Compressed Sensing in Astronomy

Jérôme Bobin, Jean-Luc Starck, and Roland Ottensamer

Abstract—Recent advances in signal processing have focused on the use of sparse representations in various applications. A new field of interest based on sparsity has recently emerged: *compressed sensing*. This theory is a new sampling framework that provides an alternative to the well-known Shannon sampling theory. In this paper, we investigate how *compressed sensing* (CS) can provide new insights into astronomical data compression. We first give a brief overview of the compressed sensing theory which provides very simple coding process with low computational cost, thus favoring its use for real-time applications often found onboard space mission. In practical situations, owing to particular observation strategies (for instance, raster scans) astronomical data are often redundant; in that context, we point out that a CS-based compression scheme is flexible enough to account for particular observational strategies. Indeed, we show also that CS provides a new fantastic way to handle multiple observations of the same field view, allowing us to recover low level details, which is impossible with standard compression methods. This kind of CS data fusion concept could lead to an elegant and effective way to solve the problem ESA is faced with, for the transmission to the earth of the data collected by PACS, one of the instruments onboard the Herschel spacecraft which will launched in late 2008/early 2009. We show that CS enables to recover data with a spatial resolution enhanced up to 30% with similar sensitivity compared to the averaging technique proposed by ESA.

Index Terms—Astronomy, compressed sensing, remote sensing, sparsity, wavelets.

I. INTRODUCTION

FROM year to year, the quantity of astronomical data increases at an ever-growing rate. In part, this is due to very large digital sky surveys in the optical and near infrared, which in turn has been made possible by the development of digital imaging arrays such as charge-coupled devices (CCDs). The size of digital arrays is continually growing, pushed by the demands of astronomical research for ever larger quantities of data in ever shorter time periods. As a result, the astronomical community is also confronted with a rather desperate need for data compression techniques.

Several techniques have in fact been used, or even developed, for astronomical data compression [1], [2]. For some projects, we need to achieve huge compression ratios, which cannot be obtained by current methods without introducing unacceptable distortions. Furthermore, for most astronomical data compression problems, three main properties must be under control: resolution (point spread function), sensitivity (ability to detect low level signals), and photometry.

The Herschel satellite,[1] which will be launched in late 2008/early 2009, is faced with a similar problem. Indeed, the photometer data need to be compressed by a factor of approximately 6 to be transferred. As the CPU load has to be extremely small, conventional compression methods cannot be used.

Recently, an alternative sampling theory has emerged which shows that signals can be recovered from far fewer samples (measurements) than what the Nyquist/Shannon sampling theory states. This new theory coined *compressed sensing* or (compressive sensing) introduced in the seminal papers [3], [4] relies on the compressibility of signals or more precisely on the property for some signals to be sparsely represented. From the compressed sensing viewpoint, sparse signals could be acquired "economically" (from a few samples) *without loss of information*. It introduces new conceptions in data acquisition and sampling.

Scope of the Paper: We propose a new alternative approach for the transmission of astronomical images, based on CS. Section II reviews the principle of the CS theory. We will see that CS can be used in different ways for data compression purposes: i) sensor design or ii) as a classical "compression-decompression" two-stage scheme. In this paper, we will particularly focus on the latter. Indeed, in practical situations (more particularly for onboard applications), CS provides a simple coding or compression stage that only requires a low computational burden. Most of the computational complexity is in the decoding step. In Section III, we show how CS offers us a flexible data compression framework as i) compression and decompression are decoupled and ii) CS is able to account for the redundancy of the data due to some particular observational strategies to enhance the decoding process. It is particularly profitable when multiple observations of the same sky area are available. This happens very often in astronomical imaging when we need to build a large map from a micro-scan or a raster-scan. Section IV highlights the effectiveness of the proposed CS-based compression for solving the Herschel data compression problem. Indeed, we show the advantage of CS over the averaging approach (proposed by the European Spatial Agency-ESA) which has been considered so far.

II. A BRIEF INTRODUCTION TO COMPRESSED SENSING

In this section, we give a brief and non-exhaustive review of compressed sensing and show how this new sampling theory will probably lead to a "revolution" in signal processing and communication theory. For more exhaustive tutorials in this

J. Bobin and J.-L. Starck are with the CEA, Saclay, IRFU, SEDI/Service d'Astrophysique, Laboratoire AIM, CEA/DSM-CNRS-Université Paris Diderot, 91191 Gif-sur-Yvette, France (e-mail: jerome.bobin@cea.fr; jstarck@cea.fr).

R. Ottensamer is with the Institute of Astronomy, University of Vienna, A-1180 Wien, Austria (e-mail: ottensamer@astro.univie.ac.at).

[1]See http://www.esa.int/science/herschel.

field, we refer the reader to the review papers [5], [6]. In this paper, we will assume that the signal x belongs to $\mathbb{R}^t$ (written as a column vector with t entries or samples). We will also assume that x is **compressible**.

A. Gist of Compressed Sensing

Compressibility: The content of most astronomical images if often well structured: diffuse gas clouds, point sources ... to name only a few. Recent advances in harmonic analysis have provided tools that efficiently represent such structures (wavelets, curvelets, ... to name a few). In this context, efficient representations mean sparse representations. Let us consider a signal x of size t. Let $\mathbf{\Phi}$ be an orthonormal basis (e.g. classically, an orthogonal wavelet basis for astronomical data processing) and let us consider the projection of x in $\mathbf{\Phi}$

$$\alpha = \mathbf{\Phi}^T x. \tag{1}$$

The signal x is said to be **sparse** in $\mathbf{\Phi}$ if most entries of the so-called coefficient vector α are zero or close to zero and thus only a few have significant amplitudes. In other words, the signal x can be efficiently approximated (with low ℓ_2 approximation error or distortion) from only a few significant coefficients. Then such sparse signal is said to be *compressible*. Note that, in the last decade, sparsity has emerged as one of the leading concepts in a wide range of signal processing applications. More formally, we will distinguish two categories of compressible signals.

— **Strict sparsity**: Only $K \ll t$ entries of α are different form zero. x is said to be K-sparse in $\mathbf{\Phi}$.
— **Wide sense compressibility**: A more realistic definition of compressibility consists in describing how the entries of α behave. Let us consider that $\alpha[1] \geq \cdots \geq \alpha[t]$. Then x is said to be compressible in $\mathbf{\Phi}$, if there exists $p \in]0,1]$ such that: $|\alpha[i]| \leq Ci^{-1/p}$. Here p defines a kind of sparsity or compressibility degree. Real-world data are more akin to be wide sense compressible.

