

Enhanced Correlation Estimators for Distributed Source Coding in Large Wireless Sensor Networks

Joan Enric Barceló-Lladó, *Student Member, IEEE*, Antoni Morell Pérez,
and Gonzalo Seco-Granados, *Senior Member, IEEE*

Abstract—In this paper, we propose two estimators based on correlation parameters for the two key steps of a practical *distributed source coding* (DSC) scheme, namely: 1) the computation of the side-information at the receiver side and 2) the estimation of the required number of bits to compress the readings in order to guarantee a certain symbol error probability. We show that using the proposed estimators, the DSC algorithm performs better in terms of the compression rate and the symbol error rate. In particular, this improvement is especially significant when the number of snapshots used in the training phase is only slightly larger than the observation vector. However, when the number of snapshots is much higher than the observation dimension, our proposed estimators perform similarly to the classical estimators.

Index Terms—Distributed source coding (DSC), energy efficiency, generalized statistical analysis (GSA), random matrix theory (RMT), wireless sensor networks (WSN).

I. INTRODUCTION

SAMPLE estimators are widely used in statistical signal processing and it is well-known that their performance is highly conditioned to the number of samples considered [1]. In particular, sample correlation estimators perform the best when the number of samples is sufficiently large in comparison with the dimension of the observation vector. However, when both magnitudes are similar, the performance may be severely degraded and other techniques should be addressed.

In this paper, we propose two enhanced correlation estimators derived from *Generalized Statistical Analysis* (GSA) (introduced by V. L. Girko [2] and extended in [3]). This discipline comes from *Random Matrix Theory* (RMT) [4] and provides consistent estimators when both the number of snapshots of the training phase N and the observa-

tion dimension M are arbitrarily large and comparable in magnitude.

Following this approach, the main motivation is to include such derived estimators for *Distributed Source Coding* (DSC) applied to a large *Wireless Sensor Network* (WSN) framework, which typically are formed by a large number of high space-time correlated sources (e.g. fire control in forests monitoring the temperature or humidity levels, or tracking the location of the products in large stores), where DSC may be used to remove the inherent redundancy in such a correlated readings [5], [6] and hence send compressed messages with the subsequent energy savings.

Surprisingly, existing results from information theory (precisely, from the work of Slepian and Wolf [7]) show that this compression can be executed in a fully blind manner, i.e., only with the knowledge of the local data. It means that sensors compress the data without the knowledge of the signals of the other sensors, and interestingly, without any loss of performance in comparison with the *centralized* approach. Theoretically, the DSC achieves the maximum sum rate, however, practical algorithms still perform far from the theoretical limits [6].

However, practical (and suboptimal) solutions can be found in the literature. For a star-topology WSN, the authors in [8] propose a DSC scheme divided in two phases: the training phase and the compression phase. During the training phase, the correlation parameters are estimated. Hence, the duration of this phase depends on the network configuration and the requirements of the application. In particular, they consider a network composed only of two source nodes and one sink. For higher number of sources, the number of snapshots used in the correlation estimation notably increases. The authors extend their results in [9] to a cluster-based WSN, where each cluster manages a total of four nodes and acts separately to the rest of the clusters. However, in both [8] and [9], the estimation of the correlation parameters is not detailed.

For a relay WSN scenario, the authors in [10] also present a two-phase DSC algorithm. As in [8] and [9], they assume that the training phase is large enough to achieve the desired accuracy in the correlation estimation for an arbitrary number of source nodes. Even so, the sources are managed into smaller groups or clusters. Following this grouping approach, the DSC algorithm cannot fully exploit all the spatial correlations within

Manuscript received January 12, 2012; revised March 30, 2012; accepted May 18, 2012. Date of publication June 4, 2012; date of current version July 24, 2012. This work was supported by the Spanish Government under Project TEC2011-28219 and by the Catalan Government under Grant 2009 SGR 298. The Associate Editor coordinating the review of this paper and approving it for publication was Prof. Kiseon Kim.

The authors are with the Telecommunications and Systems Engineering Department, Universitat Autònoma de Barcelona, Barcelona 08193, Spain (e-mail: joanenric.barcelo@uab.es; antoni.morell@uab.cat; gonzalo.seco@uab.es).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JSEN.2012.2201936

the network, since only the correlations among the sensors of a cluster are used. Therefore, a lot of useful information is missed.

For a multi-hop WSNs, the scheme proposed in [11] exploits the “redundancy free” nature of the DSC to optimize jointly the DSC and the routing paths in order to increase the lifetime of the network. More recent results in [12] extend the DSC algorithm for a multistage scheme. The authors particularize the results for two sources and two layers.

Many other interesting DSC algorithms are also being actively studied, with new alternatives that continuously improve many aspects of DSC. In this paper, we extend the work in [13], where the authors propose a simple DSC algorithm in order to compress the signal from multiple space-time correlated sources. Although their analysis is for an arbitrarily large number of nodes, they particularize for the case of two nodes.

In the literature of DSC, there exists a lack of study regarding the performance drop in the correlation estimation due to a large number of correlated sources.

In order to overcome this limitation, we address the case of a DSC algorithm applied to a large WSN scenario, where the observation dimension M is typically large (since it depends on the number of sensors that composes the network) and classical sample estimators may fail unless a very long training phase is considered (becoming in most cases unpractical). On the contrary, our proposed estimators improve the trade-off between the training phase duration N and the accuracy of the estimation of the correlation parameters. Our contributions are listed as follows:

- 1) We analyze the performance of the DSC algorithm for large WSNs, and in particular, we study the correlation estimation problem in such a scenario.
- 2) We propose two enhanced estimators to mitigate the performance drop of DSC algorithm when the number of sources is arbitrarily large and conventional estimators are used.
- 3) We numerically compare the performance of conventional sample estimators with our enhanced estimators. Our enhanced DSC algorithm turns out to decrease largely the training phase duration and allows us to reduce the number of transmitted bits in comparison with the conventional approach.

Furthermore, this paper completes a preliminary work started in [14] with:

- 1) an expanded background section;
- 2) improved derivations for the proposed correlation estimators by dropping many assumptions, making them simpler and rigorous;
- 3) improved simulation results that present a complete study of the behavior of the DSC algorithm in terms of MSE and SER;
- 4) a comparison between the performance of our proposed estimators with some of the most popular estimation techniques (like *diagonal loading* or *principal component analysis*) providing the corresponding details of the results.

