

# Distributed Source Coding in Large Wireless Sensor Networks

J.E. Barceló-Lladó, Antoni Morell, Gonzalo Seco-Granados  
Dpt. Telecom. and Syst. Engineering  
Universitat Autònoma de Barcelona, Bellaterra 08193 - Barcelona, Spain  
{joanenric.barcelo,antoni.morell,gonzalo.seco}@uab.es

**Abstract**—We study the performance of *distributed source coding* in large wireless sensor networks obtained with enhanced correlation estimators. *Distributed source coding* is especially useful when data correlation exists since it tries to remove the redundancy in the information; and dense sensor networks are rich in correlations. Existing results from information theory show that this compression can be executed in a distributed fashion and without any performance loss in comparison with the centralized approach. However, there is still performance gap between the theoretical bounds and the results achieved with practical implementations. In order to mitigate this, we propose the use of enhanced correlation estimators. Simulation results show a performance improvement in the energy consumption by reducing the number of transmitted bits compared to classical methods.

*Topics*— E-6, C-2.

## I. INTRODUCTION

Currently, *Wireless Sensor Networks* (WSNs) have become a hot research topic with a promising impact. In this field, novel results have been derived in many areas, such as signal processing and information theory. In contrast with the devices of other networks (e.g. laptops, cellular phones, PDAs), WSNs consist of many tiny, low-powered and un-expensive wireless nodes. Hence, they are strongly energy-limited, and in most cases, their battery cannot be recharged.

Following with this motivation, one can find many strategies to mitigate the effect of energy-limited networks by applying energy-saving mechanisms. Without the aim of being exhaustive, we point out some examples; *i*) energy-aware routing for cooperative wireless sensor networks and ad-hoc networks [1], *ii*) signal processing techniques for minimum-power distributed beamforming [2], and *iii*) *data-aware* techniques to reduce energy by using efficient information processing [3]. Our study falls in the third category and may be complementary to the other approaches. In this paper, we propose to study the energy savings that we can obtain using *Distributed Source Coding* (DSC) mechanisms. DSC removes the inherent redundancy in correlated readings [3], [4]. Typically, dense WSNs are formed by a large number of high space-time correlated sources (e.g. fire control in forests or monitoring the location of the products in large stores). Surprisingly, existing

The work has been supported by the Spanish Government under projects TEC2008-06305 and CENIT-SINTONIA, the Catalan Government under grant 2009 SGR 298, and the Chair of Knowledge and Technology Transfer Parc de Recerca UAB - Santander.

results from information theory (precisely, from the work of Slepian and Wolf [5]) 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 carry out the data compression without the knowledge of the other signals, 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 [4].

One of the causes is the inexact estimation of the correlation parameters. In large WSNs this effect is aggravated because the estimated parameters tend to be high-dimensional, and classical methods may not be feasible in the sense that they may need very long training phases. Instead, we propose enhanced estimators based on *Random Matrix Theory* (RMT) [6] to reduce the number of snapshots required to guarantee a certain level of estimation accuracy.

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 provided in Section IV. Simulation results are given in Section V, and conclusions are drawn in Section VI.

## II. SYSTEM MODEL

We assume a large WSN configured in a star topology and composed of two types of nodes: *i*) *S sensing nodes* that transmit the environmental measures when requested, and *ii*) *one fusion center* that receives and processes the collected data from the *sensing nodes*.

Let  $x_s(n) \in \mathbb{R}$  denote the scalar reading from the  $s$ th sensor and  $n$  is the discrete time index. Matrix  $\mathbf{X} \in \mathbb{R}^{S \times T}$  assembles the time sequence of measurements for every *sensing node* (where  $T$  is the duration of the observation window). We assume that the spatial observations present a column covariance matrix denoted by  $\mathbf{R}_s$ , and  $\mathbf{R}_t$  is the row (temporal) covariance matrix.