Acquiring Incoherent Measurements: Assuming that x is compressible (i.e., x has a sparse representation in a particular basis $\mathbf{\Phi}$), x can be efficiently approximated (with low distortion) from only a few entries of $\alpha = \mathbf{\Phi}^T x$. In the Compressed Sensing framework, the signal x is not acquired directly; it is *observed* or *measured* from $M < t$ measurements $\{y_k\}_{k=1,\cdots,M}$. These measurements are obtained by projecting the signal x on a set of so-called measurement vectors $\{\theta_k\}_{k=1,\cdots,M}$ as follows:

$$y_k = \langle x, \theta_k \rangle. \tag{2}$$

Each sample y_k is then the scalar product of x with a specific *measurement* vector θ_k. The gist of compressed sensing relies on two major concepts: i) the data to be compressed are indeed compressible; more precisely the data x have a "structured" content so that they can be sparsely represented in some basis $\mathbf{\Phi}$; ii) the measurement vectors $\{\theta_k\}_{k=1,\cdots,M}$ are non adaptive (they should not depend on x) and incoherent with the basis in which x is assumed to be sparse.

As stated in [5], two main categories of measurement ensembles can be used for CS coding.

— **Random measurements**: $\mathbf{\Theta}$ is not explicitly used; the measurements $\{y_k\}_{k=1,\cdots,M}$ are random linear combinations of the entries of x. Fourier, Binary or Gaussian measurements are widely used. In the CS framework, incoherent measurements can be obtained by using random ensembles (see [3], [7] and references therein). Randomness is likely to provide *incoherent* projections.
— **Incoherent measurements**: In that case, $\mathbf{\Theta}$ is a deterministic basis which is assumed to be *incoherent* with $\mathbf{\Phi}$. More quantitatively, incoherence between $\mathbf{\Theta}$ and $\mathbf{\Phi}$ is measured by their mutual coherence: $\mu = \max_{i,k} |\langle \phi_i, \theta_k \rangle|$. The lower μ is, the more incoherent $\mathbf{\Theta}$ and $\mathbf{\Phi}$ are. In practical situations, typical astronomical data are compressible in a wavelet basis $\mathbf{\Phi}$; a good choice for $\mathbf{\Theta}$ is the noiselet basis [8].

In this paper, measurement vectors are designed by selecting at random a set (indexed by Λ) of vectors from a deterministic orthonormal basis $\mathbf{\Theta}$ as suggested in [9] $y = \mathbf{\Theta}_\Lambda x$.

An empirical interpretation: In the CS framework, the signal to be transferred is $y = \mathbf{\Theta}\mathbf{\Phi}\alpha$. Let recall that the backbone of CS is twofold.

— **Data are compressible**: Only a few entries of α have a significant amplitude; x is then almost entirely determined from only a few entries α.
— **Measurements are incoherent**: The measurement matrix $\mathbf{\Theta}$ and $\mathbf{\Phi}$ are incoherent. From a empirical point of view, the incoherence of $\mathbf{\Theta}$ and $\mathbf{\Phi}$ means that the information carried by a few entries of α is spread all over the M entries of $y = \mathbf{\Theta}\mathbf{\Phi}\alpha$. Each sample y_k is likely to contain a piece of information of each significant entry of x. As $M < t$, the ratio $\rho = M/t$ is equivalent to a compression ratio.

B. Signal Recovery

1) Exact Solutions: The previous paragraph emphasized on the way the compression step should be devised. The decompression step amounts to recovering the original signal x out of the compressed signal $y = \mathbf{\Theta}_\Lambda x$. Furthermore, x is known *a priori* to be compressible in $\mathbf{\Phi}$. Then the recovery problem boils down to emphasizing on the sparsity of the vector $\alpha = x\mathbf{\Phi}$. As proposed in [3], [4], the decompression or decoding step is equivalent to solving the following optimization problem:

$$\min_\alpha \|\alpha\|_{\ell_1} \text{ s.t. } y = \mathbf{\Theta}_\Lambda \mathbf{\Phi}\alpha. \tag{3}$$

Several strong recovery results in the particular CS framework have been proved based on specific assumptions with random measurement ensembles (see [3], [10]). For instance, in the extreme strict sparsity case where only K entries of α are nonzero, conditions are given in [3] showing that the problem in (3) provides the *exact* solution α. Nevertheless, the data are often corrupted by noise. A more realistic compression model would be the following:

$$y = \mathbf{\Theta}_\Lambda(x + n) \tag{4}$$

where n is a white Gaussian noise with variance σ_n^2. As the measurement matrix $\mathbf{\Theta}_\Lambda$ is a submatrix of the orthonormal matrix $\mathbf{\Theta}$, the projected noise $n_\Lambda = \mathbf{\Theta}_\Lambda n$ is still white and Gaussian

with the same variance σ_n^2. The projected data are then recast as follows: $y = \boldsymbol{\Theta}_\Lambda x + n_\Lambda$. The recovery step then boils down to solving the next optimization problem

$$\min_\alpha \|\alpha\|_{\ell_1} \text{ s.t. } \|y - \boldsymbol{\Theta}_\Lambda \boldsymbol{\Phi} \alpha\|_{\ell_2} \leq \epsilon \qquad (5)$$

where ϵ is an upper bound of $\|n\|_{\ell_2}$. Defining $\epsilon = \sqrt{M + 2\sqrt{2M}} \sigma_n$ provides a reasonable upper bound on the noise ℓ_2 norm, with overwhelming probability. In [11] conditions are given that guarantee some optimality results from the problem in (5).

The convex program (second-order cone program) in (5) then provides an efficient and robust mechanism to provide an approximate to the signal x.

III. COMPRESSED SENSING IN ASTRONOMY

In the next sections, we focus on applying the compressed sensing framework to astronomical remote sensing. In Section II, we show that compressed sensing and more precisely its way of coding information provides alternatives to astronomical instrument design. Section II-B gives emphasis on the ability of CS decoding to easily account for the redundancy of the data (for instance, provided by specific observational strategies) thus improving the whole compression performances.

A. Compressed Sensing as a Way of Designing Sensors

The philosophy of compressed sensing (i.e., projecting onto incoherent measurement ensembles) should be directly applied in the design of the detector. Devising an optical system that directly "measures" incoherent projections of the input image would provide a compression system that encodes in the analog domain.

Interestingly, this kind of measurement paradigm is far from being science-fiction. Indeed, in the field of γ-ray imaging, the so-called coded-masks[2] (see [12] and references therein) are used since the sixties and are currently operating in the ESA/Integral space mission.[3] In γ-ray (high energy) imaging, coded masks are used as aperture masks scattering the incoming γ photons. Empirically, γ-ray data are overwhelmingly composed of point sources (i.e. already rather sparse). The coded mask scatters the photons coming from each point sources over almost all the detectors. Each detector then provide incoherent projection. The last step to compressed sensing would amount to selecting only a few detector's signal (potentially at random) to be transmitted thus completing the compression stage. The first application of compressed sensing then dates back to the 1960s! In the compressed sensing community, the coded mask concept has inspired the design of the celebrated "compressed sensing camera" [13] that provide effective image compression with a single pixel. Simulations involving coded mask for compressive imaging have been made by Willet in [14].

