Distortion-Rate Bounds for Distributed Estimation using Wireless Sensor Networks†

*Ioannis D. Schizas, Georgios B. Giannakis (contact author) and Nihar Jindal**

Abstract

We deal with centralized and distributed rate-constrained estimation of random signal vectors performed using a network of wireless sensors (encoders) communicating with a fusion center (decoder). For this context, we determine lower and upper bounds on the corresponding distortion-rate (D-R) function. The nonachievable lower bound is obtained by considering centralized estimation with a single-sensor which has available all observation data, and by determining the associated D-R function in closed-form. Interestingly, this D-R function can be achieved using an estimate first compress afterwards (EC) approach, where the sensor: i) forms the minimum mean-square error (MMSE) estimate for the signal of interest; and ii) optimally (in the MSE sense) compresses and transmits it to the FC that reconstructs it. We further derive a novel alternating scheme to numerically determine an achievable upper bound of the D-R function for general distributed estimation using multiple sensors. The proposed algorithm tackles an analytically intractable minimization problem, while it accounts for sensor data correlations. The obtained upper bound is tighter than the one determined by having each sensor performing MSE optimal encoding independently of the others. Numerical examples indicate that the algorithm performs well and yields D-R upper bounds which are relatively tight with respect to analytical alternatives obtained without taking into account the cross-correlations among sensor data.

∗ The authors are with the Dept. of Electrical and Computer Engineering, University of Minnesota, 200 Union Street SE, Minneapolis, MN 55455. Tel/fax: (612)626-7781/625-4583; Emails: {schizas,georgios,nihar}@ece.umn.edu

I. INTRODUCTION

Stringent bandwidth and energy constraints that wireless sensor networks (WSNs) must adhere to motivate efficient compression and encoding schemes when estimating random signals or parameter vectors of interest. In such networks, it is of paramount importance to determine bounds on the minimum achievable distortion between the signal of interest and its estimate formed at the fusion center (FC) using the encoded information transmitted by the sensors subject to rate constraints.

In the *reconstruction* scenario, the FC wishes to accurately reconstruct the sensor observations that are transmitted to the FC in a compressed form. In the *estimation* scenario, the FC is interested in accurately estimating an underlying random vector which is correlated with, but not equal to, the sensor observations. Thus, the FC utilizes the compressed sensor data to estimate a vector parameter which is conveyed implicitly by the sensor data. In a setup involving one sensor, single-letter characterizations of the D-R function for both scenarios are known: the reconstruction scenario is the standard distortion-rate problem [4, p. 336]; and the estimation one, also referred to as a rate-distortion problem with a *remote source*, has also been determined [1, p. 78]. In the distributed setup, involving multiple sensors with correlated observations, neither problem is well understood. The best analytical inner and outer bounds for the D-R function for reconstruction can be found in [2] and [16]; see also [17] that determines the rate-distortion region for a two-sensor setup. An iterative scheme has been developed in [6], which numerically determines an achievable upper bound for distributed reconstruction but not for signal estimation. The numerical D-R upper bound obtained by [6] is applicable when the signal to be reconstructed at the FC coincides with the sensor observations. However, this is not the case in the estimation setup considered here, where sensors observe a statistically perturbed version of the signal of interested that the FC wishes to reconstruct.

For the general problem of estimating a parameter vector with analog-amplitude entries correlated with sensor observations, most of the existing literature examines Gaussian data and Gaussian parameters. Specifically, when each sensor observes a common *scalar* random parameter contaminated with Gaussian noise, the D-R function for estimating this parameter has been determined in [3], [7]–[9], [15] to solve the so called Gaussian CEO problem. D-R bounds for a linear-Gaussian data model have been derived in [10] and [11] when the number of parameters equals the number of all scalar observations, with one scalar observation per sensor. Under a similar setup [17] determines the rate-distortion region in a *two-sensor* WSN. Another formulation was considered in [21], where each sensor has available a vector of observations having the same length as the parameter vector; see also [14] where a two-sensor

setup is considered. All existing formulations dealing with vectors of parameters and observations are special cases of the general *vector* Gaussian CEO problem. In this paper, we pursue D-R analysis for distributed estimation with WSNs, under the vector Gaussian CEO setup, without constraining the number of observations in each sensor and/or the number of random parameters to be estimated.

We first determine in closed-form the D-R function for estimating a parameter *vector* when applying rate-constrained encoding to the observation data collected by a single-sensor (Section III). Without assuming that the number of parameters equals the number of observations, we prove that the optimal scheme achieving the D-R function amounts to first computing the minimum mean-square error (MMSE) estimate of the source at the sensor, and then optimally compressing at the sensor and reconstructing at the FC the estimate via reverse water-filling (rwf). The D-R function for the single-sensor setup serves as a non-achievable lower D-R bound for rate-constrained estimation in the multi-sensor setup. Next, we develop an alternating scheme that numerically determines an achievable D-R upper bound for the multi-sensor scenario (Section IV). Using this iterative algorithm we can tackle an analytically intractable minimization problem and determine a D-R upper bound. Different from [6], which deals with WSN-based distributed reconstruction, our approach aims at general estimation problems where the parameters of interest are not directly observed at the sensors. Combining the lower bound of Section III with the numerically determined upper bound of Section IV, we specify a region where the D-R function for distributed estimation lies in.

II. PROBLEM STATEMENT

With reference to Fig. 1 (a), consider a WSN comprising L sensors that communicate with an FC. Each sensor, say the ith, observes an $N_i \times 1$ vector $\mathbf{x}_i(t)$ which is correlated with a $p \times 1$ random signal (parameter vector) of interest $\mathbf{s}(t)$, where t denotes discrete time. Similar to [8], [11], [15], we assume that:

(a1) *No information is exchanged among sensors and the links with the FC are noise-free.*

(a2) *The random vector $\mathbf{s}(t)$ is generated by a stationary Gaussian vector memoryless source with $\mathbf{s}(t) \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{ss}\right)$; the sensor data $\{\mathbf{x}_i(t)\}_{i=1}^{L}$ adhere to the linear-Gaussian model $\mathbf{x}_i(t) = \mathbf{H}_i \mathbf{s}(t) + \mathbf{n}_i(t)$, where $\mathbf{n}_i(t)$ denotes additive white Gaussian noise (AWGN); i.e., $\mathbf{n}_i(t) \sim \mathcal{N}\left(\mathbf{0}, \sigma^2 \mathbf{I}\right)$; noise $\mathbf{n}_i(t)$ is uncorrelated across sensors, across time, and with $\mathbf{s}$; and $\mathbf{H}_i$ as well as (cross-) covariance matrices $\boldsymbol{\Sigma}_{ss}$, $\boldsymbol{\Sigma}_{sx_i}$ and $\boldsymbol{\Sigma}_{x_i x_j}$ are known $\forall i, j \in \{1, \ldots, L\}$.*

Notice that (a1) holds when sufficiently strong channel codes are employed to cope with channel effects. Further, whiteness of $\mathbf{n}_i(t)$ and the zero-mean assumptions in (a2) are made without loss of

generality. The linear model in (a2) is commonly encountered in estimation and in a number of cases it even accurately approximates non-linear mappings; e.g., via a first-order Taylor expansion in target tracking applications. Although confining ourselves to Gaussian vectors $\mathbf{x}_i(t)$ is of interest on its own, following arguments similar to those in [1, p. 134], it can be shown that the D-R functions obtained in this paper upper bound their counterparts for non-Gaussian sensor data $\mathbf{x}_i(t)$ with (cross)-covariance matrices identical to those in (a2).

Blocks $\mathbf{x}_i^{(n)} := \{\mathbf{x}_i(t)\}_{t=1}^n$, comprising n consecutive time instantiations of the vector $\mathbf{x}_i(t)$, are encoded per sensor to yield each encoder's output $\mathbf{u}_i^{(n)} = \mathbf{f}_i^{(n)}(\mathbf{x}_i^{(n)})$, $i = 1, \ldots, L$. These encoded blocks are communicated through ideal orthogonal channels to the FC. There, $\mathbf{u}_i^{(n)}$'s are decoded to obtain an estimate of $\mathbf{s}^{(n)} := \{\mathbf{s}(t)\}_{t=1}^n$, which we denote as $\hat{\mathbf{s}}_R^{(n)}(\mathbf{u}_1^{(n)}, \ldots, \mathbf{u}_L^{(n)}) = \mathbf{g}_R^{(n)}(\mathbf{x}_1^{(n)}, \ldots, \mathbf{x}_L^{(n)})$, since $\mathbf{u}_i^{(n)}$ is a function of $\mathbf{x}_i^{(n)}$. The subscript R signifies the rate constraint which is imposed through a bound on the cardinality of the range of the sensor encoding functions; namely, the cardinality of the range of $\mathbf{f}_i^{(n)}$ must be no larger than 2^{nR_i}, where R_i is the available rate at the encoder of the ith sensor. The sum rate satisfies the constraint $\sum_{i=1}^L R_i \leq R$, where R is the total rate available for the L sensors. It is worth re-iterating, that this setup is precisely the vector Gaussian CEO problem in its most general form without any restrictions on the number of observations N_i at the ith sensor, and the number of random parameters p.

Under the sum rate constraint $\sum_{i=1}^L R_i \leq R$, the ultimate goal is to determine the minimum possible MSE distortion $(1/n)\sum_{t=1}^n E[\|\mathbf{s}(t) - \hat{\mathbf{s}}_R(t)\|^2]$ for estimating $\mathbf{s}$ in the limit of infinite block-length n. Such a (so called single-letter) characterization of the D-R function is available for the single-sensor case ($L = 1$), but not for the distributed multi-sensor scenario. For this reason, our objective in this paper is to derive (preferably tight) inner and outer bounds on the D-R function of the general vector CEO setup.

III. DISTORTION-RATE FOR CENTRALIZED ESTIMATION

We will first determine in closed form the D-R function for estimating $\mathbf{s}(t)$ in a *single-sensor* setup and provide a scheme that achieves it. The single-letter characterization of the D-R function in this setup allows us to drop the time index. Here, all observation data $\{\mathbf{x}_i\}_{i=1}^L := \mathbf{x}$, whose dimensionality is N, are available to a single sensor, and are related to the $p \times 1$ parameter vector $\mathbf{s}$ according to the linear model $\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}$. The D-R function in this setting provides a lower (non-achievable) bound on the MMSE that can be achieved in a multi-sensor distributed setup, where each $\mathbf{x}_i$ is observed and encoded by a different sensor. Existing works treat the case $N = p$ [5], [13], [19] and transform the D-R function with a remote
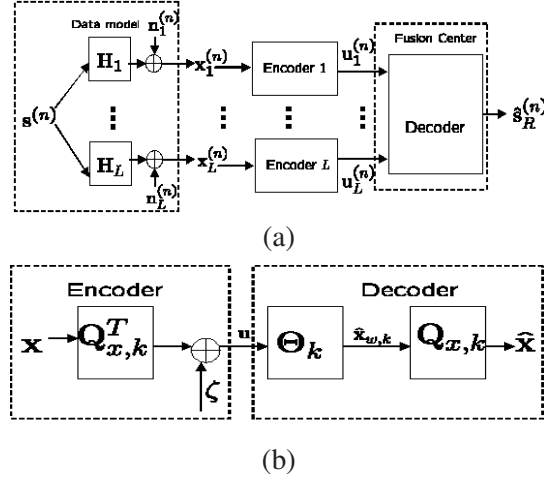
(a)

(b)

Fig. 1.   (a): Distributed setup; (b): Test channel for $\mathbf{x}$ Gaussian in a point-to-point link.

source to an ordinary reconstruction D-R problem; see also [18] which provides more general conditions under which this transformation is possible. Other works deal with practical encoding-decoding schemes using e.g., vector quantization [12]. However, here we look for the D-R function for general N and p, in the linear-Gaussian model framework.

A. Background on D-R for Reconstruction

The D-R function for encoding $\mathbf{x}$, which has probability density function (pdf) $p(\mathbf{x})$, with rate R at an individual sensor, and reconstructing it (in the MMSE sense) as $\hat{\mathbf{x}}$ at the FC, is given by [4, p. 342]

$$D_x(R) = \min_{\substack{p(\hat{\mathbf{x}}|\mathbf{x}) \\ I(\mathbf{x};\hat{\mathbf{x}}) \leq R}} E_{p(\hat{\mathbf{x}},\mathbf{x})}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2], \ \ \mathbf{x} \in \mathbb{R}^N, \ \ \hat{\mathbf{x}} \in \mathbb{R}^N \tag{1}$$

where the minimization is w.r.t. the conditional pdf $p(\hat{\mathbf{x}}|\mathbf{x})$. Let $\boldsymbol{\Sigma}_{xx} = \mathbf{Q}_x \boldsymbol{\Lambda}_x \mathbf{Q}_x^T$ denote the eigenvalue decomposition of $\boldsymbol{\Sigma}_{xx}$, where $\boldsymbol{\Lambda}_x = \text{diag}(\lambda_{x,1} \cdots \lambda_{x,N})$ and $\lambda_{x,1} \geq \cdots \geq \lambda_{x,N} > 0$.

For $\mathbf{x}$ Gaussian, $D_x(R)$ can be determined by applying rwf to the pre-whitened vector $\mathbf{x}_w := \mathbf{Q}_x^T \mathbf{x}$ [4, p. 348]. For a prescribed rate R, it turns out that $\exists\, k$ such that the first k entries $\{\mathbf{x}_w(i)\}_{i=1}^k$ of $\mathbf{x}_w$, are encoded and reconstructed independently from each other using rates $\{R_i = 0.5 \log_2 (\lambda_{x,i}/d(k,R))\}_{i=1}^k$, where $d(k,R) = \left(\prod_{i=1}^k \lambda_{x,i}\right)^{1/k} 2^{-2R/k}$; while the last $N-k$ entries of $\mathbf{x}_w$ are assigned no rate; i.e., $\{R_i = 0\}_{i=k+1}^N$ and $R = \sum_{i=1}^k R_i$. The corresponding MMSE for encoding $\mathbf{x}_w(i)$, under a rate constraint R_i, is $D_i := E[\|\mathbf{x}_w(i) - \hat{\mathbf{x}}_w(i)\|^2] = d(k,R)$ when $i = 1,\ldots,k$ and $D_i = \lambda_{x,i}$ when $i = k+1 \ldots, N$.

