Power, Spatio-Temporal Bandwidth, and Distortion in Large Sensor Networks

Michael Gastpar, *Member, IEEE,* and Martin Vetterli, *Fellow, IEEE*

***Abstract*—For a class of sensor networks, the task is to monitor an underlying physical phenomenon over space and time through an imperfect observation process. The sensors can communicate back to a central data collector over a noisy channel. The key parameters in such a setting are the fidelity (or distortion) at which the underlying physical phenomenon can be estimated by the data collector, and the cost of operating the sensor network. This is a network joint source-channel communication problem, involving both compression and communication. It is well known that these two tasks may not be addressed separately without sacrificing optimality, and the optimal performance is generally unknown.**

This paper presents a lower bound on the best achievable end-to-end distortion as a function of the number of sensors, their total transmit power, the number of degrees of freedom of the underlying source process, and the spatio-temporal communication bandwidth. Particular coding schemes are studied, and it is shown that in some cases, the lower bound is tight in a scaling-law sense. By contrast, it is shown that the standard practice of separating source from channel coding may incur an exponential penalty in terms of communication resources, as a function of the number of sensors. Hence, such code designs effectively prevent scalability. Finally, it is outlined how the results extend to cases involving missing synchronization and channel fading.

***Index Terms*—CEO problem, information theory, joint source-channel coding, sensor networks, separation theorem.**

I. Introduction

PERFORMANCE limits for various types of sensor networks are currently under investigation, see, e.g., [1]–[4]. While a majority of the efforts concerns either the sampling and compression problem, or the capacity problem, the present paper investigates the *joint* problem of compression and communication. The class of sensor networks in this study is schematically illustrated in Fig. 1 and could be termed *monitoring sensor networks*: Their goal is to observe a physical system over time and space at the highest possible fidelity. A simple example of such a sensor network was analyzed in [5], [6], and extensions thereof in [7]. This paper generalizes this analysis to multiple data sources and multiple base stations, with random parameters both in the source observation mechanism and in the communication channel. The considered network contains L (discrete-time) sources. The parameter L models the number of degrees of freedom of the underlying physical process. Much of the paper concerns the case where the source outputs are distributed according to a (joint) Gaussian law. Each of the M sensors observes a different combination of these L sources, subject to noise. The M sensors communicate to N base stations. For simplicity, we assume that communication between the base stations occurs over separate channels and is noiseless. Hence, the data collection point has access to the received values of all N base stations and must form an estimate of the underlying L sources. Moreover, we allow K channel uses for each observation. Hence, K can be interpreted as the temporal bandwidth of the communication channel, while (under appropriate conditions) N models its spatial bandwidth. The key goal of this paper is to characterize the relationship between the number of underlying sources L, the end-to-end distortion D, the total sensor power P_{tot}, and the temporal and spatial bandwidth of the communication channel, K and N, respectively.

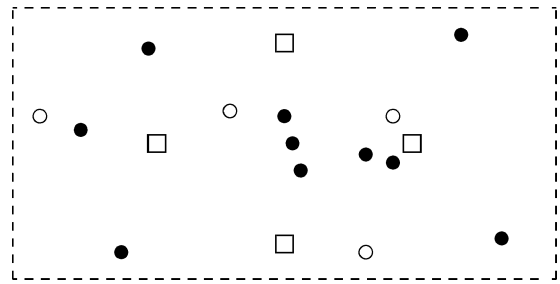

Fig. 1. Schematic rendering of the considered sensor network. There are $L = 4$ underlying physical phenomena (depicted by the empty circles), $M = 10$ sensing devices (the black disks), and $N = 4$ base stations (the squares).

The remainder of the paper is organized as follows: In Section II, we define the general sensor network topology and model studied in this paper. In Section III, we specify a particular case of the general topology, the Gaussian sensor network. Sections III, IV, and V are devoted to this special case. Using the definitions given in Section III-A, we determine in Section III-B a lower bound to the distortion, revealing two fundamentally different contributions. First, there is a distortion due to the fact that the sensor measurements are noisy. This term is independent of the communication resources and typically decreases like $1/M$, where M is the number of sensors. This is an intuitively pleasing behavior, corresponding to the decay

Manuscript received December 15, 2003; revised July 30, 2004. This paper was supported in part by the National Science Foundations under Award CCF-0347298 (CAREER) and in part by the Swiss National Science Foundation in the framework of the National Competence Center in Research for Mobile Information and Communication Systems. This paper was presented in part at the Third International Symposium on Information Processing in Sensor Networks, Berkeley, CA, April 2004.

M. Gastpar is with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720-1770 USA (e-mail: gastpar@eecs.berkeley.edu).

M. Vetterli is with the Institute of Communication Systems, Swiss Federal Institute of Technology (EPFL), CH-1015 Lausanne, Switzerland, and with the Department of Electrical Engineering and Computer Science, University of California, Berkeley, Berkeley, CA 94720 USA (e-mail: martin.vetterli@epfl.ch).

Digital Object Identifier 10.1109/JSAC.2005.843542

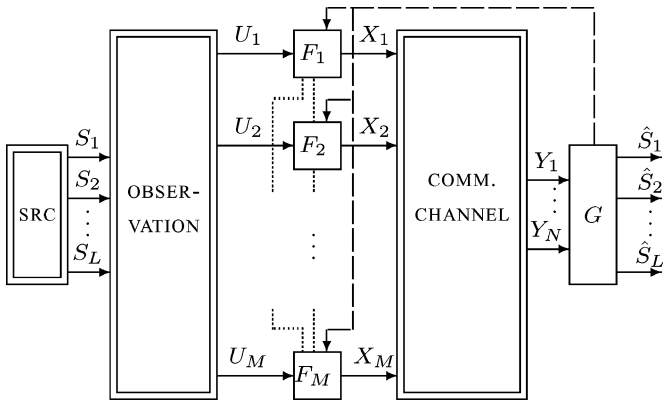

Fig. 2. Sensor network topology considered in this paper: L sources are observed in a noisy fashion by M sensors (the boxes labeled F_1 through F_M) that communicate to N base stations over a power- and bandwidth-constrained interference channel. The sensors may have (generally limited) collaboration capabilities (illustrated by the dotted lines in the figure), and there may be a certain feedback signal from the destination (illustrated by the dashed arrows in the figure).

of the mean-squared error when the sources can be estimated directly based on the observations, ignoring the communication stage. Second, there is a distortion due to the fact that the channel is noisy and subject to interference. This distortion can be decreased by allowing more power and/or bandwidth. Clearly, there is an incentive to select these parameters in such a way as to make the second distortion term decrease at the same rate as the first term (as the number of sensors increases). In Section III-D, we characterize the power and bandwidth necessary to achieve this goal. Then, in Section IV, we discuss schemes that achieve the optimum distortion scaling law for particular power-bandwidth tradeoffs, and we illustrate that separating source from channel coding can lead to exponentially suboptimal scaling behavior. Up to this point, we assume that the sensors operate with perfect synchronization. In Section V, we extend our results to the unsynchronized case. For simple examples of unsynchronized sensors, we show that our scaling laws continue to hold.