Similarly, (radio-) interferometric data are acquired in the Fourier domain (i.e., visibilities), and the observed signal is known to be sparse in the direct domain for unresolved objects or in the wavelet domain for extended sources. A posteriori, it is not surprising that the best methods which have been proposed in the past for reconstructing such images are based on sparsity. Indeed, the famous CLEAN algorithm [15] and its multiscale version [16], [17] can be seen as matching pursuit algorithms [18], and therefore enforce the sparsity of the solution, in the direct space for CLEAN and in the wavelet space for Multiresolution CLEAN. Recent l_0 and l_1 CS reconstruction methods (see, for instance, [19] and [20]) would certainly improve the computation time to derive these dirac-sparse and wavelet-sparse solutions.

B. A Flexible Compression/Decompression Scheme

In this section, we will particularly focus on using Compressed Sensing as a classical "compression/decompression" two stages technique. We will emphasize on showing that Compressed Sensing have several advantages over standard compression techniques such as the celebrated JPEG2000[4] compression standard.

Computational Complexity: The onboard computational complexity of CS compression depends on the choice of $\boldsymbol{\Theta}_\Lambda$ (for instance, noiselets in the forthcoming experiments). The only computational cost required by a CS-based compression is the computation of $\boldsymbol{\Theta}_\Lambda x$. When noiselets are used, their computational cost evolves as $\mathcal{O}(t)$ thus involving a low CPU load which is lower than the computational burden required by JPEG2000 (typically $\mathcal{O}(t \log(t))$). Furthermore, the CS projections (noiselets in the forthcoming examples), require no further encoding in contrast to classical compression methods such as JPEG2000.

Decoupling: In contrast to classical compression techniques, there is a complete decoupling between the compression and the decompression in the CS framework. Therefore the decompression step can be changed while keeping the same compressed data. This is a very nice property. Indeed, we have seen that the quality of the decompressed data is related to the sparsity of the data in a given basis $\boldsymbol{\Phi}$. If we discover in a few years a new basis which leads to a better sparsity of the data, then we can still improve the quality of the decompressed data.

Data Fusion Perspective: Accounting for the Redundancy of the Data: In astronomy, remote sensing data involving specific scanning strategies (raster scans) often provide redundant data which cannot be accounted for by standard compression techniques. In a general framework, let us consider that N observations of the same sky area are available: $\{y_i\}_{i=1,\cdots,N}$ such that

$$\forall i = 1, \cdots, N; \quad y_i = \boldsymbol{\Theta}_{\Lambda_i} x + n_i \qquad (6)$$

where $\{\boldsymbol{\Theta}_{\Lambda_i}\}_{i=1,\cdots,N}$ are N independent random submatrices of $\boldsymbol{\Theta}$ with $\text{Card}(\Lambda_i) = M$. Clearly, it would be worth recovering x from the N compressed observations $\{y_i\}_{i=1,\cdots,N}$. We then propose to substitute the decompression problem in (5) with the following:

$$\min_\alpha \|\alpha\|_{\ell_1} \text{ s.t. } \sum_{i=1}^{N} \|y_i - \boldsymbol{\Theta}_{\Lambda_i} \boldsymbol{\Phi} \alpha\|_{\ell_2} \leq \sqrt{N}\epsilon. \qquad (7)$$

---

[2]We invite the interested readers to visit the following site that is devoted to coded aperture imaging: http://astrophysics.gsfc.nasa.gov/cai/.

[3]See http://sci.esa.int/science-e/www/area/index.cfm?fareaid=21.

[4]See http://www.jpeg.org/jpeg2000/.

Owing to the convexity of this problem, it can be recast in the following Lagrangian form:

$$\min_{\alpha} \gamma \|\alpha\|_{\ell_1} + \frac{1}{2} \sum_{i=1}^{N} \|y_i - \boldsymbol{\Theta}_{\Lambda_i} \boldsymbol{\Phi} \alpha\|_{\ell_2}^2 . \qquad (8)$$

Let define $\boldsymbol{\Pi}_i$ as the diagonal matrix of size $t \times t$ the entries of which are defined as follows:

$$\forall k \in \{1, \cdots, t\}; \quad \boldsymbol{\Pi}_i[k,k] = \begin{cases} 1, & \text{if } k \in \Lambda_i \\ 0, & \text{otherwise} \end{cases} \qquad (9)$$

where $\boldsymbol{\Pi}_i[k,k]$ is the kth diagonal element of $\boldsymbol{\Pi}_i$. Let define the signal $y_i^{\sharp}$ of size t as follows:

$$y_{i\,\Lambda_i}^{\sharp} = y_i \text{ and } y_{i\,\Lambda_i^c}^{\sharp} = 0 \qquad (10)$$

where Λ_i^c is complement of Λ_i in $\{1, \cdots, t\}$. Inspired by recent iterative thresholding methods [21], [22], we propose solving the problem in (8) by means of a projected Landweber iterative algorithm. At iteration (h), the coefficients α would be updated as follows:

$$\alpha^{(h)} = \mathcal{S}_{\gamma} \left\{ \alpha^{(h-1)} + \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\Phi}^T \boldsymbol{\Theta}^T \boldsymbol{\Pi}_i \left(y_i^{\sharp} - \boldsymbol{\Pi}_i \boldsymbol{\Theta} \boldsymbol{\Phi} \alpha^{(h-1)} \right) \right\} \qquad (11)$$

where $\mathcal{S}_{\gamma}$ is the soft-thresholding operator with threshold γ. Further calculation entails appreciable simplifications

$$\alpha^{(h)} = \mathcal{S}_{\gamma} \left\{ \frac{1}{N} \boldsymbol{\Phi}^{\mathbf{T}} \boldsymbol{\Theta}^{\mathbf{T}} \sum_{i=1}^{N} \left[y_i^{\sharp} + (\mathbf{I} - \boldsymbol{\Pi}_i) \boldsymbol{\Theta} \boldsymbol{\Phi} \alpha^{(h-1)} \right] \right\} \qquad (12)$$

where $\mathbf{I}$ is the identity matrix of size $t \times t$.

Choosing the Regularization Parameter γ: The problem in (12) is a general sparse decomposition problem in a redundant dictionary (see [23], [24] and references therein). The choice of the regularization parameter γ is crucial as it balances between the sparsity constraint and the how the solution fits the data. Classical approaches would advocate the use of cross-validation to estimate an optimal value of γ. Nevertheless, cross-validation is computationally expensive and thus not appropriate for large scale problems.

In a different framework, the same kind of optimization problem has been solved using a specific iterative hard-thresholding algorithm coined Morphological Component Analysis (MCA—see [25]) for which the threshold decreases. Inspired by MCA, hard-thresholding is used and the threshold γ is decreased at each iteration. It starts from $\gamma^{(0)} = \|\boldsymbol{\Phi}^{\mathbf{T}} \boldsymbol{\Theta}^{\mathbf{T}} \sum_i y_i^{\sharp}\|_{\infty}$ and decreases towards $\gamma_{\min}$. The value of $\gamma_{\min}$ is 0 in the noiseless case. When noise corrupts the data $y^{\sharp}$, $\gamma_{\min}$ may depend on the noise level. In Section IV, numerical results are given. In these experiments, noise contamination is assumed to be white Gaussian with zero mean and variance σ_n^2. In this case, the final threshold is chosen as $\gamma_{\min} = 3\sigma_n$ which gives an upper bound for noise coefficients with overwhelming probability.