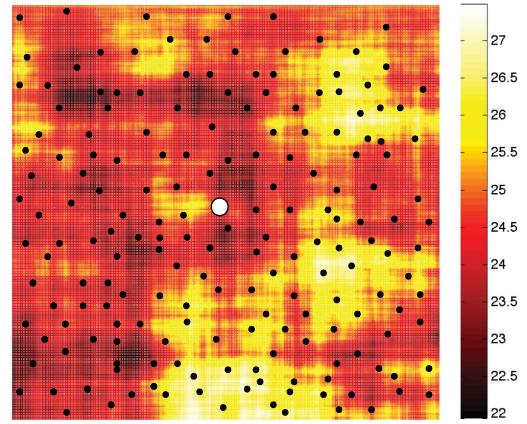


Fig. 1. Illustrative example of a correlation dominated large WSN field, composed of a set S of S sensing nodes (black dots) measuring a certain scalar magnitude and transmitting their readings to one fusion center (white dot).

The rest of the paper is organized as follows: In Section II we describe the system model. The DSC algorithm is presented in Section III. The derivation of the enhanced estimators is detailed in Section IV. Simulation results are given in Section V, and conclusions are drawn in Section VI.

II. SYSTEM MODEL

We assume a large and dense WSN scenario that measures a certain physical phenomena such as temperature or humidity. For *large* WSN we mean that the number of sensing nodes may be arbitrarily large, i.e. of hundreds or thousands of nodes, and for *dense* WSN we mean that the sensing nodes are close enough to present spatial correlations in their measured data. This scenario is graphically summarized in Fig. 1. The final interest of this paper is to study the impact when the fusion center receives the information from a large number of sensing nodes. Thus, although we assume a WSN configured in star topology for simplicity, our proposed algorithm is also compatible with multi-hop techniques or with other network configurations.

Therefore, the network is composed of two types of nodes: 1) a set S of S sensing nodes that transmit the measurements when they are requested and 2) one fusion center that manages the sensing nodes, and gathers and processes their measured data. We assume that the limitations in terms of *computing power* and *energy consumption* are in the sensing nodes, instead we assume no constraints for the fusion center.

Let¹ $x_s(n) \in \mathbb{R}$ denote the scalar reading from the s th sensor in discrete time n . It is modeled as $x_s(n) \sim \mathcal{N}(0, \sigma_{x_s}^2)$. Moreover, $\mathbf{X}_T \in \mathbb{R}^{S \times T}$ assembles the time sequence of the

¹**Notation:** Boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors, and italics denote scalars. $(\cdot)^T$, $(\cdot)^*$, $(\cdot)^H$ denote transpose, complex conjugate, and conjugate transpose (Hermitian) respectively. $[\mathbf{X}]_{i,j}$, $[\mathbf{x}]_i$ is the $(i$ th, j th) element of matrix \mathbf{X} , and i th position of vector \mathbf{x} , respectively. $[\mathbf{X}]_i$ denotes the i th column of \mathbf{X} . $(\cdot)^*$ denotes the optimal value. $\mathbf{X}^{1/2}$ denotes the Hermitian square root of the Hermitian matrix \mathbf{X} , i.e., $\mathbf{X}^{1/2}\mathbf{X}^{1/2} = \mathbf{X}$. Function $\text{Tr}(\mathbf{X})$ means the trace of \mathbf{X} . Function $\lceil \cdot \rceil$ denotes the ceil function. $\mathbb{E}[\cdot]$ is the statistical expectation. Function $\text{erfc}(\cdot)$ represents the complementary error function. Function $\mathcal{I}(a \leq b)$ is the indicator function for the condition $a \leq b$. $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2 , σ_x^2 is the variance of \mathbf{x} . The notation \hat{x} denotes the estimation of the scalar x . Symbol \asymp means almost sure convergence.

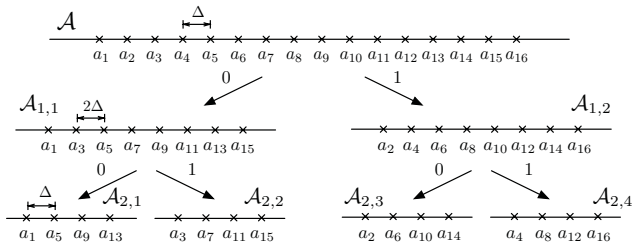


Fig. 2. Graphical representation of the sub-codebooks using a tree-based scheme. In this example, the alphabet \mathcal{A} contains 16 symbols, and a 4th level of sub-codebooks is shown.

measurements for each *sensing node* (where T names the length of the time monitoring window), namely

$$\mathbf{X}_T = \begin{bmatrix} x_1(1) & x_1(2) & \cdots & x_1(T) \\ x_2(1) & x_2(2) & \cdots & x_2(T) \\ \vdots & \vdots & & \vdots \\ x_S(1) & x_S(2) & \cdots & x_S(T) \end{bmatrix}. \quad (1)$$

The covariance matrix of the spatial observation vector $[x_1(n) \ x_2(n) \ \cdots \ x_S(n)]^T$ is given by \mathbf{R}_s , while \mathbf{R}_t denotes the covariance matrix of the time sequence of length T , i.e., $[x_s(1) \ x_s(2) \ \cdots \ x_s(T)]^T$.

III. DISTRIBUTED SOURCE CODING ALGORITHM

Since there are no practical techniques to achieve the theoretical limits of [7] and [15], suboptimal algorithms are used instead. In this paper, we follow the approach in [13], where the authors propose the construction of a codebook based on the decomposition of a given finite alphabet \mathcal{A} in several sub-codebooks. Fig. 2 gives a graphical intuition on how the codebook can be decomposed in several sublevels.

In general, the DSC algorithm is divided in two phases that involve both the *sensing nodes* and the *fusion center* (see Algorithm 1 and Algorithm 2 respectively):

- 1) The *training phase* of length N , where the *sensing node* maps its l -bit reading $x_s(n)$ according to the alphabet $\mathcal{A} = \{a_i\}_{i=1,2,\dots,2^l}$, with a quantization step of $|a_{i+1} - a_i| = \Delta$, and sends an uncompressed version of its data coded in l -bits. After collecting the N snapshots of the training phase, the *fusion center* estimates the correlation parameters for each sensor.
- 2) The *coding phase*, where a given *side-information* $y(n)$ is available at the *fusion center* and the *sensing node* can encode its reading using only $b(n) \leq l$ bits. Hence, the sensor transmits only the index B of a sub-codebook $\mathcal{A}_B \subseteq \mathcal{A}$ (B is codified in $b(n)$ bits) that contains the mapped reading $x_s(n)$. Thus, the *fusion center* receives the sub-codebook identifier B , and selects the symbol in \mathcal{A}_B closer to the *side-information* $y(n)$,

$$x_s(n) = \arg \min_{a_i \in \mathcal{A}_B} |y(n) - a_i|. \quad (2)$$

Let us concentrate on the following two steps of the *coding phase*.