II. Sensor Network Model

The sensor network model studied in this paper is shown in Fig. 2. There is a physical phenomenon, characterized by L variables, representing the degrees of freedom of the system or, equivalently, its current state. We model each degree of freedom as a random process in discrete time.[1] Generally, the degrees of freedom cannot be observed directly. Rather, in typical scenarios, each sensor measures a (different) noisy version of a combination of all of these variables. We model this observation process in a probabilistic fashion as a conditional distribution of the observations given the state. As expressed by the dotted lines in Fig. 2, the sensors may have the possibility to collaborate to some (generally limited) extent, and there may be feedback from the base stations to each of the sensors. Based on the respective sensor readings, the intersensor communication, and the feedback signals, each sensor has to produce an output to be transmitted over the communication link (e.g., a wireless link). This channel is again modeled in a probabilistic fashion by a conditional distribution. The channel outputs are received by the base stations. In this paper, we assume that the central data collection unit is ideally linked (e.g., over a backbone network) to the base stations, The goal of the data collector is to get to know, not the raw sensor readings, but the values of the underlying degrees of freedom (or state) of the physical system.

More precisely, and to fix notations, the physical phenomenon is characterized by the sequence of random vectors

$$\{S[i]\}_{i\in Z} = \{(S_1[i], S_2[i], \ldots, S_L[i])\}_{i\in Z} \tag{1}$$

where i is the time index. The analysis presented in this paper addresses the case where $\{S[i]\}_{i\in Z}$ is a sequence of independent and identically distributed (i.i.d.) random vectors. To simplify the notation in the rest of this paper, we denote sequences as $S^j \stackrel{def}{=} \{S[i]\}_{i=1}^j$. We use the upper case S to denote the random variable, and the lower case s to denote its realization. The distribution of S is denoted by $P_S(s)$. To simplify notation, we will also use the shorthand $P(s)$ when the subscript is just the capitalized version of the argument in the parentheses. The random vector $S[i]$ is not directly observed by the sensors. Rather, sensor m observes a sequence $U_m^j = \{U_m[i]\}_{i=1}^j$ which depends on the physical phenomenon according to a conditional probability distribution. For the scope of this paper, the observation process is memoryless in the sense that the observation at time i only depends on the source outputs at time i. Hence, the observation process can be characterized by $P(u_1, \ldots, u_M | s_1, \ldots, s_L)$. Sensor m may also receive information from other sensors, as well as from the destination, as illustrated by the dotted lines in Fig. 2. Denoting the totality of this information as it is available to sensor m up to time $i-1$ by V_m^{i-1}, the signal transmitted by sensor m at time n can be expressed as

$$X_m[i] = F_m^{(i)}\left(U_m^j, V_m^{i-1}\right). \tag{2}$$

The transmitted signals satisfy a power, or more generally, a cost constraint of the form

$$E\left[\rho\left(X_1^j, X_2^j, \ldots, X_M^j\right)\right] \le \Gamma. \tag{3}$$

This is a generalization of the sum power constraint for all the sensors together. In some variations of our problem, it is also interesting to consider a family of simultaneous constraints, with cost functions $\rho_l(\cdot)$ and maximum expected cost Γ_l. This is a generalization of the individual power constraints for each sensor.

The final destination uses the outputs of the communication channel $Y^j = (Y_1^j, Y_2^j, \ldots, Y_N^j)$ to construct estimates $\hat{S}^j = (\hat{S}_1^j, \hat{S}_2^j, \ldots, \hat{S}_L^j)$. The task is to design the decoder G such that $\hat{S}^j = G(Y^j)$ as close to S^j as possible, in the sense of an appropriately chosen distortion measure $d(s^j, \hat{s}^j)$. For a fixed code, composed of the encoders $F_1, F_2, \ldots, F_M$ at the sensors and the decoder G, the achieved distortion Δ is computed as follows:

$$\Delta = E\left[d(S^j, \hat{S}^j)\right]. \tag{4}$$

[1]The discrete-time model is justified by arguing that the state of the system does not change very rapidly. This may be a serious restriction for certain scenarios. The continuous-time extension is currently under investigation.

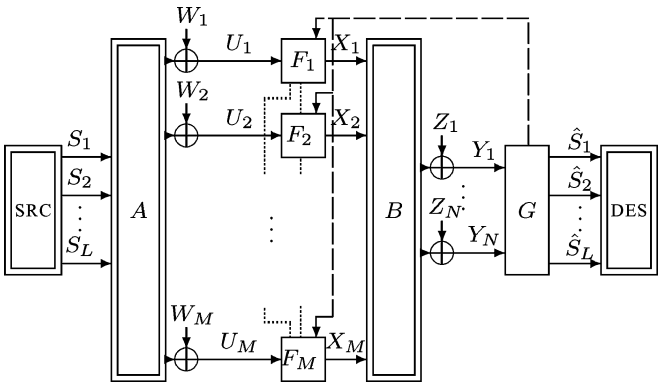

Fig. 3. The considered linear Gaussian sensor network: A and B are appropriate matrices.

The relevant figure of merit is, therefore, the tradeoff between the *cost* Γ of the transmission (3), and the achieved *distortion level* Δ (4). The problem studied in this paper is that of finding the optimal tradeoffs (Γ, Δ), where optimal is understood in an information-theoretic sense, i.e., irrespective of delay and complexity, as $j \to \infty$.

Scaling Law Notation: In this paper, we establish *scaling laws*, denoted by the symbol $\sim$, which here is taken to mean "asymptotic equivalence." More precisely, we write scaling laws as

$$f_1(M) \sim f_2(M) \tag{5}$$

which simply means that $\lim_{M\to\infty} f_1(M)/f_2(M) = c$, for some constant $c > 0$. The special case when $c = 1$ will be called a *strong scaling law*, since it correctly reports *both* the scaling behavior *and* the important constants, and will be denoted as $f_1(M) \overset{=}{\sim} f_2(M)$.