C. Example: Joint Recovery of Multiple Observations

For instance, let us consider that the data are made of $N = 100$ images $\{x_i\}_{i=1,\cdots,N}$ such that each image x_i is a noise-
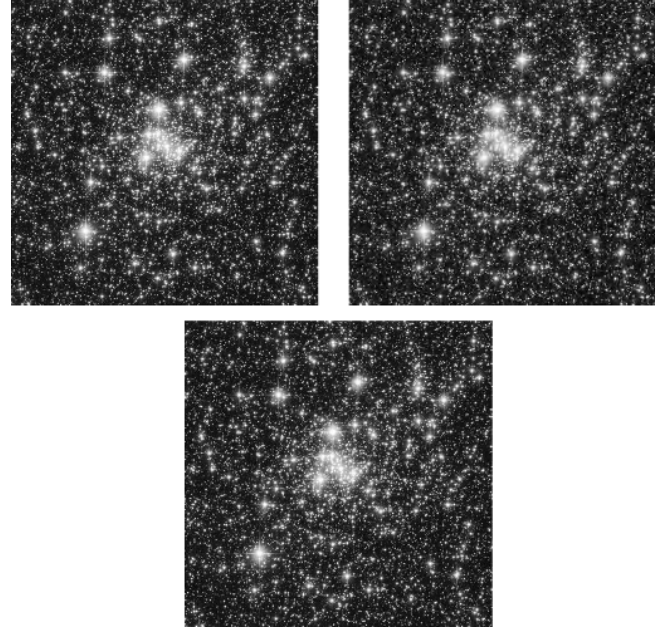

Fig. 1. **Top-left**: Input image x of size 512×512. **Top-right**: Estimate from the average of 100 images compressed by JPEG2000 with a compression rate $\rho = 0.1$. **Bottom**: Estimate from 100 pictures compressed by CS with a compression rate $\rho = 0.1$.
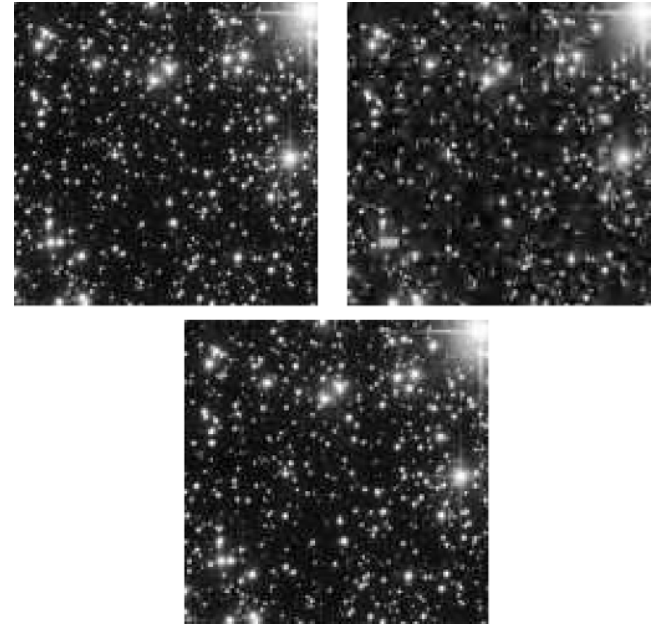

Fig. 2. **Zoomed versions—Top-left**: Input image x of size 512×512. **Top-right**: Estimate from the average of 100 images compressed by JPEG2000 with a compression rate $\rho = 0.1$. **Bottom**: Estimate from 100 pictures compressed by CS with a compression rate $\rho = 0.1$.

less observation of the same sky area x. We propose to decompress the set of observations $\{x_i\}_{i=1,\cdots,N}$ in a joint recovery scheme. Each observation is compressed using CS such that $\rho = M/t = 0.1$. Compression is made by solving the problem in (8) using the iterative thresholding algorithm described previously. Fig. 1 provides the compression/decompression results of x using CS and JPEG2000. At first sight, both techniques behave similarly. Fig. 2 depicts the zoomed versions of the previous pictures. Clearly, CS provides better visual results. Recall

TABLE I
CS VERSUS JPEG2000: RECOVERY SNR IN DECIBELS
AND RELATIVE INTENSITY ERROR

	Recovery SNR	Rel. intensity error
JPEG2000	9.77dB	$2.6 10^{-3}$
CS	46.8dB	$1.82 10^{-6}$

that in astronomy, the main properties to be under control are: i) the resolution, ii) the sensitivity, and iii) the photometry. Let us have a look at photometry by defining the intensity of x as: $f = \sum_{j=1}^{t} x[j]$. Both compression techniques can be compared in terms of the their relative intensity error defined has the ratio: $(f - \hat{f})/f$ where $\hat{f}$ is the intensity of the recovered image. Quantitative results are presented in Table I.

Where the $\text{SNR} = -20 \log((x - \tilde{x})/x)$ where $\tilde{x}$ is the recovered signal. This huge difference between both compression strategies is the consequence of a fundamental property of CS: **the linearity of the compression**. In contrast to standard compression techniques (such as JPEG2000), the CS-based compression is linear. The data to transmit are indeed simple linear projections: $y_i = \mathbf{\Theta}_{\Lambda_i}(x_i + n_i)$ where n_i models instrumental noise. Whatever the compression rate (i.e., $\rho = M/t$), the incoherence between the measurement vectors $\mathbf{\Theta}_{\Lambda_i}$ and the data is likely to guarantee that x does not belong to the null space of $\mathbf{\Theta}_{\Lambda_i}$. As a consequence, the compressed data always contain a piece of information belonging to x. Standard compression methods (which are nonlinear) do not verify this crucial property. For low level details, a standard compression method will kill its high frequencies and they will never be recovered whatever the number of times this source is observed.

Need for Joint Decompression: Assume that x is compressible. More precisely, for a fixed distortion ϵ, there exists $K(\epsilon)$ such that for $K \geq K(\epsilon) \ll t$, $\|x - x_K\|_{\ell_2} \leq \epsilon$ where x_K is the K-term approximation of x (i.e., synthesized from the K entries of α that have the most significant amplitudes). The CS theory tells us that the same error ϵ (i.e., distortion) can be obtained by solving the problem in (7) from $M \geq M(\epsilon)$ incoherent measurements. Let us assume that, the set $\{\Lambda_i\}_{i=1,\cdots,N}$ have been generated at random such that, with high probability, $\forall i \neq i'$; $\Lambda_i \cap \Lambda_{i'} = \emptyset$ then, the set of measurement vector $\{\mathbf{\Theta}_{\Lambda_i}\}$ is equivalent to a global measurement matrix $\mathbf{\Theta}_{\Lambda}$ where $\text{Card}(\Lambda) = NM$. To summarize, recovering x from the set of measurements $\{y_i\}_{i=1,\cdots,N}$ is likely to provide the same reconstruction that would be able to get from NM measurements thus leading to better recovery performances.

