

# Estimation of Spatially-Correlated Random Fields with Compressed Observations

Javier Matamoros and Carles Antón-Haro, *Senior Member, IEEE*

**Abstract**—In this paper, we focus on the use of wireless sensor networks for the estimation of spatially-correlated random fields. We explicitly take communication aspects into account and propose a number of distributed pre-coding (beamforming) schemes allowing for an over-the-air *compressed* representation of the set of spatially correlated observations. Those observations are then encoded in a number of consecutive sensor-to-Fusion Center (FC) transmissions. The ultimate goal is to minimize the distortion in the reconstructed random field and, simultaneously, keep the number of transmissions low (i.e. the compression ratio high). Specifically, we first propose a family of distributed pre-coding schemes based on the Karhunen-Loève (KL) and partial KL transforms and we derive closed form expressions of the optimal power allocation strategies. Aspects such as residual phase synchronization errors or the fact that some sensors might be idle (inactive) are specifically taken into account in the analysis. Next, we also propose an iterative and greedy scheme by which the pre-coder design and power allocation problems can be jointly solved for the *general* case. We assess the performance of the proposed pre-coding schemes by means of computer simulations. Other *uncompressed* transmission schemes are used as a benchmark.

## I. INTRODUCTION

Typically, a Wireless Sensor Network (WSN) consists of one Fusion Center (FC) and a potentially large number of sensing devices capable of conveying their measurements over wireless links. A non-exhaustive list of applications of WSNs encompasses parameter estimation (e.g. for environmental monitoring applications), event detection (such as accidental release of pollutants), localization [1], or asset positioning and tracking (see the survey in [2] and references therein).

In the last decade, the problem of decentralized estimation with WSNs has received considerable attention (see e.g. [3], [4], [5]). For scenarios where sensor observations can be modeled as noisy versions of a *single* underlying Gaussian source, cooperative beamforming turns out to be the optimal strategy [6] in terms of estimation error at the FC (for this setup the source and channel coding separation principle does not hold [7]). In other words, sensors must *simultaneously* transmit scaled (i.e. uncoded) versions of their observations. Cooperative beamforming approaches, however, require accurate phase synchronization over sensors (see e.g. the synchronization scheme presented in [8]) which, in general, is difficult to achieve. To alleviate this, a number of authors [9], [10], [11],

[12] have proposed sensor-to-FC transmissions to take place over *orthogonal* channels. Although suboptimal, such schemes merely require pair-wise synchronization for each sensor and the FC.

In other works, a more realistic signal model with spatially correlated sensor observations is adopted [13], [14], [15]. This is often referred to as a spatially-correlated random field with correlation typically being a function of their spatial locations (see e.g. [16]). In this scenario, the authors in [17] show that for spatially bandlimited processes, the bit budget per Nyquist-period can be arbitrarily reallocated along the quantization precision and/or the space (by adding more sensor nodes) axes, while retaining the same decay profile of the reconstruction. In [18], and again considering bandlimited spatial fields, the authors analyze the impact of random sampling patterns on the resulting estimation distortion.

Regarding source and channel coding aspects, the authors in [19] derived the scaling laws of distortion vs. transmit power, number of sensors or bandwidth. This is accomplished for a number of encoding schemes. Interestingly, it is shown that for some scenarios uncoded transmission is asymptotically optimal. A similar (i.e. asymptotic) analysis is conducted in [20]. Here, sensor observations are assumed to be compressible in some basis (e.g. Fourier, Wavelets, Haar, etc). Sensor observations are *simultaneously* conveyed to the FC by means of a number of projections onto (i) a known compressible basis (*known subspace* method), this also requiring a priori knowledge on the specific coefficient ordering; or (ii) an incoherent and random basis (*unknown subspace* method). In both cases, transmit power is evenly allocated to those projections.

### A. Contributions

In this paper, we address the problem of reconstructing a spatially-correlated random field from the observations collected by a WSN. The goal is to estimate the field itself rather than the underlying sources, as in [19]. Inspired by [20], [19], we allow sensors to simultaneously transmit their observations since, in general, this results into reduced transmission latency (when compared with orthogonal transmission schemes [9]). Our focus is on pre-coder design and on the computation of the optimal power allocation strategies. This is in stark contrast with [20], [19] where priority is given to the derivation of scaling laws for a given pre-coder and uniform power allocation.

The main contributions of this paper are as follows:

- 1) We propose a family of distributed pre-coding schemes based on the Karhunen-Loève (KL) and partial KL

This work is partly supported by the Catalan Government (2014 SGR 1567), the EC-funded project Newcom # (318306), and by the Spanish Government under projects TEC2013-44591-P (INTENSIV) and PCIN-2013-027 (E-CROPS).

Authors are with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Parc Mediterrani de la Tecnologia, 08860 Castelldefels, Barcelona (Spain), emails: first.last@cttc.es.

transforms [21] allowing us to obtain an over-the-air *compressed* representation of the set of observations. Clearly, this goes one step beyond other uncompressed transmission schemes such as [9], [10], [11], [12]). To that aim, we assume a priori knowledge on the spatial correlation of the field<sup>1</sup>. As a result, and similarly to the *known subspace* method in [20], KL-based schemes avoid sending redundant information by allocating the power over a limited number of signal projections.

- 2) For the aforementioned pre-coding schemes, we derive closed-form expressions of the optimal<sup>2</sup> power allocation strategies (vs. uniform power allocation per transmission/channel in [20]). We address two problems of interest, namely, the minimization of the estimation error at the FC and the minimization of the transmit power at the sensor nodes. Interestingly, such closed-form expressions allow us to determine *analytically* the optimal number of sensor-to-FC transmissions rather than its scaling law as in [20]. The analysis also encompasses scenarios with residual, yet not necessarily small, phase synchronization errors or idle (inactive) sensors. In such scenarios, numerical results have revealed a number of interesting findings. For instance, in the case of phase synchronization errors, it turns out to be more robust to allocate power to a reduced number of transmissions. In contrast, if sensors experience low duty-cycles and adopt a KL precoding scheme, the optimal power allocation tends to active more transmissions in an attempt to compensate the missing data.
- 3) Finally, we propose an iterative and greedy scheme by which the pre-coder design and power allocation problems can be jointly (and effectively) solved for the general case. By doing so, we avoid using the per-sensor channel equalizers that KL-based (and also the schemes proposed in [20]) require. This, in turn, allows avoiding the inefficiencies that per-sensor channel equalizers entail in fading scenarios. For the particular case of Gaussian channels, the iterative algorithm turns out to find the optimal solution, for which we also derive a closed-form expression. For arbitrary channels, in which case the pre-coder design and power allocation problems are inter-twined, the proposed scheme iteratively finds a solution which is virtually identical to the optimal one (which can only be computed numerically).

The rest of the paper is organized as follows. In Section II, we introduce the signal and communication models. Next, in Section III we present a family of distributed pre-coding schemes based on the Karhunen-Loève (KL) transform, and conduct a distortion analysis for the so-called (i) baseline scenario, and scenarios with (ii) phase synchronization errors and (iii) idle sensors. Section IV is devoted to compute, in a unified manner, the optimal power allocation for the three aforementioned scenarios. In Section V, we go one

<sup>1</sup>In practice, correlation properties, which typically depend on sensors' spatial locations, would be determined in an initial training phase upon deployment.

<sup>2</sup>Optimal power allocation is potentially beneficial in e.g. low transmit power regimes.

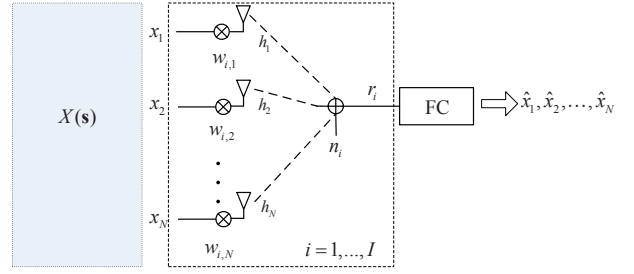


Fig. 1. Signal and communication model for a network comprising  $N$  sensors and 1 FC.

step beyond and propose an iterative and greedy scheme for joint pre-coder design and power allocation. Next, we present selected computer simulation results in Section VI. Section VII closes the paper by summarizing the main results and drawing some conclusions.

## II. SIGNAL AND COMMUNICATION MODEL

Let  $X(\mathbf{s})$  be a spatially-correlated random field defined over the two-dimensional space  $\mathbb{R}^2$ . We assume that  $X(\mathbf{s})$  is stationary, zero-mean, and Gaussian-distributed with variance  $\sigma_x^2$ . The random field is sampled by a set of  $N$  sensors located at  $\mathbf{s}_1, \dots, \mathbf{s}_N$  (locations are assumed to be known), this yielding

$$x_j \triangleq X(\mathbf{s}_j); \quad j = 1, \dots, N. \quad (1)$$

Consequently, the vector of observations  $\mathbf{x} = [x_1, \dots, x_N]^T$  is zero-mean and jointly Gaussian, namely,  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_\mathbf{x})$ . For a specific set of locations, the elements of the covariance matrix  $\mathbf{C}_\mathbf{x} = \mathbb{E}[\mathbf{x}\mathbf{x}^T]$  read  $[\mathbf{C}_\mathbf{x}]_{j,j'} = k(\mathbf{s}_j, \mathbf{s}_{j'})$ , where  $k(\cdot, \cdot)$  denotes the spatial covariance function of the field.

In order to enhance network lifetime, a pre-defined duty cycle of  $p \times 100\%$  is set. Therefore, sensors are in active mode (i.e. transmitting information to the FC) or idle mode (i.e. saving power) with probability  $p$  and  $1-p$ , respectively. In the sequel, we denote by  $\mathcal{S}_A$  and  $\mathcal{S}_A^c$  the subsets of active and idle sensors (with  $\mathcal{S} = \mathcal{S}_A \cup \mathcal{S}_A^c$  and  $\mathcal{S}_A \cap \mathcal{S}_A^c = \emptyset$ ), and their corresponding cardinalities by  $N_A = |\mathcal{S}_A|$ , and  $N_A^c = |\mathcal{S}_A^c|$ . As a result, the observation vector  $\mathbf{x}$  can be decomposed into  $\mathbf{x}_A$  and  $\mathbf{x}_A^c$  that denote the  $N_A \times 1$  and  $N_A^c \times 1$  vectors of observations collected by the subsets of active and idle sensors, respectively.

At a given time instant, the subset of *active* sensors *simultaneously* transmit (i.e. beamform) their observations to the single FC (see Fig. 1). For the  $i$ -th transmission ( $i = 1 \dots I$ ), the received signal  $r_i$  reads<sup>3</sup>

$$r_i = \sum_{j=1}^N w_{i,j}^* \cos(\xi_j) h_j b_j x_j + n_i; \quad i = 1 \dots I, \quad (2)$$

where  $\{w_{i,j}^*\}_{i,j=1}^{I,N}$  denote the complex conjugates of the transmit weights to be designed. For notational convenience, we gather those weights into a number of pre-coding vectors

<sup>3</sup>Here we assume that the spatially-correlated random field is static for at least  $I$  consecutive transmissions.

denoted by  $\mathbf{w}_i = [w_{i,1}, w_{i,2}, \dots, w_{i,N}]^T$  which allow to obtain an *over-the-air* compressed representation of the observations. From this expression, it becomes apparent that a different pre-coder will be used in each transmission. The set of i.i.d. random variables  $\xi_1, \dots, \xi_N$  in (2) model the residual (yet not necessarily small) phase synchronization errors in the sensor oscillators due to e.g. practical limitations in the closed-loop synchronization schemes [8], with  $\xi_j \sim \mathcal{U}(-\epsilon, +\epsilon)$  and  $\epsilon > 0$  (i.e. uniform distribution) [20]. Furthermore,  $h_1, h_2, \dots, h_N$  stand for the known complex sensor-to-FC channel coefficients;  $b_1, \dots, b_N$  are i.i.d. Bernoulli random variables accounting for sensor activity:  $b_j \in \{1, 0\}$  with probability  $p$  and  $1 - p$ , respectively; and, finally,  $n_i$  denotes additive white Gaussian noise of variance  $\sigma_n^2$ , that is,  $n_i \sim \mathcal{CN}(0, \sigma_n^2)$ . In addition, we assume slow fading conditions and, hence, the channel coefficients remain unchanged for the  $I$  consecutive transmissions.