Algorithm 1 Sensing Node

1. Training phase

Get l -bit reading from A/D converter

Transmit l -bit symbol.

2. coding phase

Get l -bit reading from A/D converter

Encode and transmit $b(n)$ -bit codeword.

Step A. Compute the Side-Information $y(n)$: First, let us define the observation vector $\mathbf{x}(n) \in \mathbb{R}^M$ with covariance matrix $\mathbf{R} \in \mathbb{R}^{M \times M}$ as the information available at the *fusion center* and \mathbf{r}_x is the cross-correlation vector, $\mathbf{r}_x = \mathbb{E}[\mathbf{x}(n)x_s(n)]$. The vector $\mathbf{x}(n)$ collects: 1) the K past readings of the sensor and 2) the readings of the set \mathcal{S}' of already-decoded sensors in time slot n (where $\mathcal{S}' \subset \mathcal{S}$ with cardinality S'), hence $M = K + S'$. Note also that the covariance matrix \mathbf{R} will be constructed from the corresponding entries of \mathbf{R}_s and \mathbf{R}_t . Then, the *side-information* $y(n)$ is a linear prediction of $x_s(n)$ and it is computed as a linear combination of the entries of $\mathbf{x}(n)$

$$y(n) = \mathbf{w}^H \mathbf{x}(n) \quad (3)$$

following the *Linear Wiener Filter* (LWF) solution. The LWF solution, $\mathbf{w}^* = \mathbf{R}^{-1} \mathbf{r}_x$, is known to be optimal in the *Mean Square Error* (MSE) sense [1]. Mathematically

$$\text{MSE}(\mathbf{w}) = \sigma_{x_s}^2 - 2 \text{Re}[\mathbf{w}^H \mathbf{r}_x] + \mathbf{w}^H \mathbf{R} \mathbf{w} \quad (4)$$

$$\frac{\partial \text{MSE}(\mathbf{w})}{\partial \mathbf{w}^H} = -\mathbf{r}_x + \mathbf{w}^H \mathbf{R} = 0$$

$$\mathbf{w}^* = \mathbf{R}^{-1} \mathbf{r}_x \quad (5)$$

and then, the MSE achieved is minimum and is given by

$$\text{MSE}(\mathbf{w}^*) = \sigma_{x_s}^2 - \mathbf{r}_x^H \mathbf{R}^{-1} \mathbf{r}_x. \quad (6)$$

However, to compute \mathbf{w}^* the perfect knowledge of \mathbf{R}^{-1} and \mathbf{r}_x is necessary but not available. Classical methods replace \mathbf{R}^{-1} and \mathbf{r}_x directly by their sample estimators denoted by $\hat{\mathbf{R}}^{-1}$ and $\hat{\mathbf{r}}_x$, respectively. Although when $N \gg M$ this classical approach provides good results, better estimators can be used instead when N has the same order of magnitude as M , but still $N > M$.

Step B. Compute the Number of Bits in Transmission $b(n)$: In order to determine the number of bits $b(n)$ to encode $x_s(n)$ without decoding error, one must guarantee that $|x_s(n) - y(n)| < 2^{b(n)-1} \Delta$. However, since the reading $x_s(n)$ is not yet available at the *fusion center*, we compute the number of bits to encode $x_s(n)$ in order to guarantee a given *Symbol Error Rate* threshold, SER_t .

Assuming $x_s(n) - y(n) \sim \mathcal{N}(0, \text{MSE}(\mathbf{w}))$, the SER can be expressed as

$$\text{SER} = \text{erfc} \left(\frac{2^{b(n)-1} \Delta}{\sqrt{2 \text{MSE}(\mathbf{w})}} \right). \quad (7)$$

We have focused on the particular case of Gaussian prediction errors. For a general case, other approaches can be used, as e.g., the Chebychev's inequality in [13].

Algorithm 2 Fusion Center**1. Training phase****for** $n = 1$ to N **do** **for** $s = 1$ to S **do** Request sth sensor for a l -bit reading (i.e., uncoded). **end for****end for**Estimate the correlation parameters for each sensor, i.e., \mathbf{R} and \mathbf{r}_x and compute $\hat{\mathbf{w}}$ and $\widehat{\text{MSE}}(\hat{\mathbf{w}})$ as in (19) and (29), respectively.**2. Coding phase****for** $n > N$ to *end do* **for** $s = 1$ to S **do** **Step A.** Compute *side-information* as $y(n) = \hat{\mathbf{w}}^H \mathbf{x}(n)$. **Step B.** Compute $b(n)$ following (8). Request sth sensor for a $b(n)$ -bit reading (i.e., encoded). Decode $x_s(n)$ using (2). **end for****end for.**Solving for $b(n)$ in (7) for a given SER_t , we get

$$b(n) \geq \left\lceil \log_2 \left(\frac{\sqrt{2\text{MSE}(\mathbf{w})}}{\Delta} \text{erfc}^{-1}(\text{SER}_t) \right) + 1 \right\rceil. \quad (8)$$

It should be iteratively repeated for every *sensing node* as it is represented in Algorithm 2.

In large WSNs, the number of already-decoded sensors S' (and hence M) is typically large. Therefore, maintaining a training phase such that $N \gg M$ may become inefficient in most cases. On the other hand, eq. (8) requires an accurate estimation of $\text{MSE}(\mathbf{w})$ in order to obtain the smallest $b(n)$ possible, while SER_t is guaranteed. Thus, our aim is to look for *enhanced estimators* for both \mathbf{w} and $\text{MSE}(\mathbf{w})$ that improve the *classical* estimators when N and M are large and comparable in magnitude.

IV. ENHANCED CORRELATION ESTIMATORS

First, let us consider a collection of N random observations of a certain M -dimensional stochastic process, denoted by $\mathbf{X}_N = [\mathbf{x}(1) \mathbf{x}(2) \dots \mathbf{x}(N)]$. We assume, without loss of generality, that these observations have zero mean $\mathbb{E}[\mathbf{x}(n)] = 0$, and $\mathbb{E}[|\mathbf{x}(n)|^2] = 1$, and covariance matrix \mathbf{R} .