IV. REAL-WORLD APPLICATION: THE HERSCHEL PROJECT

Herschel is one of the cornerstone missions of the European Space Agency (ESA). This space telescope has been designed to observe in the far-infrared and sub-millimeter wavelength range. Its launch is scheduled for the fall of 2008/spring of 2009. The shortest wavelength band, 57–210 μm, is covered by PACS (Photodetector Array Camera and Spectrometer) [26], which provides low to medium resolution spectroscopy and dual-band photometry. When PACS is used as a photometer, it will simultaneously image with its two bolometer arrays, a 64×32 and a

32×16 matrix, both read out at 40 Hz. The ESA is faced with a challenging problem: conventional low-cost compression techniques cannot achieve a satisfactory compression rate. In this Section, we propose a new CS-based compression scheme for the Herschel/PACS data that yield an elegant and effective way to overcome the Herschel compression dilemma.

A. Herschel Dilemma

The Herschel space telescope is partially hampered by the narrowness of the transmission band compared to the large amount of data to be transferred. This handicap stems from the limitation of conventional compression techniques to provide adequate compression rate with low computational cost, given the high readout noise. More quantitatively, the data have to be compressed in real time by a factor of 6 with very low CPU power.

Problem Statement: The Herschel spacecraft is about to be launched; ESA is faced with a compression problem as the data need to be compressed by a factor of $N = 6$. Up to now, the only acceptable solution for the ESA (with respect to computational cost and quality) to overcome this need for a higher compression rate is the average of $N = 6$ consecutive images [27]. Indeed, the compression code has no information about the scan speed or the scan direction and a shift-and-add averaging solution is not possible. Other compression techniques such JPEG or JPEG2000 are also not acceptable because of computation time constraints.

Herschel will observe wide sky areas thus requiring fast scanning strategies. Hershel/PACS will provide sets of consecutive 64×32 images that will be shifted with a typical shift value $\lambda = 1$ pixel in fast scanning mode. Unfortunately, the shift value λ is comparable to the full width at half maximum (FWHM) of the instrumental point spread function (PSF) is $\delta \simeq 3$ pixels. As a consequence, averaging N consecutive images will entail a catastrophic loss of spatial resolution. This can be catastrophic for some scientific programs. Furthermore, averaging is far less optimal for noise reduction as averaging shifted signals does not yield a $\sqrt{N}$ noise variance reduction.

An effective compression scheme would have to balance between the following performance criteria.

— **Spatial resolution**: Averaging fast scanned data entails a lower spatial resolution. An effective compression scheme should provide a lower resolution loss.

— **Sensitivity**: Averaging will reduce noise but will also blur the data thus entailing a loss of sensitivity. Sensitivity (i.e., ability to detect low level details or sources) after compression/decompression must be under control.

B. Compressed Sensing for the Herschel Data

The Herschel/PACS mission needs a compression rate equal to $\rho = 1/N$ with $N = 6$. A first approach would amount to compressing independently each image. As stated earlier, accounting for the redundancy of the data is profitable to enhance the global compression/decompression performances. Then, compressing/decompressing $N = 6$ consecutive images jointly

would be more relevant. If we consider a stack of $N = 6$ consecutive images $\{x_i\}_{i=1,\cdots,N}$, the simplest generative model is the following:

$$\forall i \in \{1, \cdots, N\}; \quad x_i = \mathcal{T}_{\lambda_i}(x) + n_i \tag{13}$$

where $\mathcal{T}_{\lambda_i}$ is an operator that shifts the original image x with a shift λ_i. In practice, $x = x_1$ and $\lambda_1 = 0$. The term n_i models instrumental noise or model imperfections. According to the compressed sensing framework, each signal is projected onto the subspace ranged by a subset of columns of $\boldsymbol{\Theta}$. Each compressed observation is then obtained as follows:

$$\forall i \in \{1, \cdots, N\}; \quad y_i = \boldsymbol{\Theta}_{\Lambda_i} x_i \tag{14}$$

where the sets $\{\Lambda_i\}$ are such that:

$$\sum_i \boldsymbol{\Pi}_i = \mathbf{I} \tag{15}$$

$$\forall i \neq i'; \Lambda_i \cap \Lambda_{i'} = \emptyset \tag{16}$$

where the cardinality of each subset is $M = \lfloor t/N \rfloor$. When there is no shift between consecutive images, these conditions guarantee that the signal x can be reconstructed univocally from $\{y_i\}_{i=1,\cdots,N}$, up to noise. Furthermore, x is assumed to be positive. The decoding step amounts to seeking the signal x as follows:

$$\min_x \gamma \|\boldsymbol{\Phi}^T \alpha\|_{\ell_1} + \frac{1}{2} \sum_{i=1}^N \|y_i - \boldsymbol{\Theta}_{\Lambda_i} \mathcal{T}_{\lambda_i}(x)\|_{\ell_2}^2. \tag{17}$$

We propose solving this problem by using an adapted version of the iterative algorithm we introduced in Section II-B. Furthermore, the content of astronomical data is often positive. Constraining the solution to be positive would help solving the recovery problem. Assuming that the shifting operator $\mathcal{T}_{\lambda_i}$ is invertible,[5] we substitute (12) by

$$x^{(h)} = \frac{1}{N} \boldsymbol{\Phi} \mathcal{S}_\gamma$$
$$\times \left\{ \boldsymbol{\Phi}^T \sum_{i=1}^N \mathcal{T}_{-\lambda_i} \left(\boldsymbol{\Theta}^{\mathbf{T}} \left[y_i^\sharp + (\mathbf{I} - \boldsymbol{\Pi}_i) \mathcal{T}_{\lambda_i} \left(x^{(h-1)} \right) \right] \right) \right\}. \tag{18}$$

The positivity constraint is accounted for by projecting at each iteration the solution of the previous update equation on the cone generated by the vectors having positive entries: $x^{(h)} \leftarrow P_C(x^{(h)})$ where the projector P_C is defined as follows:

$$\forall i = 1, \cdots, t; \quad P_C(x)[i] = \begin{cases} x[i] & \text{if } x[i] \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

where $P_C(x)[i]$ is the ith entry of $P_C(x)$. Convergence is guaranteed as long as shifting the image does note deteriorate the original signal. In practice, this condition is not valid, only a