III. Gaussian Sensor Networks

A simple case of the sensor network defined in Section II and illustrated in Fig. 2 is the scenario when all involved statistics are Gaussian, and the observation process, as well as the communication channel are characterized by linear transforms. This is illustrated in Fig. 3. In particular, there are L physical sources (which, if desired, may be interpreted as the spatial or temporal bandwidth of the source), M sensors, and N receivers (base stations). The receivers are assumed to be ideally linked to each other: in the considered network model, the data collection point has access to the exact received value at each of the N base stations. This example is of particular interest because the scaling behavior can be cast in a simple, explicit formula.

A. Network Parameters

1) Source Bandwidth L and Observation Process: The source is characterized by L i.i.d.[2] circularly complex Gaussian random variables with mean zero and variance σ_S^2. The observation process is modeled as

$$U_m = W_m + \sum_{\ell=1}^{L} a_{m,\ell} S_\ell \tag{6}$$

for $m = 1, 2, \ldots, M$, where W_m is i.i.d. circularly complex Gaussian with mean zero and variance σ_W^2. We collect the coefficients $a_{m,\ell}$ into the matrix $A \in \mathbb{C}^{M\times L}$ defined as

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,L} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,L} \\ a_{3,1} & a_{3,2} & \cdots & a_{3,L} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M,1} & a_{M,2} & \cdots & a_{M,L} \end{pmatrix}. \tag{7}$$

The matrix A has $\min\{M, L\}$ singular values that we denote by $\alpha_1, \alpha_2, \ldots$. For the scope of this paper, the coefficients $a_{m,\ell}$ are chosen according to a given distribution, and we suppose that their values are known throughout the network. Later, we argue that under certain circumstances, the sensors need not know these values, and the destination only needs limited knowledge, without changing the scaling behavior.

In the sequel, we will use the shorthand notation $E_A[\cdot]$ to denote the expectation over the distribution of the entries of the matrix A.

2) Spatial Bandwidth $\tilde{N}$ of the Communication Channel: The communication channel is the standard additive white Gaussian multiple-access channel, modeled as

$$Y_n = Z_n + \sum_{m=1}^{M} b_{n,m} X_m \tag{8}$$

for $n = 1, 2, \ldots, N$, where Z_n is i.i.d. circularly complex Gaussian with mean zero and variance σ_Z^2. We collect the coefficients $b_{n,m}$ into the matrix $B \in \mathbb{C}^{N\times M}$ defined as

$$B = \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} & \cdots & b_{1,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{N,1} & b_{N,2} & b_{N,3} & \cdots & b_{N,M} \end{pmatrix} \tag{9}$$

where at first, the entries $b_{n,m}$ are assumed fixed for all times and known throughout the network.[3]

The matrix B has $\min\{N, M\}$ singular values that we denote by $\beta_1, \beta_2, \ldots$. Each nonzero singular value corresponds to a separate spatial channel. In this sense, the rank of B corresponds to the *spatial bandwidth* of the channel. For the scope of this paper, special interest is devoted to the case where $N \le M$ and, hence, we denote the number of nonzero singular values by

$$\tilde{N} \overset{\text{def}}{=} \text{rank}(B). \tag{10}$$

The harmonic mean of the squares of the nonzero singular values is denoted by

$$\mathcal{H}(\beta) = \left(\frac{1}{\tilde{N}} \sum_{n=1}^{\tilde{N}} \frac{1}{\beta_n^2} \right)^{-1} \tag{11}$$

[2]In the Gaussian case under mean-squared error distortion, it is without loss of generality to assume the sources $S_1, S_2, \ldots, S_L$ to be independent: Suppose instead that they have covariance matrix $\sum_S$ (and mean zero). Then, we can consider L *independent* random variables $\bar{S}_1, \bar{S}_2, \ldots, \bar{S}_L$, and simply change the matrix A in (7) below into AQ^H, where Q is the Karhunen–Loève transform of S, i.e., the unitary matrix Q such that $Q^H \sum_S Q$ is diagonal.

[3]In Section V, this assumption is relaxed; instead, we assume the coefficients $b_{n,m}$ to be randomly selected in every time slot, but known to the data collector, in line with the standard models for fading channels.

and their geometric mean by

$$\mathcal{G}(\beta) = \sqrt[\tilde{N}]{\prod_{n=1}^{\tilde{N}} \beta_n^2}. \tag{12}$$

3) Temporal Bandwidth K of the Communication Channel: The channel can be used K times for each source sample. This is equivalent to multiplying the bandwidth of the channel by a factor of K and, hence, permits to study the temporal bandwidth of the channel.

4) Power on the Communication Channel: Let $X_m[j]$ be the signal transmitted by sensor m in time slot j. The power on the communication channel is constrained as

$$\sum_{j=1}^{J} \sum_{m=1}^{M} E\left[|X_m[j]|^2\right] \leq J \frac{P_{\text{tot}}}{K} \tag{13}$$

for some suitably large J. That is, P_{tot} denotes the average total sensor power available per observation vector $(U_1, \ldots, U_M)$.

5) Target Distortion: The goal of the sensor network is to minimize the mean-squared error

$$D = \frac{1}{L} \sum_{\ell=1}^{L} E\left[|S_\ell - \hat{S}_\ell|^2\right] \tag{14}$$

where the expectation is over the distribution of the source vector S, the distribution of all the noises $W_1, \ldots, W_M$, Z, as well as the distribution of the matrices A and B.

B. Lower Bound to the Optimal Distortion Scaling Law

In [6] and [7], we presented lower bounds to the distortion that can be achieved in the Gaussian sensor network defined in Section III-A for a fixed observation matrix A, and a fixed channel matrix B. This paper extends these bounds to the case where A is randomly chosen and known to the sensors.

Theorem 1: The distortion that can be achieved in the Gaussian sensor network (defined in Section III-A) cannot be smaller than[4]

$$D_{\text{lower}}(M, P_{\text{tot}}, L, K, N) = E_\alpha\left[\frac{\sigma_S^2 \sigma_W^2}{\alpha^2 \sigma_S^2 + \sigma_W^2}\right] + c_1 \left(\frac{1}{c_2 + \frac{P_{\text{tot}}}{K\tilde{N}\sigma_Z^2}\mathcal{G}(\beta)}\right)^{\frac{K\tilde{N}}{L}} \tag{15}$$

where α is distributed according to the unordered[5] singular values of A, $\beta_1, \beta_2, \ldots, \beta_{\tilde{N}}$ are the nonzero singular values of B, σ_S^2 is the variance of the underlying sources, σ_W^2 is the variance of the observation noises, σ_Z^2 is the variance of the noises in the communication channel, P_{tot} is the total sensor transmit power for the K channel uses, $\tilde{N}$ is the rank of the matrix B, and

$$c_1 = 2^{E_A\left[\frac{1}{L}\sum_{\ell=1}^{L} \log \frac{\alpha_\ell^2 \sigma_S^4}{\alpha_\ell^2 \sigma_S^2 + \sigma_W^2}\right]} \tag{16}$$

$$c_2 = \frac{\mathcal{G}(\beta)}{\mathcal{H}(\beta)} \tag{17}$$

[4]For "large" enough total sensor power P_{tot}, see Remark 3.