The *Sample Covariance Matrix* (SCM), here denoted by $\hat{\mathbf{R}}$, is constructed from the observations as in [16]

$$\begin{aligned} \hat{\mathbf{R}} &= \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n) \mathbf{x}(n)^H \\ &= \frac{1}{N} \mathbf{X}_N \mathbf{X}_N^H = \frac{1}{N} \mathbf{R}^{1/2} \Xi^H \Xi \mathbf{R}^{1/2} \end{aligned} \quad (9)$$

where Ξ defines a $N \times M$ random matrix with i.i.d. complex entries, zero mean and unit variance. Moreover, let $\hat{\mathbf{r}}_x$ be the *sample cross-correlation vector* between the observation vector $\mathbf{x}(n)$ and the desired response $x_s(n)$, defined as

$$\hat{\mathbf{r}}_x = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n) x_s(n). \quad (10)$$

The classical estimator $\hat{\mathbf{w}}_{\text{class}}$ for the solution of the LWF (5) is given by

$$\hat{\mathbf{w}}_{\text{class}} = \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (11)$$

A. Enhanced Estimator for the LWF

It is well-known that the classical LWF estimator (11) is a N -consistent estimator of the LWF solution, i.e., $\|\hat{\mathbf{w}}_{\text{class}} - \mathbf{w}\| \rightarrow 0$, as $N \rightarrow \infty$.

In practice, $\hat{\mathbf{w}}_{\text{class}}$ provides good estimates when the training phase N is sufficiently large compared to the observation dimension M . However, when $M \rightarrow \infty$, while $M/N \rightarrow c \in (0, 1)$, it does not necessary provide N, M -consistency (indeed, [17] shows that (11) is not N, M -consistent), and better estimators can be derived. Mathematically

$$\|\hat{\mathbf{w}}_{\text{class}} - \mathbf{w}\| \not\rightarrow 0, \text{ as } N, M \rightarrow \infty; M/N \rightarrow c. \quad (12)$$

In practice, it may occur when the *training phase* is short and comparable in magnitude with the dimension of the *observation vector*.

In the literature of consistent estimation, structures of the type of (5) are usually addressed assuming that the vector \mathbf{r}_x is a non-random deterministic vector [18]. Thus, from the best of the author's knowledge, the estimation of (5) where both \mathbf{R} and \mathbf{r}_x are random and statistically *dependent* is still an open problem. However we have checked using numerical simulations that the results for the random case addressed here behaves similarly to what is expected for the case where \mathbf{r}_x is deterministic. The deterministic case is already solved in the RMT literature, e.g., [3]. Considering this, we can improve the *classical* estimator in (11) and propose an enhanced estimator for the LWF.

First, let the function $F_{\mathbf{R}}(x)$ be an instance of the *empirical distribution* of the *eigenvalues* (denoted as λ_m) and *eigenvectors* of \mathbf{R} (denoted as \mathbf{v}_m) as in [19]

$$F_{\mathbf{R}}(x) = \sum_{m=1}^M \mathbf{a}^H \mathbf{v}_m \mathbf{v}_m^H \mathbf{b} \mathcal{I}(\lambda_m \leq x) \quad (13)$$

whose *Stieltjes transform* is defined by (for both the continuous and the finite size cases)

$$\begin{aligned} m_{\mathbf{R}}(z) &= \int \frac{1}{\lambda - z} dF_{\mathbf{R}}(\lambda) \\ &= \sum_{m=1}^M \frac{\mathbf{a}^H \mathbf{v}_m \mathbf{v}_m^H \mathbf{b}}{\lambda_m - z} \\ &= \mathbf{a}^H (\mathbf{R} - z \mathbf{I}_M)^{-1} \mathbf{b}, \quad z \in \mathbb{C} \end{aligned} \quad (14)$$

where vectors \mathbf{a} and \mathbf{b} are two generic and deterministic vectors.

Hence, from the results in [17, Th. 1], we can write the Stieltjes transform of $\hat{\mathbf{R}}$ as

$$\begin{aligned} m_{\hat{\mathbf{R}}}(z) &\asymp \int \frac{dF_{\mathbf{R}}(\lambda)}{w(z)\lambda - z} = \frac{1}{M} \sum_{m=1}^M \frac{\mathbf{a}^H \mathbf{v}_m \mathbf{v}_m^H \mathbf{b}}{w(z)\lambda_m - z} \\ &= \mathbf{a}^H (w(z)\mathbf{R} - z\mathbf{I})^{-1} \mathbf{b} \end{aligned} \quad (15)$$

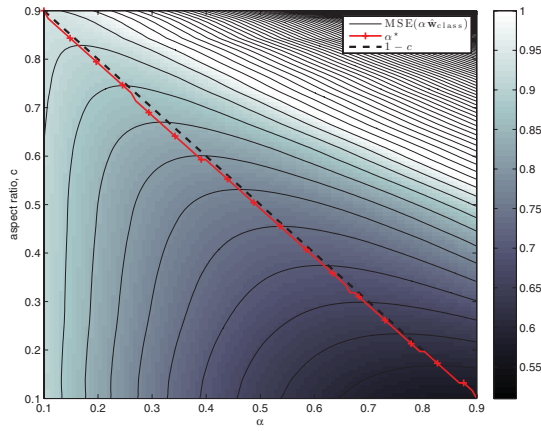


Fig. 3. MSE performance of the LWF estimator for different values of α is represented by the colormap and the contour lines. The optimal α obtained as in (21) (solid line with marker +) is also compared with the expression $1 - c$ (dashed line). The configuration is $M = 200$ and $c = [0.1, 0.9]$. This figure has been averaged over 100 realizations.

where $w(z) = 1 - c - czs_{\hat{\mathbf{R}}}(z)$, where $s_{\hat{\mathbf{R}}}(z)$ is defined in [17] as the unique solution to the following equation in the set $\{s_{\hat{\mathbf{R}}}(z) \in \mathbb{C} : -(1 - c)/z + cs_{\hat{\mathbf{R}}}(z) \in \mathbb{C}^+\}$

$$s_{\hat{\mathbf{R}}}(z) = \frac{1}{M} \sum_{m=1}^M \frac{1}{w(z)\lambda_m - z}. \quad (16)$$

Evaluating $m_{\hat{\mathbf{R}}}(z)$ for the case of $z = 0$ one can easily observe that

$$(1 - c) \mathbf{a}^H \hat{\mathbf{R}}^{-1} \mathbf{b} \asymp \mathbf{a}^H \mathbf{R}^{-1} \mathbf{b}. \quad (17)$$