The resultant overall MMSE (D-R function) is

$$D_x(R) = E[\|\mathbf{x} - \hat{\mathbf{x}}\|^2] = E[\|\mathbf{x}_w - \hat{\mathbf{x}}_w\|^2] = kd(k, R) + \sum_{i=k+1}^{N} \lambda_{x,i} \qquad (2)$$

where $\hat{\mathbf{x}}_w = \mathbf{Q}_x^T \hat{\mathbf{x}}$. Note that $d(k, R)$ can be bounded as $\max(\{\lambda_{x,i}\}_{i=k+1}^N) \leq d(k, R) < \min(\{\lambda_{x,i}\}_{i=1}^k)$. Intuitively, $d(k, R)$ is a threshold distortion determining which entries of $\mathbf{x}_w$ are assigned with nonzero rate. The first k entries of $\mathbf{x}_w$ with variance $\lambda_{x,i} > d(k, R)$ are encoded with non-zero rate, but the last $N - k$ ones with variance $\lambda_{x,i} \leq d(k, R)$ are discarded in the encoding procedure (are set to zero).

Associated with the rwf principle is the so called test channel; see Fig. 1 (b) and e.g., [4, p. 345]. The encoder's MSE optimal output is $\mathbf{u} = \mathbf{Q}_{x,k}^T \mathbf{x} + \boldsymbol{\zeta}$, where $\mathbf{Q}_{x,k}$ is formed by the first k columns of $\mathbf{Q}_x$, and $\boldsymbol{\zeta}$ models the distortion noise that results due to the rate-constrained encoding of $\mathbf{x}$. The zero-mean AWGN $\boldsymbol{\zeta}$ is uncorrelated with $\mathbf{x}$ and its diagonal covariance matrix $\boldsymbol{\Sigma}_{\zeta\zeta}$ has entries $[\boldsymbol{\Sigma}_{\zeta\zeta}]_{ii} = \lambda_{x,i}D_i/(\lambda_{x,i} - D_i)$. The part of the test channel that takes as input $\mathbf{u}$ and outputs $\hat{\mathbf{x}}$, models the decoder. The reconstruction $\hat{\mathbf{x}}$ of $\mathbf{x}$ at the decoder output is

$$\hat{\mathbf{x}} = \mathbf{Q}_{x,k}\boldsymbol{\Theta}_k\mathbf{u} = \mathbf{Q}_{x,k}\boldsymbol{\Theta}_k\mathbf{Q}_{x,k}^T\mathbf{x} + \mathbf{Q}_{x,k}\boldsymbol{\Theta}_k\boldsymbol{\zeta} \qquad (3)$$

where $\boldsymbol{\Theta}_k$ is a diagonal matrix with non-zero entries $[\boldsymbol{\Theta}_k]_{ii} = (\lambda_{x,i} - D_i)/\lambda_{x,i}$, $i = 1, \ldots, k$.

B. D-R for Estimation

The D-R function for estimating source $\mathbf{s}$ given observation $\mathbf{x}$ (where the source and observation are probabilistically drawn from the joint pdf $p(\mathbf{x}, \mathbf{s})$) with rate R at an individual sensor, and reconstructing it (in the MMSE sense) as $\hat{\mathbf{s}}_R$ at the FC is given by [1, p. 79]

$$D_s(R) = \min_{\substack{p(\hat{\mathbf{s}}_R|\mathbf{x}) \\ I(\mathbf{x};\hat{\mathbf{s}}_R) \leq R}} E_{p(\hat{\mathbf{s}}_R,\mathbf{s})}[\|\mathbf{s} - \hat{\mathbf{s}}_R\|^2], \ \ \mathbf{s} \in \mathbb{R}^p, \ \ \hat{\mathbf{s}}_R \in \mathbb{R}^p \qquad (4)$$

where the minimization is w.r.t. the conditional pdf $p(\hat{\mathbf{s}}_R|\mathbf{x})$.

In order to achieve this D-R function, one might be tempted to first compress the observation $\mathbf{x}$ by applying rwf at the sensor, without taking into account the data model relating $\mathbf{s}$ with $\mathbf{x}$, and subsequently use the reconstructed $\hat{\mathbf{x}}$ to form the MMSE estimate $\hat{\mathbf{s}}_{ce} = E[\mathbf{s}|\hat{\mathbf{x}}]$ at the FC. An alternative option would be to first form the MMSE estimate $\hat{\mathbf{s}} = E[\mathbf{s}|\mathbf{x}]$, encode the latter using rwf at the sensor by exploiting only the covariance of $\hat{\mathbf{s}}$, and after decoding at the FC, obtain the reconstructed estimate $\hat{\mathbf{s}}_{ec}$. Referring to the former option as *Compress-Estimate* (CE) and to the latter as *Estimate-Compress* (EC), we are interested in determining which one yields the smallest MSE under a rate constraint R. Another interesting question is whether any of the CE and EC schemes enjoys MMSE optimality in the sense

of achieving (4). With subscripts ce and ec corresponding to these two options, let us also define the errors $\tilde{\mathbf{s}}_{ce} := \mathbf{s} - \hat{\mathbf{s}}_{ce}$ and $\tilde{\mathbf{s}}_{ec} := \mathbf{s} - \hat{\mathbf{s}}_{ec}$.

For CE, we depict in Fig. 2 (a) the test channel for encoding $\mathbf{x}$ via rwf, followed by MMSE estimation of $\mathbf{s}$ based on $\hat{\mathbf{x}}$. Suppose that when applying rwf to $\mathbf{x}$ with prescribed rate R, the first $k_{ce} \in \{1, \ldots, N\}$ components of $\mathbf{x}_w$ are assigned with non-zero rate and the rest are discarded. The MMSE optimal encoder's output for encoding $\mathbf{x}$ is then given, as in Section III.A, by $\mathbf{u}_{ce} = \mathbf{Q}_{x,k_{ce}}^T \mathbf{x} + \boldsymbol{\zeta}_{ce}$. The covariance matrix of $\boldsymbol{\zeta}_{ce}$ has diagonal entries $[\boldsymbol{\Sigma}_{\zeta_{ce}\zeta_{ce}}]_{ii} = \lambda_{x,i} D_i^{ce}/(\lambda_{x,i} - D_i^{ce})$ for $i = 1, \ldots, k_{ce}$, where $D_i^{ce} := E[(\mathbf{x}_w(i) - \hat{\mathbf{x}}_w(i))^2]$. Recalling that $D_i^{ce} = \left(\prod_{i=1}^{k_{ce}} \lambda_{x,i}\right)^{1/k_{ce}} 2^{-2R/k_{ce}}$ when $i = 1, \ldots, k_{ce}$ and $D_i^{ce} = \lambda_{x,i}$, when $i = k_{ce} + 1, \ldots, N$, the reconstructed $\hat{\mathbf{x}}$ in CE is [c.f. (3)]

$$\hat{\mathbf{x}} = \mathbf{Q}_{x,k_{ce}} \boldsymbol{\Theta}_{ce} \mathbf{Q}_{x,k_{ce}}^T \mathbf{x} + \mathbf{Q}_{x,k_{ce}} \boldsymbol{\Theta}_{ce} \boldsymbol{\zeta}_{ce} \tag{5}$$

where $[\boldsymbol{\Theta}_{ce}]_{ii} = (\lambda_{x,i} - D_i^{ce})/\lambda_{x,i}$, for $i = 1, \ldots, k_{ce}$. Letting $\check{\mathbf{x}} := \mathbf{Q}_x^T \hat{\mathbf{x}} = [\check{\mathbf{x}}_1^T \; \mathbf{0}_{1\times(N-k_{ce})}]^T$, with $\check{\mathbf{x}}_1 := \boldsymbol{\Theta}_{ce} \mathbf{Q}_{x,k_{ce}}^T \mathbf{x} + \boldsymbol{\Theta}_{ce} \boldsymbol{\zeta}_{ce}$, we have for the MMSE estimate

$$\hat{\mathbf{s}}_{ce} = E[\mathbf{s}|\hat{\mathbf{x}}] = E[\mathbf{s}|\mathbf{Q}_x^T \hat{\mathbf{x}}] = E[\mathbf{s}|\check{\mathbf{x}}_1] = \boldsymbol{\Sigma}_{s\check{x}_1} \boldsymbol{\Sigma}_{\check{x}_1\check{x}_1}^{-1} \check{\mathbf{x}}_1 \tag{6}$$

since $\mathbf{Q}_x^T$ is unitary and the last $N - k_{ce}$ entries of $\check{\mathbf{x}}$ are useless for estimating $\mathbf{s}$. We show in Appendix A that the covariance matrix $\boldsymbol{\Sigma}_{\tilde{s}_{ce}\tilde{s}_{ce}} := E[(\mathbf{s} - \hat{\mathbf{s}}_{ce})(\mathbf{s} - \hat{\mathbf{s}}_{ce})^T]$ of the estimation error $\tilde{\mathbf{s}}_{ce}$ is

$$\boldsymbol{\Sigma}_{\tilde{s}_{ce}\tilde{s}_{ce}} = \boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s\check{x}_1} \boldsymbol{\Sigma}_{\check{x}_1\check{x}_1}^{-1} \boldsymbol{\Sigma}_{\check{x}_1 s} = \boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{sx} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\Sigma}_{xs} + \boldsymbol{\Sigma}_{sx} \mathbf{Q}_x \boldsymbol{\Delta}_{ce} \mathbf{Q}_x^T \boldsymbol{\Sigma}_{xs} \tag{7}$$

where $\boldsymbol{\Delta}_{ce} := \mathrm{diag}\left(D_1^{ce}\lambda_{x,1}^{-2} \cdots D_N^{ce}\lambda_{x,N}^{-2}\right)$. Eqs. (6) and (7) characterize fully the CE scheme.
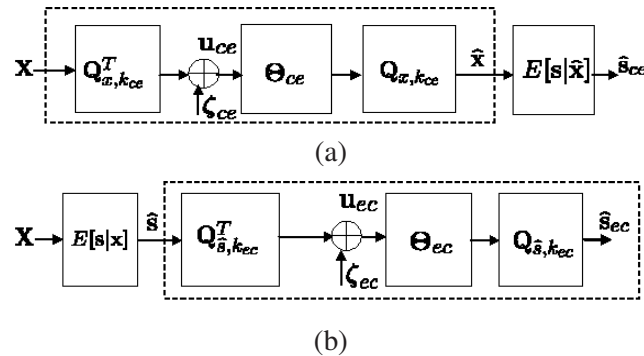

(a)

(b)

Fig. 2.   (a): Test channel for the CE scheme; (b): Test channel for the EC scheme.

In Fig. 2 (b) we depict the test channel for the EC scheme. The MMSE estimate $\hat{\mathbf{s}} = E[\mathbf{s}|\mathbf{x}]$ is followed by the test channel that results when applying rwf to a pre-whitened version of $\hat{\mathbf{s}}$, with rate R.

Let $\mathbf{\Sigma}_{\hat{s}\hat{s}} = \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xs} = \mathbf{Q}_{\hat{s}}\mathbf{\Lambda}_{\hat{s}}\mathbf{Q}_{\hat{s}}^T$ be the eigenvalue decomposition for the covariance matrix of $\hat{s}$, where $\mathbf{\Lambda}_{\hat{s}} = \text{diag}(\lambda_{\hat{s},1}\cdots\lambda_{\hat{s},p})$ and $\lambda_{\hat{s},1} \geq \cdots \geq \lambda_{\hat{s},\rho} > \lambda_{\hat{s},\rho-1} = \cdots = \lambda_{\hat{s},1} = 0$, and $\rho := \text{rank}(\mathbf{\Sigma}_{sx})$ denotes the rank of matrix $\mathbf{\Sigma}_{sx}$. Suppose now that the first $k_{ec} \in \{1,\ldots,\rho\}$ entries of $\hat{s}_w = \mathbf{Q}_{\hat{s}}^T\hat{s}$ are assigned with non-zero rate and the rest are discarded. The MSE optimal encoder's output is given by $\mathbf{u}_{ec} = \mathbf{Q}_{\hat{s},k_{ec}}^T\hat{s} + \boldsymbol{\zeta}_{ec}$, and the estimate $\hat{s}_{ec}$ is

$$\hat{\mathbf{s}}_{ec} = \mathbf{Q}_{\hat{s},k_{ec}}\mathbf{\Theta}_{ec}\mathbf{Q}_{\hat{s},k_{ec}}^T\hat{\mathbf{s}} + \mathbf{Q}_{\hat{s},k_{ec}}\mathbf{\Theta}_{ec}\boldsymbol{\zeta}_{ec} \tag{8}$$

where $\mathbf{Q}_{\hat{s},k_{ec}}$ is formed by the first k_{ec} columns of $\mathbf{Q}_{\hat{s}}$. For the $k_{ec} \times k_{ec}$ diagonal matrices $\mathbf{\Theta}_{ec}$ and $\mathbf{\Sigma}_{\zeta_{ec}\zeta_{ec}}$ we have $[\mathbf{\Theta}_{ec}]_{ii} = (\lambda_{\hat{s},i} - D_i^{ec})/\lambda_{\hat{s},i}$ and $[\mathbf{\Sigma}_{\zeta_{ec}\zeta_{ec}}]_{ii} = \lambda_{\hat{s},i}D_i^{ec}/(\lambda_{\hat{s},i} - D_i^{ec})$, where $D_i^{ec} := E[(\hat{s}_w(i) - \hat{s}_{ec,w}(i))^2]$, and $\hat{s}_{ec,w} := \mathbf{Q}_{\hat{s}}^T\hat{s}_{ec}$. Recall also, that $D_i^{ec} = \left(\prod_{i=1}^{k_{ec}}\lambda_{\hat{s},i}\right)^{1/k_{ec}}2^{-2R/k_{ec}}$ when $i = 1,\ldots,k_{ec}$ and $D_i^{ec} = \lambda_{\hat{s},i}$, for $i = k_{ec}+1,\ldots,p$. Upon defining $\mathbf{\Delta}_{ec} := \text{diag}\left(D_1^{ec}\cdots D_p^{ec}\right)$, the covariance matrix of $\tilde{s}_{ec}$ is found in Appendix B as

$$\mathbf{\Sigma}_{\tilde{s}_{ec}\tilde{s}_{ec}} = \mathbf{\Sigma}_{ss} - \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xs} + \mathbf{Q}_{\hat{s}}\mathbf{\Delta}_{ec}\mathbf{Q}_{\hat{s}}^T. \tag{9}$$