[5]That is, α is a generic random variable whose distribution is equal to the distribution of the unordered singular values of A.

where $\alpha_1, \alpha_2, \ldots, \alpha_L$ are the singular values of the matrix A, and E_A denotes the expectation over the distribution of the matrix A.

The proof of this theorem is outlined in Appendix I.

Remark 1: In typical cases (see the examples below in Section III-C), c_1 and c_2 tend to a constant limit as $M \to \infty$ and, hence, are not directly scaling-law relevant.

Remark 2: This outer bound includes the case of arbitrary collaboration between the sensors, and of arbitrary feedback signals from the data collection point to the sensors.

Remark 3: The statement of Theorem 1 is valid for "large enough" P_{tot}. The full solution is given in Appendix I, where it is shown that for some of the most interesting cases (including both examples discussed next in Section III-C), as $M \to \infty$, any $P_{\text{tot}} \geq 0$ is "large enough."

Remark 4: For the lower bound (15), the spatial and temporal bandwidth, $\tilde{N}$ and K, only appear in the shape of the product $K\tilde{N}$, which we sometimes refer to as the *spatio-temporal*, or *effective* bandwidth of the communication channel.

C. Examples

Example 1 (Idealized Case): To illustrate Theorem 1, note first that the dependence of the distortion on the number of sensors M is through the singular values α_ℓ and β_n. Suppose that L, K, and N are fixed, while M increases. For the sake of this simple example, consider the case where A has orthogonal columns, and B has orthogonal rows. In particular, suppose that all entries have the same magnitude, denoted by α_0 and β_0, respectively, but their phases are chosen such as to ensure orthogonality, like, e.g., in the Fourier matrix. Then

$$\alpha_{\ell,M}^2 = M\alpha_0^2, \qquad \text{for} \quad \ell = 1, 2, \ldots, L \tag{18}$$

$$\beta_{n,M}^2 = M\beta_0^2, \qquad \text{for} \quad n = 1, 2, \ldots, N. \tag{19}$$

Hence, $\tilde{N} = N$ and (15) evaluates to

$$D_{\text{lower}}(M, P_{\text{tot}}, L, K, N) = \frac{\sigma_S^2 \sigma_W^2}{M\alpha_0^2 \sigma_S^2 + \sigma_W^2} + \frac{\alpha_0^2 \sigma_S^4}{\alpha_0^2 \sigma_S^2 + \frac{\sigma_W^2}{M}} \left(\frac{1}{1 + \frac{MP_{\text{tot}}\beta_0^2}{(KN\sigma_Z^2)}}\right)^{\frac{KN}{L}}. \tag{20}$$

The key insight of this example is that the distortion scales at best like

$$D(M, P_{\text{tot}}, L, K, N) \sim \frac{1}{M} \tag{21}$$

irrespective of the available resources in terms of power P_{tot}, temporal bandwidth K, and spatial bandwidth N of the communication channel. We develop this analysis further in Section III-D.

As another example, in [7], a special sensor network geometry is considered that could be called *circulant* sensor network: underlying sources, sensors, and base stations all lie on concentric circles. Such a geometry gives rise to circulant matrices $A^H A$ and BB^H. The resulting behavior of the singular values can be shown to exhibit (essentially) the behavior characterized by (18) and (19), respectively, leading to a lower bound of the

type of (20). Here, we continue to consider an example where A is chosen *randomly*.

Example 2 (Semi-Circle Law): Suppose now that the entries of the matrix A are selected i.i.d.[6] according to a given distribution $p(a)$ with zero mean and variance σ_a^2. Then, we can define the scaled matrix $\tilde{A}$ as follows:

$$\tilde{A} = \frac{1}{\sqrt{M}} \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,L} \\ a_{2,1} & a_{2,2} & \dots & a_{2,L} \\ a_{3,1} & a_{3,2} & \dots & a_{3,L} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M,1} & a_{M,2} & \dots & a_{M,L} \end{pmatrix}. \tag{22}$$

Denote the singular values of the matrix $\tilde{A}$ by $\tilde{\alpha}_1, \dots$, and consider the following probability distribution:

$$P_{\tilde{A}}(x) = \frac{1}{L} \left| \{ \ell : \tilde{\alpha}_\ell^2 < x \} \right| \tag{23}$$

where $|\{\ell : \tilde{\alpha}_\ell < x\}|$ denotes the number of eigenvalues $\tilde{\alpha}_\ell^2$, for $\ell = 1, 2, \dots, L$, that are smaller than x. It is well known that as long as $L \leq M$, the quantity $P_{\tilde{A}}(x)$ converges in probability to the following version of the semi-circle law [8], [9]

$$P_{\tilde{A}}(x) = \frac{M}{2\pi L x} \sqrt{(x - d_1)(d_2 - x)} \tag{24}$$

for $d_1 < x < d_2$, where $d_1 = (1 - \sqrt{L/M})^2$ and $d_2 = (1 + \sqrt{L/M})^2$, when $M \to \infty$ and the ratio L/M is fixed. While the precise evaluation of the bound of Theorem 1 is beyond the framework of this paper, (24) reveals that as long as $L < M$, all eigenvalues $\tilde{\alpha}_\ell$ are strictly larger than zero (in probability, as $M \to \infty$). Clearly

$$\alpha_\ell^2 = M \tilde{\alpha}_\ell^2. \tag{25}$$

Hence, we find that

$$E_\alpha \left[\frac{\sigma_S^2 \sigma_W^2}{\alpha^2 \sigma_S^2 + \sigma_W^2} \right] = \frac{1}{M} E_{\tilde{\alpha}} \left[\frac{\sigma_S^2 \sigma_W^2}{\tilde{\alpha}^2 \sigma_S^2 + \frac{\sigma_W^2}{M}} \right] \stackrel{\text{def}}{=} \frac{c_0}{M} \tag{26}$$

where, as a consequence of (24), c_0 tends to a constant as $M \to \infty$. Moreover, using our findings in (16), we can rewrite in terms of $\tilde{\alpha}_\ell$ as follows:

$$\begin{aligned} c_1 &= 2^{E_A \left[\frac{1}{L} \sum_{\ell=1}^{L} \log \frac{\alpha_\ell^2 \sigma_S^4}{\alpha_\ell^2 \sigma_S^2 + \sigma_W^2} \right]} \\ &= 2^{E_{\tilde{A}} \left[\frac{1}{L} \sum_{\ell=1}^{L} \log \frac{\tilde{\alpha}_\ell^2 \sigma_S^4}{\tilde{\alpha}_\ell^2 \sigma_S^2 + \frac{\sigma_W^2}{M}} \right]} \end{aligned} \tag{27}$$

which tends to a constant as $M \to \infty$ (also a consequence of (24)). To conclude this example, suppose that the channel matrix B is chosen like in Example 1. Then, we find the following lower bound:

$$D_{\text{lower}}(M, P_{\text{tot}}, L, K, N) = \frac{c_0}{M} + c_1 \left(\frac{1}{1 + \frac{M P_{\text{tot}} \beta_0^2}{(K N \sigma_Z^2)}} \right)^{\frac{KN}{L}} \tag{28}$$

where c_0 and c_1 are the constants defined in (26) and (27), respectively. In conclusion, we remark that this is essentially the same behavior (as a function of M) as in the simple and idealized case studied in Example 1.

D. Necessary Power and Bandwidth

The lower bound to the distortion presented in this section, (15), is expressed in an intuitively pleasing fashion in terms of two separate contributions, the first of which is independent of the communication capabilities. This term is solely due to the fact that the underlying physical reality cannot be observed perfectly, but rather is always subject to measurement noise. As the number of sensors M increases, the observation process is characterized by a matrix sequence $A^{(M)}$, with singular values $\alpha_\ell^{(M)}$, and the behavior of the observation noise term is governed by the speed at which the slowest-growing singular value of the matrix sequence $A^{(M)}$ increases.

For the considered Examples 1 and 2, the smallest singular value increases like $\sqrt{M}$, and as a consequence, the first distortion term decreases like $1/M$. In general, if the slowest-growing singular value of the matrix sequence $A^{(M)}$ increases like $M^{\delta/2}$, the best distortion scaling behavior one can expect is $1/M^\delta$, and the key question becomes: What resources are necessary in order to achieve this optimum distortion scaling law? In other words, how much power and bandwidth must be invested such that the second summand in (15) decays at the same rate as the first? This question can be answered based on Theorem 1. To summarize this insight:

Corollary 2: Suppose that the slowest-growing of the singular values $\alpha_\ell^{(M)}$, $\ell = 1, 2, \dots, L$, of the observation matrix sequence $A^{(M)}$ grows like $M^{\delta/2}$. Then, the optimal distortion scaling law is at best $1/M^\delta$, and the power P_{tot}, and the bandwidth K and $\tilde{N}$, required to sustain this distortion scaling law must satisfy

$$\left(\frac{1}{c_2 + \frac{P_{\text{tot}}}{K \tilde{N} \sigma_Z^2} \sqrt[\tilde{N}]{\prod_{n=1}^{\tilde{N}} \beta_n^2}} \right)^{\frac{K\tilde{N}}{L}} \sim \frac{1}{M^\delta} \tag{29}$$

Example 1, Continued: For the idealized sensor network considered in Example 1, in order to achieve the optimal distortion scaling law of $1/M$, the resource parameters P_{tot}, K, and $\tilde{N} = N$ must be chosen such that

$$\left(1 + \frac{M P_{\text{tot}} \beta_0^2}{(K N \sigma_Z^2)} \right)^{\frac{KN}{L}} \sim M. \tag{30}$$

Hence, we find the necessary total power to sustain the optimal distortion scaling law of $1/M$ must scale at least as follows:

$$P_{\text{tot}} \sim K N M^{\frac{L}{KN} - 1}. \tag{31}$$

[6]Assuming the entries of A to be i.i.d. is certainly an inappropriate approach for strongly "geometric" phenomena such as sound or image; it may be of interest to other types of phenomena, including gas and liquid concentrations under real-world conditions.

According to this formula, provided that $L = KN$, a constant total power shared by all the sensors may be sufficient to achieve the optimal scaling behavior. In Section IV, we confirm this for a class of sensor networks. Moreover, suppose that the number of sources (i.e., the source bandwidth) L increases with M. Then, the formula says that unless the effective channel bandwidth KN also increases with M, the minimum required total power increases *exponentially* in the number of sources L.

IV. Coding Schemes

In order to assess the value of the bound developed in Section III, this section determines *achievable* (or approachable) cost-distortion tradeoffs.

First, in Section IV-A, we analyze the performance of a scheme that separates source from channel coding. In particular, we present a lower bound on the distortion for *any* coding strategy under which each sensor attempts to communicate a separate message (a bit stream) to the destination with vanishingly small error probability. The corresponding scaling behavior is exponentially worse than the fundamental lower bound of Theorem 1, raising the question whether our lower bound is too loose.

In Section IV-B, we show that this is not the case in general: we present a class of networks for which the scaling behavior as predicted by Theorem 1 can indeed be achieved. In extension of previous work [6], [7], [10], we analyze a simple joint source-channel coding strategy where the sensors transmit scaled sensor readings directly for the case $L = N \geq 1$ (while $K = 1$).

A. Separate Source and Channel Coding

It is well known that separate source and channel coding does not lead to optimal performance in general networks, but it is not known whether it leads to a different scaling behavior. To analyze the optimum performance for a scheme that separates source from channel coding, one must evaluate the optimum rate-distortion performance for the source network side, and combine it with the capacity of the channel network. This is illustrated in Fig. 4. The key characteristic of such a strategy is that each sensor attempts to convey a message (a bit sequence) across the communication channel in such a way that the base station can decode it with vanishing error probability. Unfortunately, only very few results are known for the general rate-distortion behavior of source networks and the capacity-cost behavior of channel networks. Consider the special case where A and B are fixed matrices where all entries have unit magnitude, and the columns of the matrix A, as well as the rows of B are orthogonal. Then, a lower bound to the optimum performance for the source coding problem in Fig. 4 (referred to as the CEO problem [11], [12]) has been found in [13, eq. (6)]. The rate available to communicate the resulting M bit streams to the destination cannot be larger than the capacity of the Gaussian multiple-input–multiple-output channel with inputs $X_1, \ldots, X_M$ and outputs $Y_1, \ldots, Y_n$ (see Fig. 4),