**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.,  $\mathbf{X}^{1/2}\mathbf{X}^{1/2} = \mathbf{X}$ .  $\lceil \cdot \rceil$  denotes the ceil function.  $\mathbb{E}[\cdot]$  is the statistical expectation.  $\text{erfc}(\cdot)$  represent the complementary error function.  $\mathcal{N}(\boldsymbol{\mu}, \mathbf{R})$  is a Gaussian vector distribution with mean  $\boldsymbol{\mu}$  and covariance matrix  $\mathbf{R}$ ,  $\sigma_x^2$  is the variance of  $\mathbf{x}$ . Symbol  $\asymp$  means almost sure convergence.

### III. DISTRIBUTED SOURCE CODING ALGORITHM

Although there are no practical techniques to achieve the theoretical bounds of [5], suboptimal techniques exist. In this paper, we will follow the approach in [7]. The authors proposed the construction of a codebook based on the decomposition of a given finite alphabet  $\mathcal{A}$  in several sub-codebooks.

The transmission is divided in two phases; first, during a *training phase* of length  $N$  snapshots, the *sensing node* uses a uniform  $l$ -bit A/D converter to encode its reading as  $x_s(n) \in \mathcal{A}$  (and thus  $\mathcal{A} = \{a_i\}_{i=1,2,\dots,2^l}$ , where  $|a_{i+1} - a_i| = \Delta$ ). After collecting  $N$  snapshots, the *fusion center* estimates the correlation parameters. Secondly, in the *coding phase*, a given *side-information*  $y(n)$  is available at the *fusion center* and the *sensing node* can encode its reading using only  $b(n) < l$  bits. Hence, what the sensor transmits is 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*  $y(n)$ ,

$$x_s(n) = \arg \min_{a_i \in \mathcal{A}_B} |y(n) - a_i|. \quad (1)$$

In the Algorithm 1, the following two key steps are highlighted:

Step 1. *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 node*. This vector collects both the  $K$  past readings of the sensor, and the readings of the set  $S'$  of already-decoded sensors in time slot  $n$  (where  $S' \subset S$  with cardinality  $|S'|$ ). Hence  $M = K + |S'|$ . Then, the *side-information*  $y(n)$  is computed as a linear combination of the entries of  $\mathbf{x}(n)$ , i.e.,  $y(n) = \mathbf{w}^H \mathbf{x}(n)$ , following the *Linear Wiener Filter* (LWF) solution. The LWF solution  $\mathbf{w}^*$  is known to be optimal in the *Mean Square Error* (MSE) sense. Mathematically,

$$\begin{aligned} \text{MSE}(\mathbf{w}) &= \sigma_{x_s(n)}^2 - 2 \text{Re}[\mathbf{w}^H \mathbf{r}_x] + \mathbf{w}^H \mathbf{R} \mathbf{w}. \\ \frac{\partial \text{MSE}(\mathbf{w})}{\partial \mathbf{w}^H} &= -\mathbf{r}_x + \mathbf{w}^H \mathbf{R} = 0; \\ \mathbf{w}^* &= \mathbf{R}^{-1} \mathbf{r}_x, \end{aligned} \quad (2)$$

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

$$\text{MSE}(\mathbf{w}^*) = \sigma_{x_s(n)}^2 - \mathbf{r}_x^H \mathbf{R}^{-1} \mathbf{r}_x. \quad (3)$$

However, to compute  $\mathbf{w}^*$  the knowledge of  $\mathbf{R}^{-1}$  and  $\mathbf{r}_x$  is necessary but not available. Classical methods replace  $\mathbf{R}^{-1}$  and  $\mathbf{r}_x$  by the 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  $N > M$ .

---

#### Algorithm 1 Fusion node

---

##### 1. Training phase

**for**  $n = 1$  to  $N$  **do**

**for**  $s = 1$  to  $S$  **do**

        Request sensor  $s$  for a  $l$ -bit reading (i.e., uncoded).

**end for**

**end for**

Estimate the correlation parameters, i.e.,  $\mathbf{R}$  and  $\mathbf{r}_x$  as in (15) and (28), respectively.

##### 2. Coding phase

**for**  $n > N$  to *end do*

**for**  $s = 1$  to  $S$  **do**

**Step 1.** Compute *side-information* as  $y(n) = \hat{\mathbf{w}}^H \mathbf{x}(n)$ .

**Step 2.** Compute  $b(n)$  following (5).

        Request sensor  $s$  for a  $b(n)$ -bit reading (i.e., encoded).