The MMSE associated with CE and EC is given, respectively, by [c.f. (7) and (9)]

$$D_{ce}(R) := \text{tr}(\mathbf{\Sigma}_{\tilde{s}_{ce}\tilde{s}_{ce}}) = J_o + \epsilon_{ce}(R), \quad \text{and} \quad D_{ec}(R) := \text{tr}(\mathbf{\Sigma}_{\tilde{s}_{ec}\tilde{s}_{ec}}) = J_o + \epsilon_{ec}(R) \tag{10}$$

where $\epsilon_{ce}(R) := \text{tr}(\mathbf{\Sigma}_{sx}\mathbf{Q}_x\mathbf{\Delta}_{ce}\mathbf{Q}_x^T\mathbf{\Sigma}_{xs})$, $\epsilon_{ec}(R) := \text{tr}(\mathbf{Q}_{\hat{s}}\mathbf{\Delta}_{ec}\mathbf{Q}_{\hat{s}}^T)$, and $J_o := \text{tr}(\mathbf{\Sigma}_{ss} - \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xs})$ is the MMSE achieved when estimating $\mathbf{s}$ based on $\mathbf{x}$, without source encoding ($R \to \infty$). Since J_o is common to both EC and CE it is important to compare $\epsilon_{ce}(R)$ with $\epsilon_{ec}(R)$ in order to determine which estimation scheme achieves the smallest MSE. The following theorem, proved in Appendix C, provides such an asymptotic comparison.

Theorem 1: *If under (a1) and (a2), $R > R_{th} := 0.5\max\left\{\log_2\left(\left(\prod_{i=1}^{\rho}\lambda_{x,i}\right)/\sigma^{2\rho}\right), \log_2\left(\left(\prod_{i=1}^{\rho}\lambda_{\hat{s},i}\right)/(\lambda_{\hat{s},\rho})^{\rho}\right)\right\}$, then $\epsilon_{ce}(R) = \gamma_1 2^{-2R/N}$ and $\epsilon_{ec}(R) = \gamma_2 2^{-2R/\rho}$, where γ_1, γ_2 are constants not dependent on R.*

An immediate consequence of Theorem 1 is that the MSE distortion for EC, namely $D_{ec}(R)$, converges as $R \to \infty$ to J_o with rate $O(2^{-2R/\rho})$. The MSE distortion of CE converges likewise, but with rate $O(2^{-2R/N})$. Typically sensors acquire more observations, namely N, than the number of parameters of interest p. Having $N > p$ enables identifiability and improved MSE performance in estimating $\mathbf{s}$. With $N > p$ it clearly holds that $\rho \leq \min(N,p) < N$. Then, the EC scheme approaches the lower bound J_o faster than CE, implying a more efficient usage of the available rate R. This is intuitively reasonable since CE compresses $\mathbf{x}$, taking into account only the covariance matrix $\mathbf{\Sigma}_{xx}$ which can result in using part of the rate to compress components of $\mathbf{x}$ that are irrelevant (e.g., noise) to the estimation of $\mathbf{s}$. On the

contrary, the MMSE estimator $\hat{\mathbf{s}}_{ec}$ in EC first extracts from $\mathbf{x}$ all the information pertinent to estimating $\mathbf{s}$, and then performs compression. In that way EC suppresses significant part of the noise and the rate is allocated more efficiently.

Let us now examine now some special cases to gain more insight about Theorem 1.

Scalar model ($p = 1$, $N = 1$): Let $x = hs + n$, where h is fixed, while s, n are uncorrelated with $s \sim \mathcal{N}\left(0, \sigma_s^2\right)$, $n \sim \mathcal{N}\left(0, \sigma_n^2\right)$, and $\sigma_x^2 = h^2\sigma_s^2 + \sigma_n^2$. With $\sigma_{\tilde{s}_{ce}}^2$ and $\sigma_{\tilde{s}_{ec}}^2$ denoting the variances of $\tilde{s}_{ce}$ and $\tilde{s}_{ec}$, respectively, we prove in Appendix D that

Proposition 1: *If (a1), (a2) hold and $N = p = 1$, then $\sigma_{\tilde{s}_{ce}}^2 = \sigma_{\tilde{s}_{ec}}^2$ and hence the D-R functions for EC and CE are identical; i.e., $D_{ec}(R) = D_{ce}(R)$.*

Vector model ($p = 1$, $N > 1$): With $\mathbf{x} = \mathbf{h}s + \mathbf{n}$, we establish in Appendix E that:

Proposition 2: *If (a1), (a2) hold and $R \leq R_{th} := 0.5\log_2\left(1 + \sigma_s^2\|\mathbf{h}\|^2/\sigma^2\right)$, then $\epsilon_{ce}(R) = \epsilon_{ec}(R)$, and thus $D_{ec}(R) = D_{ce}(R)$. For $R > R_{th}$, we have $\epsilon_{ce}(R) > \epsilon_{ec}(R)$; therefore, $D_{ce}(R) > D_{ec}(R)$ which implies that EC uses more efficiently the available rate.*

Matrix-vector model ($N > 1$, $p > 1$ and $\mathbf{\Sigma}_{ss} = \sigma_s^2\mathbf{I}_p$): For this setup, we have $\mathbf{\Sigma}_{sx} = \sigma_s^2\mathbf{H}^T$ and $\mathbf{\Sigma}_{xx} = \sigma_s^2\mathbf{H}\mathbf{H}^T + \sigma^2\mathbf{I}$. Letting $\mathbf{H} = \mathbf{U}_h\mathbf{\Sigma}_h\mathbf{V}_h^T$ be the SVD of $\mathbf{H}$, where $\mathbf{\Sigma}_h$ is an $N \times p$ diagonal matrix $\mathbf{\Sigma}_h = \mathrm{diag}(\sigma_{h,1} \cdots \sigma_{h,p})$, we show in Appendix F that

Proposition 3: *If (a1), (a2) hold, $N > \rho$ and $R > R_{th}$ with*

$$R_{th} := \frac{1}{2}\max\left\{\log_2\left(\prod_{i=1}^{\rho}\left(1 + \frac{\sigma_s^2\sigma_{h,i}^2}{\sigma^2}\right)\right), \log_2\left(\frac{\prod_{i=1}^{\rho}\sigma_{h,i}^2/(\sigma_{h,i}^2\sigma_s^2 + \sigma^2)}{(\sigma_{h,\rho}^{2\rho}/(\sigma_{h,\rho}^2\sigma_s^2 + \sigma^2)^{\rho})}\right)\right\} \tag{11}$$

then $\epsilon_{ce}(R) > \epsilon_{ec}(R)$, implying that the EC is more rate efficient than CE. If $N = \rho$, and $\exists\ i, j \in [1, \rho]$ with $i \neq j$ such that $\sigma_{h,i} \neq \sigma_{h,j}$, then $\epsilon_{ce}(R) > \epsilon_{ec}(R)$ and consequently $D_{ce}(R) > D_{ec}(R)\ \forall R \in [0, \infty)$. Finally, if for $N = \rho$, it holds that $\sigma_{h,1} = \ldots = \sigma_{h,\rho}$, then $D_{ce}(R) = D_{ec}(R)\ \forall R \in [0, \infty)$.

Defining the signal-to-noise ratio (SNR) as $\mathrm{SNR} = \mathrm{tr}(\mathbf{H}\mathbf{\Sigma}_{ss}\mathbf{H}^T)/(N\sigma^2)$, we compare in Fig. 3 (a,b) the MMSE when estimating $\mathbf{s}$ using the CE and EC schemes. With $\mathbf{\Sigma}_{ss} = \sigma_s^2\mathbf{I}_p$, $p = 4$ and $N = 40$, we observe that beyond a threshold rate, the distortion of EC converges to J_o faster than that of CE, which corroborates Theorem 1. Notice also that the gap between the EC and CE curves for $\mathrm{SNR} = 2$ is larger than the gap for $\mathrm{SNR} = 4$. This is true because as the noise power increases, the portion of the
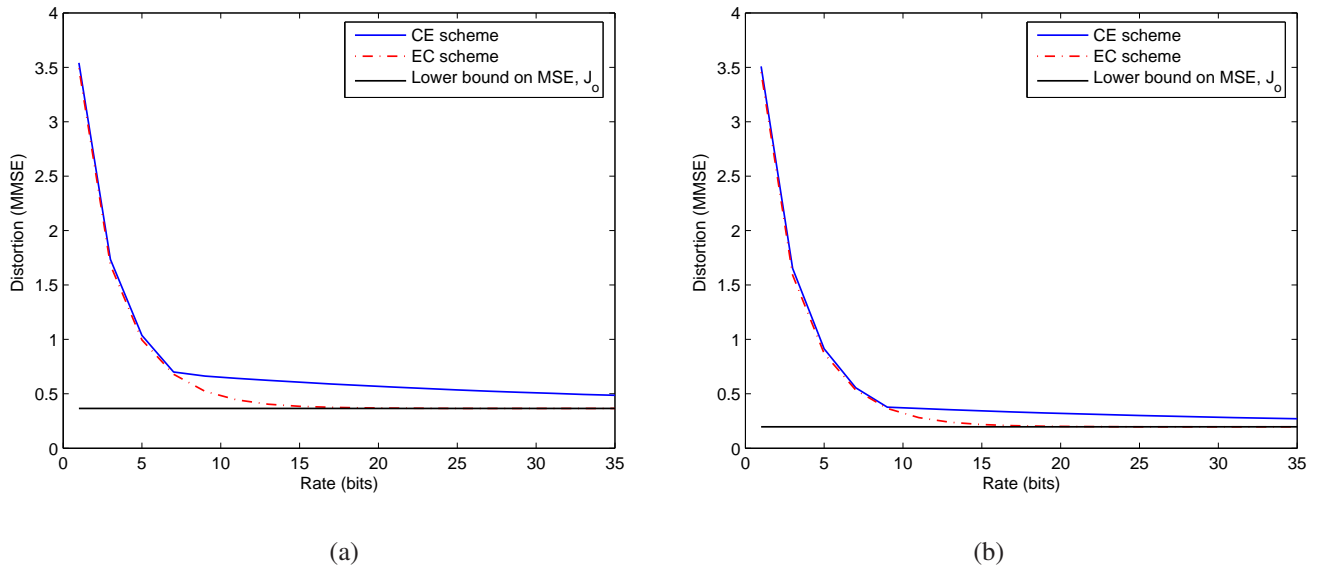
Fig. 3. D-R region for EC and CE at SNR $= 2$ (a) and SNR $= 4$ (b).

rate allocated to noise terms in CE increases accordingly. However, thanks to the MMSE estimator, EC cancels part of the noise and utilizes the available rate more efficiently.

Our analysis so far raises the question whether EC is MSE optimal. We have shown that this is the case when estimating $\mathbf{s}$ with a given rate R without forcing any relationship between N and p. A related claim has been reported in [13], [19] for $N = p$, but the extension to $N \neq p$ is not obvious. To this end, we prove in Appendix G that:

Theorem 2: *Under (a1) and (a2), the D-R function when estimating* $\mathbf{s}$ *based on* $\mathbf{x}$ *can be expressed as*

$$D_s(R) = \min_{\substack{p(\hat{\mathbf{s}}_R|\mathbf{x}) \\ I(\mathbf{x};\hat{\mathbf{s}}_R) \leq R}} E[\|\mathbf{s} - \hat{\mathbf{s}}_R\|^2] = E[\|\mathbf{s} - \hat{\mathbf{s}}\|^2] + \min_{\substack{p(\hat{\mathbf{s}}_R|\hat{\mathbf{s}}) \\ I(\hat{\mathbf{s}};\hat{\mathbf{s}}_R) \leq R}} E[\|\hat{\mathbf{s}} - \hat{\mathbf{s}}_R\|^2] \tag{12}$$

where $\hat{\mathbf{s}} = \mathbf{\Sigma}_{sx}\mathbf{\Sigma}_{xx}^{-1}\mathbf{x}$ *is the MMSE estimator, and* $\mathbf{s} - \hat{\mathbf{s}}$ *is the corresponding MMSE.*

Theorem 2 reveals that the optimal means of estimating $\mathbf{s}$ is to first form the optimal MMSE estimate $\hat{\mathbf{s}}$ and then apply optimal D-R encoding to this estimate. The lower bound on this distortion when $R \to \infty$, is $J_o = E[\|\mathbf{s} - \hat{\mathbf{s}}\|^2]$, which is intuitively appealing. The D-R function in (12) is achievable, because the rightmost term in (12) corresponds to the D-R function for reconstructing the MMSE estimate $\hat{\mathbf{s}}$ which is known to be achievable using random coding; see e.g., [1, p. 66]. Theorem 2 implies an important

separation result regarding estimation of (remote) Gaussian sources. Optimal estimation can be performed by separately estimating the source $\mathbf{s}$ based on the observation $\mathbf{x}$, and then compressing the estimate $\hat{\mathbf{s}}$ based only on the covariance of $\hat{\mathbf{s}}$. The important consequence of this result is that the total distortion $D_s(R)$ can be minimized after minimizing separately: i) the MSE distortion associated the estimation of $\mathbf{s}$ based on $\mathbf{x}$; and ii) the MSE distortion related to the compression/reconstruction task that is given by the second term in the right hand side (RHS) of (12).