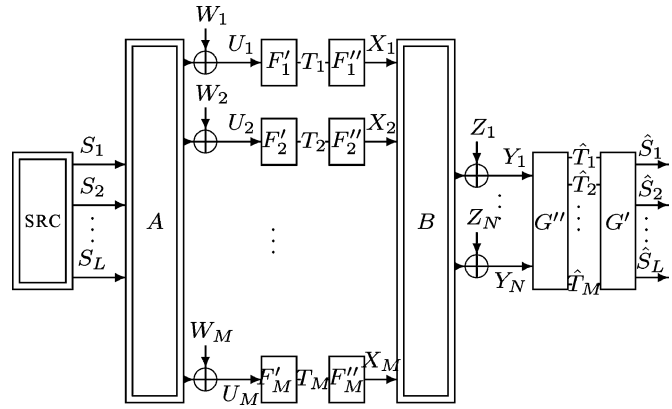

Fig. 4. Schematic rendering of separate source and channel coding for the single-source Gaussian sensor network. The bit sequence that sensor m needs to communicate to the base stations is denoted by T_m. F_m' and F_m'' denote source and channel encoding at sensor m, respectively, and G'' and G' denote channel and source decoding at the base stations, respectively. Note that for simplicity, we assume that the sensors cannot collaborate (see Remark 6), and that no feedback is available.

which is well known [see, e.g., [14], [19], and (50)–(52)]. For sufficiently large M, this evaluates to

$$D(M, P_{\text{tot}}, L, K, N) \geq \frac{\sigma_S^2 \sigma_W^2}{\sigma_S^2 \frac{KN}{L} \log_2 \left(1 + \frac{MP_{\text{tot}}}{KN\sigma_Z^2}\right) + \sigma_W^2}. \tag{32}$$

To compare with the lower bound (Theorem 1), suppose for example that $P_{\text{tot}} = MP$, and that L, K, and N are constant. Then, the resulting distortion behaves like

$$D(M, P) \sim \frac{1}{\log_2 MP} \tag{33}$$

or exponentially worse than the lower bound of Theorem 1, see (21).

Remark 5: In the special case where there are $N = M$ base stations and the matrix B is full rank, i.e., $\tilde{N} = M$, the distortion scaling behavior of separate source and channel coding, given in (32) becomes $1/M$ also, closing the gap. The simple case where B is the (M-dimensional) identity matrix, and error-free messages are transmitted (rather than Gaussian sources, as in this paper) has been addressed in [4].

Remark 6: When the sensors can collaborate perfectly and at no cost, it can be shown easily that the bound (32) no longer applies. Rather, in that case, the overall scenario becomes a point-to-point communication system for which the optimum performance is known to coincide with Theorem 1.

B. Joint Source-Channel Coding

In this section, *joint* source and channel coding strategies are analyzed. The goal is to reduce the exponential gap between the performance established in (32), and the lower bound of (21). From (31), we see that if the number of degrees of freedom of the source L is equal to the spatio-temporal channel bandwidth $K\tilde{N}$, then the total power P_{tot}, as a function of M, must be at least constant. In this subsection, we present a class of sensor networks for which a constant total power P_{tot} is also sufficient to achieve the distortion scaling law given by the lower bound determined in Section III-B.

1) Single Base Station: The first special case we consider is the scenario when $N = \tilde{N} = 1$. Hence, in order for source and channel bandwidth to be equal, $L = K$. To keep notation simple, the following theorem determines an achievable distortion for the special case $b_m = 1$, for $m = 1, 2, \ldots, M$. It is straightforward to extend the result to the case of general b_m, at the expense of extra notation.

Theorem 3: For the Gaussian sensor network defined in Section III-A with $L = K$ and $N = 1$, and with $b_m = 1$, for $m = 1, 2, \ldots, M$, the following distortion can be achieved:

$$D_1(M, P_{\text{tot}}(M)) = E_\alpha\left[\frac{\sigma_S^2\sigma_W^2}{\alpha^2\sigma_S^2 + \sigma_W^2}\right] + \frac{1}{P_{\text{tot}}(M)} E_A\left[\sum_{\ell=1}^{L} \frac{\alpha_\ell^2 u_{[\ell]}\sigma_Z^2}{\left(\frac{\alpha_\ell^2+\sigma_W^2}{\sigma_S^2}\right)^2}\right] \quad (34)$$

where $\alpha_1, \alpha_2, \ldots, \alpha_L$ denote the singular values of A, and α is distributed according to their unordered distribution, σ_S^2 is the variance of the underlying sources, σ_W^2 is the variance of the observation noises, σ_Z^2 is the variance of the noises in the communication channel, P_{tot} is the total sensor transmit power for the K channel uses, and $u_{[\ell]}$ denotes the ℓth largest of the sensed powers u_m, $u_m = \sum_{\ell=1}^{L} |a_{m,\ell}|^2\sigma_S^2 + \sigma_W^2$.

To establish this theorem, it suffices to consider a scheme that dedicates one of the L channel uses for each of the L sources. In the dedicated channel use ℓ, the sensors apply the filtering coefficients appropriate to estimate the source S_ℓ from the observations $U_1, U_2, \ldots, U_M$.

Interestingly, this achievable distortion *does* coincide (in a scaling sense) with the lower bound of Thm. 1 in some cases:

Example 1, Continued: For the idealized sensor network considered in Example 1, the bound evaluates to

$$D_1(M, P_{\text{tot}}, L) = \frac{\sigma_S^2\sigma_W^2}{M\alpha_0^2\sigma_S^2 + \sigma_W^2} + \tilde{c}_1 \frac{\alpha_0^2\sigma_S^4}{\alpha_0^2\sigma_S^2 + \frac{\sigma_W^2}{M}} \cdot \frac{1}{\frac{MP_{\text{tot}}}{(L^2\sigma_Z^2)}} \quad (35)$$

where $\tilde{c}_1 = (\alpha_0^2\sigma_S^2 + \sigma_W^2)/(\alpha_0^2\sigma_S^2 + \sigma_W^2/M)$ tends to a constant as $M \to \infty$. Clearly, this describes the same scaling behavior as the lower bound, (20) with $\beta_0^2 = 1$, $K = L$, and $N = 1$, and it also shows that a *constant* total power is sufficient to achieve the optimum $1/M$ distortion scaling law, in line with (31).

2) Matched Observation and Channel Matrices, A and B: When there are many base stations N, a more efficient strategy can be implemented: from each channel use, up to $\tilde{N}$ sources can be efficiently estimated, based on the N received values at the base stations. Such a strategy works whenever the matrices A and B are appropriately matched.

To keep matters simple, suppose that $L = N \geq 1$ and $K = 1$, and that the matrix A is fixed, rather than randomly chosen. Then, a strong notion of "matched" matrices A and B can be defined as follows.

