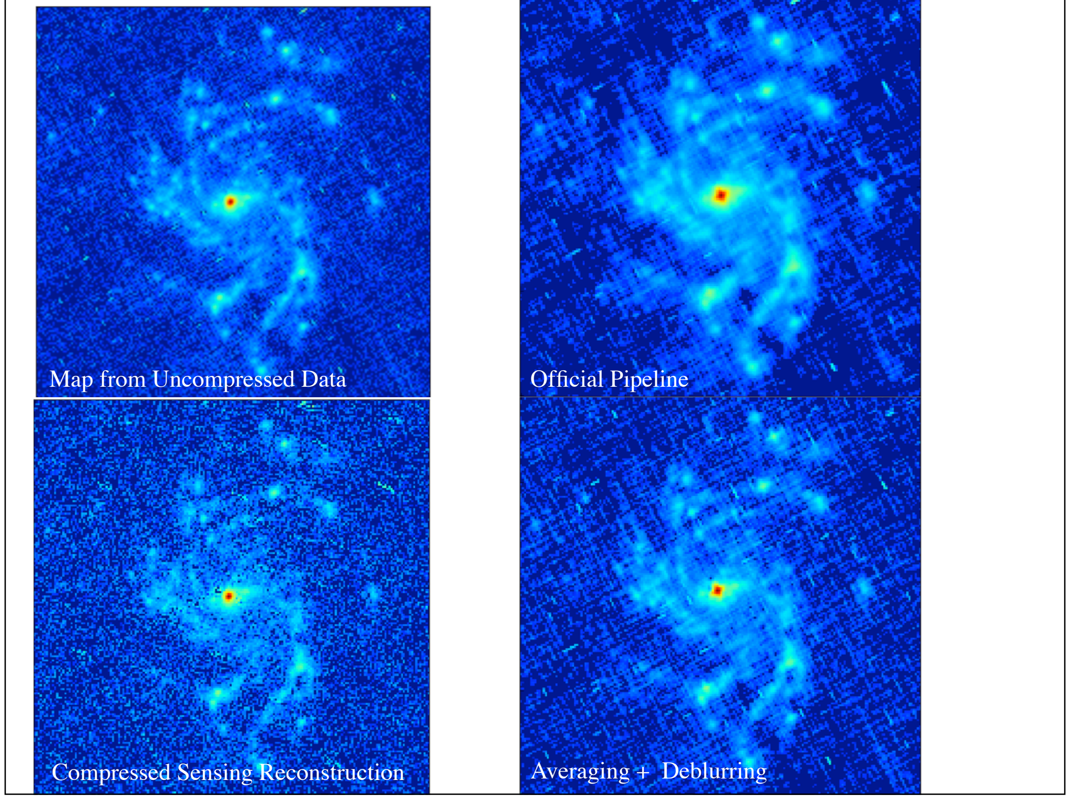


Compressed Sensing Reconstruction



Averaging + Deblurring




Map from Uncompressed Data
Official Pipeline
Compressed Sensing Reconstruction
Averaging + Deblurring

Conclusions on CS for Herschel

- CS works.
- CS is clearly better than Averaging, as predicted from the toy model simulations.
- But only slightly better than Averaging + Deblurring
  ==> at this point, the improvement does not justify to use the CS mode as the standard mode.
- The Averaging-Deblurring solution has been developed thanks to the CS spirit.
- It can however be very useful for some scientific programs, where the resolution is the key of the success.

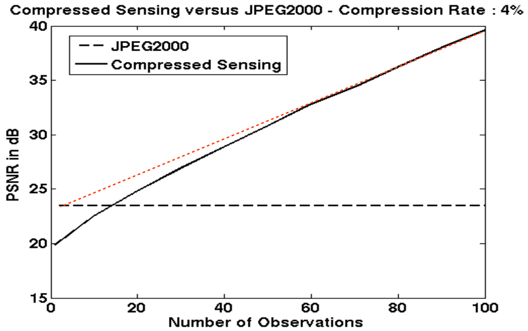
Maybe, it is not the end of the story. Possible improvement:

- Better drift removal.
- Matrix choice (Hadamard, noiselet, etc)
- Dictionary choice.
- Deconvolution would be possible with CS, and further improve the resolution.