In our case, the vector \mathbf{a} is selected as an all-zero vector with a one at the i th position (usually represented as \mathbf{e}_i) and $\mathbf{b} = \hat{\mathbf{r}}_x$, then

$$[\mathbf{w}^*]_i \asymp (1 - c) \mathbf{e}_i^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (18)$$

Hence, an enhanced estimator of the LWF solution is given by

$$\hat{\mathbf{w}} = (1 - c) \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (19)$$

The estimator in (19) can be seen as a scaled version of the classical LWF estimator as

$$\hat{\mathbf{w}} = \alpha^* \hat{\mathbf{w}}_{\text{class}} \quad (20)$$

where α is a scaling factor and α^* is its optimal value in terms of MSE and computed as

$$\alpha^* = \underset{\alpha}{\operatorname{argmin}} \{ \operatorname{MSE}(\alpha \hat{\mathbf{w}}_{\text{class}}) \}. \quad (21)$$

We test by simulation that the minimum MSE is obtained when α is actually $\alpha^* = (1 - c)$ (see Fig. 3).

The intuition behind the estimator in (19) can be seen as follows: The parameter $\alpha \in (0, 1)$ represents the confidence in the *classical* estimator. If $\hat{\mathbf{w}}_{\text{class}}$ has been estimated with a large number of samples in comparison with M , the degree of confidence will be high and $\hat{\mathbf{w}} \simeq \hat{\mathbf{w}}_{\text{class}}$ for $i = 1, \dots, M$. Otherwise, when $N > M$ but comparable in magnitude, $\hat{\mathbf{w}}_{\text{class}}$ is not expected to be the best weighting vector. In order to mitigate the performance reduction due to the missadjustment in $\hat{\mathbf{w}}_{\text{class}}$, the vector is attenuated.

B. Enhanced Estimator for the MSE

A traditional approach to estimate the MSE is by simply replacing the true correlations by their sample estimators. From (4), one can derive an estimator of the MSE given $\hat{\mathbf{w}}_{\text{class}}$ as

$$\widehat{\operatorname{MSE}}_{\text{class}}(\hat{\mathbf{w}}_{\text{class}}) = \hat{\sigma}_{x_s}^2 - \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (22)$$

When $\hat{\mathbf{w}}$ is given, the theoretical expression of the MSE is

$$\operatorname{MSE}(\hat{\mathbf{w}}) = \sigma_{x_s}^2 - 2 \operatorname{Re}[\hat{\mathbf{w}}^H \mathbf{r}_x] + \hat{\mathbf{w}}^H \mathbf{R} \hat{\mathbf{w}}. \quad (23)$$

Using the *classical* approach, one can estimate $\operatorname{MSE}(\hat{\mathbf{w}})$ as

$$\widehat{\operatorname{MSE}}_{\text{class}}(\hat{\mathbf{w}}) = \hat{\sigma}_{x_s}^2 - 2(1 - c) \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x + (1 - c)^2 \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x \quad (24)$$

where $\hat{\sigma}_{x_s}^2$ is the sample estimator of the signal variance $\sigma_{x_s}^2$, defined as

$$\hat{\sigma}_{x_s}^2 = \frac{1}{N} \sum_{n=1}^N x_s(n)^2. \quad (25)$$

The estimator $\widehat{\operatorname{MSE}}_{\text{class}}(\hat{\mathbf{w}})$ is proved to be N -consistent (one can directly check the case when $c \rightarrow 0$), but indeed it is not consistent when the observation dimension M increases without bound and at the same rate as N .

In order to overcome this problem, we proposed an enhanced estimator of (23). The first two terms of $\operatorname{MSE}(\hat{\mathbf{w}})$ are directly estimated by their sample estimators, i.e., $\hat{\sigma}_{x_s}^2 - 2 \operatorname{Re}[\hat{\mathbf{w}}^H \hat{\mathbf{r}}_x]$, since they do not involve unknown matrices in the estimation. Hence, the critical part resides in the estimation of the last term $(1 - c)^2 \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{R}} \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$ which is a function of the *true covariance matrix*. Hence we define the function

$$\beta(z) = \mathbf{h}_1^H (\hat{\mathbf{R}} - z\mathbf{I})^{-1} \mathbf{R} \mathbf{h}_2 \quad (26)$$

where $\mathbf{h}_1 = \hat{\mathbf{r}}_x$, and $\mathbf{h}_2 = \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$. Using the result in (15) we can rewrite $\beta(z)$ as

$$\beta(z) \asymp \mathbf{h}_1^H (w(z) \mathbf{R} - z\mathbf{I})^{-1} \mathbf{R} \mathbf{h}_2 \quad (27)$$

and evaluating $\beta(z)$ for $z = 0$, one can estimate $\beta(0)$ as

$$\hat{\beta} = (1 - c)^{-1} \mathbf{h}_1^H \mathbf{h}_2. \quad (28)$$

Once we have the enhanced estimator $\hat{\beta} = \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$ for the term $\beta(0) = \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{R}} \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$, we substitute each term of (21) for its estimate. Hence, an enhanced estimator of $\operatorname{MSE}(\hat{\mathbf{w}})$ is given by

$$\widehat{\operatorname{MSE}}(\hat{\mathbf{w}}) = \hat{\sigma}_{x_s}^2 - (1 - c) \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (29)$$

Note that the approach taken in this paper is slightly different to the MSE estimator in [16], where the authors give an N, M -consistent estimator for the optimal MMSE. Otherwise, in this paper we are interested in estimating the practical MSE obtained by using a certain weighting vector (in our case $\hat{\mathbf{w}}$ in (19)), which not necessarily provides the MMSE lower bound.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
Number of fusion centers:	$F = 1$
Already-decoded sensing nodes:	$S' = 200$
Number of past samples:	$K = 200$, hence $M = 400$
Length of the training phase:	$N = 1000$
Aspect ratio (M/N):	$c = 0.4$
Correlation model, $[\mathbf{R}_s]_{i,i+k} = [\mathbf{R}_r]_{i,i+k} = \rho^{ k }$:	$\rho = 0.9$
SER threshold:	$\text{SER}_t = 10^{-2}$
A/D converter depth:	$l = 12$ bits

V. NUMERICAL RESULTS

The simulated scenario is composed of 200 sensing nodes and one fusion center configured in star topology. Their measurements are assumed to be space-time correlated following the correlation model $[\mathbf{R}_s]_{i,i+k} = [\mathbf{R}_r]_{i,i+k} = \rho^{|k|}$, where $\rho = 0.9$. Although any real application measurement will be corrupted by at least a small amount of noise, we have considered noise-free communication paths in order to better evaluate the system performance.