Definition 1 (Matched Observation and Channel Matrices): Denote the singular value decomposition of the matrix $A \in \mathbb{C}^{M\times L}$ by $A = U_a D_a V_a^H$, and of the matrix $B \in \mathbb{C}^{N\times M}$ by $B = U_b D_b V_b^H$. The matrices A and B are called *matched* if $L = N$ and there exists a diagonal matrix $\Lambda_F \in \mathbb{C}^{M\times M}$ such that

$$V_b^H \Lambda_F U_a = Q\,\text{diag}(\gamma_1, \gamma_2, \ldots, \gamma_L) Q^H \quad (36)$$

for some constants $\gamma_\ell \in \mathbb{C}$, for $\ell = 1, 2, \ldots, L$, and some permutation matrix Q.

It is appropriate to point out that while this matching requirement enables simple proofs, it is *not* a necessary condition to achieve the scaling law predicted by Theorem 1.

Example 3: The following are examples of matched matrices according to Definition 1.

1) When A and B are simply vectors ($L = N = 1$) with enough nonzero entries, they are matched. This is a special case of the analysis of Section IV-B1.
2) Circulant case (see [7]): When both $A^H A$ and BB_H are circulant matrices, then A and B are matched with $\Lambda_F = I_M$.

For matched matrices A and B, it is easy to establish the following achievable distortion.

Theorem 4: If the matrices A and B are matched, then the following distortion is achievable:

$$D_1(M, P_{\text{tot}}, L = N, K = 1) = \frac{1}{L}\sum_{\ell=1}^{L} \frac{\sigma_S^2\sigma_W^2}{|\gamma_\ell|^2\alpha_\ell^2\sigma_S^2 + \sigma_W^2} + \frac{\nu LM}{\frac{P_{\text{tot}}}{\sigma_Z^2}} \cdot \frac{1}{L}\sum_{\ell=1}^{L} \frac{|\gamma_\ell|^2\alpha_\ell^2\sigma_S^2}{\beta_\ell^2\left(|\gamma_\ell|^2\alpha_\ell^2\sigma_S^2 + \sigma_W^2\right)^2} \quad (37)$$

where $\nu = 1/LM(\|\Lambda_F A\|_F^2\sigma_S^2 + \|\Lambda_F\|_F^2\sigma_W^2)$ and $\|\cdot\|_F$ denotes the Frobenius norm.

The proof of this theorem is outlined in Appendix II.

Comparing this to Theorem 1, we find a scaling law whenever the $|\gamma_\ell|^2$ remain strictly larger than zero as $M \to \infty$. This can be phrased as follows.

Theorem 5: If the matrix sequences $A^{(M)}$ and $B^{(M)}$ are matched for every M, the corresponding values of $|\gamma_\ell|$, $\ell = 1, 2, \ldots, L$, remain strictly larger than zero as $M \to \infty$, $1/LM(\|\Lambda_F A\|_F^2\sigma_S^2 + \|\Lambda_F\|_F^2\sigma_W^2)$ is bounded, the slowest-growing of the singular values $\alpha_\ell^{(M)}$ increases at least like $\sqrt{M}$, and, for $n = 1, 2, \ldots, N$, $\beta_n^{(M)} = \tilde{\beta}_n f_\beta(M)$, then the distortion scaling law is given by

$$D(M, P_{\text{tot}}, L = N, K = 1) \sim \frac{1}{L}\sum_{\ell=1}^{L} \frac{\sigma_S^2\sigma_W^2}{\alpha_\ell^2\sigma_S^2 + \sigma_W^2} + \frac{L}{P_{\text{tot}} f_\beta^2(M)}. \quad (38)$$

For an example of Theorem 5, see [10].

Remark 7: Since our lower bound (Theorem 1) includes both feedback and collaboration, Theorem 5 establishes that in the matched case, neither feedback nor collaboration are scaling-law relevant.

C. Extensions

Matchable observation and channel matrices. When $L > N$, the matrices A and B are not matched according to Definition 1. But suppose that the channel can be used $K = L/N$ times. This situation can clearly be handled by combining the arguments of Sections IV-B1 and IV-B2.

When the matrices are fundamentally mismatched, it becomes interesting for the sensors to collaborate: that way, they can implement more general overall transforms. This leads to a study of matrices for which there exists a coding scheme that matches them favorably.

Feedback. Another way to incorporate unmatched matrices is through the use of feedback. Such coding strategies are analyzed in [15].

Limited knowledge of the parameters. An interesting feature of the strategy used to establish Theorem 4 is its robustness. For example, limited knowledge of the values of the observation parameters can be shown not to be scaling-law relevant for a nontrivial class of sensor networks.

V. Synchronization

The lower bound derived in Section III applies whether or not perfect synchronization is available to the sensors. However, the "uncoded forwarding" coding scheme presented in [6] and [10] (and extended to more general cases in Section IV) seems to require perfect synchronization and, therefore, appears to be impractical. In this section, we study the case when the channel is subject to fading effects. More precisely, each sensor's transmitted signal is multiplied by an independent complex random variable b_m, for $m = 1, 2, \ldots, M$, i.i.d. over time. The precise value of this random variable is unknown to the sensors, but their distribution is known. This may model the situation where the sensors transmit modulated signals, but the carriers are not in phase. It may also model a pulsed (e.g., ultra-wideband) communication system, where the pulses do not arrive exactly at the same time, but are distributed over an interval. We show how our scheme performs under such conditions. In particular, we establish that the optimal scaling law is achieved as soon as the distribution of b_m has a nonzero mean for enough of the sensors. In the pulsed example above, this means that the distribution of arrival times over the given interval is not uniform over the entire interval. Rather, the pulse of sensor m is more likely to arrive, say, around the center of the interval.

For the purpose of this exposition, and because it suffices to illustrate the principles, we consider the case $L = K = N = 1$. The destination receives

$$Y = Z + \sum_{m=1}^{M} b_m X_m \tag{39}$$

where the coefficients b_m are complex-valued and not known to the sensors. This models the fact that the sensors are not synchronized. Clearly, the properties of the coefficients b_m determine the scaling behavior of the network performance. A lower bound can be obtained by assuming that the destination knows the coefficients b_m. Then, a slight extension of the lower bound of Theorem 1 leads to the following lower bound:

$$D_{\text{lower}}(M, P_{\text{tot}}) = \frac{\sigma_S^2 \sigma_W^2}{a(M)\sigma_S^2 + \sigma_W^2} + \frac{1}{1 + \frac{P_{\text{tot}}}{\sigma_Z^2} \sum_{m=1}^{M} E\left[|b_m|^2\right]} \frac{\sigma_S^4 a(M)}{a(M)\sigma_S^2 + \sigma_W^2} \tag{40}$$