IV. DISTORTION-RATE FOR DISTRIBUTED ESTIMATION

Let us now consider the D-R function for estimating $\mathbf{s}$ in a multi-sensor setup, under a total available rate R which has to be shared among all sensors. Because analytical specification of the D-R function in this case remains intractable, we will develop an alternating algorithm that numerically determines an achievable upper bound for it. Combining this upper bound with the non-achievable lower bound corresponding to an equivalent single-sensor setup, when applying the MMSE optimal EC scheme, will provide a (hopefully tight) region where the D-R function lies in. For simplicity in exposition, we confine ourselves to a two-sensor setup, but our results can be extended readily to any $L > 2$.

To this end, we consider the following single-letter characterization of the upper bound on the D-R function

$$\bar{D}(R) = \min_{\substack{p(\mathbf{u}_1|\mathbf{x}_1), p(\mathbf{u}_2|\mathbf{x}_2), \hat{\mathbf{s}}_R \\ I(\mathbf{x};\mathbf{u}_1,\mathbf{u}_2) \leq R}} E_{p(\mathbf{s},\mathbf{u}_1,\mathbf{u}_2)}[\|\mathbf{s} - \hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)\|^2] \tag{13}$$

where the minimization is w.r.t. $\{p(\mathbf{u}_i|\mathbf{x}_i)\}_{i=1}^2$ and $\hat{\mathbf{s}}_R := \hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)$. Achievability of $\bar{D}(R)$ can be established by readily extending to the vector case the scalar results in [3]. To carry out the minimization in (13), we develop an alternating scheme whereby $\mathbf{u}_2$ is treated as side information that is available at the decoder when optimizing (13) w.r.t. $p(\mathbf{u}_1|\mathbf{x}_1)$ and $\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)$. The minimization is carried within the class of Gaussian auxiliaries $\mathbf{u}_1,\mathbf{u}_2$. As a starting point, we assume that the side information $\mathbf{u}_2$ is the output of an optimal D-R encoder applied to $\mathbf{x}_2$ for estimating $\mathbf{s}$, without taking into account $\mathbf{x}_1$. This initialization for $\mathbf{u}_2$ is motivated by the Gaussianity of $\mathbf{s}$ and $\mathbf{x}_2$, as well as the single-sensor D-R results in Section III-B. Since $\mathbf{x}_2$ is Gaussian, the side information will have the form (c.f. Section III.B) $\mathbf{u}_2 = \mathbf{Q}_2\mathbf{x}_2 + \boldsymbol{\zeta}_2$, where $\mathbf{Q}_2 \in \mathbb{R}^{k_2 \times N_2}$ and $k_2 \leq N_2$, due to the rate constrained encoding of $\mathbf{x}_2$. Recall also that the $k_2 \times 1$ vector $\boldsymbol{\zeta}_2$ is uncorrelated with $\mathbf{x}_2$ and Gaussian; i.e., $\boldsymbol{\zeta}_2 \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}_{\zeta_2\zeta_2}\right)$.

Based on $\boldsymbol{\psi} := [\mathbf{x}_1^T \ \mathbf{u}_2^T]^T$, which is the information that the decoder can have assuming infinite rate at the first encoder, the optimal estimator for $\mathbf{s}$ is the MMSE one: $\hat{\mathbf{s}} = E[\mathbf{s}|\mathbf{x}_1,\mathbf{u}_2] = \boldsymbol{\Sigma}_{s\psi}\boldsymbol{\Sigma}_{\psi\psi}^{-1}\boldsymbol{\psi} = \mathbf{L}_1\mathbf{x}_1 + \mathbf{L}_2\mathbf{u}_2$, where $\mathbf{L}_1, \mathbf{L}_2$ are $p \times N_1$ and $p \times k_2$ matrices such that $\boldsymbol{\Sigma}_{s\psi}\boldsymbol{\Sigma}_{\psi\psi}^{-1} = [\mathbf{L}_1 \ \mathbf{L}_2]$. If $\tilde{\mathbf{s}}$ is

the corresponding MSE, then $\mathbf{s} = \hat{\mathbf{s}} + \tilde{\mathbf{s}}$, where $\tilde{\mathbf{s}} := \mathbf{s} - \hat{\mathbf{s}}$ is uncorrelated with ψ and $\hat{\mathbf{s}}$ due to the orthogonality principle. Noticing also that $\hat{\mathbf{s}}_R(\mathbf{u}_1, \mathbf{u}_2)$ is uncorrelated with $\tilde{\mathbf{s}}$ because it is a function of $\mathbf{x}_1$ and $\mathbf{u}_2$, we obtain $E[\|\mathbf{s} - \hat{\mathbf{s}}_R(\mathbf{u}_1, \mathbf{u}_2)\|^2] = E[\|\hat{\mathbf{s}} - \hat{\mathbf{s}}_R(\mathbf{u}_1, \mathbf{u}_2)\|^2] + E[\|\tilde{\mathbf{s}}\|^2]$, or,

$$E[\|\mathbf{s} - \hat{\mathbf{s}}_R(\mathbf{u}_1, \mathbf{u}_2)\|^2] = E[\|\mathbf{L}_1\mathbf{x}_1 - (\hat{\mathbf{s}}_R(\mathbf{u}_1, \mathbf{u}_2) - \mathbf{L}_2\mathbf{u}_2)\|^2] + E[\|\tilde{\mathbf{s}}\|^2]. \tag{14}$$

Since $\mathbf{x}_1$ and $\mathbf{x}_2$ are correlated, and $\mathbf{u}_1$ is stochastically related with $\mathbf{x}_1$ through the conditional pdf $p(\mathbf{u}_1|\mathbf{x}_1)$ we have the Markov chain (MC) $(\mathbf{x}_2, \mathbf{u}_2) \to \mathbf{x}_1 \to \mathbf{u}_1$. Using MC properties, we obtain after some simple algebra that $I(\mathbf{x}; \mathbf{u}_1, \mathbf{u}_2) = R_2 + I(\mathbf{x}_1; \mathbf{u}_1) - I(\mathbf{u}_2; \mathbf{u}_1)$, where $R_2 := I(\mathbf{x}; \mathbf{u}_2)$ is the rate consumed to form the side information $\mathbf{u}_2$; while the rate constraint in (13) becomes $I(\mathbf{x}; \mathbf{u}_1, \mathbf{u}_2) \leq R \Leftrightarrow I(\mathbf{x}_1; \mathbf{u}_1) - I(\mathbf{u}_2; \mathbf{u}_1) \leq R - R_2 := R_1$. The new signal of interest that we wish to reconstruct in (14) is $\mathbf{L}_1\mathbf{x}_1$. Continuing, we prove in Appendix H that

$$I(\mathbf{x}_1; \mathbf{u}_1) = I(\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) \tag{15}$$

Using (15), we obtain $I(\mathbf{x}_1; \mathbf{u}_1) - I(\mathbf{u}_2; \mathbf{u}_1) = I(\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) - I(\mathbf{u}_2; \mathbf{u}_1)$, and from the RHS of the last equation, we deduce the equivalent rate constraint $I(\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) - I(\mathbf{u}_2; \mathbf{u}_1) \leq R_1$. Combining the latter with (14) and (13), we arrive at the D-R upper bound

$$\bar{\bar{D}}(R_1) = \min_{\substack{p(\mathbf{u}_1|\mathbf{L}_1\mathbf{x}_1), \hat{\mathbf{s}}_R \\ I(\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) - I(\mathbf{u}_1; \mathbf{u}_2) \leq R_1}} E[\|\mathbf{L}_1\mathbf{x}_1 - (\hat{\mathbf{s}}_R(\mathbf{u}_1, \mathbf{u}_2) - \mathbf{L}_2\mathbf{u}_2)\|^2] + E[\|\tilde{\mathbf{s}}\|^2] \tag{16}$$

through which we can determine an achievable D-R region, having available rate R_1 at the encoder and side information $\mathbf{u}_2$ at the decoder. Since $\mathbf{x}_1$ and $\mathbf{u}_2$ are jointly Gaussian, we can apply the Wyner-Ziv result [20], which allows us to consider that $\mathbf{u}_2$ is available both at the decoder and the encoder. This, in turn, permits re-writing the first expectation in (16) as

$$\min_{\substack{p(\hat{\mathbf{s}}_R|\mathbf{L}_1\mathbf{x}_1, \mathbf{u}_2) \\ I(\mathbf{L}_1\mathbf{x}_1; \hat{\mathbf{s}}_R|\mathbf{u}_2) \leq R_1}} E[\|\mathbf{L}_1\mathbf{x}_1 - (\hat{\mathbf{s}}_R(\mathbf{u}_1, \mathbf{u}_2) - \mathbf{L}_2\mathbf{u}_2)\|^2]. \tag{17}$$

If $\hat{\mathbf{s}}_1 := E[\mathbf{L}_1\mathbf{x}_1|\mathbf{u}_2] = \mathbf{L}_1\Sigma_{x_1u_2}\Sigma_{u_2u_2}^{-1}\mathbf{u}_2$ and $\tilde{\mathbf{s}}_1$ is the corresponding MSE, then we can write $\mathbf{L}_1\mathbf{x}_1 = \hat{\mathbf{s}}_1 + \tilde{\mathbf{s}}_1$. For the rate constraint in (17), we have

$$I(\mathbf{L}_1\mathbf{x}_1; \hat{\mathbf{s}}_R|\mathbf{u}_2) = I(\mathbf{L}_1\mathbf{x}_1 - \hat{\mathbf{s}}_1; \hat{\mathbf{s}}_R - \mathbf{L}_2\mathbf{u}_2 - \hat{\mathbf{s}}_1|\mathbf{u}_2) = I(\tilde{\mathbf{s}}_1; \hat{\mathbf{s}}_R - \mathbf{L}_2\mathbf{u}_2 - \hat{\mathbf{s}}_1) \tag{18}$$

where the first equality holds because $\mathbf{u}_2$ is given; the second one holds since $\mathbf{u}_2$ is uncorrelated with $\tilde{\mathbf{s}}_1$, due to the orthogonality principle; and likewise, $\mathbf{u}_2$ can be uncorrelated with $\hat{\mathbf{s}}_{R,12}(\mathbf{u}_1, \mathbf{u}_2) := \hat{\mathbf{s}}_R(\mathbf{u}_1, \mathbf{u}_2) - \mathbf{L}_2\mathbf{u}_2 - \hat{\mathbf{s}}_1$, since $\hat{\mathbf{s}}_{R,12}$ is the reconstructed version of $\tilde{\mathbf{s}}_1$ which is uncorrelated with $\mathbf{u}_2$.

Utilizing (17) and (18), we arrive at

$$\bar{\bar{D}}(R_1) = E[\|\tilde{\mathbf{s}}\|^2] + \min_{\substack{p(\hat{\mathbf{s}}_{R,12}|\tilde{\mathbf{s}}_1) \\ I(\tilde{\mathbf{s}}_1;\hat{\mathbf{s}}_{R,12}) \leq R_1}} E[\|\tilde{\mathbf{s}}_1 - \hat{\mathbf{s}}_{R,12}(\mathbf{u}_1,\mathbf{u}_2)\|^2]. \tag{19}$$

Notice that the minimization term in (19) is the D-R function for reconstructing the MSE $\tilde{\mathbf{s}}_1$ with rate R_1. Since $\tilde{\mathbf{s}}_1$ is Gaussian, we can readily apply rwf to the pre-whitened vector $\mathbf{Q}_{\tilde{s}_1}^T \tilde{\mathbf{s}}_1$ for determining $\bar{\bar{D}}(R_1)$ and the corresponding test channel that achieves $\bar{\bar{D}}(R_1)$ (c.f. Section III.A). Through the latter, and considering the eigenvalue decomposition $\boldsymbol{\Sigma}_{\tilde{s}_1\tilde{s}_1} = \mathbf{Q}_{\tilde{s}_1}\text{diag}(\lambda_{\tilde{s}_1,1}\cdots\lambda_{\tilde{s}_1,p})\mathbf{Q}_{\tilde{s}_1}^T$, we find that the first encoder's output that minimizes (13) given side information $\mathbf{u}_2$ has the form

$$\mathbf{u}_1 = \mathbf{Q}_{\tilde{s}_1,k_1}^T \mathbf{L}_1 \mathbf{x}_1 + \boldsymbol{\zeta}_1 = \mathbf{Q}_1 \mathbf{x}_1 + \boldsymbol{\zeta}_1 \tag{20}$$

where $\mathbf{Q}_{\tilde{s}_1,k_1}$ denotes the first k_1 columns of $\mathbf{Q}_{\tilde{s}_1}$, k_1 is the number of $\mathbf{Q}_{\tilde{s}_1}^T \tilde{\mathbf{s}}_1$ entries that are assigned with non-zero rate, and $\mathbf{Q}_1 := \mathbf{Q}_{\tilde{s}_1,k_1}^T \mathbf{L}_1$. The $k_1 \times 1$ AWGN $\boldsymbol{\zeta}_1 \sim \mathcal{N}(\mathbf{0},\boldsymbol{\Sigma}_{\zeta_1\zeta_1})$ is uncorrelated with $\mathbf{x}_1$. Additionally, we have $[\boldsymbol{\Sigma}_{\zeta_1\zeta_1}]_{ii} = \lambda_{\tilde{s}_1,i} D_i^1/(\lambda_{\tilde{s}_1,i} - D_i^1)$, where $D_i^1 = \left(\prod_{i=1}^{k_1}\lambda_{\tilde{s}_1,i}\right)^{1/k_1} 2^{-2R_1/k_1}$, for $i = 1,\ldots,k_1$, and $D_i^1 = \lambda_{\tilde{s}_1,i}$ when $i = k_1+1,\ldots,p$. This way, we are able to determine also $p(\mathbf{u}_1|\mathbf{x}_1)$. The reconstruction function has the form