From the  $I \times 1$  received vector  $\mathbf{r} = [r_1, \dots, r_I]^T$ , the FC will provide the Bayesian estimate of the spatially-correlated random field at the set of sampled locations, namely,  $\hat{\mathbf{x}} = [\hat{x}_1, \dots, \hat{x}_N]^T$ . In the sequel, we assume  $I \leq N$  and, hence,  $\mathbf{r}$  can be regarded as an *over-the-air compressed* representation of the observations vector  $\mathbf{x}$  (this assumption is analyzed in detail in Section V-B ahead). Due to the various impairments, noise and compression, the resulting estimates are unavoidably subject to some distortion given by the normalized Mean Square Error (MSE) of those estimates, namely

$$D \triangleq \frac{1}{N\sigma_x^2} \sum_{j=1}^N \mathbb{E} [|\hat{x}_j - x_j|^2]. \quad (3)$$

### III. COMPRESSED TRANSMISSION VIA THE KARHUNEN-LOÈVE (KL) TRANSFORM

Our goal here is to design the set of precoding vectors  $\{\mathbf{w}_i\}_{i=1}^N$  capable of exploiting the spatial correlation properties of the random field given by (1) and, by doing so, allow for a more energy-efficient (i.e. compressed) transmission. To that aim, we resort to the well-known KL and partial KL transforms. In addition, we find closed-form expressions of the distortion given by (3) in the reconstructed field for a number of scenarios of interest.

#### A. Review of the Karhunen-Loève (KL) Transform

Let  $\mathbf{x} = [x_1, \dots, x_N]^T$  be a set of jointly Gaussian-distributed random variables. The eigendecomposition of its covariance matrix is given by

$$\mathbf{C}_x = \Phi \Lambda \Phi^T, \quad (4)$$

where  $\Phi = [\phi_1, \dots, \phi_N]$  and  $\Lambda = \text{diag}[\lambda_1, \dots, \lambda_N]$  are the unitary and diagonal matrices containing the corresponding eigenvectors and eigenvalues, respectively (with  $\lambda_i \in \{\mathbb{R}^+, 0\}$  and  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ ). The KL transform of vector  $\mathbf{x}$  can thus be defined as

$$\mathbf{y} = \Phi^T \mathbf{x} = \begin{bmatrix} \phi_1^T \mathbf{x} \\ \vdots \\ \phi_N^T \mathbf{x} \end{bmatrix}. \quad (5)$$

The elements of the transform vector  $\mathbf{y} = [y_1, \dots, y_N]^T$  turn out to be independent random variables, that is,  $\mathbf{y} \sim \mathcal{N}(0, \Lambda)$ . Since  $\Phi$  is a unitary matrix, the inverse transform can be readily expressed as

$$\mathbf{x} = \Phi \mathbf{y} = \sum_{i=1}^N \phi_i y_i. \quad (6)$$

Besides, the first  $I$  components in  $\mathbf{y}$  are known to provide the best  $I$ -length representation of vector  $\mathbf{x}$ , that is, the one that minimizes the mean square error (MSE) in the estimate given by  $\hat{\mathbf{x}} = \sum_{i=1}^I \phi_i y_i$ .

#### B. Pre-coder Design

Inspired by [20], we impose the following structure for the design of the corresponding transmit weights

$$w_{i,j}^* = \sqrt{\rho_i} \frac{h_j^*}{|h_j|^2} w'_{i,j}; \quad i, j = 1 \dots N \quad (7)$$

with  $\{\rho_i\}_{i=1}^N$  denoting a set of scalar factors to be determined later, in Section IV<sup>4</sup>. The rationale behind is that this factorization facilitates the design of the pre-coder by decoupling the channel equalization and data compression tasks. More in detail, (i) the term  $h_j^*$  allows sensors (each sensor) to coherently combine their transmissions at the FC; (ii) the term  $\frac{1}{|h_j|^2}$  carries out a per-sensor channel equalization (inversion); and (iii) the  $w'_{i,j}$  coefficients are in charge of the compression task. Since (ii) makes the system oblivious to the channel gains, we simply let  $w'_{i,j}$  be the coefficients of the (centralized) KL transform, that is,  $w'_{i,j} = \phi_{i,j}$ . In other words, in the  $i$ -th transmission we attempt to convey the observations over the  $i$ -th eigenmode of the covariance matrix  $\mathbf{C}_x$ . This is strictly true in the absence of residual phase synchronization errors *and* when all sensors are active (i.e.,  $\xi_i = 0$  and  $b_i = 1$  for  $i = 1 \dots N$ ). Otherwise, phase errors and/or the fact that only a *subset* of the sensors effectively transmit data results into some degradation of the generated beam pattern. Ultimately, this leads to an increased distortion in the reconstructed random field. This effect will be analyzed more in detail later on.

In such idealized scenarios, the total transmit power at the  $i$ -th transmission reads

$$P_i = \mathbb{E} \left[ \sum_{j=1}^N |w_{i,j} x_j|^2 \right] = \left( \sum_{j=1}^N \frac{1}{|h_j|^2} \phi_{i,j}^2 \sigma_x^2 \right) \rho_i \quad (8)$$

$$= \alpha_i \rho_i; \quad i = 1, \dots, N, \quad (9)$$

where we have defined  $\alpha_i \triangleq \sum_{j=1}^N \frac{1}{|h_j|^2} \phi_{i,j}^2 \sigma_x^2$ . Hence, the scalar factor  $\rho_i$  is directly related with the power allocation to the  $i$ -th transmission, while  $\alpha_i$  accounts for the impact of the pre-coding vector and the spatially-correlated random field on the actual transmit power. In the presence of deep fades (i.e. when some  $h_j \rightarrow 0$ ) the term  $\alpha_i$  increases and, thus,  $\rho_i$  decreases. This has a negative impact on the quality with which the information on the random field is conveyed over the  $i$ -th eigenmode (transmission).

<sup>4</sup>Note that we include the *entire* set of pre-coders  $\{\mathbf{w}_i\}_{i=1}^N$  in the problem definition. Convenience will be discussed in the next subsection.

### C. Distortion Analysis: Baseline Scenario

Here, we consider a scenario with (i) no phase synchronization errors (i.e.,  $\xi_i = 0$  for  $i = 1 \dots N$ ); and where (ii) all sensors are always active (i.e., activity factor of 100% which leads to  $b_i = 1$  for  $i = 1 \dots N$ ). From (2), the received signal in the  $i$ -th transmission thus reads:

$$r_i = \sqrt{\rho_i} \sum_{j=1}^N \phi_{i,j} x_j + n_i; \quad i = 1, \dots, N. \quad (10)$$

After  $N$  consecutive transmissions, the received signal vector at the FC can be readily expressed as

$$\mathbf{r} = \mathbf{\Gamma} \mathbf{y} + \mathbf{n}, \quad (11)$$

where  $\mathbf{y} = [y_1, \dots, y_N]^T$  stands for the length- $N$  KL transform of  $\mathbf{x}$  given in (5),  $\mathbf{\Gamma} = \text{diag}(\sqrt{\rho_1}, \dots, \sqrt{\rho_N})$ ; and  $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_N)$ . It is worth noting that, for the case of correlated observations, the total number of non-zero  $\rho_i$  coefficients  $I$  (i.e. the number of transmissions) will be smaller than  $N$ . Hence, only a subset of transmissions will actually take place, this resulting into a *compressed* representation of the set of  $N$  observations (see Section IV for details). The (optimal) Minimum Mean Squared Error (MMSE) estimator of  $\mathbf{x}$  given  $\mathbf{r}$  turns out to be the posterior mean [22], namely,

$$\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}|\mathbf{r}] = \mathbf{C}_{\mathbf{x}\mathbf{r}} \mathbf{C}_{\mathbf{r}}^{-1} \mathbf{r}, \quad (12)$$

with  $\mathbb{E}[\cdot]$  denoting the statistical expectation with respect to  $\mathbf{x}$ ,  $\mathbf{y}$  and  $\mathbf{n}$ ; and where  $\mathbf{C}_{\mathbf{x}\mathbf{r}} = \mathbb{E}[\mathbf{x}\mathbf{r}^T]$  and  $\mathbf{C}_{\mathbf{r}} = \mathbb{E}[\mathbf{r}\mathbf{r}^T]$ . Hence, the average distortion in the estimate of  $\mathbf{x}$  at the FC reads

$$\begin{aligned} D = \sum_{j=1}^N \text{Var}(x_j|\mathbf{r}) &= \frac{1}{N} \text{Tr}(\mathbf{C}_{\mathbf{x}|\mathbf{r}}) \\ &= \frac{1}{N} \text{Tr}(\mathbf{C}_{\mathbf{x}} - \mathbf{C}_{\mathbf{x}\mathbf{r}} \mathbf{C}_{\mathbf{r}}^{-1} \mathbf{C}_{\mathbf{r}\mathbf{x}}) \end{aligned} \quad (13)$$

where  $\text{Tr}(\cdot)$  denotes the trace operator, and  $\mathbf{C}_{\mathbf{x}|\mathbf{r}} = \mathbf{C}_{\mathbf{x}} - \mathbf{C}_{\mathbf{x}\mathbf{r}} \mathbf{C}_{\mathbf{r}}^{-1} \mathbf{C}_{\mathbf{r}\mathbf{x}}$  [22].

First of all, we need to compute matrix  $\mathbf{C}_{\mathbf{r}}$ . Since  $y_1, \dots, y_N$  are zero-mean independent random variables, it follows from (11) that  $\mathbf{C}_{\mathbf{r}}$  is a diagonal matrix with entries given by  $[\mathbf{C}_{\mathbf{r}}]_{i,i} = \rho_i \lambda_i + \sigma_n^2$ . Next, in order to compute matrix  $\mathbf{C}_{\mathbf{x}\mathbf{r}}$ , we recall from (6) that

$$x_j = \sum_{i=1}^N \phi_{i,j} y_i. \quad (15)$$

From this and (10), the  $\mathbb{E}[x_j r_i]$  terms yield

$$\mathbb{E}[x_j r_i] = \mathbb{E}\left[\sum_{i=1}^N \phi_{i,j} y_i r_i\right] = \sqrt{\rho_i} \phi_{i,j} \lambda_i \quad (16)$$

for  $i, j = 1, \dots, N$ . Consequently, matrix  $\mathbf{C}_{\mathbf{x}\mathbf{r}}$  reads:

$$\mathbf{C}_{\mathbf{x}\mathbf{r}} = \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Gamma}. \quad (17)$$

From all the above, it follows that

$$\mathbf{C}_{\mathbf{x}|\mathbf{r}} = \mathbf{C}_{\mathbf{x}} - \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Gamma} \mathbf{C}_{\mathbf{r}}^{-1} \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Phi}^T \quad (18)$$

and, finally, the average distortion reads:

$$D = \frac{1}{N} \text{Tr}(\mathbf{C}_{\mathbf{x}|\mathbf{r}}) \quad (19)$$

$$= \sigma_x^2 - \frac{1}{N} \text{Tr}(\mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Gamma} \mathbf{C}_{\mathbf{r}}^{-1} \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Phi}^T) \quad (20)$$

$$= \sigma_x^2 - \frac{1}{N} \sum_{i=1}^N \frac{\rho_i \lambda_i^2}{\rho_i \lambda_i + \sigma_n^2} \quad (21)$$

where, in the first and second equality, we exploited that  $\mathbf{\Phi}$  is an orthogonal matrix and  $\mathbf{\Lambda} \mathbf{\Gamma} \mathbf{C}_{\mathbf{r}}^{-1} \mathbf{\Gamma} \mathbf{\Lambda}$  is a diagonal matrix.