        Decode  $x_s(n)$  using (1).

**end for**

**end for.**

---

Step 2. *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  $\text{SER}_t$ .

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

$$\text{SER} = \text{erfc} \left( \frac{2^{b(n)-1} \Delta}{\sqrt{2 \text{MSE}(\mathbf{w})}} \right). \quad (4)$$

Although we are assuming gaussianity of the estimation error, more general approaches can be used, as e.g., the Chebychev's inequality in [7].

Solving for  $b(n)$  in (4) for a given  $\text{SER}_t$ , we get

$$b(n) \geq \left\lceil \log_2 \left( \frac{\sqrt{2 \text{MSE}(\mathbf{w})}}{\Delta} \text{erfc}^{-1}(\text{SER}_t) \right) + 1 \right\rceil. \quad (5)$$

Thus equation (5) requires an accurate estimation of  $\text{MSE}(\mathbf{w})$  to obtain the smallest  $b(n)$  that guarantees  $\text{SER}_t$ .

Either the LWF estimator or the MSE estimator should perform the best possible when  $M$  is large. It comes from the assumption that in large WSNs where the number of already-decoded sensors  $S'$  can be very large and maintain a training phase such that  $N \gg M$  could not be efficient.

### IV. ENHANCED CORRELATION ESTIMATORS

#### A. RMT Concepts and Definitions

The following definitions can be found in [8]. In this section, let  $\mathbf{A}$  denote a generic positive semidefinite  $M \times M$  matrix.

*Definition 1:* Let the function  $F_{\mathbf{A}} : \mathbb{R} \rightarrow [0, 1]$  be the *empirical spectral distribution* of the *eigenvalues* of  $\mathbf{A}$ , here

denoted as  $\lambda_m$ :

$$F_{\mathbf{A}}(x) = \frac{1}{M} \sum_{m=1}^M \mathcal{I}(\lambda_m \leq x), \quad (6)$$

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

$$s_{\mathbf{A}}(z) = \int \frac{1}{\lambda - z} dF_{\mathbf{A}}(\lambda) = \frac{1}{M} \sum_{m=1}^M \frac{1}{\lambda_m - z}. \quad (7)$$

*Definition 2:* Let the function  $H_{\mathbf{A}} : \mathbb{R} \rightarrow [0, 1]$  be an instance of the *empirical distribution* of the *eigenvalues* and *eigenvectors* of  $\mathbf{A}$ , denoted as  $\boldsymbol{\nu}_m$ :

$$H_{\mathbf{A}}(x) = \sum_{m=1}^M \mathbf{a}^H \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H \mathbf{b} \mathcal{I}(\lambda_m \leq x), \quad (8)$$

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

$$\begin{aligned} m_{\mathbf{A}}(z) &= \int \frac{1}{\lambda - z} dH_{\mathbf{A}}(\lambda) = \sum_{m=1}^M \frac{\mathbf{a}^H \boldsymbol{\nu}_m \boldsymbol{\nu}_m^H \mathbf{b}}{\lambda_m - z} \\ &= \mathbf{a}^H (\mathbf{A} - z \mathbf{I}_M)^{-1} \mathbf{b}, \quad z \in \mathbb{C}. \end{aligned} \quad (9)$$

*Definition 3:* Let  $\hat{f}(\mathbf{x})$  be a  $N$ -consistent estimator of  $f(\mathbf{x})$ , then

$$|\hat{f}(\mathbf{x}) - f(\mathbf{x})| \rightarrow 0, \text{ as } N \rightarrow \infty. \quad (10)$$

*Definition 4:* Let  $\hat{g}(\mathbf{x})$  be a  $N, M$ -consistent estimator of  $g(\mathbf{x})$ , then

$$|\hat{g}(\mathbf{x}) - g(\mathbf{x})| \rightarrow 0, \text{ as } N, M \rightarrow \infty; M/N \rightarrow c. \quad (11)$$

### B. $N, M$ -Consistent Solution for the Linear Wiener Filter

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  $\mathbb{E}[|\mathbf{x}(n)|^2] = 1$ , with covariance matrix  $\mathbf{R}$ .

