POLITECNICO DI TORINO
Repository ISTITUZIONALE

Operational Rate–Distortion Performance of Single–source and Distributed Compressed Sensing

Original
Operational Rate–Distortion Performance of Single–source and Distributed Compressed Sensing / Coluccia, Giulio; A., Roumy; Magli, Enrico. - In: IEEE TRANSACTIONS ON COMMUNICATIONS. - ISSN 0090-6778. - STAMPA. - 62:6(2014), pp. 2022-2033. [10.1109/TCOMM.2014.2316176]

Availability:
This version is available at: 11583/2546940 since:

Publisher:
IEEE - INST ELECTRICAL ELECTRONICS ENGINEERS INC

Published
DOI:10.1109/TCOMM.2014.2316176

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 October 2024

Operational Rate–Distortion Performance of Single–source and Distributed Compressed Sensing

Giulio Coluccia, *Member, IEEE,* Aline Roumy, *Member, IEEE,* and Enrico Magli, *Senior Member, IEEE*

***Abstract*—We consider correlated and distributed sources without cooperation at the encoder. For these sources, we derive the best achievable performance in the rate-distortion sense of any distributed compressed sensing scheme, under the constraint of high–rate quantization. Moreover, under this model we derive a closed–form expression of the rate gain achieved by taking into account the correlation of the sources at the receiver and a closed–form expression of the average performance of the oracle receiver for independent and joint reconstruction. Finally, we show experimentally that the exploitation of the correlation between the sources performs close to optimal and that the only penalty is due to the missing knowledge of the sparsity support as in (non distributed) compressed sensing. Even if the derivation is performed in the large system regime, where signal and system parameters tend to infinity, numerical results show that the equations match simulations for parameter values of practical interest.**

***Index Terms*—Compressed sensing, rate–distortion function, distributed source coding, Slepian–Wolf coding.**

I. Introduction

DISTRIBUTED sources naturally arise in wireless sensor networks, where sensors may acquire over time several readings of the same natural quantity, *e.g.*, temperature, in different points of the same environment. Such data must be transmitted to a fusion center for further processing. However, since radio access is the most energy–consuming operation in a wireless sensor network, data transmission among sensors needs to be minimized in order to maximize sensors' battery life. This calls for lossy compression techniques to find a cost–constrained representation in order to exploit data redundancies. In particular, following the example above, sensor readings may vary slowly over time, and hence consecutive readings have similar values, because of the slow variation of the underlying physical phenomenon. Moreover, inter–sensor correlations also exist, as the sensors may be located in the same environment, in which the temperature is rather uniform, leading to compressibility of each single signal and of the set of signals as an ensemble. The question hence arises of how to exploit such correlations in a distributed way, *i.e.,* without communication among the sensors, and employing a low–complexity signal representation in order to minimize energy consumption.

G. Coluccia and E. Magli are with Politecnico di Torino, Dipartimento di Elettronica e Telecomunicazioni (email: giulio.coluccia@polito.it, enrico.magli@polito.it). Their research has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement number 279848.

A. Roumy is with INRIA, campus de Beaulieu, Rennes, France (email: aline.roumy@inria.fr).

In this framework, compressed sensing (CS) [1], [2] has emerged in past years as an efficient technique for sensing a signal with fewer coefficients than dictated by classic Shannon/Nyquist theory. The hypothesis underlying this approach is that the signal to be sensed must have a sparse – or at least compressible – representation in a convenient basis. In CS, sensing is performed by taking a number of linear projections of the signal onto pseudorandom sequences. Therefore, the acquisition presents appealing properties. For example, it has low encoding complexity, since no sorting of the sparse signal coefficients is required. Moreover, the choice of the sensing matrix is blind to the source distribution.

Using CS as signal representation requires to cast the representation/coding problem in a rate–distortion (RD) framework, particularly regarding the rate necessary to encode the measurements. For single sources, this problem has been addressed by several authors. In [3], a RD analysis of CS reconstruction from quantized measurements was performed, when the observed signal is sparse. Instead, [4] considered the RD behavior of strictly sparse or compressible memoryless sources in their own domain. [5], [6] considered the cost of encoding the random measurements for single sources. More precisely, RD analysis was performed and it was shown that adaptive encoding, taking into account the source distribution, outperforms scalar quantization of random measurements at the cost of higher computational complexity. However, in the distributed context, adaptive encoding may loose the inter-correlation between the sources since it is adapted to the distribution of each single source or even the realization of each source.

On the other hand, the distributed case is more sophisticated. Not only one needs to encode a source, but also to design a scheme capable of exploiting the correlation among different sources. Therefore, distributed CS (DCS) was proposed in [7] and further analyzed in [8]. In those papers, an architecture for separate acquisition and joint reconstruction was defined, along with three different joint sparsity models (which where merged into a single formulation in the latter paper). For each model, necessary conditions were posed on the number of measurement to be taken on each source to ensure perfect reconstruction. An analogy between DCS and Slepian–Wolf distributed source coding was depicted, in terms of the necessary conditions about the number of measurements, depending on the sparsity degree of sources, and the necessary conditions on encoding rate, depending on conditional and joint entropy, typical of Slepian–Wolf theory. Moreover, it was shown that a distributed system based on CS could save up to 30% of measurements with respect to separate CS encoding/decoding

of each source. On the other hand, [7] extended CS in the acquisition part, but, like CS [1], [2], was mainly concerned with the performance of perfect reconstruction, and did not consider the representation/coding problem, which is one of the main issues of a practical scheme and a critical aspect of CS.

In [9], we proposed a DCS scheme that takes into account the encoding cost, exploits both inter- and intra–correlations, and has low complexity. The main idea was to exploit the knowledge of side information (SI) not only as a way to reduce the encoding rate, but also in order to improve the reconstruction quality, as is common in the Wyner–Ziv context [10]. The proposed architecture applies the same Gaussian random matrix to information and SI sources, then quantizes and encodes the measurements with a SW source code.

In this paper, we study analytically the best achievable RD performance of any single–source and distributed CS scheme, under the constraint of high–rate quantization, providing simulation results that perfectly match the theoretical analysis. In particular, we provide the following contributions. First, we derive the asymptotic (in the rate and in the number of measurements) distribution of the measurement vector. Even if the analysis is asymptotic, we show that the convergence to a Gaussian distribution occurs with parameter values of practical interest. Moreover, we provide an analytical expression of the rate gain obtained exploiting inter–source correlation at the decoder. Second, we provide a closed–form expression of the average reconstruction error using the oracle receiver, improving the results existing in literature, consisting only in bounds hardly comparable to the results of numerical simulations [11], [12]. The proof relies on recent results on random matrix theory [13]. Third, we provide a closed–form expression of the rate gain due to joint reconstruction from the measurements of multiple sources. We compare the results obtained by theory both with the ideal oracle receiver and with a practical algorithm [9], showing that the penalty with respect to the ideal receiver is due to the lack of knowledge of the sparsity support in the reconstruction algorithm. Despite this penalty, the theoretically–derived rate gain matches that obtained applying distributed source coding followed by joint reconstruction to a practical reconstruction scheme. With respect to [7], [8], we use information theoretic tools to provide an analytical characterization of the performance of CS and DCS, for a given number of measurements and set of system parameters.

This paper is organized as follows. Some background information about source coding with side information at the decoder and CS is given in section II. Novel analytical results are presented in sections III and IV. These results are validated via numerical simulations that are presented in section V. Finally, concluding remarks are given in section VI.

II. Background

A. Notation and definitions

We denote (column-) vectors and matrices by lowercase and uppercase boldface characters, respectively. The (m,n)-th element of a matrix $\mathbf{A}$ is $(\mathbf{A})_{m,n}$. The m-th row of matrix $\mathbf{A}$ is $(\mathbf{A})_m$. The n-th element of column vector $\mathbf{v}$ is $(\mathbf{v})_n$. The transpose of a matrix $\mathbf{A}$ is $\mathbf{A}^\mathsf{T}$.

The notation $\|\mathbf{v}\|_0$ denotes the number of nonzero elements of vector $\mathbf{v}$. The notation $\|\mathbf{v}\|_1$ denotes the ℓ_1-norm of the vector $\mathbf{v}$ and is defined as $\|\mathbf{v}\|_1 \triangleq \sum_i |(\mathbf{v})_i|$. The notation $\|\mathbf{v}\|_2$ denotes the Euclidean norm of the vector $\mathbf{v}$ and is defined as $\|\mathbf{v}\|_2 \triangleq \sqrt{\sum_i |(\mathbf{v})_i|^2}$. The notation $A \sim \mathcal{N}(\mu, \sigma^2)$ means that the random variable A is Gaussian distributed, its mean is μ, and its variance is σ^2. Additional notation will be defined throughout the paper where appropriate.

B. Source Coding with Side Information at the decoder

Source Coding with SI at the decoder refers to the problem of compressing a source X when another source Y, correlated to X, is available at the decoder only. It is a special case of distributed source coding, where the two sources have to be compressed without any cooperation at the encoder.

For lossless compression, if X is compressed without knowledge of Y at its conditional entropy, *i.e.*, $R_X > H(X|Y)$, it can be recovered with vanishing error rate exploiting Y as SI. This represents the asymmetric setup, where source Y is compressed in a lossless way ($R_Y > H(Y)$) or otherwise known at the decoder. Therefore, the lack of SI at the encoder does not incur any compression loss with respect to joint encoding, as the total rate required by DSC is equal to $H(Y)+H(X|Y) = H(X,Y)$. The result holds for i.i.d. finite sources X and Y [14] but also for ergodic discrete sources [15], or when X is i.i.d. finite, Y is i.i.d. continuous and is available at the decoder [16, Proposition 19].

For lossy compression of i.i.d sources, [17] shows that the lack of SI at the encoder incurs a loss except for some distributions (Gaussian sources, or more generally Gaussian correlation noise). Interestingly, [16] shows that uniform scalar quantization followed by lossless compression incurs a sub-optimality of 1.53 dB, in the high–rate regime. Therefore, practical solutions (see for example [18]) compress and decompress the data relying on an *inner lossless distributed codec*, usually referred to as Slepian–Wolf Code (SWC), and an *outer quantization–plus–reconstruction filter*.

C. Compressed Sensing

In the standard CS framework, introduced in [1], [2], a signal $\mathbf{x} \in \mathbb{R}^{N\times 1}$ which has a sparse representation in some basis $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$, *i.e.*:

$$\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}, \quad \|\boldsymbol{\theta}\|_0 = K, \quad K \ll N$$

can be recovered by a smaller vector of linear measurements $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^{M\times 1}$ and $K < M < N$, where $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ is the *sensing matrix*. The optimum solution, requiring at least $M = 2K$ measurements, would be

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_0 \quad \text{s.t.} \quad \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta} = \mathbf{y} \ .$$

Since the ℓ_0 norm minimization is an NP-hard problem, one can resort to a linear programming reconstruction by minimizing the ℓ_1 norm

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta} = \mathbf{y} \tag{1}$$

and $\widehat{\mathbf{x}} = \mathbf{\Psi}\widehat{\boldsymbol{\theta}}$, provided that $M = O(K\log(N/K))$ [2].