In particular, we study the behavior of the following figures as a function of c and how they affect to the SER performance.

- 1) $\text{MSE}(\mathbf{w})$: the MSE for a given filter \mathbf{w} in (4).
- 2) $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}}_{\text{class}})$: the classical MSE estimator in (22).
- 3) $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}})$: the classical MSE estimator in (24).
- 4) $\widehat{\text{MSE}}(\hat{\mathbf{w}})$: the proposed MSE estimator in (29).

The 2–4) have been computed using the corresponding mathematical expressions. On the other hand, the $\text{MSE}(\mathbf{w})$ has been computed experimentally. We have developed a WSN simulation environment in Matlab and we have implemented in it our proposed DSC algorithm.

Table I summarizes the parameters that configure the basic setup of the simulation environment.

A. Performance of the Proposed LWF Estimator, $\hat{\mathbf{w}}$

In this subsection, we evaluate the MSE performance obtained by simulation of the *proposed* LWF estimator involved in the *side-information* $y(n)$.

Following the approach exposed in Subsection IV-A, Fig. 3 draws the simulation results for the MSE obtained with the LWF as a function of the parameter α and for different configurations of M and N . From this simulation experiment, we can compute the optimal α as in (21), which is represented in Fig. 3 as solid line with markers $+$. Moreover, we show that α^* actually fits with the theoretical limit $\alpha^* = 1 - c$ (dashed line), as predicted in Section IV. However, there is still a gap between the MSE obtained with the *proposed* method and the one obtained assuming full correlation knowledge of \mathbf{R} and \mathbf{r}_x .

Fig. 4 compares the performance in terms of MSE of our *proposed* estimator with some of the most popular estimation techniques, i.e., the *classical sample estimator* (11), the sample estimator with *diagonal loading* (DL), and three instances of the *Principal Component Analysis* (PCA) method.

1) *Classical Sample Estimator*: The behavior is clear; for low values of c^{-1} , the *proposed* LWF estimator outperforms the classical method. On the other hand, when we let c^{-1} increase, both estimators perform similarly.

2) *DL Estimator*: Namely

$$\hat{\mathbf{w}}_{\text{DL}} = (\hat{\mathbf{R}} + \gamma \mathbf{I})^{-1} \hat{\mathbf{r}}_x. \quad (30)$$

Although for low values of c^{-1} DL presents lower MSE, our *proposed* method shows two important advantages; 1) DL is not consistent when $c^{-1} \rightarrow \infty$ and 2) the optimum loading factor γ^* that minimizes the MSE may vary according to the scenario, and in the literature there is not a clear expression to obtain γ^* analytically but only iteratively or by simulation. We use $\gamma = 0.8$, which gives the minimum MSE for $c = 0.4$.

3) *PCA Estimator*: Keeping only the $M' < M$ largest eigenvalues of $\hat{\mathbf{R}}$ (because the smallest are more difficult to be estimated and hence they may introduce higher errors), the MSE can be improved [20]. Therefore, $\hat{\mathbf{R}}_{\text{PCA}}$ is a lower rank projection onto the subspace generated by the M' larger eigenvalues of $\hat{\mathbf{R}}$. Thus

$$\hat{\mathbf{w}}_{\text{PCA}} = \hat{\mathbf{R}}_{\text{PCA}}^{-1} \hat{\mathbf{r}}_x. \quad (31)$$

PCA presents the same limitations as DL but with the difference that this trade-off is balanced changing M' . In addition, when $M, N \rightarrow \infty$, the eigendecomposition may become hard to handle by practical small sensors.

In order to analyze the impact of the *proposed estimator* $\hat{\mathbf{w}}$ on the system performance, we compare in Fig. 5 the experimental SER for each of the LWF estimators as a function of the compression level $b(n)/l$.

One can observe that for large values of $b(n)/l$, e.g., $b(n)/l = 0.75$, the SER obtained is eight times smaller for $\hat{\mathbf{w}}$ than for $\hat{\mathbf{w}}_{\text{class}}$. Even so, one may make the following argument: If we want to achieve a certain SER_t (e.g. $\text{SER}_t = 10^{-2}$), we can compress up to 0.75 using the *classical* estimator, and 0.72 using the *proposed*. At first glance, it seems that the gain is quite moderate. However, we show next in Example 1 that it has an important impact on the total system performance when both the *proposed* estimators are combined.

B. Performance of the Proposed MSE Estimator, $\widehat{\text{MSE}}(\hat{\mathbf{w}})$

The MSE is involved in the computation of $b(n)$ in (8). Hence, a good estimation of the MSE is required in order to not overestimate (getting a too conservative result) or underestimate (inducing potential errors) the parameter $b(n)$, and thus maintain the system requirements, such as the SER_t .

Fig. 6 plots the MSE curve obtained experimentally (solid line). It is compared with our *proposed* MSE estimator (29) and the *classical* approach of (24). It is easy to see that our *proposed* estimator fits considerably better with the experimental results, while the *classical* estimator is clearly underestimating, especially for low values of c^{-1} . In fact, the *classical* approach is underestimating the MSE. Following (8), the DSC algorithm will be stingy with the number of bits used, and almost certainly, the SER requirements will not be achieved.

From a user point of view, the experimental SER curves of Fig. 5 are not available *a priori*, so the user should use a

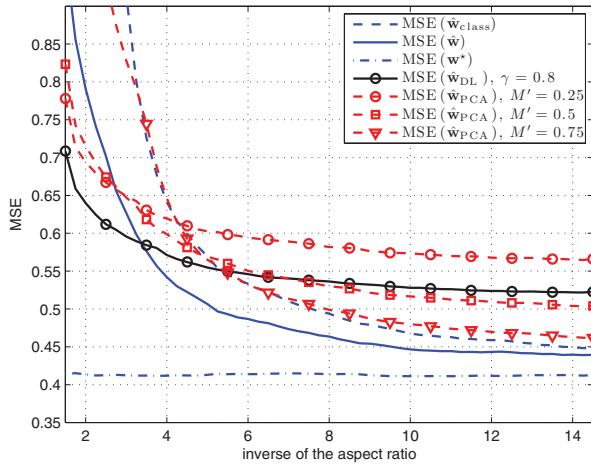


Fig. 4. MSE performance of the *classical*, DL, PCA, and the *proposed* estimators as a function of the inverse of the aspect ratio c^{-1} .