### D. Distortion Analysis: Scenario with Phase Synchronization Errors

Here, we continue to assume that all sensors are always active ( $b_i = 1$  for all  $i$ ) but, unlike in the previous section, we let phase synchronization errors in the sensor oscillators take non-zero values. This results into a more realistic scheme since, in practice, some phase synchronization error cannot be avoided. By defining  $z_j \triangleq \cos(\xi_j) - 1$ , equation (2) can be particularized and conveniently re-written as

$$r_i = \sqrt{\rho_i} \sum_{j=1}^N \phi_{i,j} x_j + \sqrt{\rho_i} \sum_{j=1}^N \phi_{i,j} z_j x_j + n_i \quad (22)$$

$$= \sqrt{\rho_i} y_i + \sqrt{\rho_i} \sum_{j=1}^N \phi_{i,j} z_j x_j + n_i \quad (23)$$

where, clearly, the second term in (23) accounts for the impact of phase errors. It is worth noting that the newly-defined random variable  $z_j$  is neither Gaussian-distributed, nor zero-mean. As we will see next, this has a number of implications and makes the derivation of a closed-form expression of the distortion far more involved (see Appendix).

Again, in order to compute the average distortion resulting from the linear MMSE<sup>5</sup> estimator of  $\mathbf{x}$  given  $\mathbf{r}$  we start by computing matrix  $\mathbf{C}_{\mathbf{r}}$ . Note that, since  $x_j$  and  $y_i$  are zero mean and  $x_j$  is independent of  $z_j$ ,  $r_i$  turns out to be a zero-mean random variable and, hence, we can write  $[\mathbf{C}_{\mathbf{r}}]_{i,i'} = \mathbb{E}[r_i r_{i'}]$ . In the Appendix, we prove that the diagonal elements (i.e. for  $i = i'$ ) can be expressed as

$$\mathbb{E}[r_i^2] = \rho_i \lambda_i (1 + \mathbb{E}[z])^2 + \rho_i \sigma_z^2 \sigma_x^2 + \sigma_n^2 \quad (24)$$

whereas, for the off-diagonal elements, we have  $\mathbb{E}[r_i r_{i'}] = 0$ . In other words, the random variables  $\{r_1, r_2, \dots, r_N\}$  are uncorrelated and, hence, matrix  $\mathbf{C}_{\mathbf{r}}$  turns out to be diagonal. Next, in order to compute matrix  $\mathbf{C}_{\mathbf{x}\mathbf{r}}$ , we must realize first that the random variables  $x_j$  and  $z_j$  are statistically independent. After some derivations, matrix  $\mathbf{C}_{\mathbf{x}\mathbf{r}}$  yields

$$\mathbf{C}_{\mathbf{x}\mathbf{r}} = (1 + \mathbb{E}[z]) \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Gamma}. \quad (25)$$

Finally, from (14), (24) and (25), the average distortion reads

$$D = \sigma_x^2 - \frac{1}{N} \sum_{i=1}^N \frac{\rho_i (1 + \mathbb{E}[z])^2 \lambda_i^2}{\rho_i (1 + \mathbb{E}[z])^2 \lambda_i + \rho_i \sigma_x^2 \sigma_z^2 + \sigma_n^2} \quad (26)$$

<sup>5</sup>Note that the linear MMSE estimator is no longer optimal since, unlike in previous section,  $\{r_i\}$  is not Gaussian distributed [22].

where  $\mathbb{E}[z]$  and  $\sigma_z^2$  denote, respectively, the mean and the variance of the random variable  $z = \cos(\xi) - 1$ , namely

$$\begin{aligned}\mathbb{E}[z] &= \frac{\sin(\epsilon)}{\epsilon} - 1 \\ \sigma_z^2 &= \frac{1}{2} \frac{\epsilon \cos(\epsilon) \sin(\epsilon) + \epsilon^2 - 2 \sin^2(\epsilon)}{\epsilon^2}.\end{aligned}\quad (27)$$

Note that in the case of perfect phase synchronization, we have  $\mathbb{E}[z] = 0$  and  $\sigma_z^2 = 0$  and, hence, (26) reverts to (21).

#### E. Distortion Analysis: Scenario with Idle sensors

For this scenario, we have  $\xi_i = 0$  for  $i = 1 \dots N$  (i.e., no phase synchronization errors) whereas the Bernoulli random variables  $b_i$  take different values over time according to sensor activity. A naive communication scheme for this scenario would consist in simply letting  $w'_{i,j}$  be the coefficients of the KL transform associated to the covariance matrix of the *entire* set of sensors  $\mathcal{S}$  (namely,  $w'_{i,j} = \phi_{i,j}$ ). Yet computationally efficient, this approach results into some mismatch between the designed and effective beampatterns since only a subset of sensors will actually transmit data. By defining  $z_j \triangleq b_j - 1$  and along the lines of (14)–(21), one can find that the distortion for this scenario is given by

$$D = \sigma_x^2 - \frac{1}{N} \sum_{i=1}^N \frac{\rho_i p^2 \lambda_i^2}{\rho_i p^2 \lambda_i + \rho_i \sigma_x^2 p (1-p) + \sigma_n^2} \quad (28)$$

which, as expected, is a monotonically decreasing function in  $p$ .

Alternatively, we ask ourselves whether it pays off to specifically design the  $w'_{i,j}$  weights for *each* realization of the subset of active sensors<sup>6</sup>. To answer that, we resort to the *partial* KL transform (pKLT) of [21] that we particularize for the problem at hand. From [21], the best  $N$ -length representation of the observation vector associated to the subset of active sensors  $\mathbf{x}_A$  is given by its pKLT, namely,

$$\mathbf{y} = \Psi^T \begin{pmatrix} \mathbf{I}_{N_A} \\ \mathbf{C}_{\mathbf{x}_A^c \mathbf{x}_A} \mathbf{C}_{\mathbf{x}_A}^{-1} \end{pmatrix} \mathbf{x}_A, \quad (29)$$

with  $\mathbf{I}_{N_A}$  standing for the  $N_A \times N_A$  identity matrix,  $\mathbf{C}_{\mathbf{x}_A^c \mathbf{x}_A} = \mathbb{E}[\mathbf{x}_A^c \mathbf{x}_A^T]$  and  $\mathbf{C}_{\mathbf{x}_A \mathbf{x}_A} = \mathbb{E}[\mathbf{x}_A \mathbf{x}_A^T]$ . Besides,  $\Psi$  is the  $N \times N$  unitary matrix of the eigenvectors of the covariance matrix  $\mathbf{C}_{\tilde{\mathbf{x}}}$ , namely,

$$\mathbf{C}_{\tilde{\mathbf{x}}} \triangleq \begin{pmatrix} \mathbf{C}_{\mathbf{x}_A} & \mathbf{C}_{\mathbf{x}_A \mathbf{x}_A^c} \\ \mathbf{C}_{\mathbf{x}_A^c \mathbf{x}_A} & \mathbf{C}_{\mathbf{x}_A^c \mathbf{x}_A^c} \end{pmatrix} \quad (30)$$

$$= \Psi \Delta \Psi^T \quad (31)$$

and  $\Delta = \text{diag}[\delta_1, \dots, \delta_N]$  is a diagonal matrix with the corresponding eigenvalues. Interestingly, from the product of the last two terms on the left handside of (29) we can define a new vector  $\tilde{\mathbf{x}}$

$$\tilde{\mathbf{x}} \triangleq \begin{bmatrix} \mathbf{x}_A \\ \tilde{\mathbf{x}}_A^c \end{bmatrix} = \begin{pmatrix} \mathbf{I}_{N_A} \\ \mathbf{C}_{\mathbf{x}_A^c \mathbf{x}_A} \mathbf{C}_{\mathbf{x}_A}^{-1} \end{pmatrix} \mathbf{x}_A, \quad (32)$$

which can be regarded as the best estimate of  $\mathbf{x}$  given  $\mathbf{x}_A$  and the *known* spatial covariance matrix  $\mathbf{C}_{\mathbf{x}}$  (or, equivalently,  $\mathbf{C}_{\tilde{\mathbf{x}}}$ ).

<sup>6</sup>Sensor activity should be properly notified to the FC by means of a signaling channel.

In order to implement the partial KL transform in a distributed fashion, it suffices to define matrix  $\Upsilon = [v_1, \dots, v_N]$  as

$$\Upsilon^T \triangleq \Psi^T \begin{pmatrix} \mathbf{I}_{N_A} \\ \mathbf{C}_{\mathbf{x}_A^c \mathbf{x}_A} \mathbf{C}_{\mathbf{x}_A}^{-1} \end{pmatrix} \quad (33)$$

and then let the transmit weights be  $w'_{i,j} = v_{i,j}$  for  $i, j = 1, \dots, N$ , where  $v_{i,j}$  stands for the  $j$ -th element of vector  $v_i$ . Following a similar derivation to that in Section III-C, the distortion for the optimal MMSE estimator can be readily expressed as:

$$D = \sigma_x^2 - \frac{1}{N} \sum_{i=1}^N \frac{\rho_i \delta_i^2}{\rho_i \delta_i + \sigma_n^2}. \quad (34)$$

Interestingly, equation (34) is formally identical to (21) except for the fact that the eigenvalues used now are those of matrix  $\mathbf{C}_{\tilde{\mathbf{x}}}$  rather than  $\mathbf{C}_{\mathbf{x}}$ .

#### IV. COMPUTATION OF THE OPTIMAL POWER ALLOCATION

In order to compute the set of transmit weights  $w_{i,j}$  given in (7), we still need to determine the optimal power allocation  $\rho_i$  to the transmissions (eigenmodes). This will be done for two problems of interest, namely, (i) the minimization of the estimation distortion for a given sum-power constraint; and, (ii) the minimization of the transmit power for a given distortion target.

The distortion analysis conducted in Section III resulted into a number of analytical expressions of the reconstruction distortion for each scenario of interest. Interestingly, expressions (21), (26), (28) and (34) can be cast into the following general framework:

$$D = \sigma_x^2 - \frac{1}{N} \sum_{i=1}^N \frac{\rho_i a_i}{\rho_i b_i + \sigma_n^2}. \quad (35)$$

where the constants  $\{a_i\}_{i=1}^N$  and  $\{b_i\}_{i=1}^N$  for each scenario are specified in Table I.

As we will see next, this framework allows us to pose and solve the power allocation problem in a unified manner.

##### A. Minimization of Distortion

Here, our interest lies in finding the power allocation which minimizes the overall distortion for a given sum-power constraint. From (35) and (9), this problem can be posed as:

$$\min_{\rho_1, \dots, \rho_N} - \sum_{i=1}^N \frac{\rho_i a_i}{\rho_i b_i + \sigma_n^2} \quad (36)$$

$$\text{s.t.} \quad \sum_{i=1}^N \alpha_i \rho_i \leq P_t \quad (37)$$

$$-\rho_i \leq 0 \quad (38)$$

Note that the problem is convex since the Hessian of the objective function is semidefinite positive and the constraints are linear functions. The associated Lagrangian reads:

$$\begin{aligned}\mathcal{L}(\{\rho_i\}_{i=1}^N, \{\mu_i\}_{i=1}^N, \nu) \\ = - \sum_{i=1}^N \frac{\rho_i a_i}{\rho_i b_i + \sigma_n^2} - \sum_{i=1}^N \mu_i \rho_i + \nu \left( \sum_{i=1}^N \alpha_i \rho_i - P_t \right)\end{aligned}$$

TABLE I  
VALUES OF  $a_i$  AND  $b_i$  FOR THE SCENARIOS OF INTEREST.