When the measurements are noisy, i.e. when $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{e}$, where $\mathbf{e} \in \mathbb{R}^{M\times 1}$ is the vector representing additive noise such that $\|\mathbf{e}\|_2 < \varepsilon$, ℓ_1 minimization with inequality constraints is used for reconstruction:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\|_1 \quad \text{s.t.} \quad \|\mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta} - \mathbf{y}\|_2 < \varepsilon \tag{2}$$

and $\widehat{\mathbf{x}} = \mathbf{\Psi}\widehat{\boldsymbol{\theta}}$, known as Basis Pursuit DeNoising (BPDN), provided that $M = O(K\log(N/K))$ and that each submatrix consisting of K columns of $\mathbf{\Phi}\mathbf{\Psi}$ is (almost) distance preserving [19, Definition 1.3]. The latter condition is the *Restricted Isometry Property* (RIP). Formally, the matrix $\mathbf{\Phi}\mathbf{\Psi}$ satisfies the RIP of order K if $\exists \delta_K \in (0,1]$ such that, for any $\boldsymbol{\theta}$ with $\|\boldsymbol{\theta}\|_0 \leq K$:

$$(1-\delta_K)\|\boldsymbol{\theta}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta}\|_2^2 \leq (1+\delta_K)\|\boldsymbol{\theta}\|_2^2, \tag{3}$$

where δ_K is the RIP constant of order K. It has been shown in [20] that when $\mathbf{\Phi}$ is an i.i.d. random matrix drawn from any subgaussian distribution and $\mathbf{\Psi}$ is an orthogonal matrix, $\mathbf{\Phi}\mathbf{\Psi}$ satisfies the RIP with overwhelming probability.

III. Rate–Distortion functions of Single–source Compressed Sensing

In this section, we derive the best achievable performance in the RD sense over all CS schemes, under the constraint of high–rate quantization. The novel result about the distribution of the measurements derived in Theorem 4 allows to write a closed–form expression of the RD functions of the measurement vectors. In Theorem 5, we derive a novel closed–form expression of the average reconstruction error of the oracle receiver, which will use the results from Theorem 4 to present the RD functions of the reconstruction.

A. System Model

Definition 1. (Sparse vector). *The vector $\mathbf{x} \in \mathbb{R}^N$ is said to be* $(K, N, \sigma_\theta^2, \mathbf{\Psi})$-sparse *if $\mathbf{x}$ is sparse in the domain defined by the orthogonal matrix $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$, namely: $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$, with $\|\boldsymbol{\theta}\|_0 = K$, and if the nonzero components of $\boldsymbol{\theta}$ are modeled as i.i.d. centered random variables with variance $\sigma_\theta^2 < \infty$. $\mathbf{\Psi}$ is independent of $\boldsymbol{\theta}$.*

The sparse $\mathbf{x}$ vector is observed through a smaller vector of Gaussian measurements defined as

Definition 2. (Gaussian measurement). *The vector $\mathbf{y}$ is called the* $(M, N, \sigma_\Phi^2, \mathbf{\Phi})$-Gaussian measurement *of $\mathbf{x} \in \mathbb{R}^N$, if $\mathbf{y} = \frac{1}{\sqrt{M}}\mathbf{\Phi}\mathbf{x}$, where the sensing matrix $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$, with $M < N$, is a random matrix*[1] *with i.i.d. entries drawn from $\mathcal{N}(0, \sigma_\Phi^2)$ with $\sigma_\Phi^2 < \infty$.*

We denote as $\mathbf{y}_q$ the quantized version of $\mathbf{y}$. To analyze the RD tradeoff, we consider the large system regime defined below.

[1]This definition complies with the usual form $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$ where the variance σ_Φ^2 of the elements of $\mathbf{\Phi}$ depends on M. Here, we wanted to keep σ_Φ^2 independent of system parameters.

Definition 3. (Large system regime, overmeasuring and sparsity rates). *Let $\mathbf{x}$ be $(K, N, \sigma_\theta^2, \mathbf{\Psi})$-sparse. Let $\mathbf{y}$ be the $(M, N, \sigma_\Phi^2, \mathbf{\Phi})$-Gaussian measurement of $\mathbf{x}$. The system is said to be in the* large system regime *if N goes to infinity, K and M are functions of N and tend to infinity as N does, under the constraint that the rates K/N and M/K converge to constants called* sparsity rate *(γ) and* overmeasuring rate *(μ) i.e.:*

$$\lim_{N\to+\infty} \frac{K}{N} = \gamma, \quad \lim_{N\to+\infty} \frac{M}{K} = \mu > 1 \tag{4}$$

The sparsity rate is a property of the signal. Instead, the overmeasuring rate is the ratio of the number of measurements to the number of non zero components and is therefore a property of the system [3].

B. Rate–distortion functions of measurement vector

The *information RD function* of an i.i.d. source X defines the minimum amount of information per source symbol R needed to describe the source under the distortion constraint D. For an i.i.d. Gaussian source $X \sim \mathcal{N}(0, \sigma_x^2)$, choosing as distortion metric the squared error between the source and its representation on R bits per symbol, the distortion satisfies

$$D_x(R) = \sigma_x^2 2^{-2R}. \tag{5}$$

Interestingly, the *operational RD function* of the same Gaussian source, with uniform scalar quantizer and entropy coding satisifies, in the high–rate regime:

$$\lim_{R\to+\infty} \frac{1}{\sigma_x^2} 2^{2R} D_x^{\mathsf{EC}}(R) = \frac{\pi e}{6} \tag{6}$$

where EC stands for *entropy–constrained scalar quantizer*. This leads to a 1.53 dB gap between the information and the operational RD curves. (6) can be easily extended to other types of quantization adapting the factor $\frac{\pi e}{6}$ to the specific quantization scheme.

Theorem 4. (CS: Asymptotic distribution of Gaussian measurements and measurement RD function). *Let $\mathbf{x}$ be $(K, N, \sigma_\theta^2, \mathbf{\Psi})$-sparse. Let $\mathbf{y}$ be the $(M, N, \sigma_\Phi^2, \mathbf{\Phi})$-Gaussian measurement of $\mathbf{x}$, s.t. $K < M < N$. Consider the large system regime with finite sparsity rate $\gamma = \lim_{N\to\infty} \frac{K}{N}$ and finite overmeasuring rate $\mu = \lim_{N\to\infty} \frac{M}{K} > 1$.*
The Gaussian measurement converges in distribution to an i.i.d. Gaussian, centered random sequence with variance

$$\sigma_y^2 = \frac{1}{\mu}\, \sigma_\Phi^2 \sigma_\theta^2. \tag{7}$$

Therefore, the information RD function satisifies

$$\lim_{N\to+\infty} \frac{1}{\sigma_y^2} 2^{2R} D_y(R) = 1, \tag{8}$$

where R is the encoding rate per measurement sample, and the entropy–constrained scalar quantizer achieves a distortion $D_{y_1}^{\mathsf{EC}}$ that satisfies

$$\lim_{R\to+\infty} \lim_{N\to+\infty} \frac{1}{\sigma_y^2} 2^{2R} D_y^{\mathsf{EC}}(R) = \frac{\pi e}{6}\ . \tag{9}$$

Sketch of proof. The distribution of the Gaussian matrix $\mathbf{\Phi}$ is invariant under orthogonal transformation. Thus, we obtain

$\mathbf{y} = \frac{1}{\sqrt{M}}\mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta} = \frac{1}{\sqrt{M}}\mathbf{U}\boldsymbol{\theta}$, where $\mathbf{U}$ is an i.i.d. Gaussian matrix with variance σ_Φ^2. Then, we consider a finite length subvector $\mathbf{y}^{(m)}$ of $\mathbf{y}$. From the multidimensional Central Limit theorem (CLT) [21, Theorem 7.18], $\mathbf{y}^{(m)}$ converges to a Gaussian centered vector with independent components. Then, as $m \to \infty$, the sequence of Gaussian measurements converges to an i.i.d. Gaussian sequence. See Appendix A for the complete proof. □

Theorem 4 generalizes [3, Theorem 2], which derives the marginal distribution of the measurements, when the observed signal is directly sparse. Instead, we derive the *joint* distribution of the measurements and consider *transformed* sparse signals.

We stress the fact that, even if the RD curves for measurement vectors do not have any "practical" direct use, they are required to derive the RD curves for the reconstruction of the sources, which can be found later in this section.

C. Rate–distortion functions of the reconstruction

We now evaluate the performance of CS reconstruction with quantized measurements. The performance depends on the amount of noise affecting the measurements. In particular, the distortion $\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2$ is upper bounded by the noise variance up to a scaling factor [22], [23].

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2 \leq c^2\varepsilon^2 \ , \tag{10}$$

where the constant c depends on the realization of the measurement matrix, since it is a function of the RIP constant. Since we consider the average[2] performance, we need to consider the worst case c and this upper bound will be very loose [19, Theorem 1.9].

Here, we consider the *oracle* estimator, which is the estimator knowing exactly the sparsity support $\Omega = \{i|\boldsymbol{\theta}_i \neq 0\}$ of the signal $\mathbf{x}$. For the oracle estimator, upper and lower bounds depending on the RIP constant can be found, for example in [11] when the noise affecting the measurements is white and in [12] when the noise is correlated. Unlike [11], [12], in this paper the average performance of the oracle, depending on system parameters only, is derived exactly.

As we will show in the following sections, the characterization of the ideal oracle estimator allows to derive the reconstruction RD functions with results holding also when non ideal estimators are used.

Theorem 5. (CS: Reconstruction RD functions). *Let* $\mathbf{x}$ *be* $(K, N, \sigma_\theta^2, \mathbf{\Psi})$*-sparse. Let* $\mathbf{y}$ *be the* $(M, N, \sigma_\Phi^2, \mathbf{\Phi})$*-Gaussian measurement of* $\mathbf{x}$*, s.t.* $K + 3 < M < N$*. Consider the large system regime with finite sparsity rate* $\gamma = \lim_{N\to\infty} \frac{K}{N}$ *and finite overmeasuring rate* $\mu = \lim_{N\to\infty} \frac{M}{K} > 1$*.* R *denotes the encoding rate per measurement sample.*
Assume reconstruction by the oracle estimator, when the support Ω *of* $\mathbf{x}$ *is available at the receiver. The operational RD function of any CS reconstruction algorithm is lower bounded by that of the oracle estimator that satisfies*

$$D_x^{\mathsf{CS}}(R) \geq D_x^{\text{oracle}}(R) = \gamma\frac{\mu}{\mu-1}\frac{1}{\sigma_\Phi^2}D_y(R) = \frac{\gamma}{\mu-1}\sigma_\theta^2 2^{-2R}. \tag{11}$$

Similarly, the entropy-constrained RD function satisfies in the high-rate regime

$$D_x^{\mathsf{EC-CS}}(R) \geq D_x^{\mathsf{EC}\ \text{oracle}}(R) = \frac{\gamma}{\mu-1}\sigma_\theta^2\frac{\pi e}{6}2^{-2R}. \tag{12}$$

Sketch of proof. We use a novel result about the expected value of a matrix following a generalized inverse Wishart distribution [13, Theorem 2.1]. This result can be applied to the distortion of the oracle estimator for finite–length signals, depending on the expected value of the pseudo inverse of Wishart matrix [24]. The key consequence is that the distortion of the oracle only depends on the variance of the quantization noise and not on its covariance matrix. Therefore, our result holds even if the noise is correlated (for instance if vector quantization is used). Hence, this result applies to any quantization algorithm. This result improves those in [11, Theorem 4.1] and [12], where upper and lower bounds depending on the RIP constant of the sensing matrix are given, and it also generalizes [3, section III.C], where a lower bound is derived whereas we derive the exact average performance. See Appendix B for the complete proof. □

It must be noticed that the condition $M > K + 3$ is not restrictive since in all cases of practical interest, $M > 2K$.