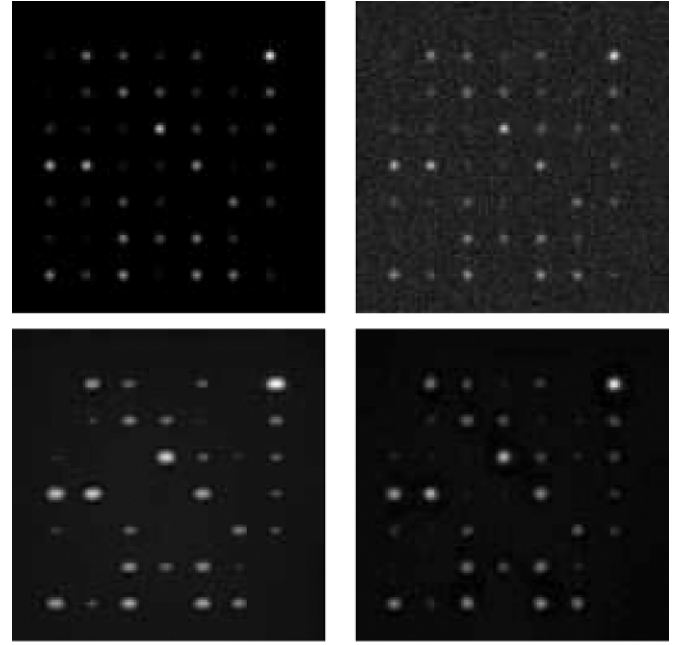[5]This assumption is true when shifting the image does note deteriorate the original signal.

Fig. 3. **Top left**: Original image of size 128×128 the total intensity of which is $f = 3500$. **Top right**: First input noisy map (out of 6). White Gaussian with variance $\sigma_n^2 = 1$ was added. **Bottom left**: Mean of the 6 input images. **Bottom right**: Reconstruction from noiselet-based CS projections. The iterative algorithm described in Section III has been used with 100 iterations.

portion of the image can be recovered with precision. Similarly to the discussion in Section III, joint decompression should be profitable as sets of consecutive shifted images provide redundant data. In the next section, we illustrate the good performances of the proposed decoding scheme.

Notations: In the next experiments, the data will made of pointwise sources; it is worth defining some useful notations. Recall that we assume the telescope's PSF to have a FWHM equal to δ. The shift between the original datum x and the ith datum x_i is λ_i. The intensity f of the datum x (defined in Section III). We also assume the x has positive entries.

C. A Toy-Example

In the following experiments, the datum x is a 128×128 image. The instrument is assumed to have a FWHM $\delta = 3$ pixels. For the sake of simplicity, each shift $\lambda_i = i$ pixels. White Gaussian noise is added to account for the instrumental noise. As we stated earlier, three main properties must be under control: i) spatial resolution, ii) sensitivity, and iii) photometry. Concerning the last point, it was shown in Section III-C, that CS was very efficient compared to JPEG2000. In this section, we will particularly focus on sensitivity and spatial resolution.

1) Sensitivity: In this experiment, the datum contains 49 point sources that have been uniformly scattered. The amplitude of each point source is generated at random with a Gaussian distribution. The top-left picture of Fig. 3 shows the input data x. The additive Gaussian noise has a fixed unit variance. The top-right panel of Fig. 3 features the data x contaminated with noise. Comparisons between the MO6 ("Mean of 6 images") and CS methods are made by evaluating for varying intensity value (from 700 to 140 000; it is equivalent to a SNR varying from -13.2 to 33 dB) the rate of detected point sources. To
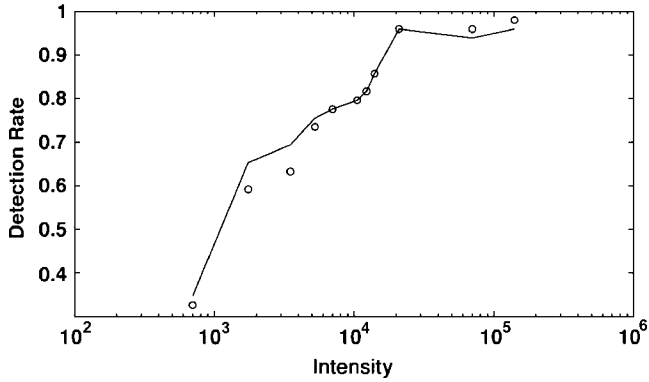
Fig. 4. **Detection rate when the intensity of the input data varies**: *Solid line* Resolution defined by the Rayleigh criterion of the CS-based reconstruction. ○: Resolution of the solution provided by the mean of 6 images.

avoid false detection, the same pre-processing step is performed: i) "à trous" bspline wavelet transform (see [28]), ii) $5\sigma_M$ hard-thresholding[6] where σ_M is the residual standard deviation estimated by a Median Absolute Deviation (MAD) at each wavelet scale, and iii) reconstruction. The bottom-left panel of Fig. 3 features such filtered decoded image using the MO6 strategy. The bottom-right picture in Fig. 3 shows the filtered CS-based solution. In this experiment the total intensity of the point sources is set to 3500. At first sight, both methods provide similar detection performances. As expected, the CS-based solution has a better spatial resolution.

Fig. 4 shows the detection rate (with no false detection) of each method for intensities varying from $f = 700$ to $f = 140\,000$. At high intensity (higher than $f = 10^4$), both MO6 and CS provide rather similar detection performances. Interestingly, at low intensity, CS provides slightly better results. This unexpected phenomenon is partly due to the spread that results from the average of shifted images.

MO6 is theoretically (for low shifts) near-optimal for point source detection. In contrast, this experiment shows that CS can provide similar or better detection performances than MO6.

2) Resolution: Spatial resolution is a crucial instrumental feature. Averaging shifted images clearly deteriorates the final spatial resolution of Hershel/PACS. In this experiment, the original datum x is made of a couple of point sources. In the worst case, these point sources are aligned along the scan direction. The top-left picture of Fig. 5 features the original signal x. In the top-right panel of Fig. 5, the intensity of the point sources is set to $f = 1000$ while the noise variance is $\sigma_n^2 = 1$. The SNR of the data to compress is equal to 2.7 dB. The MO6 solution (*resp.* the CS-based solution) is shown on the left (*resp.* right) at the bottom of Fig. 5. As expected, the spatial resolution of the MO6 is clearly worse than the resolution of the input datum x. Visually, the CS-based solution mitigate the resolution loss.

For different intensity of the datum x (from 100 to 2000), the spatial resolution is evaluated according to the Rayleigh criterion. The Rayleigh criterion is the generally accepted criterion for the minimum resolvable detail: two point sources are resolved when the first minimum is lower than the amplitude at

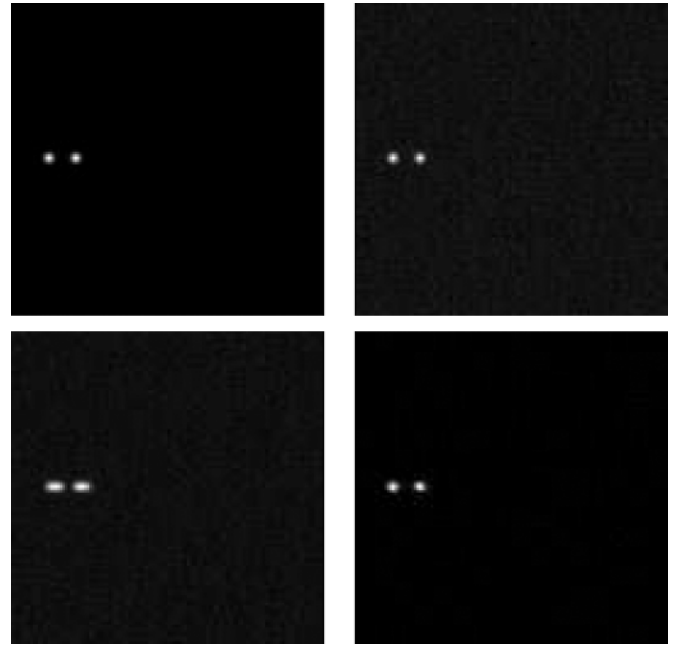[6]Such $5\sigma_M$ is likely to avoid false detection as it defines a rather conservative threshold.