$$\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2) = \mathbf{Q}_{\tilde{s}_1,k_1}\boldsymbol{\Theta}_1\mathbf{u}_1 - \mathbf{Q}_{\tilde{s}_1,k_1}\boldsymbol{\Theta}_1\mathbf{Q}_{\tilde{s}_1,k_1}^T\mathbf{L}_1\boldsymbol{\Sigma}_{x_1 u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\mathbf{u}_2 + \mathbf{L}_1\boldsymbol{\Sigma}_{x_1 u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\mathbf{u}_2 + \mathbf{L}_2\mathbf{u}_2 \tag{21}$$

where $[\boldsymbol{\Theta}_1]_{ii} = (\lambda_{\tilde{s}_1,i} - D_i^1)/\lambda_{\tilde{s}_1,i}$, and the corresponding MMSE is $\bar{\bar{D}}(R_1) = \sum_{j=1}^{p} D_j^1 + E[\|\tilde{\mathbf{s}}\|^2]$. Notice that due to the uncorrelatedness of $\mathbf{u}_2$ with $\mathbf{x}_1 - E[\mathbf{x}_1|\mathbf{u}_2]$, the vector $\tilde{\mathbf{s}}_1$ can also be expressed as $\tilde{\mathbf{s}}_1 = E[\mathbf{s}|\mathbf{x}_1 - E[\mathbf{x}_1|\mathbf{u}_2]]$, and the MMSE estimate $\hat{\mathbf{s}}$ can be rewritten as $\hat{\mathbf{s}} = E[\mathbf{s}|\mathbf{u}_2] + E[\mathbf{s}|\mathbf{x}_1 - E[\mathbf{x}_1|\mathbf{u}_2]]$. Interestingly, it can be seen from (19) and the last expressions for $\tilde{\mathbf{s}}_1$ and $\hat{\mathbf{s}}$, that the MSE optimal approach for estimating $\mathbf{s}$ with side information $\mathbf{u}_2$ is exactly the EC scheme with the difference that in the compression step we apply rwf to the part of the MMSE estimate $\hat{\mathbf{s}}$ that is formed by the "innovation" signal $\mathbf{x}_1 - E[\mathbf{x}_1|\mathbf{u}_2]$, namely $E[\mathbf{s}|\mathbf{x}_1 - E[\mathbf{x}_1|\mathbf{u}_2]]$. Note that the optimal encoder for sensor 1 in (20) has the same structure as the one assumed for the side information $\mathbf{u}_2$ at the initialization step. Thus, we can proceed as described earlier to determine the optimal encoder for sensor 2 after treating $\mathbf{u}_1$ in (20) as side information.

The approach in this subsection can be applied in an alternating fashion from sensor to sensor in order to determine appropriate $p(\mathbf{u}_i|\mathbf{x}_i)$, for $i = 1, 2$, and $\hat{\mathbf{s}}_R(\mathbf{u}_1,\mathbf{u}_2)$ that at best globally minimize (16). The importance of the algorithm lies on the fact that it provides a way to numerically tackle (16) and determine an achievable D-R upper bound when estimating $\mathbf{s}$ at the FC based on compressed sensor observations. The conditional pdfs can be determined by finding the appropriate covariances $\boldsymbol{\Sigma}_{\zeta_i\zeta_i}$. Furthermore, by

specifying the optimal $\mathbf{Q}_1$ and $\mathbf{Q}_2$, we have a complete characterization of the encoders' structure. Relative to [6], the algorithm here can be applied to derive D-R upper bounds in general estimation setups where the parameter vector $\mathbf{s}$ that the FC wishes to estimate-reconstruct based on compressed sensor data, is observed at the sensors via $\mathbf{x}_j$'s. The scheme in [6] can be viewed as a special case of the present one corresponding to $\mathbf{x}_j = \mathbf{s}$. The resultant algorithm is summarized next.

Algorithm:

Initialize $\mathbf{Q}_1^{(0)}, \mathbf{Q}_2^{(0)}, \mathbf{\Sigma}_{\zeta_1\zeta_1}^{(0)}, \mathbf{\Sigma}_{\zeta_2\zeta_2}^{(0)}$ by applying optimal D-R encoding to each sensor's test channel independently. For a total rate R, generate M random increments $\{r(m)\}_{m=0}^M$, such that $0 \leq r(m) \leq R$ and $\sum_{m=0}^M r(m) = R$. Set $R_1(0) = R_2(0) = 0$, and for $j = 1, \ldots, M$, set $R(j) = \sum_{l=0}^j r(l)$.

> for $i = 1, 2$
>
>> $\bar{i} = \mod(i, 2) + 1$ %The complementary index
>>
>> $R_0(j) = I(\mathbf{x}; \mathbf{u}_{\bar{i}}^{(j)})$ % $\mathbf{u}_{\bar{i}}^{(j)}$ is the side information provided by the $\bar{i}$th sensor
>>
>> *Use $\mathbf{Q}_{\bar{i}}^{(j-1)}, \mathbf{\Sigma}_{\zeta_{\bar{i}}\zeta_{\bar{i}}}^{(j-1)}, R(j), R_0(j)$ to determine $\mathbf{Q}_i^{(j)}, \mathbf{\Sigma}_{\zeta_i\zeta_i}^{(j)}$ and distortion $\bar{\bar{D}}(R_i(j))$*
>>
>> from (20) and (21).
>
> end

Update the matrices $\mathbf{Q}_l^{(j)}, \mathbf{\Sigma}_{\zeta_i\zeta_i}^{(j)}$ that result the smallest distortion $\bar{\bar{D}}(R_l(j))$, with $l \in [1, 2]$

Set $R_l(j) = R(j) - I(\mathbf{x}; \mathbf{u}_{\bar{l}}^{(j)})$ and $R_{\bar{l}}(j) = I(\mathbf{x}; \mathbf{u}_{\bar{l}}^{(j)})$.

In Fig. 4, we plot the non-achievable lower bound which corresponds to one sensor having available the entire $\mathbf{x}$ and using the optimal EC scheme. The same figure also depicts an achievable D-R upper bound determined by letting the i-th sensor form its local estimate $\hat{\mathbf{s}}_i = E[\mathbf{s}|\mathbf{x}_i]$, and then apply optimal D-R encoding to $\hat{\mathbf{s}}_i$. If $\hat{\mathbf{s}}_{R,1}$ and $\hat{\mathbf{s}}_{R,2}$ are the reconstructed versions of $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$, respectively, then the decoder at the FC forms the final estimate as $\hat{\mathbf{s}}_R = E[\mathbf{s}|\hat{\mathbf{s}}_{R,1}, \hat{\mathbf{s}}_{R,2}]$. We refer to this approach as the decoupled EC scheme. We also plot the achievable D-R region determined numerically by the alternating algorithm. For each rate, we keep the smallest distortion returned after 500 executions of the algorithm simulated with $\mathbf{\Sigma}_{ss} = \mathbf{I}_p$, $p = 4$, and $N_1 = N_2 = 20$, at SNR $= 2$. We observe that the proposed algorithm provides a tighter upper bound for the achievable D-R region than the one obtained using the decoupled EC strategy. This is expected since the proposed algorithm takes into account the cross-correlations among the sensor data when determining the encoders, whereas the decoupled EC approach does not. This way the rate wasted to encode redundant information is reduced. Using also the non-achievable lower bound (solid line), we have effectively reduced the 'uncertainty region' where the D-R function lies.
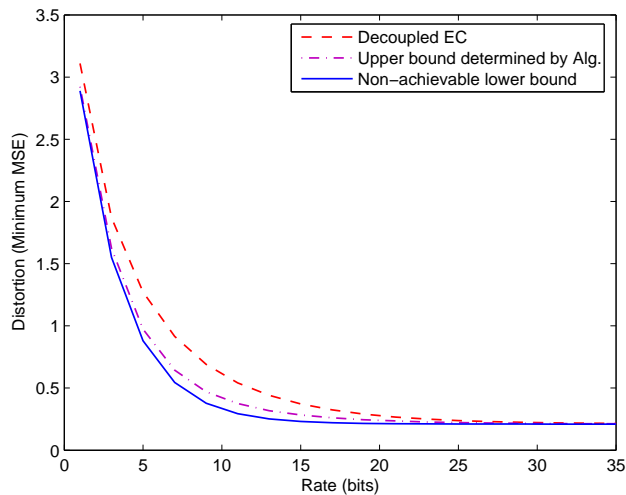
Fig. 4. Distortion-rate bounds for estimating $\mathbf{s}$ in a two-sensor setup, where SNR $= 2$, $N_1 = N_2 = 20$ and $p = 4$.

V. CONCLUSIONS

We derived inner and outer D-R bounds for the generalized Gaussian CEO problem. Specifically, we determined the D-R function for estimating a random vector in a single-sensor setup and established optimality of an estimate-first compress-afterwards (EC) approach along with the (sub)optimality of a compress-first estimate-afterwards (CE) alternative. When it comes to estimation using multiple sensors, the corresponding D-R function can be bounded from below using the single-sensor D-R function achieved using the EC scheme. An alternating algorithm was also derived for determining numerically an achievable D-R upper bound in the distributed multi-sensor setup. Simulations demonstrated that the numerically determined upper bound is more tight than analytically found alternatives (cf. the decoupled EC scheme), which is expected since the novel algorithm accounts for the cross-correlations among sensor data during the design of the encoders.

Issues of interest not accounted by this paper's analysis include general (possibly non-linear) dynamical data models where the distribution of the observation data is no longer stationary or Gaussian.[1]

REFERENCES

[1] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice Hall, 1971.

---

[1]The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies of the Army Research Laboratory or the U. S. Government.

[2] T. Berger, "Multiterminal Source Coding," *Lectures Presented at CISM Summer School on the Info. Theory Approach to Comm.*, July 1977.

[3] J. Chen, X. Zhang, T. Berger and S. B. Wicker, "An Upper Bound on the Sum-Rate Distortion Function and Its Corresponding Rate Allocation Schemes for the CEO Problem," *IEEE Journal on Sel. Areas in Comm.*, vol. 50, pp. 406–411, August 2004.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Ed., John Wiley and Sons, 1991.

[5] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise ," *IEEE Trans. on Info. Theory*, vol. 8, pp. 293–304, September 1962.

[6] M. Gastpar, P. L. Draggoti and M. Vetterli, "The Distributed Karhunen-Loève Transform," *IEEE Trans. on Info. Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006

[7] P. Ishwar, R. Puri, K. Ramchadran and S.S. Pradhan, "On Rate-Constrained Distributed Estimation in Unreliable Sensor Networks," *IEEE Journal on Sel. Areas in Comm.*, vol. 23, pp. 765–775, April 2005.

[8] Y. Oohama, "The Rate-Distortion Function for the Quadratic Gaussian CEO Problem," *IEEE Trans. On Info. Theory*, vol. 44, pp. 1057–1070, May 1998.

[9] Y. Oohama, "Rate-Distortion Theory for Gaussian Multiterminal Source Coding Systems With Several Side Informations at the Decoder," *IEEE Trans. On Info. Theory*, vol. 51, pp. 2577–2593, July 2005.

[10] Y. Oohama, "Rate Distortion Region for Separate Coding of Correlated Gaussian Remote Osbervations," *Proc. of the 43rd Allerton Conference*, Allerton, IL, September 2005.

[11] A. Pandya, A. Kansal, G. Pottie and M. Srivastava, "Fidelity and Resource Sensitive Data Gathering," *Proc. of the 42nd Allerton Conference,* Allerton, IL, September 2004.

[12] A. Rao, D. Miller, K. Rose and A. Gersho, "A Generalized VQ Method for Combined Compression and Estimation," *Proc. of Intl. Conf. on Acoust., Speech and Sig. Proc.,* pp. 2032–2035, Atlanta, GA, 1996.

[13] D. J. Sakrison, "Source encoding in the presence of random disturbance," *IEEE Trans. on Info. Theory*, vol. 14, pp. 165–167, January 1968.

[14] S. Tavildar and P. Viswanath, "On the Sum-Rate of the Vector Gaussian CEO Problem," *Proc. of 41st Asilomar Conf. on Sig., Systems, and Computers*, pp. 3–7, Pacific Grove, CA, Oct.-Nov., 2005.

[15] H. Viswanathan and T. Berger, "The Quadratic Gaussian CEO Problem," *IEEE Trans. on Info. Theory*, vol. 43, pp. 1549–1559, September 1997.

[16] A. B. Wagner and V. Anantharam, "An Improved Outer Bound for the Multiterminal Source-Coding Problem," *Proc. of the Intl. Symp. on Info. Theory*, pp. 1406–1410, Adelaide, Sep. 2005.

[17] A. B. Wagner, S. Tavildar, and P. Viswanath, "Rate Region of the Quadratic Gaussian Two-Encoder Source-Coding Problem," *Proc. of the Intl. Symp. on Info. Theory*, Seattle, WA, pp. 1404–1408, July 2006.

[18] H. S. Witsenhausen, "Indirect Rate Distortion Problems," *IEEE Trans. on Info. Theory*, vol. 26, pp. 518–521, September 1980.

[19] J. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. on Info. Theory*, vol. 16, pp. 406–411, July 1970.

[20] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder," *IEEE Trans. on Info. Theory*, vol. 26, pp. 1–10, January 1976.

[21] X. Zhang and S. B. Wicker, "On the Rate Region of the Vector Gaussian CEO Problem," *Proc. of the Conference on Info. Sciences and Systems*, The John Hopkins University, March 2005.