IV. Rate–Distortion functions of Distributed Compressed Sensing

In this section, we derive the best achievable performance in the RD sense over all DCS schemes, under the constraint of high–rate quantization. Note that [9] (see Fig. 1) is one instance of such a scheme. Novel results about the distribution of the measurements in the distributed case are presented in Theorem 8. Hence, Theorem 9, will combine the results of Theorem 8 and previous section to derive the RD functions of the reconstruction in the distributed case.

A. Distributed System Model

Definition 6. (Correlated Sparse vectors).
J vectors $\mathbf{x}_j \in \mathbb{R}^{N\times 1}, j \in \{1, \ldots, J\}$ are said to be $(\{K_{\mathsf{I},j}\}_{j=1}^J, K_{\mathsf{C}}, \{K_j\}_{j=1}^J, N, \sigma_{\theta_{\mathsf{C}}}^2, \{\sigma_{\theta_{\mathsf{I},j}}^2\}_{j=1}^J, \mathbf{\Psi})$ -sparse if:
i) *Each vector $\mathbf{x}_j = \mathbf{x}_{\mathsf{C}} + \mathbf{x}_{\mathsf{I},j}$ is the sum of a* common *component $\mathbf{x}_{\mathsf{C}}$ shared by all signals and an* innovation *component $\mathbf{x}_{\mathsf{I},j}$, which is unique to each signal $\mathbf{x}_j$.*
ii) *Both $\mathbf{x}_{\mathsf{C}}$ and $\mathbf{x}_{\mathsf{I},j}$ are sparse in the same domain defined by the orthogonal matrix $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$, namely: $\mathbf{x}_{\mathsf{C}} = \mathbf{\Psi}\boldsymbol{\theta}_{\mathsf{C}}$ and $\mathbf{x}_{\mathsf{I},j} = \mathbf{\Psi}\boldsymbol{\theta}_{\mathsf{I},j}$, with $\|\boldsymbol{\theta}_{\mathsf{C}}\|_0 = K_{\mathsf{C}}$, $\|\boldsymbol{\theta}_{\mathsf{I},j}\|_0 = K_{\mathsf{I},j}$ and $K_{\mathsf{C}}, K_{\mathsf{I},j} < N$.*
iii) *The global sparsity of $\mathbf{x}_j$ is K_j, with $\max\{K_{\mathsf{C}}, K_{\mathsf{I},j}\} \leq K_j \leq K_{\mathsf{C}} + K_{\mathsf{I},j}$.*
iv) *The nonzero components of $\boldsymbol{\theta}_{\mathsf{C}}$ and $\boldsymbol{\theta}_{\mathsf{I},j}$ are i.i.d. centered random variables with variance $\sigma_{\theta_{\mathsf{C}}}^2 < \infty$ and $\sigma_{\theta_{\mathsf{I},j}}^2 < \infty$, respectively.*

The correlation between the sources is modeled through a *common* component and their difference through an individual

[2]The average performance is obtained averaging over all random variables i.e. the measurement matrix, the non-zero components $\boldsymbol{\theta}$ and noise, as for example in [12].

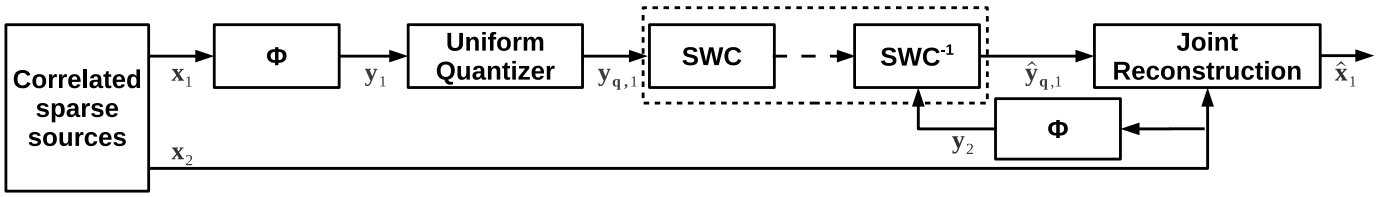


Fig. 1. The Distributed Compressed Sensing scheme in [9]

innovation component. This is a good fit for signals acquired by a group of sensors monitoring the same physical event in different spacial positions, where local factors can affect the innovation component of a more global behavior taken into account by this common component. Note that the Joint Sparsity Model-1 (JSM-1) [7] and the ensemble sparsity model (ESM) in [8] are deterministic models. Instead, the sparse model (Def. 6) is probabilistic, since we look for the performance averaged over all possible realizations of the sources.

Focusing without loss of generality on the case $J = 2$, we assume that $\mathbf{x}_1$ and $\mathbf{x}_2$ are $(K_{\mathsf{I},1}, K_{\mathsf{I},2}, K_{\mathsf{C}}, K_1, K_2, N, \sigma^2_{\theta_{\mathsf{C}}}, \sigma^2_{\theta_{\mathsf{I},1}}, \sigma^2_{\theta_{\mathsf{I},2}}, \mathbf{\Psi})$-sparse. $\mathbf{x}_1$ is the source to be compressed whereas $\mathbf{x}_2$ serves as SI. $\mathbf{y}_1$ and $\mathbf{y}_2$ are the $(M, N, \sigma^2_{\Phi}, \mathbf{\Phi})$-Gaussian measurements of $\mathbf{x}_1$ and $\mathbf{x}_2$, and $\mathbf{y}_{q,j}$ is the quantized version of $\mathbf{y}_j$.

The large system regime becomes in the distributed case:

Definition 7. (Large system regime, sparsity andovermeasuring rates, and overlaps).
*Let the J vectors $\mathbf{x}_j \in \mathbb{R}^{N\times 1}$, $\forall j \in \{1, \ldots, J\}$ be $(\{K_{\mathsf{I},j}\}_{j=1}^J, K_{\mathsf{C}}, \{K_j\}_{j=1}^J, N, \sigma^2_{\theta_{\mathsf{C}}}, \{\sigma^2_{\theta_{\mathsf{I},j}}\}_{j=1}^J, \mathbf{\Psi})$ -sparse. For each j, let $\mathbf{y}_j \in \mathbb{R}^{M\times 1}$ be the $(M, N, \sigma^2_{\Phi}, \mathbf{\Phi})$-*Gaussian measurement *of $\mathbf{x}_j$. The system is said to be in the* large system regime *if:*
i) *N goes to infinity.*
ii) *The other dimensions $K_{\mathsf{I},j}, K_{\mathsf{C}}, K_j, M$ are functions of N and tend to infinity as N does.*
iii) *The following rate converges to a constant called* sparsity rate *as N goes to infinity:*

$$\frac{K_j}{N} \to \gamma_j \ . \tag{13}$$

iv) *The following rate converges to a constant called* overmeasuring rate *as N goes to infinity:*

$$\frac{M}{K_j} \to \mu_j > 1. \tag{14}$$

v) *All following rates converge to constants called* overlaps *of the common and innovation components as N goes to infinity:*

$$\frac{K_{\mathsf{C}}}{K_j} \to \omega_{\mathsf{C},j}, \ \frac{K_{\mathsf{I},j}}{K_j} \to \omega_{\mathsf{I},j} \ . \tag{15}$$

Note that $\max\{\omega_{C,j}, \omega_{I,j}\} \leq 1 \leq \omega_{C,j} + \omega_{I,j} \leq 2$.

B. Rate–distortion functions of measurement vector

The information RD function can also be derived for a pair $(X, Y) \sim \mathcal{N}(0, \mathbf{K}_{xy})$ of i.i.d. jointly Gaussian distributed random variables with covariance matrix

$$\mathbf{K}_{xy} = \begin{pmatrix} \sigma_x^2 & \rho_{xy}\sigma_x\sigma_y \\ \rho_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix}. \tag{16}$$

Interestingly, when the SI is available at both encoder and decoder or at the decoder only, the information RD function is the same:

$$D_{x|y}(R) = \sigma_x^2(1-\rho_{xy}^2)2^{-2R} = D_x(R + R^*), \tag{17}$$

where $R^* = \frac{1}{2}\log_2 \frac{1}{1-\rho_{xy}^2} \geq 0$ is the *rate gain*, measuring the amount of rate we save by using the side information Y to decode X. This result holds for optimal vector quantizer [17] but also for scalar uniform quantizers [16, Theorem 8 and Corollary 9] by replacing D_x in (17) by the entropy constrained distortion function $D_x^{\mathsf{EC}}(R)$, defined in (6).

To derive the RD curves for the reconstruction of the sources, we first generalize Theorem 4 and derive the asymptotic distribution of pairs of measurements.