Fig. 5. **Top left**: Original image of size 128×128 the total intensity of which is $f = 1000$. **Top right**: First input noisy map (out of 6). White Gaussian with variance $\sigma_n^2 = 1$ was added. **Bottom left**: Mean of the 6 input images. **Bottom right**: Reconstruction from noiselet-based CS projections. The iterative algorithm has been used with 100 iterations.

half maximum of a single point source as illustrated in Fig. 6. For a fixed intensity f, the resolution limit is evaluated by seeking the minimal distance between the point sources for which the Rayleigh criterion is verified. For intensities varying from $f = 100$ to $f = 2000$, the resolution limit is reported in Table II.

The CS-based compression scheme provides a solution with better spatial resolution. At high intensity, the resolution gain (in comparison with MO6) is equal to a third of the instrumental FWHM (1 pixel). At low intensity, the resolution gain provided by the CS-based method slightly decreases.

This experiment shows that CS mitigates the resolution loss resulting from the joint compression of 6 consecutive images.

D. Realistic Data

1) Data: Real Herschel/PACS data are more complex than those we simulated in the previous experiments. The original datum x is contaminated with a slowly varying "flat field" component c_f. In a short sequence of 6 consecutive images, the flat field component is almost fixed. In this context, the data $\{x_i\}_{i=0,\cdots,1}$ can then be modeled as follows:

$$x_i = \mathcal{T}_{\lambda_i}(x) + n_i + c_f. \tag{20}$$

If c_f is known (which will be the case in the forthcoming experiments), $\mathcal{T}_{\lambda_i}(x^{(h-1)})$ is replaced by $\mathcal{T}_{\lambda_i}(x^{(h-1)}) + c_f$ in (18). If c_f is unknown, it can be estimated within the iterative algorithm. The next section focuses on the resolution gain provided by the CS-based method in the scope of real Herschel/PACS data. The data have been designed by adding realistic pointwise sources to real calibration measurements performed in mid-2007.
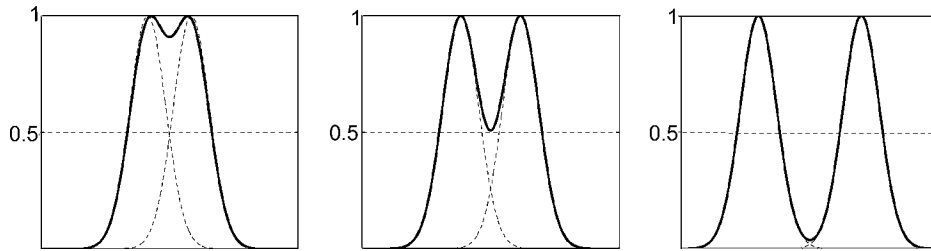
Fig. 6. **Rayleigh criterion—Left**: The point sources are not resolved. **Middle**: Resolution limit. **Right**: Fully resolved point sources.

TABLE II
Spatial resolution in pixels: FOR VARYING DATUM FLUX, THE RESOLUTION LIMIT OF EACH COMPRESSION TECHNIQUE IS REPORTED. THE CS-BASED COMPRESSION ENTAILS A RESOLUTION GAIN EQUAL TO A 30% OF THE SPATIAL RESOLUTION PROVIDED BY MO6

SNR (in dB)	−17.3	−9.35	−3.3	0.21	2.7
Intensity (a.u.[7])	100	250	500	750	1000
MO6	2.7	2.7	2.7	2.7	2.7
CS	2	2	1.7	1.7	1.7
SNR	4.7	6.2	7.6	8.7	
Intensity	1250	1500	1750	2000	
MO6	2.7	2.7	2.7	2.7	
CS	1.7	1.7	1.7	1.7	

TABLE III
Spatial resolution in pixels: FOR VARYING DATUM FLUX, THE RESOLUTION LIMIT OF EACH COMPRESSION TECHNIQUE IS REPORTED. THE CS-BASED COMPRESSION ENTAILS A RESOLUTION GAIN EQUAL TO A 30% OF THE SPATIAL RESOLUTION PROVIDED BY MO6

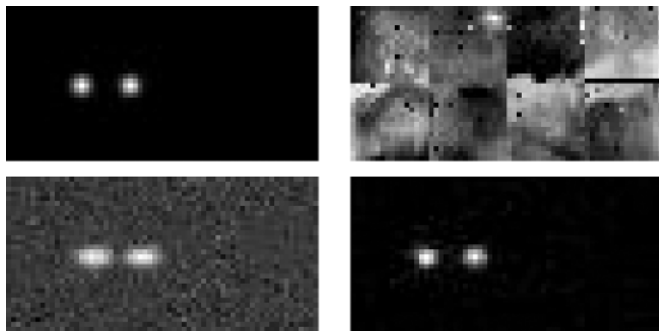| SNR (in dB)	−17.3	−9.35	−3.3	0.21	2.7
Intensity (a.u.)	900	2250	4500	6750	9000
MO6	3	3	3	3	3
CS	2.33	2.33	2	2	2
SNR	4.7	6.2	7.6	8.7	
Intensity	11250	13500	15750	18000	
MO6	3	3	3	3	
CS	2	2	2	2	

Fig. 7. **Top left**: Original image of size 32×64 with a total intensity of $f = 4500$. **Top right**: First input noisy map (out of 6). The PACS data already contains approximately Gaussian noise. **Bottom left**: Mean of the 6 input images. **Bottom right**: Reconstruction from noiselet-based CS projections. The iterative algorithm has been used with 100 iterations.

2) Resolution: Similarly to the experiments performed in Section IV-C-2, we added a couple of point sources to Herschel/PACS data. The top-left picture of Fig. 7 features the original signal x. In the top-right panel of Fig. 7, the intensity of the point sources is set to $f = 4500$. The "flat field" component overwhelms the useful part of the data so that x has at best a level that is 30 times lower than the "flat field" component. The MO6 solution (*resp.* the CS-based solution) is shown on the left (*resp.* right) and at the bottom of Fig. 7 and all the results are presented in Table III. Similarly to the previous fully simulated experiment, the CS-based algorithm provides better resolution performances. The resolution gain can reach 30% of the FWHM of the instrument's PSF for a wide range of signal intensities. This experiment illustrates the reliability of the CS-based compression to deal with real-world data compression.