APPENDIX

A. Proof of Equation (7)

Using (6) we find that the covariance matrix of $\tilde{\mathbf{s}}_{ce}$ is given by $\boldsymbol{\Sigma}_{\tilde{s}_{ce}\tilde{s}_{ce}} = \boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s\check{x}_1}\boldsymbol{\Sigma}_{\check{x}_1\check{x}_1}^{-1}\boldsymbol{\Sigma}_{\check{x}_1 s}$. From the definition of $\check{\mathbf{x}}_1$, it also follows that

$$\boldsymbol{\Sigma}_{s\check{x}_1} = \boldsymbol{\Sigma}_{sx}\mathbf{Q}_{x,k_{ce}}\boldsymbol{\Theta}_{ce}, \quad \boldsymbol{\Sigma}_{\check{x}_1\check{x}_1} = \boldsymbol{\Theta}_{ce}^2(\boldsymbol{\Lambda}_{x,k_{ce}} + \boldsymbol{\Sigma}_{\zeta_{ce}\zeta_{ce}}) \tag{22}$$

where $\boldsymbol{\Lambda}_{x,k_{ce}}$ denotes the first k_{ce} diagonal entries of $\boldsymbol{\Lambda}_x$. Apparently, the $k_{ce} \times k_{ce}$ matrix $\boldsymbol{\Sigma}_{\check{x}_1\check{x}_1}$ is diagonal with entries $[\boldsymbol{\Sigma}_{\check{x}_1\check{x}_1}]_{ii} = [\boldsymbol{\Theta}]_{ii}^2([\boldsymbol{\Lambda}]_{ii} + [\boldsymbol{\Sigma}_{\zeta_{ce}\zeta_{ce}}]_{ii}) = \lambda_{x,i} - D_i^{ce}$. Let us define the diagonal matrix $\mathbf{D}^{ce} := \text{diag}(D_1^{ce}, \cdots, D_N^{ce})$ and let $\mathbf{D}_{k_{ce}}^{ce}$ denote the upper left $k_{ce} \times k_{ce}$ submatrix of $\mathbf{D}^{ce}$. We then have $\boldsymbol{\Sigma}_{\check{x}_1\check{x}_1} = \boldsymbol{\Lambda}_{x,k_{ce}} - \mathbf{D}_{k_{ce}}^{ce}$ using which we can write

$$\boldsymbol{\Sigma}_{s\check{x}_1}\boldsymbol{\Sigma}_{\check{x}_1\check{x}_1}^{-1}\boldsymbol{\Sigma}_{\check{x}_1 s} = \boldsymbol{\Sigma}_{sx}\mathbf{Q}_{x,k_{ce}}\boldsymbol{\Theta}_{ce}(\boldsymbol{\Lambda}_{x,k_{ce}} - \mathbf{D}_{k_{ce}}^{ce})^{-1}\boldsymbol{\Theta}_{ce}\mathbf{Q}_{x,k_{ce}}^T\boldsymbol{\Sigma}_{xs}$$

$$= \boldsymbol{\Sigma}_{sx}\mathbf{Q}_{x,k_{ce}}\boldsymbol{\Lambda}_{x,k_{ce}}^{-2}(\boldsymbol{\Lambda}_{x,k_{ce}} - \mathbf{D}_{k_{ce}}^{ce})\mathbf{Q}_{x,k_{ce}}^T\boldsymbol{\Sigma}_{xs}. \tag{23}$$

Let $\mathbf{D}_{N-k_{ce}}^{ce}$ and $\boldsymbol{\Lambda}_{x,N-k_{ce}}$ denote the lower right submatrix of $\mathbf{D}^{ce}$ and $\boldsymbol{\Lambda}_x$, respectively; and similarly, let $\mathbf{Q}_{x,N-k_{ce}}$ be formed by the last $N - k_{ce}$ columns of $\mathbf{Q}_x$. Because the last $N - k_{ce}$ entries of $\mathbf{Q}_x^T\mathbf{x}$ are not assigned any rate, we have $\mathbf{D}_{N-k_{ce}}^{ce} = \boldsymbol{\Lambda}_{x,N-k_{ce}}$. Adding and subtracting from (23) the matrix $\boldsymbol{\Sigma}_{sx}\mathbf{Q}_{x,N-k_{ce}}\boldsymbol{\Lambda}_{x,N-k_{ce}}^{-2}\mathbf{D}_{N-k_{ce}}^{ce}\mathbf{Q}_{x,N-k_{ce}}^T\boldsymbol{\Sigma}_{xs}$ and also adding the matrix $\boldsymbol{\Sigma}_{ss}$, we arrive at the RHS of (7). $\square$

B. Proof of Equation (9)

Starting from the definition of the covariance matrix of $\tilde{\mathbf{s}}_{ec}$, we have $\boldsymbol{\Sigma}_{\tilde{s}_{ec}\tilde{s}_{ec}} := E[(\mathbf{s} - \hat{\mathbf{s}}_{ec})(\mathbf{s} - \hat{\mathbf{s}}_{ec})^T] = \boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s\hat{s}_{ec}} - \boldsymbol{\Sigma}_{\hat{s}_{ec}s} + \boldsymbol{\Sigma}_{\hat{s}_{ec}\hat{s}_{ec}}$. An expression that will prove useful result for our subsequent analysis is

$$\boldsymbol{\Sigma}_{s\hat{s}} = E[\mathbf{s}\hat{\mathbf{s}}^T] = \boldsymbol{\Sigma}_{sx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xs} = \boldsymbol{\Sigma}_{\hat{s}\hat{s}} = \mathbf{Q}_{\hat{s}}\boldsymbol{\Lambda}_{\hat{s}}\mathbf{Q}_{\hat{s}}^T. \tag{24}$$

Furthermore, we will use the (cross-)covariance matrices

$$\boldsymbol{\Sigma}_{s\hat{s}_{ec}} = \boldsymbol{\Sigma}_{s\hat{s}}\mathbf{Q}_{\hat{s},k_{ec}}\boldsymbol{\Theta}_{ec}\mathbf{Q}_{\hat{s},k_{ec}}^T$$

$$= \mathbf{Q}_{\hat{s},k_{ec}}\boldsymbol{\Lambda}_{\hat{s},k_{ec}}\boldsymbol{\Theta}_{ec}\mathbf{Q}_{\hat{s},k_{ec}}^T \tag{25}$$

$$\boldsymbol{\Sigma}_{\hat{s}_{ec}\hat{s}_{ec}} = \mathbf{Q}_{\hat{s},k_{ec}}\boldsymbol{\Theta}_{ec}^2(\boldsymbol{\Lambda}_{\hat{s},k_{ec}} + \boldsymbol{\Sigma}_{\zeta_{ec}\zeta_{ec}})\mathbf{Q}_{\hat{s},k_{ec}}^T. \tag{26}$$

Substituting (25) and (26) into $\mathbf{\Sigma}_{\tilde{s}_{ec}\tilde{s}_{ec}}$, we obtain

$$\mathbf{\Sigma}_{\tilde{s}_{ec}\tilde{s}_{ec}} = \mathbf{\Sigma}_{ss} + \mathbf{Q}_{\hat{s},k_{ec}}\mathbf{\Omega}\mathbf{Q}_{\hat{s},k_{ec}}^T \qquad (27)$$

where $\mathbf{\Omega} := \mathbf{\Theta}_{ec}^2(\mathbf{\Lambda}_{\hat{s},k_{ec}} + \mathbf{\Sigma}_{\zeta_{ec}\zeta_{ec}}) - 2\mathbf{\Lambda}_{\hat{s},k_{ec}}\mathbf{\Theta}_{ec} = \mathbf{\Delta}_{ec,k_{ec}} - \mathbf{\Lambda}_{\hat{s},k_{ec}}$, and $\mathbf{\Delta}_{ec,k_{ec}}$ denotes the $k_{ec} \times k_{ec}$ upper left submatrix of $\mathbf{\Delta}_{ec}$. Since $D_i^{ec} = \lambda_{\hat{s},i}$ for $i = k_{ec}+1, \ldots, p$, we have $\mathbf{\Lambda}_{\hat{s},p-k_{ec}} = \mathbf{\Delta}_{ec,p-k_{ec}}$, where $\mathbf{\Lambda}_{\hat{s},p-k_{ec}}$ and $\mathbf{\Delta}_{ec,p-k_{ec}}$ denote the $(p-k_{ec}) \times (p-k_{ec})$ lower right submatrices of $\mathbf{\Lambda}_{\hat{s}}$ and $\mathbf{\Delta}_{ec}$, respectively. Adding and subtracting the matrix $\mathbf{Q}_{\hat{s},p-k_{ec}}\mathbf{\Lambda}_{\hat{s},p-k_{ec}}\mathbf{Q}_{\hat{s},p-k_{ec}}^T$ from (27), where $\mathbf{Q}_{\hat{s},p-k_{ec}}$ is formed by the last $p-k_{ec}$ columns of $\mathbf{Q}_{\hat{s}}$, we arrive at (9).

C. Proof of Theorem 1

Consider first the CE scheme with $k_{ce} = N$. In this case, the rwf threshold is given by

$$d_{ce}(N,R) = D_i^{ce} = \left(\prod_{i=1}^{\rho} \lambda_{x,i}\right)^{1/N} (\sigma^2)^{(N-\rho)/N} 2^{-2R/N}, \quad i = 1, \ldots, N. \qquad (28)$$

Since for $k_{ce} = N$ all entries of $\mathbf{Q}_x^T \mathbf{x}$ are assigned with non-zero rate, we infer that $d_{ce}(N,R) < \sigma^2$, or, equivalently

$$R > (1/2)\log_2\left(\left(\prod_{i=1}^{\rho} \lambda_{x,i}\right)/(\sigma^2)^{\rho}\right) := R_{ce}. \qquad (29)$$

Focusing on the EC scheme for $k_{ec} = \rho$, we have

$$d_{ec}(\rho,R) = D_i^{ec} = \left(\prod_{i=1}^{\rho} \lambda_{\hat{s},i}\right)^{1/\rho} 2^{-2R/\rho}, \quad i = 1, \ldots, \rho. \qquad (30)$$

When $k_{ec} = \rho$, all entries of $\mathbf{Q}_{\hat{s}}^T \hat{s}$ are assigned with non-zero rate. The latter implies that $d_{ec}(\rho,R) < \lambda_{\hat{s},\rho}$, which translates into

$$R > (1/2)\log_2\left(\left(\prod_{i=1}^{\rho} \lambda_{\hat{s},i}\right)/(\lambda_{\hat{s},\rho})^{\rho}\right) := R_{ec}. \qquad (31)$$

If $R > \max(R_{ce}, R_{ec})$, then we have $k_{ce} = N$ and $k_{ec} = \rho$. Additionally, we can easily obtain

$$\mathbf{\Delta}_{ce} = 2^{-2R/N}\alpha_1 \text{diag}(\lambda_{x,1}^{-2}, \cdots, \lambda_{x,N}^{-2}), \quad \mathbf{\Delta}_{ec} = \text{diag}(2^{-2R/\rho}\alpha_2 \mathbf{I}_\rho, \mathbf{0}) \qquad (32)$$

where $\alpha_1 := (\prod_{i=1}^{\rho} \lambda_{x,i})^{1/N} (\sigma^2)^{(N-\rho)/N}$ and $\alpha_2 := (\prod_{i=1}^{\rho} \lambda_{\hat{s},i})^{1/\rho}$. Since $\alpha_1, \alpha_2, \mathbf{Q}_x, \mathbf{Q}_{\hat{s}}$, and $\{\lambda_{x,i}\}_{i=1}^N$ do not depend on R, it follows readily that $\epsilon_{ce}(R) = \text{tr}(\mathbf{\Sigma}_{sx}\mathbf{Q}_x\mathbf{\Delta}_{ce}\mathbf{Q}_x^T\mathbf{\Sigma}_{xs}) = \gamma_1 2^{-2R/N}$ and $\epsilon_{ec}(R) = \text{tr}(\mathbf{Q}_{\hat{s}}\mathbf{\Delta}_{ec}\mathbf{Q}_{\hat{s}}^T) = \gamma_2 2^{-2R/\rho}$, where γ_1, γ_2 are constants, not dependent on R. $\square$

D. Proof of Proposition 1

For the CE scheme, $\Delta_{ce} = \sigma_x^{-2} 2^{-2R} = (h^2\sigma_s^2 + \sigma_n^2)^{-1} 2^{-2R}$ while the variance of $\tilde{s}_{ce}$ is $\sigma_{\tilde{s}_{ce}}^2 = J_o + (E[sx])^2\Delta_{ce} = \sigma_s^2 - h^2\sigma_s^4(\sigma_x^2)^{-1} + h^2\sigma_s^4\Delta_{ce}$; or, equivalently

$$\sigma_{\tilde{s}_{ce}}^2 = \sigma_s^2 - h^2\sigma_s^4\sigma_x^{-2} + h^2\sigma_s^4(h^2\sigma_s^2 + \sigma_n^2)^{-1} 2^{-2R}. \qquad (33)$$

Likewise, we have for the EC scheme that $\Delta_{ec} = \sigma_{\hat{s}}^2 2^{-2R} = h^2\sigma_s^4\sigma_x^{-2}2^{-2R}$, while $\sigma_{\tilde{s}_{ec}}^2 = J_o + \Delta_{ec}$, or,

$$\sigma_{\tilde{s}_{ec}}^2 = \sigma_s^2 - h^2\sigma_s^4\sigma_x^{-2} + h^2\sigma_s^4(h^2\sigma_s^2 + \sigma_n^2)^{-1}2^{-2R}. \tag{34}$$

It then follows readily from (33) and (34) that $\sigma_{\tilde{s}_{ce}}^2 = \sigma_{\tilde{s}_{ec}}^2$. $\square$

E. Proof of Proposition 2

Using the vector model $\mathbf{x} = \mathbf{h}s + \mathbf{n}$, we can easily verify that $\mathbf{\Lambda}_x = \mathrm{diag}\left(\sigma^2 + \sigma_s^2\|\mathbf{h}\|^2, \sigma^2, \cdots, \sigma^2\right)$ and $\mathbf{Q}_x = [\mathbf{q}_{x,1}, \cdots, \mathbf{q}_{x,N}]$, where $\mathbf{q}_{x,1} = (\mathbf{h}/\|\mathbf{h}\|)$. In the CE scheme, if $d_{ce}(k_{ce}, R) \geq \sigma^2$ then only the first entry of $\mathbf{Q}_x^T\mathbf{x}$ is assigned with positive rate; while if $d_{ce}(k_{ce}, R) < \sigma^2$, then all the elements of $\mathbf{Q}_x^T\mathbf{x}$ are assigned with non-zero rate. Thus, k_{ce} can be either 1, or, N. When $k_{ce} = 1$, the rwf threshold is given by $d_{ce}(1, R) = (\sigma_s^2\|\mathbf{h}\|^2 + \sigma^2)2^{-2R}$. Because for $k_{ce} = 1$ we must have $d_{ce}(1, R) \geq \sigma^2$, we deduce that

$$R \leq (1/2)\log_2\left(1 + (\sigma_s^2\|\mathbf{h}\|^2)/\sigma^2\right) := R_{th}. \tag{35}$$