Theorem 8. (Distributed CS: Asymptotic distribution of the pair of Gaussian measurements and measurement RD functions). *Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be $(K_{\mathsf{I},1}, K_{\mathsf{I},2}, K_{\mathsf{C}}, K_1, K_2, N, \sigma^2_{\theta_{\mathsf{C}}}, \sigma^2_{\theta_{\mathsf{I},1}}, \sigma^2_{\theta_{\mathsf{I},2}}, \mathbf{\Psi})$-sparse. $\mathbf{x}_2$ serves as SI for $\mathbf{x}_1$ and is available at the decoder, only. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be the $(M, N, \sigma^2_{\Phi}, \mathbf{\Phi})$-Gaussian measurements of $\mathbf{x}_1$ and $\mathbf{x}_2$. Let (Y_1, Y_2) be the pair of random processes associated to the random vectors $(\mathbf{y}_1, \mathbf{y}_2)$. In the large system regime, (Y_1, Y_2) converges to an i.i.d. Gaussian sequence with covariance matrix*

$$\mathbf{K}_{12} = \begin{pmatrix} \sigma_{y_1}^2 & \rho_{12}\sigma_{y_1}\sigma_{y_2} \\ \rho_{12}\sigma_{y_1}\sigma_{y_2} & \sigma_{y_2}^2 \end{pmatrix}, \tag{18}$$

$$\sigma_{y_j}^2 = \frac{\sigma_{\Phi}^2}{\mu_j}\left[\omega_{\mathsf{C},j}\sigma^2_{\theta_{\mathsf{C}}} + \omega_{\mathsf{I},j}\sigma^2_{\theta_{\mathsf{I},j}}\right] \tag{19}$$

$$\rho_{12} = \left[\left(1 + \frac{\omega_{\mathsf{I},1}}{\omega_{\mathsf{C},1}}\frac{\sigma^2_{\theta_{\mathsf{I},1}}}{\sigma^2_{\theta_{\mathsf{C}}}}\right)\left(1 + \frac{\omega_{\mathsf{I},2}}{\omega_{\mathsf{C},2}}\frac{\sigma^2_{\theta_{\mathsf{I},2}}}{\sigma^2_{\theta_{\mathsf{C}}}}\right)\right]^{-\frac{1}{2}}. \tag{20}$$

Let R be the encoding rate per measurement sample. When the SI is not used, the information RD function satisifies

$$\lim_{N\to+\infty} \frac{1}{\sigma_{y_1}^2} 2^{2R} D_{y_1}(R) = 1, \tag{21}$$

and the entropy–constrained scalar quantizer achieves a distortion $D_{y_1}^{\mathsf{EC}}$ that satisfies

$$\lim_{R\to+\infty} \lim_{N\to+\infty} \frac{1}{\sigma_{y_1}^2} 2^{2R} D_{y_1}^{\mathsf{EC}}(R) = \frac{\pi e}{6} \ . \tag{22}$$

When the measurement $\mathbf{y}_2$ of the SI is used at the decoder, the information RD function satisifies

$$\lim_{N\to+\infty} \frac{1}{\sigma_{y_1}^2} 2^{2(R+R^*)} D_{y_1|y_2}(R) = 1 \ , \tag{23}$$

while the entropy–constrained scalar quantizer achieves a distortion $D^{\mathsf{EC}}_{y_1|y_2}$ that satisfies

$$\lim_{R\to+\infty}\lim_{N\to+\infty}\frac{1}{\sigma^2_{y_1}}2^{2(R+R^*)}D^{\mathsf{EC}}_{y_1|y_2}(R)=\frac{\pi e}{6}\ , \tag{24}$$

where

$$R^*=\frac{1}{2}\log_2\frac{1}{1-\rho^2_{12}}\ . \tag{25}$$

Therefore, in the large system regime (and in the high–rate regime for entropy constrained scalar quantizer), the measurement $\mathbf{y}_2$ of the SI helps reducing the rate by R^* (25) bits per measurement sample:

$$D_{y_1|y_2}(R)=D_{y_1}(R+R^*). \tag{26}$$

Sketch of proof. We consider a vector of finite length $2m$, which contains the first m components of $\mathbf{y}_1$ followed by the first m components of $\mathbf{y}_2$. The vector can be seen as a sum of three components, where each component converges to a Gaussian vector from the multidimensional CLT [21, Theorem 7.18]. Finally, we obtain that (Y_1, Y_2) converges to an i.i.d. Gaussian process. Therefore, classical RD results for i.i.d. Gaussian sources apply. See Appendix C for the complete proof. □

Theorem 8 first states that the measurements of two sparse vectors converge to an i.i.d. Gaussian process in the large system regime. Then, lossy compression of the measurements is considered and the information and entropy constrained rate distortion functions are derived. It is shown that if one measurement vector is used as side information at the decoder, some rate can be saved, depending on the sparse source characteristics, only (see (25) and (20)).

C. Rate–distortion functions of the reconstruction

We now derive the RD functions after reconstruction of the DCS scheme.

Theorem 9. (Distributed CS: Reconstruction RD functions). *Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be $(K_{\mathsf{I},1}, K_{\mathsf{I},2}, K_{\mathsf{C}}, K_1, K_2, N, \sigma^2_{\theta_{\mathsf{C}}}, \sigma^2_{\theta_{\mathsf{I},1}}, \sigma^2_{\theta_{\mathsf{I},2}}, \mathbf{\Psi})$-sparse. $\mathbf{x}_2$ serves as SI for $\mathbf{x}_1$ and is available at the decoder, only. Let $\mathbf{y}_1$ and $\mathbf{y}_2$ be the $(M, N, \sigma^2_\Phi, \mathbf{\Phi})$-Gaussian measurements of $\mathbf{x}_1$ and $\mathbf{x}_2$, s.t. $K_1+3<M<N$. Let R be the encoding rate per measurement sample. The distortion[3] of the source $\mathbf{x}_1$ is denoted as $D^{\mathsf{IR}}_{x_1}$ when the SI is not available at the receiver, $D^{\mathsf{IR}}_{x_1|y_2}$ when the measurements of the SI are available at the SWC decoder (*IR *stands for* independent reconstruction*), and $D^{\mathsf{JR}}_{x_1|x_2}$ when the SI is used not only to reduce the encoding rate but also to improve the reconstruction fidelity (*JR *stands for* joint reconstruction*).*
Then, when independent reconstruction is performed, the RD functions for $\mathbf{x}_1$ satisfy, in the large system regime:

$$D^{\mathsf{IR}}_{x_1}(R)\geq D^{\mathsf{IR}\ \text{oracle}}_{x_1}(R)=\gamma_1\frac{\mu_1}{\mu_1-1}\frac{1}{\sigma^2_\Phi}D_{y_1}(R), \tag{27}$$

$$D^{\mathsf{IR}}_{x_1|y_2}(R)\geq D^{\mathsf{IR}\ \text{oracle}}_{x_1|y_2}(R)=\gamma_1\frac{\mu_1}{\mu_1-1}\frac{1}{\sigma^2_\Phi}D_{y_1|y_2}(R), \tag{28}$$

Therefore, in the large system regime, the operational RD functions satisfy

$$D^{\mathsf{IR}}_{x_1}(R)\geq D^{\mathsf{IR}\ \text{oracle}}_{x_1}(R)=\gamma_1\frac{\omega_{\mathsf{C},1}\sigma^2_{\theta_{\mathsf{C}}}+\omega_{\mathsf{I},1}\sigma^2_{\theta_{\mathsf{I},1}}}{\mu_1-1}2^{-2R}, \tag{29}$$

$$D^{\mathsf{IR}}_{x_1|y_2}(R)\geq D^{\mathsf{IR}\ \text{oracle}}_{x_1|y_2}(R)=\gamma_1\frac{\omega_{\mathsf{C},1}\sigma^2_{\theta_{\mathsf{C}}}+\omega_{\mathsf{I},1}\sigma^2_{\theta_{\mathsf{I},1}}}{\mu_1-1}2^{-2(R+R^*)}, \tag{30}$$

$$D^{\mathsf{IR}}_{x_1|y_2}(R)=D^{\mathsf{IR}}_{x_1}(R+R^*), \tag{31}$$

where R^* is defined in (25). In the large system regime and in the high–rate regime, the entropy constrained RD functions satisfy:

$$D^{\mathsf{IR}\ \mathsf{EC}}_{x_1}(R)\geq\gamma_1\frac{\omega_{\mathsf{C},1}\sigma^2_{\theta_{\mathsf{C}}}+\omega_{\mathsf{I},1}\sigma^2_{\theta_{\mathsf{I},1}}}{\mu_1-1}\frac{\pi e}{6}2^{-2R}, \tag{32}$$

$$D^{\mathsf{IR}\ \mathsf{EC}}_{x_1|y_2}(R)\geq\gamma_1\frac{\omega_{\mathsf{C},1}\sigma^2_{\theta_{\mathsf{C}}}+\omega_{\mathsf{I},1}\sigma^2_{\theta_{\mathsf{I},1}}}{\mu_1-1}\frac{\pi e}{6}2^{-2(R+R^*)}. \tag{33}$$

When joint reconstruction is performed, the RD functions for $\mathbf{x}_1$ satisfy:

$$D^{\mathsf{JR}}_{x_1|x_2}(R)\geq D^{\mathsf{JR}\ \text{oracle}}_{x_1|x_2}(R)=\omega_{\mathsf{I},1}\gamma_1\frac{\mu_1}{\mu_1-\omega_{\mathsf{I},1}}\frac{1}{\sigma^2_\Phi}D_{y_1|y_2}(R), \tag{34}$$

where, in the large system regime,

$$D^{\mathsf{JR}\ \text{oracle}}_{x_1|x_2}(R)=\omega_{\mathsf{I},1}\gamma_1\frac{\omega_{\mathsf{C},1}\sigma^2_{\theta_{\mathsf{C}}}+\omega_{\mathsf{I},1}\sigma^2_{\theta_{\mathsf{I},1}}}{\mu_1-\omega_{\mathsf{I},1}}2^{-2(R+R^*)}, \tag{35}$$

and in the high rate regime

$$D^{\mathsf{JR}\ \mathsf{EC}\ \text{oracle}}_{x_1|x_2}(R)=\omega_{\mathsf{I},1}\gamma_1\frac{\omega_{\mathsf{C},1}\sigma^2_{\theta_{\mathsf{C}}}+\omega_{\mathsf{I},1}\sigma^2_{\theta_{\mathsf{I},1}}}{\mu_1-\omega_{\mathsf{I},1}}\frac{\pi e}{6}2^{-2(R+R^*)}. \tag{36}$$

Finally,

$$D^{\mathsf{JR}}_{x_1|x_2}(R)\geq D^{\mathsf{IR}}_{x_1}(R+R^*+R^{\mathsf{JR}}), \tag{37}$$

$$\text{where } R^{\mathsf{JR}}=\frac{1}{2}\log_2\left[\frac{1}{\omega_{\mathsf{I},1}}\frac{\mu_1-\omega_{\mathsf{I},1}}{\mu_1-1}\right] \tag{38}$$

and where R^* has been defined in (25). Therefore, when the SI is available at the decoder, it helps reducing the rate by R^*+R^{JR} bits per measurement sample.

Sketch of proof. An oracle is considered in order to derive lower bounds. More precisely, it is assumed that the sparsity support of $\mathbf{x}_1$ is known if independent reconstruction is performed and that also the support of the common component $\mathbf{x}_{\mathsf{C}}$ is known if joint reconstruction is performed. The exact distortion of the oracles are derived, from which a closed–form expression of the rate gains are given. See the complete proof in Appendix D. □

As one would expect, when there is no innovation component ($\omega_{\mathsf{I},j}\to 0$), the distortion of the oracle is zero and the rate gain R^{JR} is largest (tends to infinity). On the contrary, when there is no common component ($\omega_{\mathsf{I},j}\to 1$), R^{JR} tends to zero. Moreover, even if the SI could be exploited also to enhance the quality of the dequantization of the unknown source, the gain due to Joint Dequantization becomes negligible in the high–rate region [9]. For this reason we neglected Joint Dequantization in this analysis.