Scenario	$a_i$	$b_i$
Baseline	$\lambda_i^2$	$\lambda_i$
Phase synchronization errors	$\lambda_i^2 \frac{\sin^2 \epsilon}{\epsilon^2}$	$\lambda_i \frac{\sin^2 \epsilon}{\epsilon^2} + \frac{1}{2} \frac{\epsilon \cos(\epsilon) \sin(\epsilon) + \epsilon^2 - 2 \sin^2(\epsilon)}{\epsilon^2} \sigma_x^2$
Active and idle sensors - KLT	$\lambda_i^2 p^2$	$\lambda_i p^2 + p(1-p) \sigma_x^2$
Active and idle sensors - Partial KLT	$\delta_i^2$	$\delta_i$

where  $\nu$  and  $\{\mu_i\}_{i=1}^N$  stand for the Lagrange multipliers associated to the problem constraints (37) and (38), respectively. From all this, the KKT conditions [23] for this problem can be expressed as

$$\frac{\partial \mathcal{L}(\{\rho_i\}_{i=1}^N, \{\mu_i\}_{i=1}^N, \nu)}{\partial \rho_k} = 0 \quad ; \quad k = 1, \dots, N \quad (39)$$

$$\sum_{i=1}^N \alpha_i \rho_i^* - P_t \leq 0 \quad ; \quad k = 1, \dots, N \quad (40)$$

$$-\rho_i^* \leq 0 \quad ; \quad k = 1, \dots, N \quad (41)$$

$$\mu_i^* \rho_i^* = 0 \quad ; \quad k = 1, \dots, N \quad (42)$$

$$\nu^* \left( \sum_{i=1}^N \alpha_i \rho_i^* - P_T \right) = 0 \quad (43)$$

$$\mu_i^* \geq 0 \quad ; \quad k = 1, \dots, N \quad (44)$$

$$\nu^* \geq 0 \quad ; \quad k = 1, \dots, N \quad (45)$$

From (42), for positive values of the optimal solution ( $\rho_i^* > 0$ ) we have necessarily that  $\mu_i^* = 0$ . Bearing this in mind, from (39) and (41) we conclude that the optimal power allocation to transmissions is given by a waterfilling-like solution, namely,

$$\rho_k^* = \left[ \frac{\sqrt{a_k \sigma_n^2}}{b_k \sqrt{\nu^* \alpha_k}} - \frac{\sigma_n^2}{b_k} \right]^+ \quad ; \quad k = 1, \dots, N, \quad (46)$$

where we have defined  $[x]^+ \triangleq \max\{x, 0\}$ . The optimal Lagrange multiplier  $\nu^*$  has to be computed numerically in such a way that the total power constraint (37) is satisfied with equality (see [24, Algorithm 2], for further details). Interestingly, from equation (46) not only the optimal power allocation follows but also the optimal number of consecutive sensor-to-FC transmissions (i.e. no need for an explicit optimization on the number of transmissions).

### B. Minimization of Transmit Power

Next, we address the dual problem of minimizing the total transmit power for a prescribed distortion target, namely,

$$\min_{\rho_1, \dots, \rho_N} \sum_{i=1}^N \alpha_i \rho_i \quad (47)$$

$$\text{s.t.} \quad N \sigma_x^2 - \sum_{i=1}^N \frac{\rho_i a_i}{\rho_i b_i + \sigma_n^2} \leq D_T \quad (48)$$

$$\rho_i \geq 0 \quad \text{for } i = 1, \dots, N \quad (49)$$

where  $D_T$  stands for such distortion target. Again, this problem is convex since the cost function is linear and the

constraints are convex functions. Analogously to Section IV-A, the solution to this problem reads

$$\rho_k^* = \left[ \frac{\sigma_n \sqrt{\nu^* a_k \sigma_n^2}}{b_k \sqrt{\alpha_k}} - \frac{\sigma_n^2}{b_k} \right]^+ \quad ; \quad k = 1, \dots, N. \quad (50)$$

with  $\nu^*$  now denoting the Lagrange multiplier associated to the distortion target constraint (49).

### V. COMPRESSED TRANSMISSION VIA GREEDY ITERATIVE PRE-CODER DESIGN

In order to avoid the detrimental effects of per-sensor channel equalization in the presence of deep fades (see Section III-B), here we depart from the structure imposed to transmit weights in (7). This allows us to find a more general solution. Unfortunately, the problem to be solved is far more complex. For this reason, in the sequel we restrict our analysis to the baseline scenario of Section III-C. Specifically, in this section we adopt an iterative approach to pre-coder design in which distortion is minimized transmission after transmission by allocating transmit power in a greedy manner.

For the baseline scenario, the received signal (10) can be expressed in matrix notation as

$$\mathbf{r}_i = \mathbf{w}_i^H \mathbf{H} \mathbf{x} + n_i \quad ; \quad i = 1, \dots, I, \quad (51)$$

where, to recall,  $\mathbf{w}_i = [w_{i,1}, w_{i,2}, \dots, w_{i,N}]^T$  stands for the pre-coder to be designed; and the diagonal matrix  $\mathbf{H} = \text{diag}[h_1, h_2, \dots, h_N]$  gathers the corresponding sensor-to-FC channel coefficients which are assumed to be known. Besides, we let  $\mathbf{r}_{1:i-1} = [r_1, r_2, \dots, r_{i-1}]$  denote the vector with the first  $i-1$  elements (transmissions) in  $\mathbf{r}$ . From  $\mathbf{r}_{1:i-1}$ , the FC provides an MMSE estimate of the observations vector which is given by the posterior mean, namely,

$$\begin{aligned} \hat{\mathbf{x}}^{(i-1)} &= \mathbb{E}\{\mathbf{x} | r_1, r_2, \dots, r_{i-1}\} \\ &= \mathbf{C}_{\mathbf{x} \mathbf{r}_{1:i-1}} \mathbf{C}_{\mathbf{r}_{1:i-1}}^{-1} \mathbf{r}_{1:i-1}, \end{aligned} \quad (52)$$

The normalized average distortion after the  $(i-1)$ -th transmission thus reads:

$$D^{(i-1)} = \frac{1}{N} \text{Tr}(\mathbf{C}_{\mathbf{x} | \mathbf{r}_{1:i-1}}) \quad (53)$$

By increasing the number of transmissions by one, the current estimate of the random field can be successively refined as follows:

$$\hat{\mathbf{x}}^{(i)} = \mathbb{E}\{\mathbf{x} | r_1, r_2, \dots, r_{i-1}, r_i\} \quad (54)$$

$$= \mathbf{C}_{\mathbf{x} \mathbf{r}_{1:i}} \mathbf{C}_{\mathbf{r}_{1:i}}^{-1} \mathbf{r}_{1:i}. \quad (55)$$

Since  $\mathbf{x}$  and  $\mathbf{r}_{1:i}$  are jointly Gaussian, the following identity holds<sup>7</sup>[22]

$$\mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i}} = \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} - \frac{\text{Cov}(\mathbf{x}, r_i^* | \mathbf{r}_{1:i-1}) \text{Cov}(r_i^*, \mathbf{x}^H | \mathbf{r}_{1:i-1})}{\text{Var}(r_i | \mathbf{r}_{1:i-1})} \quad (56)$$

where

$$\text{Var}(r_i | \mathbf{r}_{1:i-1}) = \mathbf{w}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H \mathbf{w}_i + \sigma_n^2, \quad (57)$$

$$\text{Cov}(r_i^*, \mathbf{x}^H | \mathbf{r}_{1:i-1}) = \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H \mathbf{w}_i. \quad (58)$$

From (56) again, the distortion after the  $i$ -th transmission,  $D^{(i)} = \frac{1}{N} \text{Tr}(\mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i}})$ , can be recursively expressed as:

$$\begin{aligned} D^{(i)} &= D^{(i-1)} - \frac{1}{N} \text{Tr} \left( \frac{\mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H \mathbf{w}_i \mathbf{w}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}}{\mathbf{w}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H \mathbf{w}_i + \sigma_n^2} \right) \\ &= D^{(i-1)} - \frac{1}{N} \frac{\mathbf{w}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}^2 \mathbf{H}^H \mathbf{w}_i}{\mathbf{w}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H \mathbf{w}_i + \sigma_n^2}. \end{aligned} \quad (59)$$

As we will see next, this last expression allows us to find the  $i$ -th precoding vector such that it *successively* (and optimally) *refines* the previous estimate of the random field. In other words, the one which results into the lowest possible distortion  $D^{(i)}$  given  $D^{(i-1)}$ .

#### A. Optimal pre-coder design

From (59), the  $i$ -th pre-coding vector is given by the solution to the following optimization problem:

$$\max_{\mathbf{w}_i} \frac{\mathbf{w}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}^2 \mathbf{H}^H \mathbf{w}_i}{\mathbf{w}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H \mathbf{w}_i + \sigma_n^2} \quad (60)$$

$$\text{s.t.} \quad \|\mathbf{w}_i\|_2^2 \leq \rho_i \quad (61)$$

with  $\rho_i$  denoting the amplifying scaling factor for the  $i$ -th transmission. The transmit power at the  $i$ -th transmission then verifies:

$$\sum_{j=1}^N \mathbb{E} \left\{ |w_{i,j}^* x_j|^2 \right\} = \sigma_x^2 \|\mathbf{w}_i\|_2^2 \leq \rho_i \sigma_x^2. \quad (62)$$

Clearly, the optimal solution will satisfy the above power constraint with equality (see Appendix B) and, thus, the optimization problem can be re-written as

$$\max_{\tilde{\mathbf{w}}_i} \frac{\tilde{\mathbf{w}}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}^2 \mathbf{H}^H \tilde{\mathbf{w}}_i}{\tilde{\mathbf{w}}_i^H \left( \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H + \frac{\sigma_n^2}{\rho_i} \mathbf{I}_N \right) \tilde{\mathbf{w}}_i} \quad (63)$$

$$\text{s.t.} \quad \|\tilde{\mathbf{w}}_i\|_2^2 = 1 \quad (64)$$

where  $\tilde{\mathbf{w}}_i \triangleq \sqrt{\frac{1}{\rho_i}} \mathbf{w}_i$  is the *normalized* pre-coder and  $\mathbf{I}_N$  stands for the identity matrix of size  $N$ . The problem above corresponds to the maximization of the so-called *generalized* Rayleigh quotient [25] and, thus, the optimal normalized pre-coder is given by

$$\tilde{\mathbf{w}}_i^* = \mathbf{v}_{\max} \left( \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}^2 \mathbf{H}^H, \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H + \frac{\sigma_n^2}{\rho_i} \mathbf{I}_N \right) \quad (65)$$

<sup>7</sup>For  $i = 1$ , the term  $\mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}$  in (56) must be replaced by  $\mathbf{C}_{\mathbf{x}}$ .

where  $\mathbf{v}_{\max}(\mathbf{A}, \mathbf{B})$  stands for the generalized eigenvector associated to the largest generalized eigenvalue of matrices  $\mathbf{A}$  and  $\mathbf{B}$ . This last expression reveals that the pre-coder design and power allocation problems (to be addressed in the next subsection) are, in general, inter-twined:  $\tilde{\mathbf{w}}_i^*$  depends not only on the transmit power allocated to the  $i$ -th transmission (through  $\rho_i$ ) but, also, on the power allocated to all previous transmissions (through  $\mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}$ ).

#### B. Greedy power allocation

The optimal power allocation strategy  $\boldsymbol{\rho} = \{\rho_1, \dots, \rho_I\}$  can be found by solving

$$\min_{\rho_1, \dots, \rho_I, I} D^{(I)} \quad (66)$$

$$\text{s.t.} \quad \sigma_x^2 \sum_{i=1}^I \rho_i = P_t \quad (67)$$

with  $P_t$  denoting the total transmit power<sup>8</sup>. It is worth noting that the minimization is over the set of transmit powers  $\{\rho_i\}_{i=1}^I$  and the number of transmissions  $I$ . This, along with the coupling of the pre-coder design and power allocation problems, renders the problem not solvable analytically for the general case. However, a closed form solution exists for Gaussian channels, as the next section illustrates.

##### 1) Gaussian Channels:

A closed-form solution will be found in two steps. First, we propose an iterative (and greedy) algorithm. Not only shall we realize that this approach is optimal for Gaussian channels<sup>9</sup> but, also, the insights gained will allow us to propose an extension (and some justification) for the case with arbitrary channels addressed in Section V-B-2.

For Gaussian channels, we have  $\mathbf{H} = \mathbf{I}_N$  and, thus, equation (65) can be re-written as<sup>10</sup>

$$\tilde{\mathbf{w}}_l = \mathbf{v}_{\max} \left( \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:l-1}}^2, \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:l-1}} + \frac{\sigma_n^2}{\rho_l} \mathbf{I}_N \right) \quad (68)$$

$$= \mathbf{v}_{\max} (\mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:l-1}}) \quad (69)$$

where the second equality follows from elementary properties of matrix algebra. Unlike in the case with arbitrary channels, the design of the normalized pre-coder  $\tilde{\mathbf{w}}_l^*$  here is no longer coupled with the power to be allocated to the  $l$ -th transmission itself. This considerably simplifies the problem at hand. In order to simultaneously solve the pre-coder design and power allocation problems, we propose to *iteratively* allocate transmit power in a greedy manner. To that aim, we define a *power token*  $\epsilon$  as an indivisible and (sufficiently) small fraction of the total transmit power, namely  $\epsilon \triangleq P_t \sigma_x^2 / L$ , where  $L \gg 1$  stands for the total number of power tokens or iterations. For the first iteration ( $l = 1$ ), it follows from (69) that  $\tilde{\mathbf{w}}_1^* = \phi_1$ , that is, the eigenvector associated to  $\lambda_1$ , the largest eigenvalue

<sup>8</sup>Again, the optimal solution will satisfy the power constraint with equality and the proof is along the lines of one in Appendix B.