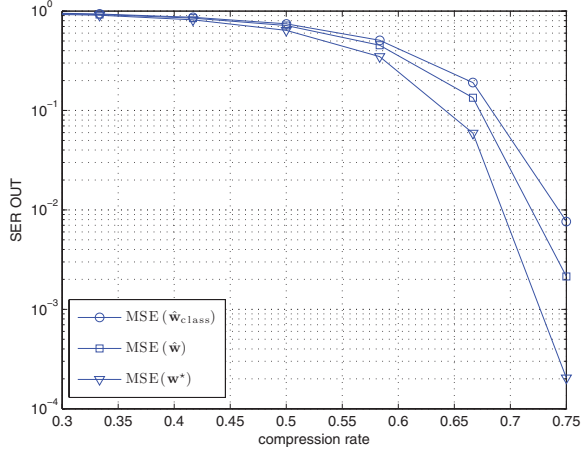


Fig. 5. Performance of the experimental SER using the *classical* $\hat{\mathbf{w}}_{\text{class}}$ and the *proposed* $\hat{\mathbf{w}}$ estimators as a function of the compression rate for $c = 0.4$.

predicted version of the SER instead to determine which is the maximum compression rate that one can apply in order to guarantee a given SER_t .

In Fig. 7 we compare the experimental SER with the following:

- 1) Predicted SER when the *classical* MSE estimator is used and $\hat{\mathbf{w}}$ is given, i.e., $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}})$.
- 2) Predicted SER when the *proposed* MSE estimator is used and $\hat{\mathbf{w}}$ is given, i.e., $\widehat{\text{MSE}}(\hat{\mathbf{w}})$.
- 3) Predicted SER when the *classical* MSE estimator is used and $\hat{\mathbf{w}}_{\text{class}}$ is given, i.e., $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}}_{\text{class}})$.

They are calculated using the formula (7) replacing the $\text{MSE}(\mathbf{w})$ of the denominator by their respective estimators.

In Fig. 7 we observe that the *proposed* estimator curve fits the best with the experimental SER (which is also shown in Fig. 5). However, all the estimators are indeed underestimating. The consequences of this fact are illustrated in the following example.

Example 1: Let us take as a system requirement $\text{SER}_t = 10^{-2}$. Observing the predicted SER curves in Fig. 7, we may decide to compress our messages with a ratio of 0.66 if we

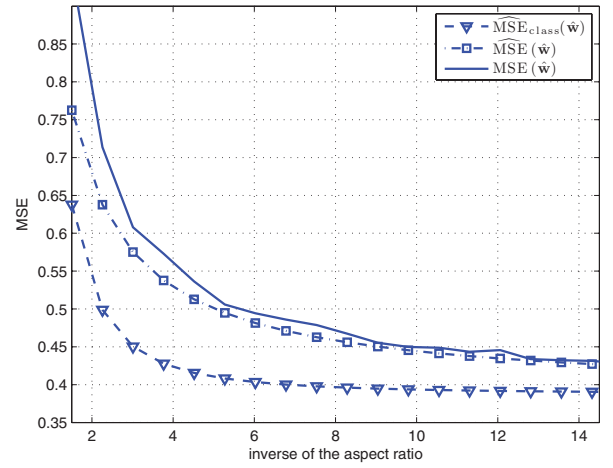


Fig. 6. Performance of the *classical* $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}})$ and the *proposed* $\widehat{\text{MSE}}(\hat{\mathbf{w}})$ estimators compared to the *experimental* reference $\text{MSE}(\hat{\mathbf{w}})$ (solid line) as a function of the inverse of the aspect ratio c^{-1} .

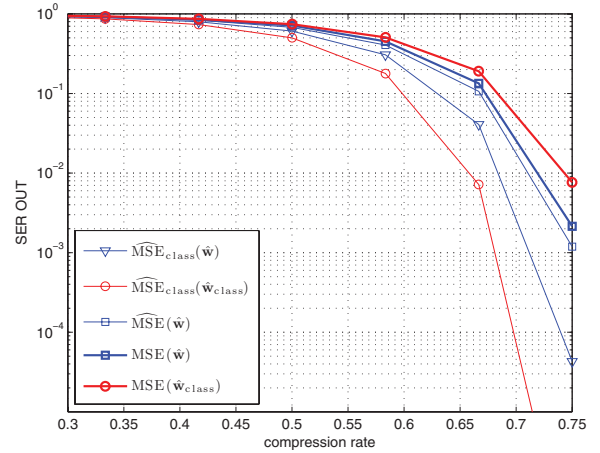


Fig. 7. Performance of the predicted SER using different MSE estimators. It is compared to the experimental SER for $c = 0.4$.

are using $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}}_{\text{class}})$, 0.68 if $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}})$, and 0.71 if $\widehat{\text{MSE}}(\hat{\mathbf{w}})$. Now, we map these three points to their respective *experimental* curves, i.e., $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}}_{\text{class}})$ to $\text{MSE}(\hat{\mathbf{w}}_{\text{class}})$, and $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}})$ and $\widehat{\text{MSE}}(\hat{\mathbf{w}})$ to $\text{MSE}(\hat{\mathbf{w}})$. The real output SER of the system would be $2 \cdot 10^{-1}$ and about 10^{-1} for the first and second option respectively, which is one order of magnitude larger than the expected SER. On the other hand, we get $1.2 \cdot 10^{-2}$ (instead of 10^{-2} , so it is still slightly underestimated), obtaining a more accurate solution.

Hence, using the *proposed* estimators the gain in front of the *classical* methods is twofold; on the one hand we can obtain higher compression (thus higher energy savings) since $\text{MSE}(\hat{\mathbf{w}}) < \text{MSE}(\hat{\mathbf{w}}_{\text{class}})$, and on the other hand the *proposed* estimators adjust substantially better to the system requirements than the *classical* estimators do. Moreover, the more stringent the SER_t the higher the gain.

VI. CONCLUSION

This paper has proposed two enhanced correlation estimators for the *Linear Wiener Filter* and the *Mean Square Error* to operate when the number of snapshots N and the observation

dimension M are large and comparable in magnitude, or equivalently for short training phases. This scenario is very suitable for large WSNs due to its large number of sensors. Concretely, the enhanced estimators have been designed to carry out the two key steps in a *Distributed Source Coding* algorithm, i.e., the computation of the *side-information* $y(n)$ based on the existing space-time correlations, and the computation of the minimum number of bits to encode the readings in order to guarantee a certain *Symbol Error Rate*. Numerical results show that our proposed estimators perform far better for values of the aspect ratio M/N close to one. Furthermore, they perform as the corresponding sample estimators when $M/N \rightarrow 0$ (i.e., for very long training phases). In practice, it allows us to reduce the number of transmitted bits (and hence reduce the energy consumption) at the same time that it allows us to decrease largely the training phase in *Distributed Source Coding* schemes.