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

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n) \mathbf{x}(n)^H = \frac{1}{N} \mathbf{X}_N \mathbf{X}_N^H = \frac{1}{N} \mathbf{R}^{1/2} \boldsymbol{\Xi}^H \boldsymbol{\Xi} \mathbf{R}^{1/2}, \quad (12)$$

where  $\boldsymbol{\Xi}$  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

$$\hat{\mathbf{r}}_x = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n) x_s(n). \quad (13)$$

The *minimum* MSE (MMSE) solution for the LWF satisfies the Wiener-Hopf equations, i.e.,  $\mathbf{R} \mathbf{w} = \mathbf{r}_x$ , and, consequently, the optimal weighting vector (in the MMSE sense) is (2). Then, the classical estimator  $\hat{\mathbf{w}}_{\text{class}}$  for the solution of the LWF (2) is given by

$$\hat{\mathbf{w}}_{\text{class}} = \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (14)$$

It is well-known that the classical estimator in (14) is a  $N$ -consistent estimator of the LWF solution. In practice, it provides good estimates when the training phase  $N$  is sufficiently high compared to the observation dimension  $M$ . However, when  $M \rightarrow \infty$ , while  $M/N \rightarrow c$ , it does not necessarily provide  $N, M$ -consistency (e.g., [9] shows that (14) is not  $N, M$ -consistent), and better estimators can be derived.

*Theorem 1:* An  $N, M$ -consistent estimator of the LWF solution is given by

$$\hat{\mathbf{w}} = (1 - c) \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (15)$$

*Proof:* We focus on the estimation of scalar functionals of the inverse of  $\hat{\mathbf{R}}$  i.e.,  $\varphi(\hat{\mathbf{R}}^{-1}) : \mathbb{R}^{M \times M} \rightarrow \mathbb{R}$  of the type:

$$\varphi(\hat{\mathbf{R}}^{-1}) = \mathbf{a}^H \mathbf{R}^{-1} \mathbf{b}. \quad (16)$$

Furthermore, for this proof, we make use of the *Marčenko-Pastur Theorem* [10, Theorem 1], for matrices of the form  $\boldsymbol{\Phi} = \boldsymbol{\Upsilon} + \frac{1}{N} \boldsymbol{\Xi} \mathbf{R} \boldsymbol{\Xi}^H$ , where:

- $\boldsymbol{\Upsilon}$  is an arbitrary Hermitian  $N \times N$  matrix.
- $\boldsymbol{\Xi}$  is an  $N \times M$  matrix such that its entries are iid complex random variables with zero mean and variance 1, i.e.  $[\boldsymbol{\Xi}]_{i,j} \in \mathbb{C}$ ,  $\mathbb{E}[\boldsymbol{\Xi}_{i,j}] = 0$  and  $\mathbb{E}[|\boldsymbol{\Xi}_{i,j}|^2] = 1$ .
- $\mathbf{R}$  is the true covariance matrix, and the empirical distribution function of its eigenvalues  $\{\lambda_1, \lambda_2, \dots, \lambda_M\}$  converges almost surely in distribution to a nonrandom cumulative distribution function  $F_{\mathbf{R}}(\lambda)$  as  $N \rightarrow \infty$ .

Then,

$$s_{\boldsymbol{\Phi}}(z) = s_{\boldsymbol{\Upsilon}} \left( z - c \int \frac{\lambda dF_{\mathbf{R}}(\lambda)}{1 + \lambda s_{\boldsymbol{\Phi}}(z)} \right). \quad (17)$$

Furthermore, we assume

$$\boldsymbol{\Upsilon} = \mathbf{0}_N, \text{ and } \boldsymbol{\Phi} = \frac{1}{N} \boldsymbol{\Xi} \mathbf{R} \boldsymbol{\Xi}^H = \frac{1}{N} \boldsymbol{\Xi} \mathbf{R}^{1/2} \mathbf{R}^{1/2} \boldsymbol{\Xi}^H. \quad (18)$$