<sup>9</sup>The scenario of Gaussian channels refers to the case where sensors-to-FC channels do not experience fading.

<sup>10</sup>For notational convenience, the transmission index  $i$  is replaced here by the iteration index  $l$  (see next paragraphs).



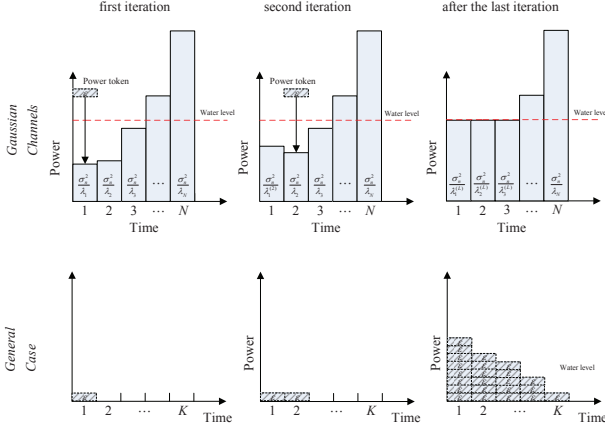


Fig. 2. Graphical representation of the greedy iterative power allocation scheme.

of  $\mathbf{C}_{\mathbf{x}}$  (to recall,  $\mathbf{C}_{\mathbf{x}} = \Phi \Lambda \Phi^T$  with  $\Phi = [\phi_1, \dots, \phi_N]$ ). From equations (56)-(57) and since  $\tilde{\mathbf{w}}_1^* = \phi_1$ , it follows that

$$\mathbf{C}_{\mathbf{x}|r_1} = \mathbf{C}_{\mathbf{x}} - \frac{\rho_1 \lambda_1^2}{\rho_1 \lambda_1 + \sigma_n^2} \phi_1 \phi_1^H \quad (70)$$

and, hence, the eigenvectors of matrices  $\mathbf{C}_{\mathbf{x}|r_1}$  and  $\mathbf{C}_{\mathbf{x}}$  are identical. Clearly, this also applies to all matrices  $\mathbf{C}_{\mathbf{x}|r_{1:l}}$  to be drawn in subsequent iterations (but not for the case with arbitrary channels, as will be discussed later). Since  $\rho_1 = \epsilon$ , from (70) we have that the eigenvalues of  $\mathbf{C}_{\mathbf{x}|r_1}$ , denoted by  $\lambda_1^{(1)}, \lambda_2^{(1)}, \dots, \lambda_N^{(1)}$  verify

$$\lambda_1^{(1)} = \lambda_1 - \frac{\epsilon \lambda_1^2}{\epsilon \lambda_1 + \sigma_n^2} \quad (71)$$

whereas  $\lambda_k^{(1)} = \lambda_k$  for all  $k \neq 1$ . The power token in the second iteration will be allocated to the eigenvector associated to the largest eigenvalue out of  $\lambda_1^{(1)} \dots \lambda_N^{(1)}$ . This iterative procedure is illustrated in Fig. 2. Note that, from (71),  $\lambda_1^{(1)}$  is not necessarily the largest eigenvalue of  $\mathbf{C}_{\mathbf{x}|r_1}$ . In this case, the power token for the second transmission would go to a so far inactive eigenvector/eigenmode (e.g.  $\lambda_2^{(1)}$  in Fig. 2). Otherwise, if  $\lambda_1^{(1)}$  continues to be the largest eigenvalue, we have  $\tilde{\mathbf{w}}_2^* = \phi_1$  again. Accordingly, it can be shown that the eigenvalue of the resulting covariance matrix  $\mathbf{C}_{\mathbf{x}|r_{1,2}}$  denoted by  $\lambda_1^{(1)}$  reads

$$\lambda_1^{(2)} = \lambda_1 - \frac{(\rho_1 + \rho_2) \lambda_1^2}{(\rho_1 + \rho_2) \lambda_1 + \sigma_n^2} \quad (72)$$

Clearly, this is equivalent to allocate a power of  $\rho_1 + \rho_2 = 2\epsilon$  and transmit just once with  $\tilde{\mathbf{w}}_1^* = \phi_1$ . After  $L$  iterations, and since the optimal pre-coders  $\tilde{\mathbf{w}}_l^*$  to be used in *any* transmission necessarily belong to the set of  $N$  eigenvectors of the *unconditional* covariance matrix  $\mathbf{C}_{\mathbf{x}}$ , this iterative scheme leads to the waterfilling solution of (46) (depicted in the rightmost plot in Figure 2). This holds true as long as the power tokens are *small* enough since this allows all the eigenmodes to accurately reach the *common* waterlevel.

## 2) Arbitrary Channels:

The iterative greedy algorithm to be presented here is largely inspired in that of Section V-B-1. However, the fact that  $\mathbf{H}$

is no longer an identity matrix has a substantial impact in the problem. Specifically,

- There is no straightforward relation between the solution to the generalized eigenvalue problem in (65) for different values of the transmission index  $i$  (or iteration index  $l$ ). Here, neither the eigenvectors in consecutive iterations are identical, nor only one of the eigenvalues changes over them. Essentially, they must be re-computed anew.
- The design of the normalized pre-coder for the  $l$ -th iteration does depend on its own power token (and preceding ones too).
- The problem requires an *explicit* optimization on  $I$  since it does not follow from the iterative power allocation or waterfilling scheme, as in the particular case for Gaussian channels. For simplicity, in the sequel we resort to an exhaustive search over  $I$ .
- The optimal number of transmissions  $I^*$  can (potentially) be larger than  $N$  since it is not upper bounded by the total number of different eigenvectors of  $\mathbf{C}_{\mathbf{x}}$ .

All this, in turn, calls for a number of adaptations in the iterative scheme. As in the Gaussian case, however, the transmit power is allocated to the set of pre-coders on a token by token basis. In addition, no changes of previously allocated tokens are allowed (not an exhaustive search over power allocations). For the sake of clarity, we introduce the shorthand notation  $\phi_1^{l-1}(\rho)$  to denote the eigenvector associated to the largest eigenvalue of the generalized eigenvalue problem in (65). The superscript  $l-1$  accounts for the number of conditioning elements in the covariance matrix  $\mathbf{C}_{\mathbf{x}|r_{1:l-1}}$  in (65), while  $\rho$  is the *cumulative* power allocated to such eigenvector (including the current iteration). So, herinafter we fix the number of transmissions  $I$  and describe the iterative scheme for the  $I = 3$  case:

**First iteration** ( $l = 1$ ): The first power token  $\epsilon$  is necessarily allocated to  $\phi_1^0(\epsilon)$ . It is retained as the best pre-coder/power allocation combination so far and, hence, will be part of all the combinations in subsequent iterations.

**Second iteration** ( $l = 2$ ): The allocation of the new power token results into two possible combinations of pre-coders and powers (i)  $\{\phi_1^0(\epsilon + \epsilon)\}$ , one transmission (pre-coder); or (ii)  $\{\phi_1^0(\epsilon), \phi_1^1(\epsilon)\}$ , two transmissions. The resulting distortion is then recursively computed for both combinations according to (59). Assume that (ii) attains the lowest distortion so far and, thus, this combination is retained.

**Third iteration** ( $l = 3$ ): There exist three possible combinations for the allocation of the new power token, namely, (i)  $\{\phi_1^0(\epsilon + \epsilon), \phi_1^1(\epsilon)\}$ , with two transmissions; or (ii)  $\{\phi_1^0(\epsilon), \phi_1^1(\epsilon + \epsilon)\}$ , two transmissions again; or (iii)  $\{\phi_1^0(\epsilon), \phi_1^1(\epsilon), \phi_1^2(\epsilon)\}$ , with three transmissions. Assume that (iii) attains the lowest distortion this causing the maximum number of transmissions ( $I=3$ ) to be reached. From now on, no additional eigenvectors will be tried in subsequent iterations. Still, the eigenvectors selected so far might need to be re-computed if any of the subsequent power tokens is allocated to them. The assumption here is that the greedy allocation of previous power tokens continues to be optimal for the re-computed eigenvectors, which is reasonable as long as  $\epsilon$  is small.



The algorithm goes on until the  $L$  power tokens have been allocated. The (at most)  $I$  eigenvectors retained in the last iteration will be used as the actual set of pre-coders  $\{\tilde{\mathbf{w}}_i\}_{i=1}^I$  along with the allocation of power tokens to such eigenvector set. Yet no optimality can be claimed for this approach, it exhibits a remarkable performance, as the next section illustrates.

## VI. NUMERICAL RESULTS

The simulation scenario consists of  $N$  sensors deployed at random locations within a  $d \times d$  rectangular area with a uniform distribution. As in [15], the spatially-correlated random field is modeled as a Gaussian Markov Ornstein-Uhlenbeck process with correlation (covariance) function given by  $k(\mathbf{s}_i, \mathbf{s}_j) = \sigma_x^2 \exp(-\theta \|\mathbf{s}_i - \mathbf{s}_j\|_2)$ . Clearly, the higher the parameter  $\theta$ , the lower the correlation among observations. In all cases, the variance of the random field and the additive noise read  $\sigma_x^2 = 1$  and  $\sigma_n^2 = 1$ , respectively. Results have been averaged over 1000 runs.

As a benchmark, we adopt an *uncompressed* transmission scheme whereby only *one* sensor sends data to the FC in each transmission period (i.e. Time Division Multiple Access, TDMA). For instance, if sensor  $i$  sends its observation to the FC at time instant  $j = i$  then, in agreement with (7), we have

$$w'_{i,j} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i, j = 1, \dots, N \quad (73)$$

instead of using the elements in the corresponding eigenvector  $\phi_i$ . Further, we let  $\rho_i = \rho = P_t / \sum_{j=1}^N \frac{1}{|h_j|^2}$  for  $i = 1, \dots, N$ , that is, we do not favor any observation in particular (to be consistent with the fact that spatial correlation is not exploited in any way). As in all previous cases, to reconstruct the random field at the FC the corresponding MMSE estimator is used.

### A. Compressed transmission via the KL Transform

In Fig. 3, we depict the average distortion in the estimated random field as a function of the available transmit power for the baseline scenario with *Gaussian* channels. In both compressed and uncompressed transmission cases, distortion is a monotonically decreasing function of the transmit power. However, the higher the correlation (i.e. smaller  $\theta$ ), the larger the gap between both curves (some 3-4 dB for  $\theta = 10^{-3}$ ). This, on the one hand, is due to the fact that KL-based scheme avoids sending redundant information by just activating a *limited* number of eigenmodes ( $I \leq N$ ) rather than forcing the total number of transmissions to be equal to  $N$  (one per sensor, as in uncompressed transmission). On the other, with uncompressed transmission, we restrict ourselves to *uniformly* allocate power over sensors which is only optimal for uncorrelated observations (see curves labeled with  $\theta \rightarrow \infty$ ).

Complementarily, in Fig. 4, we present some additional results for the baseline scenario with *Rayleigh-fading* channels. With respect to the case with Gaussian channels, distortion is higher here. This results from the definition of the transmit weights in (7) by which part of the transmit power must be used for channel equalization. In Rayleigh-fading scenarios, some channel coefficients  $|h_j|^2$  eventually vanish (while

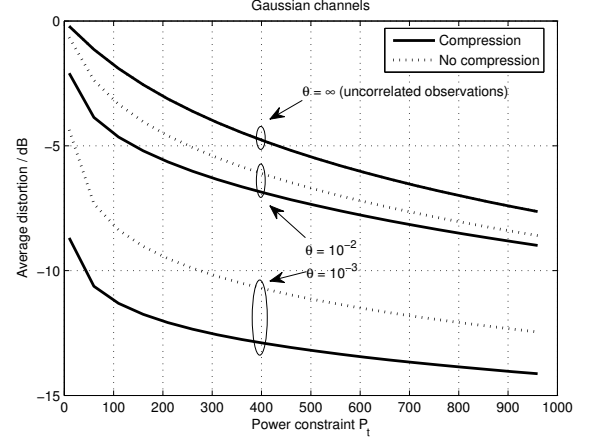


Fig. 3. Distortion vs. transmit power for Gaussian channels ( $N = 200$ ,  $d = 1000$ ,  $\sigma_n^2 = 1$ ,  $\sigma_x^2 = 1$ ).