[3] All the RD functions are operational referred to CS reconstruction algorithms, so we omit the CS superscript not to overload the notation

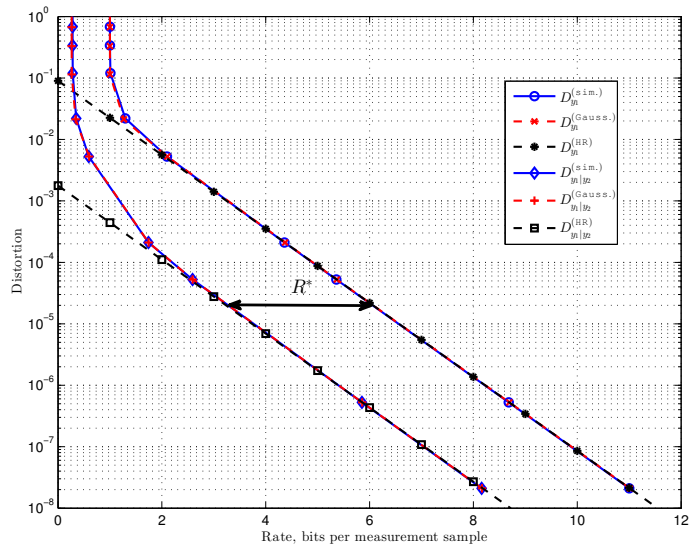
V. Numerical Results

In this section, we validate by simulations the results obtained in the previous sections, comparing numerical results to the derived equations. In particular, first we validate the RD functions of the measurement vector, both in the single–source and in the distributed case, hence validating at the same time the rate gain R^*. Then, we validate the RD functions of independent and joint reconstruction (hence the rate gain R^{JR}). Finally, we compare the results obtained using the oracle estimator with the results obtained using practical reconstruction algorithms, showing that the rate gains hold also in a practical scenario. For each test, $\mathbf{\Psi}$ is the DCT matrix, each non zero component of $\boldsymbol{\theta}$ is drawn from a normal distribution, and $\sigma^2_{\mathbf{\Phi}} = 1$.

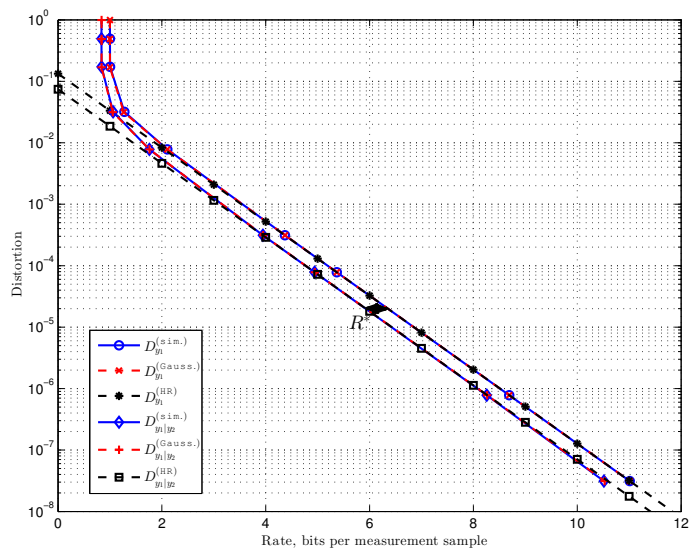
First, we test the validity of the RD functions derived for the measurements. Fig. 2 plots the quantization distortion of $\mathbf{y}_1$, *i.e.*, $\mathbb{E}\left[\frac{1}{M}\left\|\mathbf{y}_1 - \mathbf{y}_{q,1}\right\|_2^2\right]$ versus the rate R, measured in bits per measurement sample (bpms). The distortion has been averaged over 10^4 trials, and for each trial different realizations of the sources, the sensing matrix and the noise have been drawn. Fig. 2 shows two subfigures corresponding to different sets of signal and system parameters (signal length N, sparsity of the common component K_{C} and of the innovation component $K_{\mathsf{I},j}$, variance of the nonzero components $\sigma^2_{\theta_{\mathsf{C}}}$ and $\sigma^2_{\theta_{\mathsf{I},j}}$, respectively, and length of the measurement vector M). Each subfigure shows two families of curves, corresponding to the cases in which $\mathbf{y}_2$ is (respectively, is not) used as SI. Each family is composed by 3 curves. $i)$ The curve labeled as (HR) – standing for *high rate* – is the asymptote of the operational RD curves (22) (or (24)). $ii)$ The curve labeled as (Gauss.) corresponds to the distortion of a synthetic correlated Gaussian source pair with covariance matrix as in (18), where $\sigma^2_{\mathbf{y}_j}$ is defined in (19) and $\rho_{\mathbf{y}_1\mathbf{y}_2}$ in (20), and quantized with a uniform scalar quantizer. The rate is computed as the *symbol* entropy of the samples $\mathbf{y}_{q,1}$ (the conditional *symbol* entropy of $\mathbf{y}_{q,1}$ given $\mathbf{y}_2$), quantized with a uniform scalar quantizer. Entropy and conditional entropy have been evaluated computing the number of symbol occurrences over vectors of length 10^8. $iii)$ The curves labeled as (sim.) are the simulated RD for a measurement vector pair obtained generating $\mathbf{x}_1$ and $\mathbf{x}_2$ according to Def. 6, measuring them with the same $\mathbf{\Phi}$ to obtain $\mathbf{y}_1$ and $\mathbf{y}_2$ and quantizing $\mathbf{y}_1$ and $\mathbf{y}_2$ with a uniform scalar quantizer.

First, we notice that the (HR) equation perfectly matches the simulated curves when $R > 2$, showing that the high–rate regime occurs for relative small values of R. Then, it can be noticed that (Gauss.) curves perfectly overlap the (sim.) ones, validating both equations (19) and (20) and showing that the convergence to the Gaussian case occurs for low values of N, M, K_C, $K_{\mathsf{I},j}$.

After validating the RD functions derived for the measurements, we test the RD functions for the oracle reconstruction. Fig. 3 depicts the performance of the complete DCS scheme, in terms of reconstruction error, *i.e.*, $\mathbb{E}\left[\frac{1}{N}\left\|\widehat{\mathbf{x}}_1 - \mathbf{x}_1\right\|_2^2\right]$ versus the rate per measurement sample R. The figure shows two subfigures corresponding to different sets of signal and system parameters (N, K_{C}, $K_{\mathsf{I},j}$, $\sigma^2_{\theta_{\mathsf{C}}}$, $\sigma^2_{\theta_{\mathsf{I},j}}$, M). Each subfigure shows



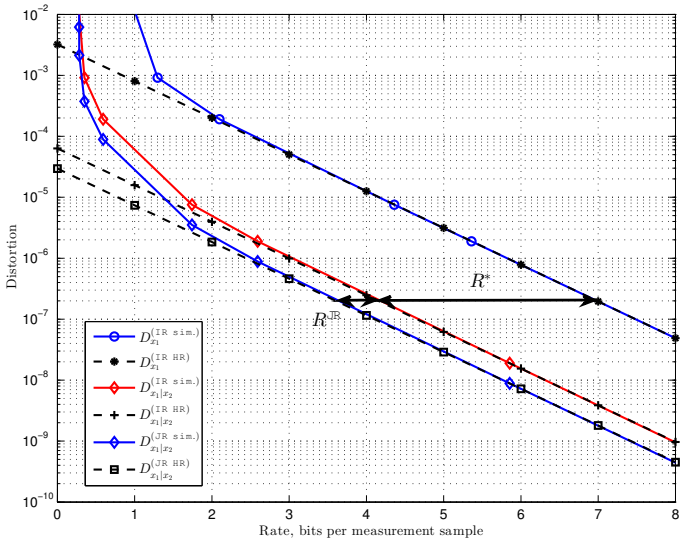
(a) $N = 512$, $K_{\mathsf{C}} = K_{\mathsf{I},j} = 8$, $\sigma^2_{\theta_{\mathsf{C}}} = 1$, $\sigma^2_{\theta_{\mathsf{I},j}} = 10^{-2}$, $M = 128$



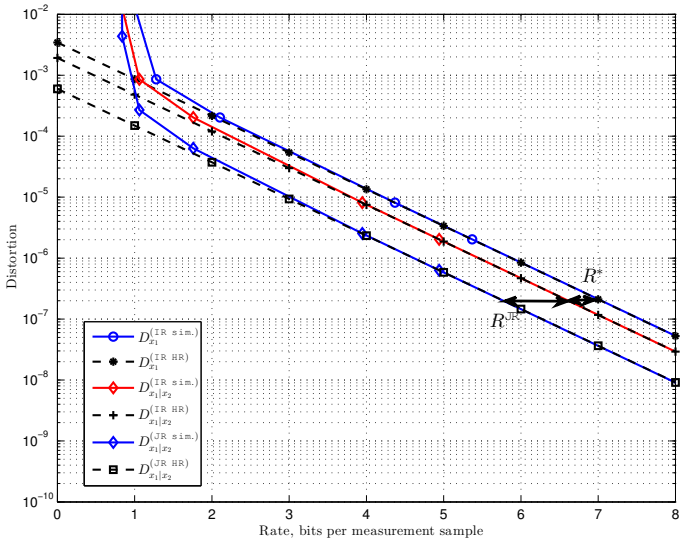
(b) $N = 1024$, $K_{\mathsf{C}} = 16$, $K_{\mathsf{I},j} = 8$, $\sigma^2_{\theta_{\mathsf{C}}} = \sigma^2_{\theta_{\mathsf{I},j}} = 1$, $M = 256$

Fig. 2. Simulated vs. theoretical Rate–Distortion functions of *measurement vectors* for two different sets of signal and system parameters. Single–source and distributed cases.

three pairwise comparisons. First, it compares the RHS of equation (32) (IR HR – standing for *independent reconstruction high rate*) with the oracle reconstruction distortion from $\mathbf{y}_{q,1}$ vs. the *symbol* entropy of the samples $\mathbf{y}_{q,1}$ (IR sim. – standing for *independent reconstruction simulated*), obtaining a match for $R > 2$. Second, it compares the RHS of equation (33) (IR HR) with the oracle reconstruction distortion from $\mathbf{y}_{q,1}$ vs. the conditional *symbol* entropy of $\mathbf{y}_{q,1}$ given $\mathbf{y}_2$ (IR sim.), obtaining a match for $R > 2.5$ and validating once more the evaluation of the rate gain R^* due to the SWC. Third, it compares the RHS of equation (36) (JR HR – standing for *joint reconstruction high rate*) with the ideal (knowing the sparsity support of the common component) oracle Joint Reconstruction distortion from $\mathbf{y}_{q,1}$ and $\mathbf{y}_2$ vs. the conditional *symbol* entropy of $\mathbf{y}_{q,1}$ given $\mathbf{y}_2$ (JR sim. –



(a) $N = 512$, $K_{\mathsf{C}} = K_{\mathsf{I},j} = 8$, $\sigma^2_{\theta_{\mathsf{C}}} = 1$, $\sigma^2_{\theta_{\mathsf{I},j}} = 10^{-2}$, $M = 128$



(b) $N = 1024$, $K_{\mathsf{C}} = 16$, $K_{\mathsf{I},j} = 8$, $\sigma^2_{\theta_{\mathsf{C}}} = \sigma^2_{\theta_{\mathsf{I},j}} = 1$, $M = 256$

Fig. 3. Simulated vs. theoretical Rate–Distortion functions of the *oracle reconstruction* for two different sets of signal and system parameters. Single–source, distributed and joint reconstruction cases.