If $R > R_{th}$, we have $k_{ce} = N$ and the threshold is $d_{ce}(N, R) = (\sigma_s^2\|\mathbf{h}\|^2 + \sigma^2)^{(1/N)}\left(\sigma^2\right)^{(1-1/N)}2^{-2R/N}$, while the distortion term $\epsilon_{ce}(R) = \mathrm{tr}\left(\mathbf{\Sigma}_{sx}\mathbf{Q}_x\mathbf{\Delta}_{ce}\mathbf{Q}_x^T\mathbf{\Sigma}_{xs}\right)$ is given by

$$\epsilon_{ce}(R) = \mathrm{tr}(\sigma_s^4\mathbf{h}^T\mathbf{Q}_x\mathbf{\Delta}_{ce}\mathbf{Q}_x^T\mathbf{h}) = \begin{cases} \beta 2^{-2R}, & R \leq R_{th} \\ \beta\left(\sigma^2(\sigma_s^2\|\mathbf{h}\|^2 + \sigma^2)^{-1}\right)^{(1-1/N)}2^{-2R/N}, & R > R_{th} \end{cases} \tag{36}$$

where $\beta = \sigma_s^4\|\mathbf{h}\|^2(\sigma_s^2\|\mathbf{h}\|^2 + \sigma^2)^{-1}$. For the EC scheme, we obtain $\sigma_{\hat{s}}^2 = \sigma_s^4\mathbf{h}^T\mathbf{Q}_x\mathbf{\Lambda}_x^{-1}\mathbf{Q}_x^T\mathbf{h} = \beta$. Since in the EC scheme we compress the MMSE estimate $\hat{s}$, we have $\epsilon_{ec}(R) = \beta 2^{-2R}, \forall\ R$. The result now follows immediately after direct comparison of $\epsilon_{ce}(R)$ with $\epsilon_{ec}(R)$ when $R \leq R_t$, and when $R > R_t$, respectively. $\square$

F. Proof of Proposition 3

From the matrix-vector model $\mathbf{x} = \mathbf{H}s + \mathbf{n}$, it follows immediately that $\mathbf{Q}_x = \mathbf{U}_h$ and $\mathbf{\Lambda}_x = \mathrm{diag}(\sigma_s^2\sigma_{h,1}^2 + \sigma^2, \cdots, \sigma_s^2\sigma_{h,\rho}^2 + \sigma^2, \sigma^2, \cdots, \sigma^2)$. The covariance of $\hat{s}$ can be written as $\mathbf{\Sigma}_{\hat{s}\hat{s}} = \sigma_s^2\mathbf{V}_h\mathbf{\Sigma}_h^T(\sigma_s^2\mathbf{\Sigma}_h\mathbf{\Sigma}_h^T + \sigma^2\mathbf{I})^{-1}\mathbf{\Sigma}_h\mathbf{V}_h^T$. Furthermore, we can easily verify that $\mathbf{Q}_{\hat{s}} = \mathbf{V}_h$, and

$$\mathbf{\Lambda}_{\hat{s}} = \mathrm{diag}\left((\sigma_s^4\sigma_{h,1}^2)(\sigma_s^2\sigma_{h,1}^2 + \sigma^2)^{-1}, \cdots, (\sigma_s^4\sigma_{h,\rho}^2)(\sigma_s^2\sigma_{h,\rho}^2 + \sigma^2)^{-1}, 0, \cdots, 0\right).$$

Focusing on the CE scheme, we have $\mathbf{\Sigma}_{sx}\mathbf{Q}_x\mathbf{\Delta}_{ce}\mathbf{Q}_x^T\mathbf{\Sigma}_{xs} = \sigma_s^4\mathbf{V}_h\mathbf{\Sigma}_h^T\mathbf{\Delta}_{ce}\mathbf{\Sigma}_h\mathbf{V}_h^T$ and the trace of the last matrix is $\epsilon_{ce}(R) = \sigma_s^4\sum_{i=1}^{\rho}(\sigma_{h,i}^2 D_i^{ce})(\sigma_s^2\sigma_{h,i}^2 + \sigma^2)^{-2}$. When $k_{ce} = N$, all the components in $\mathbf{Q}_x^T\mathbf{x}$ are assigned with non-zero rate and the rwf threshold is

$$d_{ce}(N, R) = D_i^{ce} = \left(\prod_{i=1}^{\rho}(\sigma_s^2\sigma_{h,i}^2 + \sigma^2)\right)^{1/N}(\sigma^2)^{(N-\rho)/N}2^{-2R/N}, \; i = 1, \ldots, N. \tag{37}$$

Notice also that $d_{ce}(N, R) < \sigma^2$ implies

$$R > (1/2) \log_2 \left(\prod_{i=1}^{\rho} (1 + (\sigma_s^2 \sigma_{h,i}^2) \sigma^{-2}) \right) := R_{ce}. \tag{38}$$

In the EC scheme, the trace of $\mathbf{Q}_{\hat{s}} \mathbf{\Delta}_{ec} \mathbf{Q}_{\hat{s}}^T$ is equal to $\epsilon_{ec}(R) = \mathrm{tr}(\mathbf{\Delta}_{ec}) = \sum_{i=1}^{\rho} D_i^{ec}$. When $k_{ec} = \rho$, the corresponding rwf threshold is given by

$$d_{ec}(\rho, R) = D_i^{ec} = \left(\prod_{j=1}^{\rho} (\sigma_s^4 \sigma_{h,i}^2)(\sigma_s^2 \sigma_{h,i}^2 + \sigma^2)^{-1} \right)^{1/\rho} 2^{-2R/\rho}, \; i = 1, \ldots, \rho. \tag{39}$$

The equality $k_{ec} = \rho$ is satisfied when $d_{ec}(\rho, R) < (\sigma_s^4 \sigma_{h,\rho}^2)(\sigma_s^2 \sigma_{h,\rho}^2 + \sigma^2)^{-1}$, which yields the rate constraint

$$R > (1/2) \log_2 \left(\frac{\prod_{i=1}^{\rho} \sigma_{h,i}^2 (\sigma_{h,i}^2 \sigma_s^2 + \sigma^2)^{-1}}{(\sigma_{h,\rho}^2)^{\rho} (\sigma_{h,\rho}^2 \sigma_s^2 + \sigma^2)^{-\rho}} \right) := R_{ec}. \tag{40}$$

Notice that when $R > \max(R_{ce}, R_{ec})$, we have $k_{ce} = N, k_{ec} = \rho$, while

$$\epsilon_{ce}(R) = \sigma_s^4 \sigma^2 2^{-2R/N} (\prod_{j=1}^{\rho} ((\sigma_s^2 \sigma_{h,j}^2) \sigma^{-2} + 1))^{1/N} \sum_{i=1}^{\rho} \sigma_{h,i}^2 (\sigma_s^2 \sigma_{h,i}^2 + \sigma^2)^{-2}, \tag{41}$$

$$\epsilon_{ec}(R) = \rho \sigma_s^4 2^{-2R/\rho} (\prod_{j=1}^{\rho} \sigma_{h,j}^2 (\sigma_s^2 \sigma_{h,j}^2 + \sigma^2)^{-1})^{1/\rho}. \tag{42}$$

For $N > \rho$ and after some algebraic manipulations, we conclude that for $\epsilon_{ce}(R) > \epsilon_{ec}(R)$ to hold, we must have

$$R > (1/2) \log_2 (\prod_{i=1}^{\rho} (1 + (\sigma_s^2 \sigma_{h,i}^2) \sigma^{-2})) + N\rho(2(N - \rho))^{-1} \log_2 \gamma := \bar{R}, \tag{43}$$

where

$$\gamma = (\prod_{i=1}^{\rho} \sigma_{h,i}^2 (\sigma_{h,i}^2 \sigma_s^2 + \sigma^2)^{-2})^{1/\rho} (\rho^{-1} \sum_{i=1}^{\rho} (\sigma_{h,i}^2 (\sigma_{h,i}^2 \sigma_s^2 + \sigma^2)^{-2}))^{-1}.$$

From the arithmetic mean-geometric mean inequality, we further deduce that $\gamma \leq 1$, which in turn implies that $\max(R_{ce}, R_{ec}) > \bar{R}$. Thus, $\epsilon_{ce}(R) > \epsilon_{ec}(R)$, when $N > \rho$ and $R > \max(R_{ce}, R_{ec})$. When $N = \rho$, it follows readily from (41) and (42) that if $\sigma_{h,i} = \sigma_h$ for $i = 1, \ldots, \rho$, then $\epsilon_{ce}(R) = \epsilon_{ec}(R) \; \forall R$; otherwise, $\epsilon_{ce}(R) > \epsilon_{ec}(R) \; \forall R$. $\square$

G. Proof of Theorem 2

Using the orthogonality principle, we can write $\mathbf{s} = \hat{\mathbf{s}} + \tilde{\mathbf{s}}$, where $\tilde{\mathbf{s}}$ is independent of $\mathbf{x}$; thus

$$E[\|\mathbf{s} - \hat{\mathbf{s}}_R\|^2] = E[\|\hat{\mathbf{s}} - \hat{\mathbf{s}}_R\|^2] + E[\|\tilde{\mathbf{s}}\|^2] \tag{44}$$

which stems directly from the fact that $\hat{\mathbf{s}}$ and $\hat{\mathbf{s}}_R$ are independent of $\tilde{\mathbf{s}}$ since they are functions of $\mathbf{x}$. In order to arrive at (12), it suffices to show that $I(\mathbf{x}; \hat{\mathbf{s}}_R) = I(\hat{\mathbf{s}}; \hat{\mathbf{s}}_R)$.

To this end, consider the SVD of $\mathbf{\Sigma}_{sx} = \mathbf{U}_{sx} \mathbf{S}_{sx} \mathbf{V}_{sx}^T$, where $\mathbf{V}_{sx}$ is an $N \times N$ unitary matrix. Further, recall that $\rho = \mathrm{rank}(\mathbf{\Sigma}_{sx})$, and we define the $N \times N$ matrix $\mathbf{T} := [(\mathbf{Q}_{\hat{s},\rho}^T \mathbf{\Sigma}_{sx} \mathbf{\Sigma}_{xx}^{-1})^T, \mathbf{V}_{sx,N-\rho}]^T$, where

$\mathbf{V}_{sx,N-\rho}$ contains the last $N - \rho$ columns of $\mathbf{V}_{sx}$. Note, that $\mathbf{V}_{sx,N-\rho}^T \mathbf{\Sigma}_{sx}^T = \mathbf{0}_{N-\rho \times p}$. We will prove by contradiction that $\mathbf{T}$ is invertible. Suppose that there exists an $N \times 1$ nonzero vector $\mathbf{u}$ such that $\mathbf{u}^T \mathbf{T} = \mathbf{0}^T$; i.e., assume that $\mathbf{u} := [\mathbf{u}_1^T \ \mathbf{u}_2^T]^T$ satisfies

$$\mathbf{u}_1^T \mathbf{Q}_{\hat{s},\rho}^T \mathbf{\Sigma}_{sx} \mathbf{\Sigma}_{xx}^{-1} + \mathbf{u}_2^T \mathbf{V}_{sx,N-\rho}^T = \mathbf{0}^T. \tag{45}$$

If $\mathbf{u}_1 = \mathbf{0}$, or, $\mathbf{u}_2 = \mathbf{0}$, then (45) yields $\mathbf{u}_2^T \mathbf{V}_{sx,N-\rho}^T = \mathbf{0}^T$, or, $\mathbf{u}_1^T \mathbf{Q}_{\hat{s},\rho}^T \mathbf{\Sigma}_{sx} \mathbf{\Sigma}_{xx}^{-1} = \mathbf{0}^T$ respectively. Note that $\mathbf{u}_2^T \mathbf{V}_{sx,N-\rho}^T = \mathbf{0}^T$ cannot be true for $\mathbf{u}_2 \neq \mathbf{0}$ since $\mathbf{V}_{sx,N-\rho}^T$ is a full row rank matrix. Similarly, $\mathbf{u}_1^T \mathbf{Q}_{\hat{s},\rho}^T \mathbf{\Sigma}_{sx} \mathbf{\Sigma}_{xx}^{-1} = \mathbf{0}^T$ requires $\mathbf{u}_1^T \mathbf{Q}_{\hat{s},\rho}^T \mathbf{\Sigma}_{sx} = \mathbf{0}^T$ which is impossible for $\mathbf{u}_1^T \neq \mathbf{0}^T$ because the columns of $\mathbf{Q}_{\hat{s},\rho}$ are orthogonal to the nullspace of $\mathbf{\Sigma}_{sx}^T$. Consider next the case where both $\mathbf{u}_1$ and $\mathbf{u}_2$ are nonzero. Upon defining $\mathbf{u}_1' := \mathbf{\Sigma}_{sx}^T \mathbf{Q}_{\hat{s},\rho} \mathbf{u}_1$ and $\mathbf{u}_2' := \mathbf{V}_{sx,N-\rho} \mathbf{u}_2$, it can be readily seen that $\mathbf{u}_1'$ and $\mathbf{u}_2'$ are orthogonal; while $\mathbf{u}_1'$ cannot be zero for $\mathbf{u}_1 \neq \mathbf{0}$, since the columns of $\mathbf{Q}_{\hat{s},\rho}$ are orthogonal to the nullspace of $\mathbf{\Sigma}_{sx}^T$. Thus, from (45) we arrive at

$$(\mathbf{u}_1')^T \mathbf{\Sigma}_{xx}^{-1} + (\mathbf{u}_2')^T = \mathbf{0} \tag{46}$$

which further implies that $(\mathbf{u}_1')^T \mathbf{\Sigma}_{xx}^{-1} \mathbf{u}_1' = 0$. Since $\mathbf{\Sigma}_{xx}$ is full rank the latter leads to a contradiction which establishes that $\mathbf{T}$ is invertible.