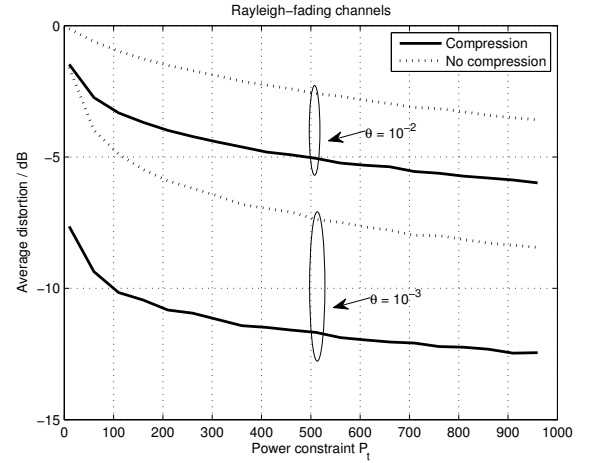


Fig. 4. Distortion vs. transmit power for Rayleigh-fading channels ( $N = 200$ ,  $d = 1000$ ,  $\sigma_n^2 = 1$ ,  $\sigma_x^2 = 1$ ).

$|h_j|^2 = 1$  for all  $j$  for Gaussian channels), the parameter  $\alpha_i = \sum_{j=1}^N \frac{1}{|h_j|^2} \phi_{i,j}^2 \sigma_x^2$  increases which, in turn, forces  $\rho_i$  to take smaller values (see e.g. equation (37)). Being distortion a monotonically decreasing function in  $\rho_i$ , this translates into a larger distortion.

Figure 5 provides further insights on the average number of active eigenmodes (or, equivalently, consecutive transmissions needed, or transmission latency) with the KLT-based transmission scheme. The *compression* level can be defined as the ratio between the number of active eigenmodes, and the number of observations  $N$ , which equals the maximum number of transmissions. Unsurprisingly, as the transmit power  $P_t$  increases, so does the number of active eigenmodes. Conversely, when transmit power is scarce, one cannot but compress the observations in a few eigenmodes (see curves around  $P_t = 0$ ). For a given transmit power, the higher the correlation (i.e. smaller  $\theta$ ), the lower the number of active eigenmodes. In other words, it is better to allocate some power to the eigenmodes that actually convey non-redundant information (i.e. the ones with larger eigenvalues  $\lambda_i$ ), which for  $\theta = 10^{-3}$  amount to a

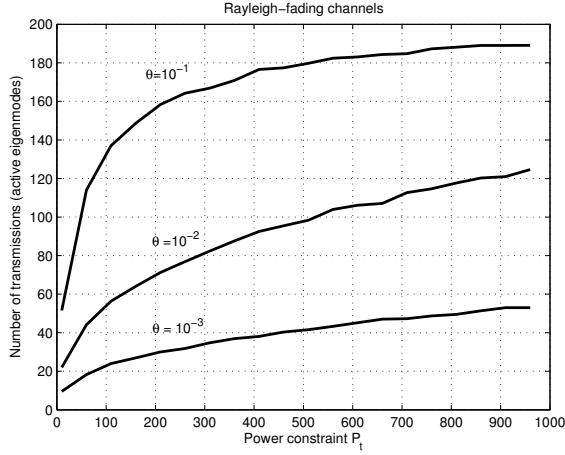


Fig. 5. Number of consecutive transmission vs. transmit power ( $N = 200$ ,  $d = 1000$ ,  $\sigma_n^2 = 1$ ,  $\sigma_x^2 = 1$ ).

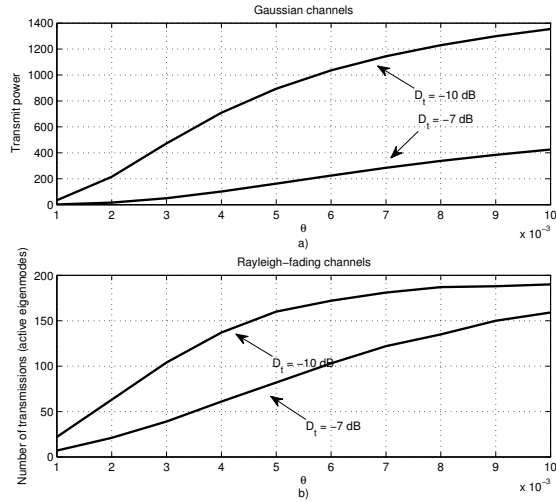


Fig. 6. a) Transmit power vs.  $\theta$  and b) Number of transmissions vs.  $\theta$  ( $N = 200$ ,  $d = 1000$ ,  $\sigma_n^2 = 1$ ,  $\sigma_x^2 = 1$ ).

mere  $\frac{55}{200} = 27.5\%$  ( $P_t = 1000$ ).

Next, in Fig. 6, we show some additional results for the dual problem of minimizing the transmit power for a given distortion target (baseline scenario). The upper plot reveals that a 3 dB reduction in distortion (from  $D_T = -7$  to  $D_T = -10$  dB) entails a substantial increase of the transmit power needed (5-fold for  $\theta = 5 \cdot 10^{-3}$ , 3-fold for  $\theta = 10^{-2}$ ). This suggests that distortion decreases logarithmically in the transmit power (see also Fig. 3). The lower plot, instead, illustrates the mechanisms used to meet the distortion target. For the range of high-to-moderate correlations ( $\theta = 2 \sim 6 \cdot 10^{-3}$ ), the number of active eigenmodes is substantially increased. Conversely, for low correlation values ( $10^{-2}$ ) the gap between the curves diminishes. Since most of the eigenmodes were already active for  $D_T = 7$  dB, here distortion can only be further reduced by increasing the SNR in the sensor-to-FC channels.

Regarding scenarios *with* phase synchronization errors, in

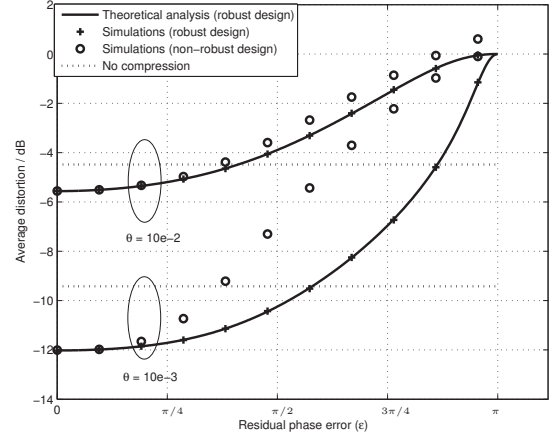


Fig. 7. Distortion vs. residual phase errors in Gaussian channels ( $N = 200$ ,  $d = 1000$ ,  $P_t = 200$ ,  $\sigma_n^2 = 1$ ,  $\sigma_x^2 = 1$ ).

Fig. 7, we depict the average distortion in the estimated random field as a function of the dynamic range of the residual phase error (i.e.  $\xi \in (-\epsilon, \epsilon)$ ), for Gaussian channels. For benchmarking purposes, we also show results for (i) a system with *uncompressed* transmissions in which, as discussed above, only *one* sensor sends data to the FC in each transmission period<sup>11</sup>; and (ii) a *non-robust* design for which, erroneously, we assume that sensors are perfectly synchronized and, thus, we let  $\xi = 0$  at the time of computing the power allocation. Several comments are in line. First, there is a perfect match between theoretical results (i.e. the distortion in (26)) and computer simulation results. Unsurprisingly, distortion is a monotonically increasing function in  $\epsilon$  since residual phase errors always have a negative impact in the beamforming gain. Still, the proposed (robust) design leverages on the statistical knowledge on such phase errors, allocates power to the various eigenmodes accordingly and, by doing so, it partly mitigates their effects. As a result, the robust scheme outperforms the non-robust one in all cases (up to 4 dB distortion gap for intermediate values of  $\epsilon$ ). Interestingly, such statistical knowledge allows us to preserve the beamforming/compression gains (i.e. crossing points with dotted lines) for large values of the dynamic margin of residual phase errors (e.g.  $\epsilon = 7\pi/12$  for  $\theta = 10^{-3}$ ). On the contrary, performance rapidly degrades for the non-robust design, as the gap between the curve with circles and the solid one illustrates (in particular, for small  $\theta$ ).

Figure 8 provides further insights on the average number of active eigenmodes (transmissions) in scenarios with phase synchronization errors. Interestingly, when phase errors increase, it is better to allocate power to a reduced number of eigenmodes (i.e. the ones with larger eigenvalues  $\lambda_i$ ), rather than spreading power thinner to weaker eigenmodes the performance of which will be seriously impaired by phase uncertainties. Same thing applies to scenarios where transmit power is scarce (curves for  $P_t = 50$ ).

<sup>11</sup>Here, we assume that, in the absence of beamforming,  $\xi_j$  can be perfectly estimated for each sensor.

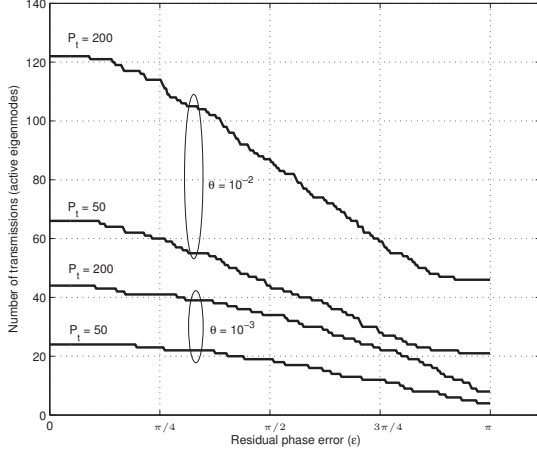


Fig. 8. Number of consecutive transmission vs. residual phase errors in Gaussian channels ( $N = 200$ ,  $d = 1000$ ,  $\sigma_n^2 = 1$ ,  $\sigma_x^2 = 1$ ).

Next, we focus on scenarios with active and idle sensors. In Fig. 9a, we depict the average distortion in the estimated random field as a function of the sensor duty-cycle  $p$ . We show results both for the KLT- and pKLT-based schemes. Unsurprisingly, distortion is a monotonically decreasing function in  $p$ . By increasing  $p$  the active number of sensors increases on average and, hence, the beamforming gain is larger. Besides, the pKLT-based scheme outperforms the KLT-based one for all  $p$ . This stems from the fact that pKLT exploits the information on the subset of active sensors in each realization not only for power allocation purposes but also for the computation of the *compression* term ( $w_{ij}^c$ ) in the beamforming weights (and, consequently, the design of the corresponding estimator in the FC). Despite of the computational burden that pKLT entails, the gain can be regarded as moderate for highly correlated fields (2 dB for 30% duty cycle when  $\theta = 10^{-3}$ ) or negligible in fields with lower correlation values ( $\theta = 10^{-2}$ ).

Complementarily, Fig. 9b shows some results on the average number of active eigenmodes. The impact of duty cycle in the KLT- and pKLT-based approaches is radically different. In the KLT case, the power allocation scheme activates more eigenmodes for small  $p$ . In this way, it attempts to partly compensate for the increasing distortion in the beampatterns by sending more (in principle redundant) information. On the contrary, the pKLT scheme tends to activate less eigenmodes. This is attributed to the fact that, when  $p$  diminishes, the attempt to infer the observations of idle sensors by conditioning on those of the active ones (the  $\tilde{x}_A^c$  term in equation (32)) is less accurate. Likewise, matrix  $\mathbf{C}_{\tilde{x}}$  becomes substantially different from  $\mathbf{C}_x$  this having a direct impact on the designed eigenbeamformers. Their *combined* effects result into a reduction of the number of active modes (i.e. focusing on the more reliable ones). Interestingly, the gap is larger for scenarios with low correlation (large  $\theta$ ) since the difficulty to infer unknown data is higher.