Hence, the *Stieltjes transform* of  $\boldsymbol{\Upsilon}$  is given by

$$s_{\boldsymbol{\Upsilon}}(z) = \frac{1}{0 - z} = -z^{-1}. \quad (19)$$

Using (17), we get the equation

$$s_{\boldsymbol{\Phi}}(z) = - \left( z - c \int \frac{\lambda dF_{\mathbf{R}}(\lambda)}{1 + \lambda s_{\boldsymbol{\Phi}}(z)} \right)^{-1}. \quad (20)$$

Typically, to make  $\boldsymbol{\Phi}$  define an arbitrary SCM  $\hat{\mathbf{R}}$ , it should have dimension  $M \times M$  rather than  $N \times N$ . So, we introduce the  $M \times M$  SCM as  $\hat{\mathbf{R}} = \frac{1}{N} \mathbf{R}^{1/2} \boldsymbol{\Xi}^H \boldsymbol{\Xi} \mathbf{R}^{1/2}$ . Matrix  $\hat{\mathbf{R}}$  has the same structure than in (12).

Note that the non-zero eigenvalues of  $\boldsymbol{\Phi}$  and  $\hat{\mathbf{R}}$  are the same, however  $\boldsymbol{\Phi}$  has  $N - M$  zero eigenvalues extra. So, we can relate the eigenvalue distributions (and hence their Stieltjes transforms) for both  $\boldsymbol{\Phi}$  and  $\hat{\mathbf{R}}$  as follows,

$$\begin{aligned} \frac{dF_{\boldsymbol{\Phi}}}{d\lambda} &= \frac{M}{N} \frac{dF_{\hat{\mathbf{R}}}}{d\lambda} + \frac{(N - M)}{N} \delta(\lambda), \\ F_{\boldsymbol{\Phi}} &= \frac{M}{N} F_{\hat{\mathbf{R}}} + \frac{(N - M)}{N} u(\lambda), \\ s_{\boldsymbol{\Phi}}(z) &= c s_{\hat{\mathbf{R}}}(z) - \frac{(1 - c)}{z}. \end{aligned} \quad (21)$$

Substituting (21) in (20), and after some algebraic manipulations, we obtain (for both the continuous and the finite size approach)

$$\begin{aligned} s_{\hat{\mathbf{R}}}(z) &= \int \frac{dF_{\mathbf{R}}(\lambda)}{(1-c-czs_{\hat{\mathbf{R}}}(z))\lambda-z}, \quad (22) \\ &= \frac{1}{M} \sum_{m=1}^M \frac{1}{(1-c-czs_{\hat{\mathbf{R}}}(z))\lambda_m-z}. \quad (23) \end{aligned}$$

Following some general assumptions, the asymptotic behavior of  $s_{\mathbf{A}}(z)$  and  $m_{\mathbf{A}}(z)$  is the same [11, Theorem 1], and hence one can apply the results above for  $m_{\hat{\mathbf{R}}}$ , and evaluate it for the case of  $z = 0$ . Then

$$m_{\hat{\mathbf{R}}}(0) = \sum_{m=1}^M \frac{\mathbf{a}^H \hat{\mathbf{v}}_m \hat{\mathbf{v}}_m^H \mathbf{b}}{(1-c)\lambda_m} \asymp \frac{m_{\mathbf{R}}(0)}{1-c}. \quad (24)$$

Using Definition 2,  $m_{\hat{\mathbf{R}}}(z) \asymp \mathbf{a}^H (\hat{\mathbf{R}} - z\mathbf{I})^{-1} \mathbf{b}$ , we obtain

$$(1-c)\mathbf{a}^H \hat{\mathbf{R}}^{-1} \mathbf{b} \asymp \mathbf{a}^H \mathbf{R}^{-1} \mathbf{b}. \quad (25)$$

In our particular case,  $\mathbf{a}$  is selected as 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  $\varphi(\mathbf{R}^{-1}) = [\mathbf{w}^*]_i$ .