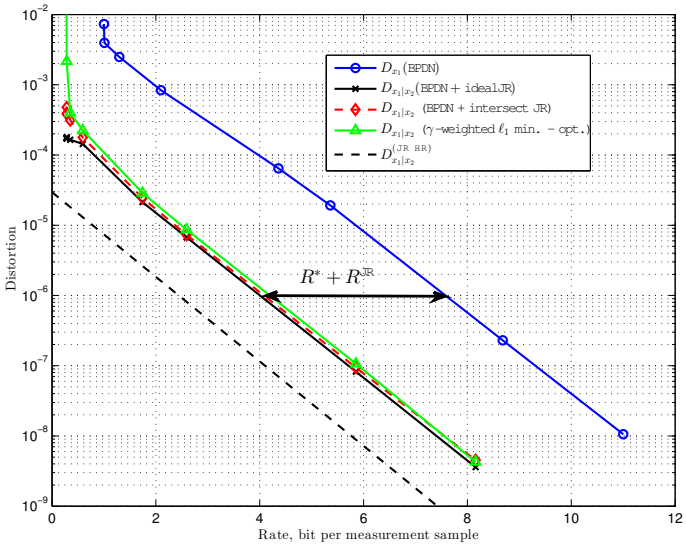


Fig. 4. Practical Reconstruction Rate–Distortion function. $N = 512$, $K_{\mathsf{C}} = K_{\mathsf{I},j} = 8$, $\sigma^2_{\mathbf{\Phi}} = 1$, $\sigma^2_{\theta_{\mathsf{C}}} = 1$, $\sigma^2_{\theta_{\mathsf{I},j}} = 10^{-2}$ and $M = 128$

standing for *joint reconstruction simulated*), obtaining a match for $R > 3$, validating the expression of the Rate Gain due to Joint Reconstruction given in (38).

Finally, we report in Fig. 4 the performance of practical reconstruction algorithms solving the optimization problem (2). The curve labelled as (BPDN) reports the RD function of the independent reconstruction. The curve labelled as (BPDN + ideal JR) reports the RD function of the Joint Reconstruction when the sparsity support of the common component is known at the decoder. The curve labelled as (BPDN + Intersect JR), instead, shows the RD performance of a Joint Reconstruction scheme in which the sparsity support of the common component is not known *a priori*, but is estimated from the measurements $\mathbf{y}_1$ and $\mathbf{y}_2$ using the JR Algorithm 1 in [9] (Intersect JR). The principle behind the Intersect JR algorithm is that the sparsity support of the common component is obtained as the intersection between the estimated sparsity supports of the information source and the SI. Comparing (BPDN + ideal JR) with the oracle performance curve (JR HR), it can be noticed that, apart from a penalty due to the missing knowledge of the sparsity support, the slope of the RD curve is the same as in the ideal oracle case. Moreover, Fig. 4 shows that the practical JR Algorithm 1 in [9] performs very close to ideal JR algorithm. Note that a fully practical scheme will use a real SWC encoder/decoder. In [9] we showed that a system implementing a real SWC encoder/decoder performs very close to the lower bound represented by the entropy and conditional entropy. Finally, in Fig. 4 it can be seen that $R^* + R^{\mathsf{JR}} = 3.49$ bpms, which roughly corresponds to the sum of $R^* = 2.83$ bpms in Figs. 2a and 3a and $R^{\mathsf{JR}} = 0.55$ bpms in Fig. 3a, proving that the Rate Gains derived for the Oracle receiver hold in a practical scenario, as well. Finally, Fig. 4 plots the performance of the γ-weighted ℓ_1-norm minimization algorithm [7, (12)], which for a fair comparison we optimized to take into account that the measurements of the SI are perfectly known whereas the measurements of the unknown source are subject to quantization noise. It can be noticed that the performance is slightly worse with respect to the Intersect JR algorithm. Plus, the complexity increase is significant since the γ-weighted ℓ_1-norm minimization algorithm needs the solution of a problem of size $3N$, which means a compexity increase of 8 times since the complexity of basis pursuit minimization algorithms is cubic with the size of the problem.

VI. Conclusions

We have studied the best achievable performance in the RD sense over all single–source and DCS schemes, under the constraint of high–rate quantization. Closed form expressions of the RD curves have been derived in the asymptotic regime, and simulations have shown that the asymptote is reached for relatively small number of measurements ($M \simeq 100$) and

small rate ($R > 2$ bits/measurement sample). The RD curve computation is based on the convergence of the measurement vector to the multidimensional standard normal distribution. This generalizes [3, Theorem 2] that derives the marginal distribution of the measurement samples when the observed signal is directly sparse. We have then derived a closed–form expression of the highest rate gain achieved exploiting all levels of source correlations at the receiver, along with a closed–form expression of the average performance of the oracle receiver, using novel results on random Wishart matrix theory. Simulations showed that the scheme proposed in [9] almost achieves this best rate gain, and that the only penalty is due to the missing knowledge of the sparsity support as in single–source compressed sensing.

APPENDIX A
PROOF OF THEOREM 4

The Gaussian random matrix $\mathbf{\Phi}$ is orthogonal invariant [25, Section 4.3], [26, Ex. 2.4]. Therefore, $\mathbf{U} = \mathbf{\Phi\Psi}$ is a random matrix with i.i.d. entries drawn from $\mathcal{N}(0, \sigma_\Phi^2)$. Without loss of generality, we assume that the K non zeros of $\boldsymbol{\theta}$ are placed at the beginning of $\boldsymbol{\theta}$. To show that the infinite length measurement random variables are asymptotically independent and Gaussian, we consider a finite length vector $\mathbf{y}^{(m)}$ that contains the m first components of $\mathbf{y}$, and show that, for each m, $\mathbf{y}^{(m)}$ converges to a Gaussian vector with diagonal covariance matrix, as N grows.

Let $\underline{Y}$ and $\underline{U}_k$denote the random vectors in $\mathbb{R}^m$ associated to the Gaussian measurement vector $\mathbf{y}^{(m)}$ and to the k-th column of the matrix $\mathbf{U}$, restricted to its m first rows. Let Θ_k denote the random variable in $\mathbb{R}$ associated to the k-th component of the vector $\boldsymbol{\theta}$. We have

$$\underline{Y} = \frac{1}{\sqrt{M}} \sum_{k=1}^{K} \underline{U}_k \Theta_k = \sqrt{\frac{K}{M}} \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \underline{U}_k \Theta_k \tag{39}$$

By definition, $\underline{U}_k$ and Θ_k are independent, and the sequences $\{\underline{U}_k\}_k$ and $\{\Theta_k\}_k$ are i.i.d. centered with covariance matrix $\sigma_\Phi^2 \mathbf{I}$ and variance σ_θ^2, respectively. Therefore, each vector $\underline{U}_k \Theta_k$ is centered with covariance matrix $\sigma_\Phi^2 \sigma_\theta^2 \mathbf{I}$, where $\mathbf{I}$ is the $m \times m$ identity matrix. From the multidimensional Central Limit theorem [21, Theorem 7.18], Y converges (in distribution) to a multidimensional Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix $\frac{1}{\mu}\sigma_\Phi^2 \sigma_\theta^2 \mathbf{I}$. Thus, the entries of the vector are independent. Letting m grow to ∞, we have that the sequence of Gaussian measurements converges to a Gaussian i.i.d. sequence with mean 0 and variance $\frac{1}{\mu}\sigma_\Phi^2 \sigma_\theta^2$. □

APPENDIX B
PROOF OF THEOREM 5

We derive a lower bound on the achievable distortion by assuming that the sparsity support Ω of $\mathbf{x}$ is known at the decoder. Let $\mathbf{U}_\Omega$ be the submatrix of $\mathbf{U}$ obtained by keeping the columns of $\mathbf{\Phi\Psi}$ indexed by Ω, and let Ω^c denote the complementary set of indexes. The optimal reconstruction is then obtained by using the pseudo–inverse of $\mathbf{U}_\Omega$, denoted by $\mathbf{U}_\Omega^\dagger$:

$$\begin{cases} \widehat{\boldsymbol{\theta}}_\Omega & = \sqrt{M}\mathbf{U}_\Omega^\dagger \mathbf{y}_q := \sqrt{M}\left(\mathbf{U}_\Omega^\mathsf{T}\mathbf{U}_\Omega\right)^{-1}\mathbf{U}_\Omega^\mathsf{T}\mathbf{y}_q \\ \widehat{\boldsymbol{\theta}}_{\Omega^c} & = \mathbf{0} \end{cases} \tag{40}$$

and $\widehat{\mathbf{x}} = \mathbf{\Psi}\widehat{\boldsymbol{\theta}}$, where $\mathbf{y}_q$ is the quantized version of $\mathbf{y}$ i.e. $\mathbf{y}_q = \mathbf{y} + \mathbf{e}$, where $\mathbf{e}$ is the quantization noise of variance σ_e^2. Note that in (40) the product by $\sqrt{M}$ is a consequence of Definition 2.

$$\mathbb{E}\left[\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2\right] = \mathbb{E}\left[\left\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|_2^2\right] = \mathbb{E}\left[\left\|\widehat{\boldsymbol{\theta}}_\Omega - \boldsymbol{\theta}_\Omega\right\|_2^2\right] \tag{41}$$

$$= M\mathbb{E}\left[\left\|\mathbf{U}_\Omega^\dagger \mathbf{e}\right\|_2^2\right] \tag{42}$$

$$= M\mathbb{E}\left[\mathbf{e}^T \mathbb{E}\left[(\mathbf{U}_\Omega \mathbf{U}_\Omega^T)^\dagger\right]\mathbf{e}\right] \tag{43}$$