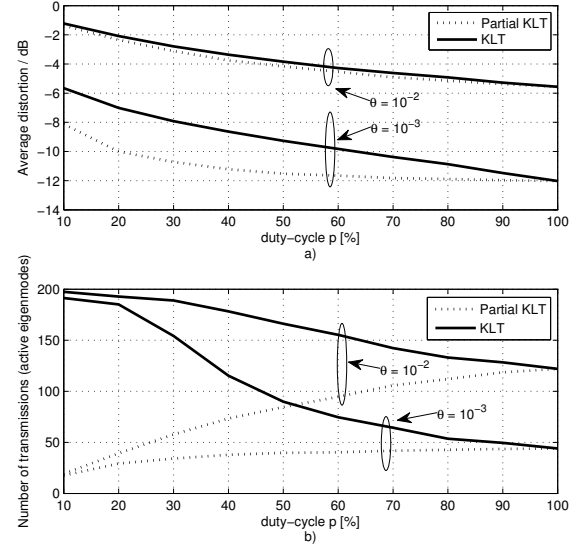


Fig. 9. a) Average distortion vs duty-cycle  $p$  [%] and b) Number of transmissions vs duty-cycle  $p$  [%] ( $N = 200$ ,  $d = 1000$ ,  $\sigma_n^2 = 1$ ,  $\sigma_x^2 = 1$ )

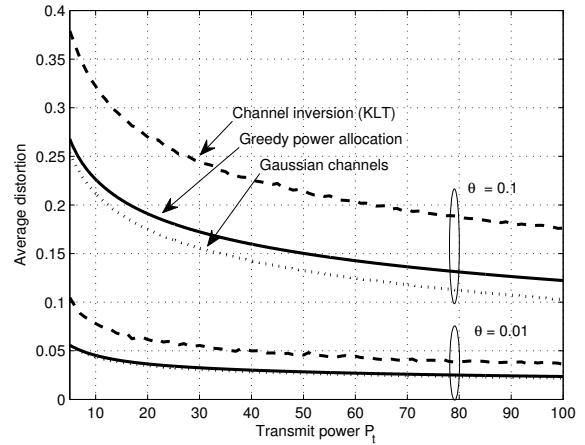


Fig. 10. Average distortion vs transmit power ( $N = 25$ ,  $d = 10$ )

### B. Compressed transmission via Greedy Pre-coder Design

Next, we show some results for the greedy (iterative) pre-coder solution presented in Section V for Rayleigh-fading channel scenarios. In Figure 10, we depict the reconstruction distortion averaged over channel realizations. Clearly, the pre-coder obtained with the greedy approach outperforms the KL-based one with per-sensor channel equalization (curves labeled with 'Channel Inversion KLT') for all values of correlation and total transmit power. As a second benchmark, we also depict the distortion attained by this scheme with Gaussian channels. Interestingly, Rayleigh fading has a marginal impact in the performance of the pre-coder obtained with the greedy approach (almost unnoticeable for  $\theta = 10^{-2}$ ). On the contrary, the performance gap with the KL-based scheme is substantial.

Finally, Figure 11 provides further insights on the behavior of the compressed transmission scheme via greedy pre-coder design. First, we observe that the performance of the proposed

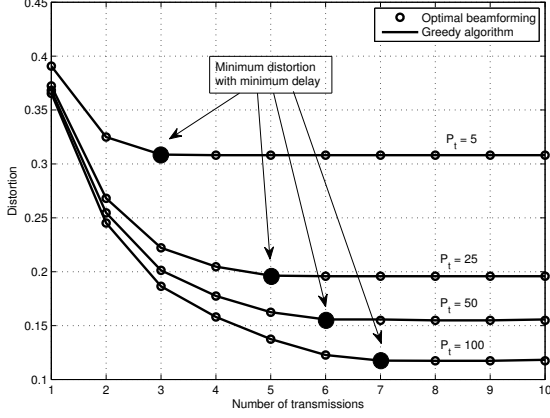


Fig. 11. Distortion vs transmission number ( $N=10$ ,  $d=10$ ,  $\theta=10^{-3}$ )

solution is virtually identical to that of the optimal one which can only be computed numerically. As it follows from (59), distortion decreases with the number of transmissions although beyond some point (big round markers on the curves) all curves saturate. This corresponds to the solution with the highest compression level (or, equivalently, lowest latency) for a given transmit power. In a practical implementation, the exhaustive search over  $I$  would be stopped as soon as e.g. the decrease in distortion with respect to the previous value would be within a prescribed margin. Yet not constrained by the problem formulation itself, the optimal (i.e. smallest) number of pre-coders was always found to be smaller than  $N$  for all practical values of  $\theta$  and  $P_t$ . In other words, the distributed pre-coding scheme actually achieves compression. Of course, by increasing the total amount of transmit power available, a larger number of "useful" transmissions which effectively convey non-redundant information to the FC *can* be afforded.

## VII. CONCLUSIONS

In this paper, we have first proposed a family of distributed pre-coding schemes based on the KL and partial KL transforms. We have derived, in a unified manner, closed form expressions of the optimal power allocation strategies for two problems of interest, namely, the minimization of the estimation error at the fusion center and the minimization of the transmit power at the sensor nodes; in three different scenarios encompassing residual phase synchronization errors or idle sensors. The optimal power allocation strategy designed in the presence of residual phase synchronization errors is *robust* since statistical knowledge on such errors is exploited.

The aforementioned KL-based techniques require per-sensor channel equalization. Hence, in Rayleigh fading channels a substantial part of the available transmit power must be used just to equalize channel fades. To avoid that, we have also proposed an iterative and greedy scheme by which the pre-coder design and power allocation problems can be simultaneously (and effectively) solved for the *general case*.

Computer simulation results revealed that, in the baseline scenario, the KL-based pre-coding scheme outperforms uncompressed transmission by some 3-4 dB ( $\theta=10^{-3}$ ,

Gaussian channels). This is due to the fact the KL transform avoids sending redundant information by activating a limited number of eigenmodes to which power is optimally allocated. Specifically, the compression level of the proposed pre-coding scheme reaches 27.5% for  $\theta=10^{-3}$ . In Rayleigh-fading channels, distortion is, as expected, higher. Interestingly, the performance of the robust version of the KL-based pre-coding scheme degrades gracefully with residual phase errors. The distortion gap with respect to non-robust designs is substantial, in particular for highly correlated observations ( $\theta=10^{-3}$ ) and intermediate values of phase uncertainties. For large residual phase errors, the robust scheme tends to reduce the number of active eigenmodes. We have also observed that diminishing the duty cycle has a negative impact in the performance of KL-based pre-coding schemes. This loss can be partly compensated by resorting to the partial KL transform, at the expense of a substantial increase in computational burden and associated signalling. Finally, we have learned that the pre-coders obtained with the greedy approach clearly outperform KL-based ones in terms of reconstruction distortion. Moreover, the performance associated to the solution given by the iterative and greedy approach is virtually identical to that of the optimal one which can only be computed numerically. Last, we have also found that distortion decreases with the number of transmissions although beyond some point it saturates. This corresponds to the solution with the highest compression level (lowest latency) for a given transmit power.

## APPENDIX A COMPUTATION OF MATRIX $\mathbf{C}_r$

In this appendix, we derive the covariance matrix  $\mathbf{C}_r$  that has been used for the computation of (28)

### A. Diagonal elements in $\mathbf{C}_r$

The diagonal elements of  $\mathbf{C}_{rr}$  can be computed as follows:

$$\begin{aligned} \mathbb{E}[r_i^2] &= \mathbb{E}_{y_i, \mathbf{z}, \mathbf{x}, n_i} \left[ \left( \sqrt{\rho_i} y_i + \sqrt{\rho_i} \sum_{j=1}^N \phi_{i,j} z_j x_j + n_i \right)^2 \right] \\ &= \underbrace{\rho_i \lambda_i + \sigma_n^2 + 2\rho_i \mathbb{E}_{y_i, \mathbf{z}, \mathbf{x}} \left[ y_i \sum_{j=1}^N \phi_{i,j} z_j x_j \right]}_{\beta_i} \end{aligned} \quad (74)$$

$$+ \underbrace{\rho_i \mathbb{E}_{\mathbf{z}, \mathbf{x}} \left[ \left( \sum_{j=1}^N \phi_{i,j} z_j x_j \right)^2 \right]}_{\theta_i} \quad (75)$$

where we have exploited, on the one hand, that  $n_i$  is statistically independent of  $y_i$  and  $x_i$ ; and, on the other, that  $x_i$  and

$n_i$  are zero-mean. As for  $\beta_i$ , we have

$$\beta_i = \mathbb{E}_{y_i, \mathbf{z}, \mathbf{x}} \left[ y_i \sum_{j=1}^N \phi_{i,j} z_j x_j \right] \quad (76)$$

$$= \mathbb{E}[z] \mathbb{E}_{y_i, \mathbf{x}} \left[ y_i \underbrace{\sum_{j=1}^N \phi_{i,j} x_j}_{y_i} \right] = \mathbb{E}[z] \lambda_i. \quad (77)$$

Next, the  $\theta_i$  term in (75) can be re-written as

$$\theta_i = \mathbb{E}_{\mathbf{z}, \mathbf{x}} \left[ \left( \sum_{j=1}^N \phi_{i,j} z_j x_j \right)^2 \right] = \mathbb{E}_{\mathbf{z}, \mathbf{x}} [\phi_i^T \mathbf{Z} \mathbf{x} \mathbf{x}^T \mathbf{Z} \phi_i]$$

with  $\mathbf{Z} \triangleq \text{diag}(z_1, z_2, \dots, z_N)$ . The random variables  $\mathbf{z}$  and  $\mathbf{x}$  are independent and, consequently,

$$\theta_i = \mathbb{E}_{\mathbf{z}, \mathbf{x}} [\phi_i^T \mathbf{Z} \mathbf{x} \mathbf{x}^T \mathbf{Z} \phi_i] = \mathbb{E}_{\mathbf{z}} [\phi_i^T \mathbf{Z} \mathbf{C}_{\mathbf{x}\mathbf{x}} \mathbf{Z} \phi_i] \quad (78)$$

From (4), we know that  $\mathbf{C}_{\mathbf{x}\mathbf{x}} = \sum_{k=1}^N \lambda_k \phi_k \phi_k^T$  and, hence,

$$\theta_i = \mathbb{E}_{\mathbf{z}} \left[ \sum_{k=1}^N \lambda_k \phi_i^T \mathbf{Z} \phi_k \phi_k^T \mathbf{Z} \phi_i \right] \quad (79)$$

$$= \sum_{k=1}^N \lambda_k \mathbb{E}_{\mathbf{z}} [\phi_i^T \mathbf{Z} \phi_k \phi_k^T \mathbf{Z} \phi_i] \quad (80)$$

Each expectation term in the last summation can be alternatively expressed as

$$\mathbb{E}_{\mathbf{z}} [\phi_i^T \mathbf{Z} \phi_k \phi_k^T \mathbf{Z} \phi_i] = \quad (81)$$

$$= \mathbb{E} \left[ \left( \sum_{l=1}^N \phi_{i,l} \phi_{k,l} z_l \right) \left( \sum_{l'=1}^N \phi_{i,l'} \phi_{k,l'} z_{l'}' \right) \right] \quad (82)$$

$$= \sum_{l=1}^N \sum_{l'=1}^N \phi_{i,l} \phi_{k,l} \phi_{i,l'} \phi_{k,l'} \mathbb{E}[z_l z_{l'}'] \quad (83)$$

$$= \sum_{l=1}^N \phi_{i,l}^2 \phi_{k,l}^2 \mathbb{E}[z^2] + \sum_{l=1}^N \sum_{\substack{l'=1 \\ l' \neq l}}^N \phi_{i,l} \phi_{k,l} \phi_{i,l'} \phi_{k,l'} \mathbb{E}^2[z]$$

where we have exploited that the random variables  $z_l$  are identically distributed. By letting  $\sigma_z^2 = \mathbb{E}[z^2] - \mathbb{E}^2[z]$  denote the variance of  $z$ , we re-write

$$\mathbb{E}_{\mathbf{z}} [\phi_i^T \mathbf{Z} \phi_k \phi_k^T \mathbf{Z} \phi_i] = \quad (84)$$

$$= \sum_{l=1}^N \sigma_z^2 \phi_{i,l}^2 \phi_{k,l}^2 + \sum_{l=1}^N \sum_{l'=1}^N \phi_{i,l} \phi_{k,l} \phi_{i,l'} \phi_{k,l'} \mathbb{E}^2[z] \quad (85)$$

$$= \sigma_z^2 \sum_{l=1}^N \phi_{i,l}^2 \phi_{k,l}^2 + \mathbb{E}^2[z] \phi_i^T \phi_k \phi_k^T \phi_i \quad (86)$$

$$= \sigma_z^2 \sum_{l=1}^N \phi_{i,l}^2 \phi_{k,l}^2 + \mathbb{E}^2[z] \delta[i - k] \quad (87)$$

with  $\delta[i - k]$  standing for the Kronecker's delta. Next, by replacing (87) into (80), it yields

$$\theta_i = \sum_{k=1}^N \lambda_k \left( \sigma_z^2 \sum_{l=1}^N \phi_{i,l}^2 \phi_{k,l}^2 + \mathbb{E}^2[z] \delta[i - k] \right) \quad (88)$$

$$= \sigma_z^2 \sum_{l=1}^N \phi_{i,l}^2 \sum_{k=1}^N \lambda_k \phi_{k,l}^2 + \mathbb{E}^2[z] \lambda_i. \quad (89)$$