This concludes the proof of Theorem 1. An alternative proof can be found in [12]. ■

### C. $N, M$ -Consistent Solution for the Mean Square Error

The MSE obtained by using  $\hat{\mathbf{w}}$  from Theorem 1 is:

$$\begin{aligned} \text{MSE}(\hat{\mathbf{w}}) &= \hat{\sigma}_{x_s} - (1-c) \left( \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \mathbf{r}_x + \mathbf{r}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x \right) \\ &+ (1-c)^2 \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{R}} \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (26) \end{aligned}$$

A traditional approach to estimate the MSE is by simply replacing the true correlation parameters by their sample estimators. In the case above (see equation (26)), one can derive an estimator for the MSE as

$$\begin{aligned} \widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}}) &= \hat{\sigma}_{x_s} - 2(1-c)\hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x \\ &+ (1-c)^2 \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x, \quad (27) \end{aligned}$$

where  $\hat{\sigma}_{x_s}$  is the sample estimation of the signal variance  $\sigma_{x_s}$ , defined in (29). The estimator  $\widehat{\text{MSE}}_{\text{class}}(\hat{\mathbf{w}})$  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  $N$ .

*Theorem 2:* An  $N, M$ -consistent estimator of  $\text{MSE}(\hat{\mathbf{w}})$  is given by

$$\widehat{\text{MSE}}(\hat{\mathbf{w}}) = \hat{\sigma}_{x_s} - (1-c)\hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (28)$$

*Remark 1:* Note that the approach taken in this paper is slightly different to the MSE estimator in [13], where the authors give an  $N, M$ -consistent estimator for the optimal MMSE. On the contrary, 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 (15)), which is not necessary to be the MMSE lower bound.

*Proof:* To estimate the signal variance  $\hat{\sigma}_{x_s}$  one can use

$$\hat{\sigma}_{x_s} = \frac{1}{N} \sum_{n=1}^N x_s(n)^2, \quad (29)$$

which is  $N, M$ -consistent [14]. Furthermore, one can see that  $2(1-c)\hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x$  is already a  $N, M$ -consistent estimator of the term  $(1-c) \left( \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \mathbf{r}_x + \mathbf{r}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x \right)$ . The reason is that  $\hat{\mathbf{r}}_x$  is already an  $N, M$ -consistent estimator of  $\mathbf{r}_x$ . Hence, the critical part resides in the estimation of the last term

$$\beta = (1-c)^2 \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{R}} \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x, \quad (30)$$

which involves both the inverse of the SCM and the *true covariance matrix*.

Using the equivalence in equation (12) in (31), we get

$$\beta = N^2(1-c)^2 \hat{\mathbf{r}}_x^H (\mathbf{R}^{1/2} \mathbf{\Xi}^H \mathbf{\Xi} \mathbf{\Xi}^H \mathbf{\Xi} \mathbf{R}^{1/2})^{-1} \hat{\mathbf{r}}_x. \quad (31)$$

Consequently, we can define the following two random matrices:  $\mathbf{\Gamma} \in \mathbb{C}^{M \times N}$ , where  $\mathbf{\Gamma} = \frac{1}{\sqrt{N}} \mathbf{R}^{1/2} \mathbf{\Xi}^H$  (note that  $\hat{\mathbf{R}} = \mathbf{\Gamma} \mathbf{\Gamma}^H$ ); and  $\hat{\mathbf{\Gamma}} = \frac{1}{N} \mathbf{\Gamma} \mathbf{\Xi} \mathbf{\Xi}^H \mathbf{\Gamma}$ . Hence

$$\begin{aligned} \beta &= N(1-c)^2 \hat{\mathbf{r}}_x^H (\mathbf{\Gamma} \mathbf{\Xi} \mathbf{\Xi}^H \mathbf{\Gamma}^H)^{-1} \hat{\mathbf{r}}_x \\ &= (1-c)^2 \hat{\mathbf{r}}_x^H \hat{\mathbf{\Gamma}}^{-1} \hat{\mathbf{r}}_x. \quad (32) \end{aligned}$$

We have observed that a good estimator for the scalar quantity  $\hat{\mathbf{r}}_x^H \hat{\mathbf{\Gamma}}^{-1} \hat{\mathbf{r}}_x$  is:

$$\hat{\mathbf{r}}_x^H \hat{\mathbf{\Gamma}}^{-1} \hat{\mathbf{r}}_x \asymp (1-c)^{-1} \hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (33)$$

Thus, using (33) in equation (31), one can estimate  $\beta$  as:

$$\hat{\beta} = (1-c)\hat{\mathbf{r}}_x^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{r}}_x. \quad (34)$$

Thus the proof of Theorem 2 is complete. ■

## V. NUMERICAL RESULTS

In this section, we provide simulation results to show the performance of the *proposed* methodology in comparison with the *classical* approach. We also plot in the figures the performance of the DSC algorithm for the case when the receiver has full access to the true correlation parameters.