The first equality in (41) follows from the orthogonality of the matrix $\mathbf{\Psi}$, whereas the second follows from the assumption that Ω is the true support of $\boldsymbol{\theta}$. (42) derives from the definition of the pseudo-inverse, and (43) from $\mathbf{U}_\Omega^{\dagger\,T}\mathbf{U}_\Omega^\dagger = (\mathbf{U}_\Omega\mathbf{U}_\Omega^T)^\dagger$. Then, if $M > K + 3$,

$$\mathbb{E}\left[\|\widehat{\mathbf{x}} - \mathbf{x}\|_2^2\right] = M\mathbb{E}\left[\mathbf{e}^T \frac{K}{M(M-K-1)} \frac{1}{\sigma_\Phi^2}\mathbf{I}\,\mathbf{e}\right] \tag{44}$$

$$= \frac{MK}{M-K-1}\frac{\sigma_e^2}{\sigma_\Phi^2} \tag{45}$$

$$\sigma_{\hat{x}}^2 = \frac{K}{N}\frac{M}{M-K-1}\frac{\sigma_e^2}{\sigma_\Phi^2} \tag{46}$$

where $\sigma_{\hat{x}}^2$ stands for the distortion of the oracle estimator at finite length N. (44) comes from the fact that, since $M > K$, $\mathbf{U}_\Omega\mathbf{U}_\Omega^T$ is rank deficient and follows a singular M-variate Wishart distribution with K degrees of freedom and scale matrix $\sigma_\Phi^2\mathbf{I}$ [24]. Its pseudo-inverse follows a generalized inverse Wishart distribution, whose distribution is given in [24] and mean in [13, Theorem 2.1], under the assumption that $M > K + 3$. Note that the distortion of the oracle only depends on the variance of the quantization noise and not on its covariance matrix. Therefore, our result holds even if the noise is correlated (for instance if vector quantization is used). As a consequence, we can apply our result to any quantization algorithm. Therefore as $N \to \infty$, if $\mu > 1$

$$D_x(R) \geq D_x^{\text{oracle}}(R) = \frac{\gamma}{1-\mu}\frac{1}{\sigma_\Phi^2}D_y(R) \tag{47}$$

$$= \gamma\frac{\mu}{1-\mu}\sigma_\theta^2 2^{-2R}$$

where the first equality is obtained by taking the limit $N \to \infty$ of (46) and the second equality follows from Theorem 4. Substituting the entropy-constrained RD function of the measurements (9) in (47) leads to the entropy-constrained reconstruction RD function (12). □

APPENDIX C
PROOF OF THEOREM 8

We first consider non-overlapping sparse components, i.e. $\boldsymbol{\theta}_\mathsf{C}$, $\boldsymbol{\theta}_{\mathsf{I},1}$ and $\boldsymbol{\theta}_{\mathsf{I},2}$ have non overlapping sparsity supports. From

Theorem 4, $\mathbf{U} = \mathbf{\Phi\Psi}$ is a random matrix with i.i.d. entries drawn from $\mathcal{N}(0, \sigma_\Phi^2)$. Without loss of generality, we assume that all non zeros of $\boldsymbol{\theta}$ are placed at the beginning of $\boldsymbol{\theta}$, with first the K_{C} common, then the $K_{\mathsf{I},1}$ and the $K_{\mathsf{I},2}$ individual components.

We build the finite length $2m$ vector $\mathbf{y}^{(2m)}$ that contains the m first components of $\mathbf{y}_1$ concatenated with the m first components of $\mathbf{y}_2$, and let N go to infinity. Let $\underline{Y}$ denote the random vector in $\mathbb{R}^{2m}$ associated to the Gaussian measurement vector $\mathbf{y}^{(2m)}$ and let $\underline{U}_k$ denote the random vector in $\mathbb{R}^m$ associated to the k-th column of the matrix U, restricted to its m first rows. Let Θ_k denote the random variable in $\mathbb{R}$ associated to the k-th component of the vector $\boldsymbol{\theta}$. By definition of the sparse vectors and their measurements, we have

$$\underline{Y} = \frac{1}{\sqrt{M}} \begin{pmatrix} \sum_{k=1}^{K_{\mathsf{C}}} \underline{U}_k \Theta_k + \sum_{k=K_{\mathsf{C}}+1}^{K_{\mathsf{C}}+K_{\mathsf{I},1}} \underline{U}_k \Theta_k \quad + \mathbf{0} \\ \sum_{k=1}^{K_{\mathsf{C}}} \underline{U}_k \Theta_k + \mathbf{0} \qquad + \sum_{k=K_{\mathsf{C}}+K_{\mathsf{I},1}+1}^{K_{\mathsf{C}}+K_{\mathsf{I},1}+K_{\mathsf{I},2}} \underline{U}_k \Theta_k \end{pmatrix}$$
$$= \sqrt{\frac{K_{\mathsf{C}}}{M}} \underline{Y}_{\mathsf{C}} + \sqrt{\frac{K_{\mathsf{I},1}}{M}} \underline{Y}_{\mathsf{I},1} + \sqrt{\frac{K_{\mathsf{I},2}}{M}} \underline{Y}_{\mathsf{I},2} \tag{48}$$

where $\mathbf{0}$ stands for the all zero vector of length m and $\underline{Y}_{\mathsf{C}}$, $\underline{Y}_{\mathsf{I},1}$and $\underline{Y}_{\mathsf{I},2}$ are defined as

$$\underline{Y}_{\mathsf{C}} := \frac{1}{\sqrt{K_{\mathsf{C}}}} \begin{pmatrix} \sum_{k=1}^{K_{\mathsf{C}}} \underline{U}_k \Theta_k \\ \sum_{k=1}^{K_{\mathsf{C}}} \underline{U}_k \Theta_k \end{pmatrix}$$
$$\underline{Y}_{\mathsf{I},1} := \frac{1}{\sqrt{K_{\mathsf{I},1}}} \begin{pmatrix} \sum_{k=K_{\mathsf{C}}+1}^{K_{\mathsf{C}}+K_{\mathsf{I},1}} \underline{U}_k \Theta_k \\ \mathbf{0} \end{pmatrix}$$
$$\underline{Y}_{\mathsf{I},2} := \frac{1}{\sqrt{K_{\mathsf{I},2}}} \begin{pmatrix} \mathbf{0} \\ \sum_{k=K_{\mathsf{C}}+K_{\mathsf{I},1}+1}^{K_{\mathsf{C}}+K_{\mathsf{I},1}+K_{\mathsf{I},2}} \underline{U}_k \Theta_k \end{pmatrix}$$

Each vector $\underline{U}_k \Theta_k$ is centered with covariance matrix

$$\begin{array}{lll} \sigma_\Phi^2 \sigma_{\theta_{\mathsf{C}}}^2 \mathbf{I} & \text{if} & 1 \le k \le K_{\mathsf{C}} \\ \sigma_\Phi^2 \sigma_{\theta_{\mathsf{I},1}}^2 \mathbf{I} & \text{if} & K_{\mathsf{C}}+1 \le k \le K_{\mathsf{C}}+K_{\mathsf{I},1} \\ \sigma_\Phi^2 \sigma_{\theta_{\mathsf{I},2}}^2 \mathbf{I} & \text{if} & K_{\mathsf{C}}+K_{\mathsf{I},1}+1 \le k \le K_{\mathsf{C}}+K_{\mathsf{I},1}+K_{\mathsf{I},2} \ , \end{array}$$

where $\mathbf{I}$ is the $m \times m$ identity matrix.

From the multidimensional CLT [21, Theorem 7.18], $\underline{Y}_{\mathsf{C}}$ converges (in distribution) to a multidimensional Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix $\sigma_\Phi^2 \sigma_{\theta_{\mathsf{C}}}^2 \left(\begin{smallmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & \mathbf{I} \end{smallmatrix}\right)$ in the large system regime. Similarly, in the large system regime, $\underline{Y}_{\mathsf{I},1}$ and $\underline{Y}_{\mathsf{I},2}$ converge (in distribution) to a multidimensional Gaussian distribution with mean vector $\mathbf{0}$ and covariance matrix $\sigma_\Phi^2 \sigma_{\theta_{\mathsf{I},1}}^2 \left(\begin{smallmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{smallmatrix}\right)$, $\sigma_\Phi^2 \sigma_{\theta_{\mathsf{I},2}}^2 \left(\begin{smallmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{smallmatrix}\right)$ respectively, where $\mathbf{0}$ stands for the all zero $m \times m$ matrix[4]. Moreover, $\underline{Y}_{\mathsf{C}}$, $\underline{Y}_{\mathsf{I},1}$ and $\underline{Y}_{\mathsf{I},2}$ are independent, since the elements of each sum are all different in (48). Therefore, $\underline{Y}$ converges to a Gaussian centered vector with covariance matrix

$$\sigma_\Phi^2 \begin{pmatrix} \left(\frac{\omega_{\mathsf{C},1}}{\mu_1}\sigma_{\theta_{\mathsf{C}}}^2 + \frac{\omega_{\mathsf{I},1}}{\mu_1}\sigma_{\theta_{\mathsf{I},1}}^2\right)\mathbf{I} & \sqrt{\frac{\omega_{\mathsf{C},1}}{\mu_1}\frac{\omega_{\mathsf{C},2}}{\mu_2}}\sigma_{\theta_{\mathsf{C}}}^2 \mathbf{I} \\ \sqrt{\frac{\omega_{\mathsf{C},1}}{\mu_1}\frac{\omega_{\mathsf{C},2}}{\mu_2}}\sigma_{\theta_{\mathsf{C}}}^2 \mathbf{I} & \left(\frac{\omega_{\mathsf{C},2}}{\mu_2}\sigma_{\theta_{\mathsf{C}}}^2 + \frac{\omega_{\mathsf{I},2}}{\mu_2}\sigma_{\theta_{\mathsf{I},2}}^2\right)\mathbf{I} \end{pmatrix} , \tag{49}$$

since $\frac{\omega_{\mathsf{C},1}}{\mu_1} = \frac{\omega_{\mathsf{C},2}}{\mu_2}$. In the case of overlapping supports, each entry of the vector $\boldsymbol{\theta}$ can either contain the contribution of only one component (common, or any innovation), or any combination of at least two components. Therefore, the support of the non zero components of $\boldsymbol{\theta}$ and therefore the random vector $\underline{Y}$, can be decomposed into 7 terms, that correspond to the number of subsets of a set of 3 elements (excluding the empty set). Note that each term is a sum of i.i.d. random vectors such that the multidimensional CLT still applies and (49) holds.