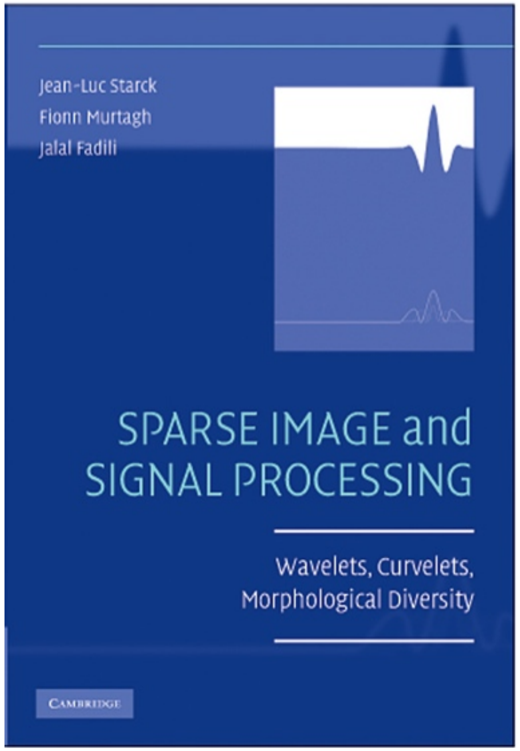
What we have learned for on-board compression:

- Thanks to CS, we clearly can increase the acquisition rate. This is important for instance for the drift background removal.

- CS works, but an Hybrid Compression Scheme (HCS) would certainly be much more appropriate.



- CS will clearly be considered seriously for the design of our future space missions.


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

Non Convex Penalties: Iterative Hard Thresholding

$$\min_{\alpha} \qquad \frac{1}{2}\|y - A\Phi\alpha\|^2 + \frac{\lambda^2}{2}\|\alpha\|_0 \ ,$$

Iterative hard thresholding (Starck et al, 2004; Elad et al 2005; Topp, 2006; Blumensath, 2008; Ramlau, 2008; Maleki and Donoho, 2009) consists of replacing soft thresholding by hard thresholding:

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

Converges to a stationary point which is a local minimizer (Blumensath, 2008).

In compressed sensing recovery, IHT was reported to perform better than IST (Maleki and Donoho, 2009). This is not necessarily true for other inverse problems.

Varying the Regularization Parameter in Iterative Thresholding

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

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

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

- The idea underlying this threshold update recipe has a flavor of homotopy continuation (Osborne, 2000). Continuation has been shown to speed up convergence, to confer robustness to initialization and better recovery of signals with high dynamic range.

- As the sparsity penalties are not convex anymore, continuation and start by a decreasing threshold is even more crucial in IHT than in IST to ensure robustness to initialization and local minima.

- When A=Id and exactly sparse solutions, a sufficient condition based on the incoherence of Φ was given (Maleki et Donoho, 2009) to ensure convergence and correct sparsity recovery by IST with decreasing threshold.

Iterative Soft Thresholding: Simplicity and Robustess

$$\min_{\alpha} \quad \|y - A\Phi\alpha\|^2 + \lambda\|\alpha\|_1 \ ,$$

The IST scheme is very easy to implement (Nowak et al, 2003; Daubechies et al 2004; Combettes et al, 2007):

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

The algorithm is therefore not only simple, but also relatively fast. Furthermore, IST is robust with regard to numerical errors.

However, its convergence speed strongly depends on the operator A, and slow convergence may be observed. Accelerated IST variants have been proposed in the literature (Vonesch et al, 2007; Elad et al 2008; Wright et al., 2008; Nesterov, 2008 and Beck-Teboulle, 2009; etc).