The parameters that configure the basic setup of the simulation environment are as follows:

- Number of already-decoded *sensing nodes*:  $S' = 200$ .
- Number of past samples:  $K = 200$ , hence  $M = 400$ .
- Length of the training phase:  $N = 1000$  snapshots.
- Aspect ratio,  $c = 0.4$ ;
- Correlation model,  $[\mathbf{R}_s]_{i,i+k} = [\mathbf{R}_t]_{i,i+k} = 0.9^{|k|}$ .
- SER threshold:  $\text{SER}_t = 10^{-2}$ .
- A/D converter depth:  $l = 12$  bits.

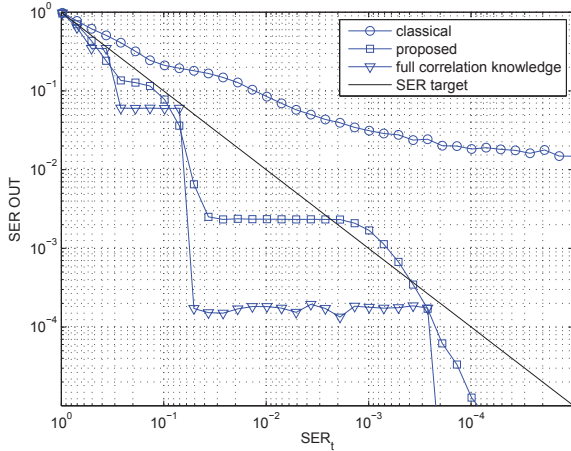


Fig. 1. Comparison of the SER as a function of the  $SER_t = 10^{-3}$  for both the *classical* and *proposed* methods for the case of  $c = 0.4$ . This figure has been averaged over 100 realizations.

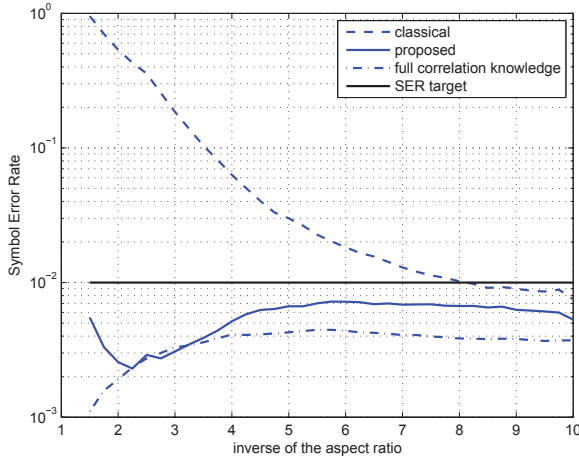


Fig. 2. Comparison of the SER performance of *classical* and the *proposed* methods imposing a  $SER_t = 10^{-2}$ . It has been averaged over 100 realizations.

### A. Symbol Error Rate as a function of $SER_t$

The final purpose of DSC is the reduction of the transmitted bits in order to reduce power consumption. However, system requirements must be taken into account in the design phase. Thus we analyze the performance in terms of the SER fidelity. In other words, we compare the SER obtained with the *proposed* and *classical* techniques as a function of the  $SER_t$ .

Graphically, Fig. 1 shows the SER performance for the case  $c = 0.4$ . Ideally, the SER performance curve should be below but as close as possible to the SER threshold (the solid line in the figure). Our proposed method is actually below the threshold, except for a small area around  $SER_t = 10^{-3}$ . The curve is staircase-shaped due to the ceiling function of (5).

On the contrary, the *classical* approach does not fit the system requirements. One possible solution to counteract this effect is to increase the training phase  $N$  (i.e., decrease  $c$ ).