Letting m grow to ∞, we have that, in the large system regime, (Y_1, Y_2) converges to an i.i.d. Gaussian sequence with covariance matrix

$$\mathbf{K}_{12} = \begin{pmatrix} \sigma_{y_1}^2 & \rho_{12}\sigma_{y_1}\sigma_{y_2} \\ \rho_{12}\sigma_{y_1}\sigma_{y_2} & \sigma_{y_2}^2 \end{pmatrix}, \text{ where}$$
$$\sigma_{y_j}^2 = \frac{\sigma_\Phi^2}{\mu_j}\left[\omega_{\mathsf{C},j}\sigma_{\theta_{\mathsf{C}}}^2 + \omega_{\mathsf{I},j}\sigma_{\theta_{\mathsf{I},j}}^2\right] \text{ and}$$
$$\rho_{12} = \left[\left(1 + \frac{\omega_{\mathsf{I},1}}{\omega_{\mathsf{C},1}}\frac{\sigma_{\theta_{\mathsf{I},1}}^2}{\sigma_{\theta_{\mathsf{C}}}^2}\right)\left(1 + \frac{\omega_{\mathsf{I},2}}{\omega_{\mathsf{C},2}}\frac{\sigma_{\theta_{\mathsf{I},2}}^2}{\sigma_{\theta_{\mathsf{C}}}^2}\right)\right]^{-\frac{1}{2}} ,$$

since $\frac{\omega_{\mathsf{C},1}}{\mu_1} = \frac{\omega_{\mathsf{C},2}}{\mu_2}$. Therefore, the RD functions are given in (6), (5), and (17). □

Appendix D
Proof of Theorem 9

Let us first consider independent reconstruction. This means that the measurements of the SI are used at the SWC decoder only and not to improve the quality of the dequantization or the reconstruction stages. We derive a lower bound on the achievable distortion by assuming that the sparsity support Ω_1 of $\mathbf{x}_1$ is known at the decoder. This receiver is called the oracle and leads to a variance of estimation $\sigma_{\hat{x}_1}^2$:

$$\sigma_{\hat{x}_1}^2 = \frac{K_1}{N}\frac{M}{M-K_1-1}\frac{\sigma_e^2}{\sigma_\Phi^2} \tag{50}$$

where the derivation is similar to the non distributed case (see Appendix B). σ_e^2 is the quantization noise variance, i.e. $D_{y_1|y_2}(R)$ if the $\mathbf{y}_2$ is used as side information at the SWC decoder and $D_{y_1}(R)$ otherwise. This leads to (27) and (28). Then, in the large system regime, the measurements are Gaussian and the RD functions of the measurements have a closed form expression, which is used to derive (29), (30), (32), and (33).

Finally, note that the gap between the oracle based lower bound and the true RD functions, in the IR case, is only due to the performance of the algorithm chosen to reconstruct $\mathbf{x}_1$ from $\mathbf{y}_1$. The performance of a deterministic CS reconstruction algorithm depends only on the density $p_{x_1|y_{q,1}}$, which in the present Gaussian case is determined by the MSE $D_{y_1}(R)$. In

[4]For the sake of brevity and when it is clear from the context, we use $\mathbf{0}$ to denote either an m length vector or an all zero $m \times m$ matrix.

the IR case, $D_{y_1}(R)$ depends on the presence of the quantization/dequantization step, since the source coding/decoding step is considered a zero–error stage. Hence, the presence or absence of a (distributed) source encoding and decoding stage does not alter $D_{y_1}(R)$ since it does not alter the output of the dequantizer. Conditioning on y_2 yields a rate shift $D_{y_1|y_2}(R) = D_{y_1}(R + R^*)$ that will thus be directly reflected in the CS reconstruction. Hence, it can be written that $D^{\mathsf{IR}}_{x_1}(R) = f(D_{y_1}(R))$ and $D^{\mathsf{IR}}_{x_1|y_2}(R) = f(D_{y_1|y_2}(R)) = f(D_{y_1}(R + R^*))$ for some $f(\cdot)$ depending on the specific reconstruction algorithm. This yields (31).

As for the joint reconstruction, the SI is used to reduce the sparsity level of the unknown signal, and hence the performance of the reconstruction. More precisely, to give a lower bound we assume that the receiver perfectly knows the common component $\mathbf{x}_{\mathsf{C}}$ and $\Omega_{\mathsf{I},1}$ of the innovation component $\mathbf{x}_{\mathsf{I},1}$. Hence, it will use the former to estimate the measurements of the common component $\mathbf{y}_{\mathsf{C}} = \mathbf{\Phi}\mathbf{x}_{\mathsf{C}}$ and then subtract them from $\mathbf{y}_1$. In this way, the vector to be reconstructed is the innovation component $\mathbf{x}_{\mathsf{I},1}$ only, which is sparser than $\mathbf{x}_1$ (given the same N and M). Then, it will use the latter information to apply the oracle, leading to a variance of estimation $\sigma^2_{\hat{x}_1}$:

$$\sigma^2_{\hat{x}_1} = \frac{K_{\mathsf{I},1}}{N} \frac{M}{M - K_{\mathsf{I},1} - 1} \frac{\sigma^2_e}{\sigma^2_\Phi} \tag{51}$$

where σ^2_e is the measurement distortion, when $\mathbf{y}_2$ is used as side information at the receiver. This leads to (34-36). □

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the anonymous Reviewers for their valuable suggestions that helped to improve the quality of the final version of this paper.

REFERENCES

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

[2] E. Candès and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[3] W. Dai and O. Milenkovic, "Information theoretical and algorithmic approaches to quantized compressive sensing," *IEEE Trans. on Communications*, vol. 59, pp. 1857–1866, 2011.

[4] C. Weidmann and M. Vetterli, "Rate distortion behavior of sparse sources," *IEEE Trans. on Information Theory*, vol. 58, no. 8, pp. 4969–4992, Aug. 2012.

[5] A. Fletcher, S. Rangan, and V. Goyal, "On the rate-distortion performance of compressed sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 3. IEEE, 2007.

[6] V. Goyal, A. Fletcher, and S. Rangan, "Compressive sampling and lossy compression," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 48–56, 2008.

[7] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed Compressive Sensing," *ArXiv e-prints*, Jan. 2009.

[8] M. Duarte, M. Wakin, D. Baron, S. Sarvotham, and R. Baraniuk, "Measurement bounds for sparse signal ensembles via graphical models," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4280–4289, 2013.

[9] G. Coluccia, E. Magli, A. Roumy, and V. Toto-Zarasoa, "Lossy compression of distributed sparse sources: a practical scheme," in *European Signal Processing Conf. (EUSIPCO)*, August 2011.

[10] Z. Liu, S. Cheng, A. D. Liveris, and Z. Xiong, "Slepian-Wolf coded nested lattice quantization for Wyner-Ziv coding: High-rate performance analysis and code design," *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4358–4379, 2006.

[11] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding vs. dynamic range," *CoRR*, vol. abs/1104.4842, 2011.

[12] J. N. Laska and R. G. Baraniuk, "Regime change: Bit-depth versus measurement-rate in compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 60, no. 7, pp. 3496–3505, 2012.

[13] R. D. Cook and L. Forzani, "On the mean and variance of the generalized inverse of a singular wishart matrix," *Electronic Journal of Statistics*, vol. 5, pp. 146–158, 2011.

[14] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[15] T. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources (Corresp.)," *IEEE Trans. on Information Theory*, vol. 21, no. 2, pp. 226–228, Sep. 1975.

[16] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, "High-rate quantization and transform coding with side information at the decoder," *Signal Process.*, vol. 86, no. 11, pp. 3160–3179, Nov. 2006.

[17] A. Wyner, "The rate-distortion function for source coding with side information at the decoder-ii: General sources," *Information and Control*, vol. 38, no. 1, pp. 60–80, 1978.

[18] F. Bassi, M. Kieffer, and C. Weidmann, "Wyner-Ziv coding with uncertain side information quality," in *European Signal Processing Conf. (EUSIPCO)*, 2010.

[19] Y. Eldar and G. Kutyniok, Eds., *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[20] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[21] D. Khoshnevisan, *Probability*, ser. Graduate Studies in Mathematics. AMS, 2007.

[22] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, 2006.

[23] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematiques*, vol. 346, no. 9-10, pp. 589–592, 2008.

[24] J. A. Díaz-García and R. Gutiérrez-Jáimez, "Distribution of the generalised inverse of a random matrix and its applications," *Journal of statistical planning and inference*, vol. 136, no. 1, pp. 183–192, 2006.

[25] A. Edelman and N. R. Rao, "Random matrix theory," *Acta Numerica*, vol. 14, pp. 233–297, 4 2005.

[26] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004. [Online]. Available: http://dx.doi.org/10.1516/0100000001

Giulio Coluccia Giulio Coluccia received the bachelor in science (cum laude) in 2003 and the master in science (cum laude) in 2005 in Telecommunications Engineering, both from the Politecnico di Torino, Torino, Italy. He was a Ph.D. student within the Electronics Department of the Politecnico di Torino, Torino, Italy, under the supervision of prof. Giorgio Taricco. During his Ph.D. program, he visited the Telecommunications Research Center Vienna (ftw.), Vienna, Austria. He received the Ph.D. degree in Electronic and Communications engineering in 2009. His research activity included MIMO communications and space-time detection.

Currently, he is a Post Doctoral Researcher within the Image Processing Lab at Politecnico di Torino, leaded by Prof. Enrico Magli. His resarch is focused on Compressed Sensing, with particular interest in its application to Image Processing, Multidimensional Signals and to Distributed Source Coding and Wireless Sensor Netowrks. He is involved in the 5-year project entitled "CRISP - Towards compressive information processing systems" funded by the European Union.

In 2011 he received the Best Paper Award for Signal Processing at the GTTI (Gruppo Telecomunicazioni e Tecnologie dell'Informazione - Telecommunications and Information Technology Group) meeting.

Aline Roumy Aline Roumy received the Engineering degree from Ecole Nationale Superieure de l'Electronique et de ses Applications (ENSEA), Cergy, France in 1996, the Master degree in 1997 and the Ph.D. degree in 2000 from the University of Cergy-Pontoise, France. During 2000-2001, she was a research associate at Princeton University, Princeton, NJ. In 2001, she joined INRIA, Rennes, France as a research scientist. She has held visiting positions at Eurecom and Berkeley University.

Her current research and study interests include the area of statistical signal and image processing, coding theory and information theory.

She has been a Technical Program Committee member and session chair at several international conferences, including ISIT, ICASSP, Eusipco. She is currently serving as a member of the French National University Council (CNU 61). She received the 2011 "Francesco Carassa" Best paper award.

Enrico Magli Enrico Magli received the Ph.D. degree in Electrical Engineering in 2001, from Politecnico di Torino, Italy. He is currently an Associate Professor at the same university, where he leads the Image Processing Lab. His research interests are in the field of multimedia signal processing and networking, compressive sensing, distributed source coding, image and video security, and compression of satellite images.

He is an associate editor of the IEEE Transactions on Circuits and Systems for Video Technology, and of the IEEE Transactions on Multimedia. He is a member of the Multimedia Signal Processing technical committee of the IEEE Signal Processing Society, and of the Multimedia Systems and Applications and the Visual Signal Processing and Communications technical committees of the IEEE Circuits and Systems Society. He has been co-editor of JPEG 2000 Part 11 - Wireless JPEG 2000. He is general co-chair of IEEE MMSP 2013, and has been TPC co-chair of ICME 2012, VCIP 2012, MMSP 2011 and IMAP 2007.

He has published about 40 papers in refereed international journals, 3 book chapters, and over 100 conference papers. He is a co-recipient of the IEEE Geoscience and Remote Sensing Society 2011 Transactions Prize Paper Award, and has received the 2010 Best Reviewer Award of IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing. He has received a 2010 Best Associate Editor Award of IEEE Transactions on Circuits and Systems for Video Technology.