V. CONCLUSION

In this paper, we overview the potential applications of compressed sensing (CS) in astronomical imaging. The CS appeal in astronomy is twofold: i) it provides a very easy and computationally cheap coding scheme for onboard astronomical remote sensing and ii) from a data fusion perspective, the decoding stage is flexible enough to account for the redundancy of the data thus leading to significant recovery enhancements. We particularly point out the huge advantage of compressed sensing over standard compression techniques in the scope of multiple scanning observations (observing the same sky area several times). We have shown that compressed sensing data fusion can lead to improvements compared to standard techniques. Preliminary numerical experiments illustrate the reliability of a CS-based compression scheme in the scope of astronomical remote sensing such as the Herschel space mission. We show that compressed sensing provides an elegant and effective compression technique that overcome the compression issue ESA is faced with. In the next step we will focus on performing more realistic experiments in the scope of the Herschel space mission by adding some physical information: calibration, statistical noise models, flat field estimation ... to name a few. This paper show that applying CS to the Herschel space mission is of major interest. Indeed, CS is the only existing alternative solution to the averaging solution, and CS enables to recover data with a spatial resolution enhanced up to 30% with similar sensitivity compared to the averaging technique. CS will probably be implemented onboard thus being the first application of compressed sensing for a real-world space mission.

REFERENCES

[1] R. White, M. Postman, and M. Lattanzi, "Compression of the guide star digitised Schmidt plates," in *Digitized Optical Sky Surveys*, H. MacGillivray and E. Thompson, Eds. Norwell, MA: Kluwer, 1992, pp. 167–175.

[2] J.-L. Starck, F. Murtagh, B. Pirenne, and M. Albrecht, "Astronomical image compression based on noise suppression," *Publ. Astron. Soc. Pacific*, vol. 108, pp. 446–455, 1996.

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] E. Candès, "Compressive sampling," in *International Congress of Mathematics*, Madrid, 2006.

[6] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, July 2007.

[7] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 5433–5480, 2006.

[8] R. Coifman, F. Geshwind, and Y. Meyer, "Noiselets," *Appl. Comput. Harmon. Anal.*, vol. 10, no. 1, pp. 27–44, 2001.

[9] E. Candès and J. Romberg, Sparsity and Incoherence in Compressive Sampling *Preprint* [Online]. Available: http://www.dsp.ece.rice.edu/cs/, 2006

[10] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[11] E. Candès, The Restricted Isometry Property and Its Implications for Compressed Sensing Caltech, Pasadena, CA, 2008.

[12] G. Skinner, "Coded-mask imaging in gamma-ray astronomy—Separating the real and imaginary parts of a complex subject," in *Proc. 22nd Moriond Astrophysics Meeting "The Gamma-Ray Universe"*, 2002.

[13] D. Takhar, J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *Proc. Computational Imaging IV*, San Jose, CA, 2006.

[14] R. Marcia and R. Willett, "Compressive coded aperture video reconstruction," in *Proc. Eur. Signal Processing Conf.*, 2008.

[15] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl. Ser.*, vol. 15, pp. 417–426, 1974.

[16] B. Wakker and U. Schwarz, "The multi-resolution Clean and its application to the short-spacing problem in interferometry," *Annu. Rev. Astron. Astrophys.*, vol. 200, p. 312, 1988.

[17] J.-L. Starck and F. Murtagh, *Astronomical Image and Data Analysis*. New York: Springer, 2006.

[18] A. Lannes, E. Anterrieu, and P. Marechal, "Clean and wipe," *AAS*, vol. 123, pp. 183–198, May 1997.

[19] D. L. Donoho, Y. Tsaig, I. Drori, and J.-L. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," *IEEE Trans. Information Theory*, to be published.

[20] E. J. Candès, M. B. Wakin, and S. P. Boyd, Enhancing Sparsity by Reweighted l1 Minimization Caltech, Pasadena, CA, 2008.

[21] M. Elad, J.-L. Starck, D. Donoho, and P. Querre, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," *ACHA*, vol. 19, no. 3, pp. 340–358, 2005.

[22] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[23] A. Bruckstein, D. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, to be published.

[24] J. Tropp, "Just relax: Convex programming methods for subset selection and sparse approximation," *IEEE Trans. Inform. Theory*, vol. 52, no. 3, pp. 1030–1051, 2006.

[25] J.-L. Starck, M. Elad, and D. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," *Adv. Imag. Electron Phys.*, vol. 132, pp. 287–348, 2004.

[26] A. Poglitscha, C. Waelkensb, O. Bauera, J. Cepac, H. Feuchtgrubera, T. Henning, C. van Hoofe, F. Kerschbaumf, D. Lemked, E. Renotteg, L. Rodriguez, P. Saracenoi, and B. Vandenbussche, "The photodetector array camera and spectrometer (pacs) for the Herschel space observatory," *Proc. SPIE*, 2006.

[27] A. N. Belbachir, H. Bischof, R. Ottensamer, F. Kerschbaum, and C. Reimers, "On-board data processing to lower bandwidth requirements on an infrared astronomy satellite: Case of herschel-pacs camera," *EURASIP J. Appl. Signal Process.*, vol. 15, pp. 2585–2594, 2005.

[28] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

Jérôme Bobin graduated from the Ecole Normale Superieure (ENS) de Cachan, France, in 2005 and received the M.Sc. degree in signal and image processing from ENS Cachan and Université Paris XI, Orsay, France. He received the Agrégation de Physique in 2004. Since 2005, he has been pursuing the Ph.D. degree with J.-L. Starck at the CEA.

His research interests include statistics, information theory, multiscale methods, and sparse representations in signal and image processing.

Jean-Luc Starck received the Ph.D. degree from the University Nice-Sophia Antipolis and the Habilitation degree from the University Paris XI.

He was a visitor at the European Southern Observatory (ESO) in 1993, at UCLA in 2004, and at Stanford University's Statistics Department in 2000 and 2005. He has been a Researcher at CEA since 1994. His research interests include image processing, statistical methods in astrophysics, and cosmology. He is an expert in multiscale methods such wavelets and curvelets, He is Leader of the project Multiresolution at CEA and he is a core team member of the PLANCK ESA project. He has published more than 100 papers in different areas in scientific journals. He is also author of two books: *Image Processing and Data Analysis: the Multiscale Approach* (Cambridge, U.K.: Cambridge University Press, 1998) and *Astronomical Image and Data Analysis* (New York: Springer, 2nd ed., 2006).

Roland Ottensamer received the M.S. degree in 2004 from the University of Vienna, Austria, where he holds a research position at the Department of Astronomy.

During the last decade, he was highly involved in the development of the data compression/decompression software for the IR photo-detector camera PACS and has presented his work at the major conferences for instrumentation in astronomy. He is currently occupied with Herschel flight software maintenance and has joined the SPICA/SAFARI consortium for assessing the necessity of onboard reduction during the study phase.