### B. Symbol Error Rate as a function of $c$

We set the SER threshold of  $10^{-2}$ . Fig. 2 shows that the tendency of the *classical* DSC approach is to fulfill the SER requirements only for high values of  $c^{-1}$  due to its  $N$ -consistency property. However for low values of  $N$  (i.e., when the ratio  $c$  increases), it cannot fit the system requirements.

One can observe that the *classical* method requires a training phase at least eight times longer than the observation dimension  $M$  to fulfill the  $SER_t$ . On the other hand, using the *proposed* estimators, one can guarantee the requirements even for values of  $c$  close to one.

## VI. CONCLUSIONS

This paper has proposed a DSC algorithm for large WSNs based on enhanced correlation estimators. It presents a full derivation of these estimators under a RMT framework. Specifically, we present two estimators for each key step of the DSC algorithm; *i*) the computation of the *side-information*, and *ii*) the bit rate required to guarantee a certain  $SER_t$ .

When we compare the resulting SER for short training phase, we observe that our *proposed* method performs far better under a given SER requirement. Moreover, it performs as the corresponding SCM estimators when the training phase increases. In practice, it allows to reduce largely the training phase in DSC schemes.

## REFERENCES

- [1] C.-K. Toh, "Maximum Battery Life Routing to Support Ubiquitous Mobile Computing in Wireless Ad Hoc Networks," *Communications Magazine, IEEE*, vol. 39, no. 6, pp. 138–147, Jun. 2001.
- [2] R. Mudumbai, D.R. Brown, U. Madhoo, and H.V. Poor, "Distributed Transmit Beamforming: Challenges and Recent Progress," *Communications Magazine, IEEE*, vol. 47, no. 2, pp. 102–110, Feb. 2009.
- [3] S.S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed Compression in a Dense Microsensor Network," *Signal Processing Magazine, IEEE*, vol. 19, no. 2, pp. 51–60, Mar. 2002.
- [4] S.S. Pradhan and K. Ramchandran, "Distributed Source Coding Using Syndromes (DISCUS): Design and Construction," *Information Theory, IEEE Transactions on*, vol. 49, no. 3, pp. 626–643, Mar. 2003.
- [5] D. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources," *Information Theory, IEEE Transactions on*, vol. 19, no. 4, pp. 471–480, Jul. 1973.
- [6] M. L. Mehta, *Random Matrices*, Academic Press, Boston, 1991.
- [7] J. Chou, D. Petrovic, and K. Ramchandran, "A Distributed and Adaptive Signal Processing Approach to Reducing Energy Consumption in Sensor Networks," in *IEEE INFOCOM 2003*, Mar. 2003, vol. 2, pp. 1054–1062.
- [8] S. Verdú A. M. Tulino, *Random matrix theory and wireless communications*, Foundations and Trends in Communications and Information Theory 1 (1), June 2004.
- [9] X. Mestre, "On the Asymptotic Behavior of the Sample Estimates of Eigenvalues and Eigenvectors of Covariance Matrices," *Signal Processing, IEEE Transactions on*, vol. 56, no. 11, pp. 5353–5368, 2008.
- [10] V. Marčenko and L. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Math USSR Sbornik*, vol. 1, pp. 457–483, 1967.
- [11] B. Q. Miao Z. D. Bai and G. M. Pan, "On Asymptotics Behavior of Eigenvectors of Large Sample Covariance Matrices," *Ann. Probab.*, vol. 35, no. 4, pp. 1532–1572, 2007.
- [12] V. L. Girko, " $G_{25}$ —Estimators of Principal Components," *Theory. Probab. Mathematical Statistics*, vol. 40, pp. 1–10, 1990.
- [13] F. Rubio and X. Mestre, "Consistent Reduced-Rank LMMSE Estimation With a Limited Number of Samples per Observation Dimension," *Signal Processing, IEEE Transactions on*, vol. 57, no. 8, pp. 2889–2902, 2009.
- [14] V. L. Girko, *An Introduction to Statistical Analysis of Random Arrays*, VSP, The Netherlands, 1998.