REFERENCES

- [1] S. Haykin, *Adaptive Filter Theory*, 4th ed. Englewood Cliffs, NJ: Prentice-Hall, Sep. 2001.
- [2] V. L. Girko, "G₂₅-estimators of principal components," *Theory Probab. Math. Stat.*, vol. 40, pp. 1–10, 1990.
- [3] V. L. Girko, *An Introduction to Statistical Analysis of Random Arrays*. Harderwijk, The Netherlands: VSP, 1998.
- [4] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, Jun. 2004.
- [5] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 51–60, Mar. 2002.
- [6] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," *IEEE Trans. Inf. Theory*, vol. 49, no. 3, pp. 626–643, Mar. 2003.
- [7] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [8] F. Oldewurtel, M. Foks, and P. Mahonen, "On a practical distributed source coding scheme for wireless sensor networks," in *Proc. IEEE Veh. Technol. Conf.*, May 2008, pp. 228–232.
- [9] F. Oldewurtel, J. Ansari, and P. Mahonen, "Cross-layer design for distributed source coding in wireless sensor networks," in *Proc. 2nd Int. Sensor Technol. Appl.*, Aug. 2008, pp. 435–443.
- [10] Z. Tang, I. A. Glover, A. N. Evans, and J. He, "Energy efficient transmission protocol for distributed source coding in sensor networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2007, pp. 3870–3875.
- [11] H. Wang, D. Peng, W. Wang, H. Sharif, and H. H. Chen, "Cross-layer routing optimization in multirate wireless sensor networks for distributed source coding based applications," *IEEE Trans. Wireless Commun.*, vol. 7, no. 10, pp. 3999–4009, Oct. 2008.
- [12] A. Saxena and K. Rose, "On scalable distributed coding of correlated sources," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2875–2883, May 2010.
- [13] J. Chou, D. Petrovic, and K. Ramchandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *Proc. IEEE INFOCOM*, vol. 2, Mar. 2003, pp. 1054–1062.
- [14] J. E. Barcelo-Llado, A. Morell, and G. Seco-Granados, "Distributed source coding in large wireless sensor networks," in *Proc. 44th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1–5.
- [15] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
- [16] F. Rubio and X. Mestre, "Consistent reduced-rank LMMSE estimation with a limited number of samples per observation dimension," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2889–2902, Aug. 2009.
- [17] X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5353–5368, Nov. 2008.
- [18] X. Mestre and M. A. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 69–82, Jan. 2006.
- [19] M. Lagunas and X. Mestre, "Diagonal loading for finite sample size beamforming: An asymptotic approach," in *Robust Adaptive Beamforming*, J. Li and P. Stoica, Eds. New York: Wiley, 2005.
- [20] I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer-Verlag, 2002.

Joan Enric Barceló-Lladó (S'09) received the M.Sc. degree in electrical engineering from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain in 2007. He is currently pursuing the Ph.D. degree at the Signal Processing for Communications and Navigation Group, Department of Telecommunication and Systems Engineering, UAB.

He was a Research Assistant with the Università degli Studi di Siena, Siena, Italy, in 2007, where he was involved in the topics of cross-layer techniques for satellite networks. From September 2008 to March 2009, he held a visiting position with UNIK-University Graduate Center, Oslo, Norway, where he was involved in resource allocation for satellite communications based on game theory. He has been involved in several national and international projects with public and private funding. His current research interests include algorithms for energy-efficient communication with correlated sources in wireless sensor networks.

Antoni Morell Pérez received the M.Sc. degree in electrical engineering and the Ph.D. degree (*cum laude*) from the Universitat Politècnica de Catalunya (UPC), Barcelona, in 2002 and 2008, respectively.

He was with the Signal Theory and Communications Department, UPC, from September 2002 to September 2005, where he was involved in several research and development projects, including the European projects MEDEA+ MARQUIS and PLANETS. He was with the Universitat Autònoma de Barcelona, first as a Research and Teaching Assistant in September 2005 and as an Assistant Professor teaching courses in communications in September 2008. He has been involved in several research and development projects, such as IST FP6 SATNEX, ESA DINGPOS, CENIT ATLANTIDA, CENIT SINTONIA, INFOREGIO XALOC, and RECERCAIXA 3SENS. He has expertise in optimization techniques applied to communications, also in the field of wireless sensor networks. He has published more than 25 papers in recognized international journals and conference proceedings. His current research interests include cooperative communication, radio resource allocation, distributed multiple access protocols, distributed source coding, and compressed sensing.

Prof. Pérez received a grant from the Spanish government (FPU) in January 2003 for his Ph.D. studies, and was with the Centre Tecnològic de Telecomunicacions de Catalunya for one year.

Gonzalo Seco-Granados (S'97–M'02–SM'08) received the Ph.D. degree in telecommunication engineering from the Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, in 2000, and the M.B.A. degree from IESE, University of Navarra, Barcelona, in 2002.

He was the Technical Staff Member with the RF Payload Division, European Space Research and Technology Center (ESTEC), European Space Agency, Noordwijk, The Netherlands, from 2002 to 2005, where he was involved in the Galileo Project. He led the activities concerning navigation receivers and indoor positioning for GPS and Galileo. Since 2006, he has been an Associate Professor with the Department of Telecommunications and Systems Engineering, Universitat Autònoma de Barcelona (UAB), Barcelona. Since May 2011, he has been the Vice-Director of the Engineering School, UAB, and was the Coordinator of the Telecommunications Engineering degree program from 2007 to 2011. He leads the Research Group on Signal Processing for Communications and Navigation (SPCOMNAV). He has been a Principal Investigator of more than 12 national, European, and international research projects. He has held visiting appointments at Brigham Young University and at the University of California, Irvine. His current research interests include signal processing for wireless communication and navigation, array signal processing, location-based communication, synchronization, estimation theory, and multicarrier signal design.

Dr. Seco-Granados was appointed as the Director of one of the six Chairs of Technology Knowledge Transfer in March 2008, granted by UAB Reserach Park and supported by Banco Santander in recognition of the research activity of the group.