Upon multiplying $\mathbf{T}$ with the observation vector $\mathbf{x}$, we obtain

$$\mathbf{Tx} = [(\mathbf{Q}_{\hat{s},\rho}^T \mathbf{\Sigma}_{sx} \mathbf{\Sigma}_{xx}^{-1} \mathbf{x})^T \ (\mathbf{V}_{sx,N-\rho}^T \mathbf{x})^T]^T = [(\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}})^T \ (\mathbf{V}_{sx,N-\rho}^T \mathbf{x})^T]^T \tag{47}$$

where the second inequality in (47) holds because $\hat{\mathbf{s}} = \mathbf{\Sigma}_{sx} \mathbf{\Sigma}_{xx}^{-1} \mathbf{x}$. Further, the cross-correlation matrix of $\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}$ with $\mathbf{V}_{sx,N-p}^T \mathbf{x}$, is given by

$$\mathbf{Q}_{\hat{s},\rho}^T E[\hat{\mathbf{s}} \mathbf{x}^T] \mathbf{V}_{sx,N-\rho} = \mathbf{Q}_{\hat{s},\rho}^T \mathbf{\Sigma}_{sx} \mathbf{V}_{sx,N-\rho} = \mathbf{0}_{\rho \times N-\rho}. \tag{48}$$

The uncorrelatedness in (48) implies that the Gaussian vectors $\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}$ and $\mathbf{V}_{sx,N-\rho}^T \mathbf{x}$ are independent. Using the invertibility of $\mathbf{T}$, we obtain

$$I(\mathbf{x}; \hat{\mathbf{s}}_R) = I(\mathbf{Tx}; \hat{\mathbf{s}}_R) = I(\mathbf{V}_{sx,N-\rho}^T \mathbf{x}; \hat{\mathbf{s}}_R) + I(\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}; \hat{\mathbf{s}}_R | \mathbf{V}_{sx,N-\rho}^T \mathbf{x}).$$

Now, the optimal estimate $\hat{\mathbf{s}}_R$ can be independent of $\mathbf{V}_{sx,N-\rho}^T \mathbf{x}$ without affecting the distortion, since the second is uncorrelated with $\mathbf{s}$ and does not contain any information relevant to the estimation of $\mathbf{s}$; thus, we have that $I(\hat{\mathbf{s}}_R; \mathbf{V}_{sx,N-\rho}^T \mathbf{x}) = 0$. Since $\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}$ and $\mathbf{V}_{sx,N-\rho}^T \mathbf{x}$ are independent, we obtain that $I(\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}; \hat{\mathbf{s}}_R | \mathbf{V}_{sx,N-p}^T \mathbf{x}) = I(\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}; \hat{\mathbf{s}}_R)$, which implies that

$$I(\mathbf{x}; \hat{\mathbf{s}}_R) = I(\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}; \hat{\mathbf{s}}_R) = I(\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}; \hat{\mathbf{s}}_R) + I(\mathbf{Q}_{\hat{s},p-\rho}^T \hat{\mathbf{s}}; \hat{\mathbf{s}}_R | \mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}) = I(\mathbf{Q}_{\hat{s}}^T \hat{\mathbf{s}}; \hat{\mathbf{s}}_R) = I(\hat{\mathbf{s}}; \hat{\mathbf{s}}_R) \tag{49}$$

where the second equality in (49) holds because $\mathbf{Q}_{\hat{s},p-\rho}^T \hat{\mathbf{s}}$ offers no information for estimating $\mathbf{s}$, and $\hat{\mathbf{s}}_R$ and $\mathbf{Q}_{\hat{s},\rho}^T \hat{\mathbf{s}}$ are independent of $\mathbf{Q}_{\hat{s},p-\rho}^T \hat{\mathbf{s}}$. $\square$

H. Proof of Equation (15)

After expressing $\boldsymbol{\Sigma}_{\psi\psi}^{-1}$ in terms of $\boldsymbol{\Sigma}_{x_1 x_1}, \boldsymbol{\Sigma}_{x_1 u_2}, \boldsymbol{\Sigma}_{u_2 u_2}$ we find that

$$\mathbf{L}_1 = \left(\boldsymbol{\Sigma}_{s x_1} - \boldsymbol{\Sigma}_{s u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 x_1}\right)\left(\boldsymbol{\Sigma}_{x_1 x_1} - \boldsymbol{\Sigma}_{x_1 u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 x_1}\right)^{-1}. \tag{50}$$

Let $\text{rank}(\mathbf{H}_1) = \rho_1$ and consider the SVD $\mathbf{H}_1 = \mathbf{U}_{h_1}\boldsymbol{\Sigma}_{h_1}\mathbf{V}_{h_1}^T$. Also, let $\mathbf{U}_{h_1} := [\mathbf{U}_{h_1,\rho_1}\ \mathbf{U}_{h_1,N_1-\rho_1}]$ and $\boldsymbol{\Sigma}_{h_1,\rho_1}$ denote the upper left $\rho_1 \times \rho_1$ diagonal submatrix of $\boldsymbol{\Sigma}_{h_1}$ which contains the ρ_1 positive singular values of $\mathbf{H}_1$. Based on these definitions, we can re-express the matrices inside the parentheses in (50) as

$$\boldsymbol{\Sigma}_{s x_1} - \boldsymbol{\Sigma}_{s u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 x_1} = (\boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 s})\mathbf{H}_1^T$$

$$= (\boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 s})\mathbf{V}_{h_1,\rho_1}\boldsymbol{\Sigma}_{h_1,\rho_1}\mathbf{U}_{h_1,\rho_1}^T \tag{51}$$

and

$$\boldsymbol{\Sigma}_{x_1 x_1} - \boldsymbol{\Sigma}_{x_1 u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 x_1} = \mathbf{U}_{h_1,\rho_1}\boldsymbol{\Sigma}_{h_1,\rho_1}\mathbf{V}_{h_1,\rho_1}^T(\boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 s})\mathbf{V}_{h_1,\rho_1}\boldsymbol{\Sigma}_{h_1,\rho_1}\mathbf{U}_{h_1,\rho_1}^T + \sigma^2\mathbf{U}_{h_1}\mathbf{U}_{h_1}^T$$

$$= \mathbf{U}_{h_1}\text{diag}(\boldsymbol{\Omega}, \sigma^2\mathbf{I}_{N_1-\rho_1})\mathbf{U}_{h_1}^T \tag{52}$$

where $\boldsymbol{\Omega} := \boldsymbol{\Sigma}_{h_1,\rho_1}\mathbf{V}_{h_1,\rho_1}^T(\boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 s})\mathbf{V}_{h_1,\rho_1}\boldsymbol{\Sigma}_{h_1,\rho_1} + \sigma^2\mathbf{I}_{\rho_1}$. Upon substituting (51) and (52) into (50), we obtain

$$\mathbf{L}_1 = (\boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 s})\mathbf{V}_{h_1,\rho_1}\boldsymbol{\Sigma}_{h_1,\rho_1}\boldsymbol{\Omega}^{-1}\mathbf{U}_{h_1,\rho_1}^T. \tag{53}$$

To proceed, we assume that $\text{rank}(\boldsymbol{\Sigma}_{ss} - \boldsymbol{\Sigma}_{s u_2}\boldsymbol{\Sigma}_{u_2 u_2}^{-1}\boldsymbol{\Sigma}_{u_2 s}) = p$. If this is not the case, we can use instead as side information the random vector $\tilde{\mathbf{u}}_2 = \mathbf{u}_2 + \tilde{\mathbf{v}}$, where $\tilde{\mathbf{v}}$ is white noise with very small power. In so doing, we ensure that $\text{rank}(\mathbf{L}_1) = \rho_1$ and $\text{range}(\mathbf{L}_1^T) = \text{span}(\mathbf{U}_{h_1,\rho_1})$. The next step is to show that $I(\mathbf{x}_1; \mathbf{u}_1) = I(\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1)$. To this end, let $\mathbf{L}_1\boldsymbol{\Sigma}_{x_1 x_1}\mathbf{L}_1^T = \mathbf{Q}_{L_1}\boldsymbol{\Lambda}_{L_1}\mathbf{Q}_{L_1}^T$ be the eigenvalue decomposition of the matrix $\mathbf{L}_1\boldsymbol{\Sigma}_{x_1 x_1}\mathbf{L}_1^T$. As in Appendix G, we will consider the $N_1 \times N_1$ invertible matrix $\mathbf{T}_1 = [(\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1)^T\ \mathbf{U}_{h_1,N_1-\rho_1}]^T$. Multiplying $\mathbf{T}_1$ with $\mathbf{x}_1$, we obtain $\mathbf{T}_1\mathbf{x}_1 = [(\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1)^T\ (\mathbf{U}_{h_1,N_1-\rho_1}^T\mathbf{n}_1)^T]^T$; and based on the latter, we obtain

$$I(\mathbf{x}_1; \mathbf{u}_1) = I(\mathbf{T}_1\mathbf{x}_1; \mathbf{u}_1)$$

$$= I(\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1, \mathbf{U}_{h_1,N_1-\rho_1}^T\mathbf{n}_1; \mathbf{u}_1)$$

$$= I(\mathbf{U}_{h_1,N_1-\rho_1}^T\mathbf{n}_1; \mathbf{u}_1) + I(\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1|\mathbf{U}_{h_1,N_1-\rho_1}^T\mathbf{n}_1)$$

$$= I(\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) \tag{54}$$

where (54) follows because: (i) vectors $\mathbf{U}_{h_1,N_1-\rho_1}^T\mathbf{n}_1$ and $\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1$ are independent; and (ii) $\mathbf{u}_1$ can also be independent of $\mathbf{U}_{h_1,N_1-\rho_1}^T\mathbf{n}_1$ without affecting the resulting distortion. Evaluating (54) in terms

of the mutual information $I(\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1)$, we find

$$I(\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) = I(\mathbf{Q}_{L_1}^T\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1)$$

$$= I(\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) + I(\mathbf{Q}_{L_1,p-\rho_1}^T\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1|\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1)$$

$$= I(\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) + h(\mathbf{u}_1|\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1) - h(\mathbf{u}_1|\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1, \mathbf{Q}_{L_1,p-\rho_1}^T\mathbf{L}_1\mathbf{x}_1)$$

$$= I(\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1) \tag{55}$$

where (55) follows since $\mathbf{Q}_{L_1,p-\rho_1}^T\mathbf{L}_1\mathbf{x}_1$ does not convey additional information about $\mathbf{u}_1$, beyond what is provided by $\mathbf{Q}_{L_1,\rho_1}^T\mathbf{L}_1\mathbf{x}_1$. Combining (54) and (55), we conclude that $I(\mathbf{x}_1; \mathbf{u}_1) = I(\mathbf{L}_1\mathbf{x}_1; \mathbf{u}_1)$. $\square$

PLACE PHOTO HERE

Ioannis D. Schizas received the Diploma in Computer Engineering and Informatics with Honors from the University of Patras, Patras, Greece, in 2004 and the M. Sc. degree in electrical and computer engineering from the University of Minnesota, Minneapolis in 2007.

Since August 2004, he has been working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis. His research interests lie in the areas of communication theory, signal processing, and networking. His current research focuses on distributed signal processing with wireless ad hoc and sensor networks, and distributed compression and source coding.

PLACE PHOTO HERE

Georgios B. Giannakis received his Diploma in Electrical Engr. from the Ntl. Tech. Univ. of Athens, Greece, 1981. From 1982 to 1986 he was with the Univ. of Southern California (USC), where he received his MSc. in Electrical Engineering, 1983, MSc. in Mathematics, 1986, and Ph.D. in Electrical Engr., 1986. Since 1999 he has been a professor with the ECE Department at the Univ. of Minnesota, where he now holds an ADC Chair in Wireless Telecommunications.

His general interests span the areas of communications, networking and statistical signal processing - subjects on which he has published more than 250 journal papers, 450 conference papers, two edited books and two upcoming research monographs on Space-Time Coding for Broadband Wireless Communications (Wiley 2006) and Ultra-Wideband Wireless Communications (Cambridge Press 2007). Current research focuses on diversity techniques, complex-field and space-time coding, multicarrier, cooperative wireless communications, cognitive radios, cross-layer designs, mobile ad hoc networks, and wireless sensor networks.

G. B. Giannakis is the (co-) recipient of six paper awards from the IEEE Signal Processing (SP) and Communications Societies including the G. Marconi Prize Paper Award in Wireless Communications. He also received Technical Achievement Awards from the SP Society (2000), from EURASIP (2005), a Young Faculty Teaching Award and the G. W. Taylor Award for Distinguished Research from the University of Minnesota. He has served the IEEE in a number of posts.

PLACE
PHOTO
HERE

Nihar Jindal received the B.S. degree in Electrical Engineering and Computer Science from U.C. Berkeley in 1999, and the M.S. and Ph.D. degrees in Electrical Engineering from Stanford University in 2001 and 2004. He is an assistant professor in the Department of Electrical and Computer Engineering at the University of Minnesota. His industry experience includes summer internships at Intel Corporation, Santa Clara, CA in 2000 and at Lucent Bell Labs, Holmdel, NJ in 2002. Dr. Jindal currently serves as an Associate Editor for IEEE Transactions on Communications. Dr. Jindal's research spans the fields of information theory and wireless communication, with specific interests in multiple-antenna/multi-user channels, dynamic resource allocation, and sensor and ad-hoc networks. Dr. Jindal was the recipient of the 2005 IEEE Communications Society and Information Theory Society Joint Paper Award, and the University of Minnesota McKnight Land-Grant Professorship Award in 2007.