In the last expression, the term  $\sum_{k=1}^N \lambda_k \phi_{k,l}^2$  equals the  $l$ -th diagonal element in  $\mathbf{C}_{\mathbf{x}\mathbf{x}}$ , namely  $\sum_{k=1}^N \lambda_k \phi_{k,l}^2 = \sigma_x^2$ . Besides,  $\sum_{l=1}^N \phi_{i,l}^2 = 1$  since  $\Phi$  is an *orthonormal* basis. Bearing all this in mind, the  $\theta_i$  term can be expressed as

$$\theta_i = \sigma_z^2 \sigma_x^2 + \mathbb{E}^2[z] \lambda_i. \quad (90)$$

Finally, by replacing (90) and (77) into (75), we have that

$$\mathbb{E}_{y_i, \mathbf{z}, \mathbf{x}} [r_i^2] = \rho_i \lambda_i (1 + \mathbb{E}[z])^2 + \rho_i \sigma_z^2 \sigma_x^2 + \sigma_n^2 \quad (91)$$

### B. Off-diagonal elements in $\mathbf{C}_{\mathbf{r}\mathbf{r}}$

The off-diagonal elements, can be computed as follows:

$$\begin{aligned} \mathbb{E}[r_i r_{i'}] &= \mathbb{E}_{y_i, y_{i'}, \mathbf{z}, \mathbf{x}, n_i, w_{i'}} [(\sqrt{\rho_i} y_i + n_i + \sqrt{\rho_i} \phi_i^T \mathbf{Z} \mathbf{x}) \\ &\quad (\sqrt{\rho_{i'}} y_{i'} + w_{i'} + \sqrt{\rho_{i'}} \phi_{i'}^T \mathbf{Z} \mathbf{x})] \\ &= \sqrt{\rho_{i'} \rho_i} \underbrace{\mathbb{E}_{y_i, \mathbf{z}, \mathbf{x}} [y_i \mathbf{x}^T \mathbf{Z} \phi_{i'}]}_{\beta_i} \\ &\quad + \sqrt{\rho_{i'} \rho_i} \underbrace{\mathbb{E}_{y_{i'}, \mathbf{z}, \mathbf{x}} [y_{i'} \mathbf{x}^T \mathbf{Z} \phi_i]}_{\beta_{i'}} \\ &\quad + \sqrt{\rho_{i'} \rho_i} \underbrace{\mathbb{E}_{\mathbf{z}, \mathbf{x}} [\phi_{i'}^T \mathbf{Z} \mathbf{x} \mathbf{x}^T \mathbf{Z} \phi_i]}_{\theta_i} \end{aligned} \quad (92)$$

By recalling that  $z_i$  are identically distributed and independent of  $y_i$  and  $\mathbf{x}$ , the term  $\beta_i$  reads

$$\beta_i = \mathbb{E}_{y_i, \mathbf{z}, \mathbf{x}} [y_i \mathbf{x}^T \mathbf{Z} \phi_{i'}] = \mathbb{E}[z] \mathbb{E}_{y_i, \mathbf{x}} [y_i \mathbf{x}^T \phi_{i'}] \quad (93)$$

$$= \mathbb{E}[z] \mathbb{E}_{y_i, y_{i'}} [y_i y_{i'}] = \mathbb{E}[z] \mathbb{E}[y_i] \mathbb{E}[y_{i'}] = 0 \quad (94)$$

since both  $y_i$  and  $y_{i'}$  are zero-mean. Likewise, we have that  $\beta_{i'} = 0$ . Next and analogously to (78)–(87), we have that

$$\theta_i = \mathbb{E}_{\mathbf{z}, \mathbf{x}} [\phi_i^T \mathbf{Z} \mathbf{x} \mathbf{x}^T \mathbf{Z} \phi_{i'}] \quad (95)$$

$$= \sigma_x^2 \sigma_z^2 \phi_i^T \phi_{i'} + \mathbb{E}[z^2] \sum_{k=1}^N \lambda_k \phi_i^T \phi_k \phi_k^T \phi_{i'} \quad (96)$$

$$= \sigma_x^2 \sigma_z^2 \delta[i - i'] + \mathbb{E}[z^2] \sum_{k=1}^N \lambda_k \delta[i - k] \delta[i' - k] \quad (97)$$

$$= 0 \quad (98)$$

### APPENDIX B LEMMA 1

This lemma proves that the optimal solution of (60)–(61) satisfies the power constraint of (61) with equality. To show

this, assume that the optimal solution  $\mathbf{w}_i^*$  is such that  $\|\mathbf{w}_i\|_2^2 = c_i < \rho_i$  and rewrite the objective function of (60) as follows:

$$f(\mathbf{w}_i^{*H}) = \frac{\mathbf{w}_i^{*H} \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}^2 \mathbf{H}^H \mathbf{w}_i^*}{\mathbf{w}_i^{*H} \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H \mathbf{w}_i^* + \sigma_n^2} \quad (99)$$

$$= \frac{c_i \tilde{\mathbf{w}}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}}^2 \mathbf{H}^H \tilde{\mathbf{w}}_i^*}{c_i \tilde{\mathbf{w}}_i^H \mathbf{H} \mathbf{C}_{\mathbf{x}|\mathbf{r}_{1:i-1}} \mathbf{H}^H \tilde{\mathbf{w}}_i^* + \sigma_n^2} \quad (100)$$

with  $\tilde{\mathbf{w}}_i^* \triangleq \sqrt{\frac{1}{c_i}} \mathbf{w}_i^*$ . Then, by noticing that (100) is increasing in  $c_i$ , we could always select  $\mathbf{w}_i' = \sqrt{\rho_i} \tilde{\mathbf{w}}_i^*$  and obtain  $f(\mathbf{w}_i') > f(\mathbf{w}_i^{*H})$  which contradicts the initial assumption. Therefore, the optimal solution of (60) will necessarily satisfy the power constraint with equality.

## REFERENCES

- [1] F. Xaver, G. Matz, P. Gerstoft, and C. Meckenbräuer, "Localization of acoustic sources using a decentralized particle filter," *EURASIP J. Wirel. Comm.*, pp. 1–14, Sep. 2011.
- [2] I. F. Akyildiz, W. Su, Y. Sankarsubramanian, and E. Cayirci, "Wireless sensor networks: A survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, Mar. 2002.
- [3] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 431–442, Feb. 2006.
- [4] S. Cui, J.-J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Energy-efficient joint estimation in sensor networks: analog vs. digital," in *IEEE ICASSP 2005*, Philadelphia, Mar. 2005, pp. 745–748.
- [5] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2210–2219, Jun. 2005.
- [6] M. Gastpar, "Uncoded transmission is exactly optimal for a simple gaussian sensor network," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 5247–5251, Nov. 2008.
- [7] M. Gastpar, M. Vetterli, and P. Dragotti, "Sensing reality and communicating bits: a dangerous liaison," *IEEE Signal Process. Mag.*, vol. 23, Jul. 2006.
- [8] R. Mudumbai, J. Hespanha, U. Madhow, and G. Barriac, "Distributed transmit beamforming using feedback control," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 411–426, Jan. 2010.
- [9] S. Cui, J.-J. Xiao, A. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4683–4695, Sep. 2007.
- [10] J.-J. Xiao and Z.-Q. Luo, "Decentralized estimation in an inhomogeneous sensing environment," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3564–3575, Oct. 2005.
- [11] —, "Multiterminal source channel communication over an orthogonal multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3255–3264, Sep. 2007.
- [12] J. Matamoros and C. Antón-Haro, "Optimal network size and encoding rate for wireless sensor network-based decentralized estimation under power and bandwidth constraints," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1121–1131, 2011.
- [13] A. Dogandžić and K. Qiu, "Decentralized random-field estimation for sensor networks using quantized spatially correlated data and fusion-center feedback," *IEEE Trans. Signal Process.*, vol. 56, no. 12, pp. 6069–6085, Dec. 2008.
- [14] M. C. Vuran and I. F. Akyildiz, "Spatial correlation-based collaborative medium access control in wireless sensor networks," *IEEE/ACM Transactions on Networking*, vol. 14, no. 2, pp. 316–329, 2006.
- [15] M. Dong, L. Tong, and B. Sadler, "Impact of data retrieval pattern on homogeneous signal field reconstruction in dense sensor networks," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4352–4364, Nov. 2006.
- [16] J. O. Berger, V. de Oliveira, and B. Sanso, "Objective bayesian analysis of spatially correlated data," *Journal of the American Statistical Association*, vol. 96, pp. 1361–1374, Dec. 2001.
- [17] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: A 'bit-conservation principle'," *Lecture Notes in Comput. Sci.*, vol. 2634/2003, pp. 17–31, Jan. 2003.
- [18] D. Dardari, A. Conti, C. Buratti, and R. Verdonesi, "Mathematical evaluation of environmental monitoring estimation error through energy-efficient wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 6, no. 7, pp. 790–802, Jul. 2007.
- [19] M. Gastpar and M. Vetterli, "Power, spatio-temporal bandwidth, and distortion in large sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 745–754, Apr. 2005.
- [20] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Joint source channel communication for distributed estimation in sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3629–3653, Oct. 2007.
- [21] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed Karhunen-Loève transform," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.
- [22] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice Hall Signal Processing Series, 1993.
- [23] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 1993.
- [24] D. P. Palomar and J. R. Fonollosa, "Practical algorithms for a family of waterfilling solutions," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 686–695, Feb. 2005.
- [25] B. Moghaddam, Y. Weiss, and S. Avidan, "Generalized spectral bounds for sparse LDA," in *Proceedings of the 23rd international conference on Machine learning - ICML '06*. ACM Press, 2006.



**Javier Matamoros** was born in Sabadell, Spain, in 1982. He received the M.Sc. and the Ph.D. degrees in Telecommunications Engineering from the Polytechnic University of Catalonia in 2005 and 2010, respectively. From February 2005 until August 2005 he was with the WMC (Wireless Mobile and Communication) group of TUDelft (Delft University of Technology). In January 2006, he was granted by the CTTC and in January 2008 by the Spanish Ministry of Science and Innovation to obtain his Ph.D degree. From August 2010, he holds a research

associate position at the CTTC in the Communications Systems Division. His primary research interests include sensor networks, wireless communications, distributed signal processing and smart grids.



**Carles Antón-Haro** (M'99-SM'03) received his Ph.D. degree in Telecommunications engineering from the Technical University of Catalonia (UPC) in 1998 (cum-laude). In 1999, he joined Ericsson Spain, where he participated in rollout projects of 2G and 3G mobile networks. Currently, he is with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) as a Director of R&D Programs and Senior Researcher. His research interests are in the area of signal processing for communications, this including MIMO, opportunistic communications, array signal processing, Smart Grids, M2M communications, and wireless sensor networks. He has published more than 20 technical papers in IEEE journals as well as over 80 papers in international and national conferences. Dr. Anton-Haro is a Senior Member of the IEEE.