where $a(M) = \sum_{m=1}^{M} |a_m|^2$. We consider a specific distribution of the coefficients b_m for which this lower bound is achievable (in the scaling sense): Suppose that b_m and b_j are independent of each other $(m \neq j)$, and have nonzero mean. Then, the distortion achieved by a simple forwarding scheme is found to be at most

$$\begin{aligned} D_1(M, P_{\text{tot}}) = {} & \frac{\sigma_W^2}{a^2(M)} \sum_{m=1}^{M} \frac{|a_m|^2 E\left[|b_m|^2\right]}{|E[b_m]|^2} + \frac{\sigma_S^2}{a^2(M)} \\ & \times \sum_{m=1}^{M} \frac{|a_m|^4 Var(b_m)}{|E[b_m]|^2} + \frac{\sigma_Z^2}{a^2(M) P_{\text{tot}}} \\ & \times \sum_{m=1}^{M} \frac{|a_m|^2 \left(|a_m|^2 \sigma_S^2 + \sigma_W^2\right)}{|E[b_m]|^2}. \end{aligned} \tag{41}$$

Clearly, this distortion does not generally coincide with the lower bound (40). However, in interesting cases, (40) and (41) describe the same scaling behavior. One of these cases is described by the following result.

Theorem 6: Suppose that $0 < a_{\min} \leq |a_m| \leq a_{\max} < \infty$ and $E[|b_m|^2] \leq d_3 < \infty$, for $m = 1, 2, \ldots, M$. If at least a fraction ϵM of the b_m, $m = 1, 2, \ldots, M$, have $|E[b_m]| \geq \delta > 0$, and the total power P_{tot} is a nondecreasing function of M, then the optimum distortion scaling law is

$$D(M, P_{\text{tot}}) \sim \frac{c_3}{M} + \frac{c_4}{M P_{\text{tot}}} \tag{42}$$

where c_3 and c_4 are independent of M and P_{tot} (but depend on $a_{\min}$, $a_{\max}$, d_3, ϵ, and δ).

An outline of the proof of this theorem is given in Appendix III.

VI. Conclusion

This paper analyzes the fundamental tradeoffs between the number of sensors, their total transmit power, the number of degrees of freedom of the source, the spatio-temporal communication bandwidth, and the end-to-end distortion. In a nutshell, it is found that for typical situations, the distortion goes down at best like $1/M$, where M is the number of sensors, *irrespective* of the communication resources available to the sensors. Hence, it is meaningful to allocate communication resources in such a way as to actually achieve this distortion behavior. It is shown that if the communication scheme is designed according to the traditional and omnipresent paradigm of separation into a compression stage followed by a channel coding stage, an *exponential* waste of communication resources (as a function of M) results. By contrast, for the specific situations studied in this paper, a very simple *joint* source-channel communication strategy can be shown to actually perform optimally. Extensions of this work include the study of more general statistical and topological sensor network models along the same lines.

Appendix I

Proof: (Outline of the Proof of Theorem 1): Our lower bound is the best achievable performance for an idealized system where the sensors can collaborate arbitrarily (and for free). This idealized system is a point-to-point system. The optimum performance for point-to-point systems can be found

from Shannon's separation theorem [16, Th. 21]. For the case at hand, there is a slight difference with respect to Shannon's scenario in that the source is not directly observed, but merely in a noisy fashion. A modified version of the separation theorem can be found in [17, p. 78] (see also [18, p. 136]). The only modification is to replace the standard rate-distortion function by the "remote" rate-distortion function. The minimum distortion $D_{\min}$ for a given total power P_{tot} satisfies

$$D_{\min} = D_{\text{remote}}\left(C(P_{\text{tot}})\right) \tag{43}$$

where $D_{\text{remote}}(R)$ is the remote distortion-rate function, given by $D_{\text{remote}}(R) = \min(1/L)\sum_{\ell=1}^{L} E[|S_\ell - \hat{S}_\ell|^2]$, where the minimization is over all distributions $P(\hat{s}_1,\ldots,\hat{s}_L|u_1,\ldots,u_M)$ that satisfy $I(U_1,U_2,\ldots,U_M;\hat{S}_1,\hat{S}_2,\ldots,\hat{S}_L) \le R$. In extension of the arguments in [17, p. 124ff,], the remote rate-distortion function for the case at hand, assuming that the matrix A is known both at the encoder and at the decoder, can be expressed in parametric form as

$$D_{\text{remote}}(R) = E_A\left[\frac{1}{L}\sum_{\ell=1}^{L}\frac{\sigma_S^2\sigma_W^2}{\alpha_\ell^2\sigma_S^2+\sigma_W^2}\right] + E_A\left[\frac{1}{L}\sum_{\ell=1}^{L}\min\left\{\frac{\alpha_\ell^2\sigma_S^4}{\alpha_\ell^2\sigma_S^2+\sigma_W^2},\nu\right\}\right] \tag{44}$$

where

$$\sum_{\ell=1}^{L} E_A\left[\left(\log\frac{\alpha_\ell^2\sigma_S^4}{(\alpha_\ell^2\sigma_S^2+\sigma_W^2)\nu}\right)^+\right] = R. \tag{45}$$

The first term in the distortion expression can be rewritten by considering the *unordered* singular values of A as

$$E_A\left[\frac{1}{L}\sum_{\ell=1}^{L}\frac{\sigma_S^2\sigma_W^2}{\alpha_\ell^2\sigma_S^2+\sigma_W^2}\right] = E_\alpha\left[\frac{\sigma_S^2\sigma_W^2}{\alpha^2\sigma_S^2+\sigma_W^2}\right]. \tag{46}$$

More interestingly, if

$$R \ge \sum_{\ell=1}^{L} E_A\left[\log\frac{\alpha_\ell^2\left(\alpha_{\min}^2\sigma_S^2+\sigma_W^2\right)}{(\alpha_\ell^2\sigma_S^2+\sigma_W^2)\,\alpha_{\min}^2}\right] \tag{47}$$

where $\alpha_{\min}^2$ denotes the square of the smallest singular value of the matrix A over the ensemble of A from which A is selected (and over which the expectation is evaluated), then the argument of the logarithm in (45) is always larger than one, hence

$$\nu = 2^{E_A\left[\log\sqrt[L]{\prod_{\ell=1}^{L}\frac{\alpha_\ell^2\sigma_S^4}{\alpha_\ell^2\sigma_S^2+\sigma_W^2}}\right]}2^{-\frac{R}{L}} \tag{48}$$

and the remote rate-distortion function is found to be

$$D_{\text{remote}}(R) = E_\alpha\left[\frac{\sigma_S^2\sigma_W^2}{\alpha_1^2\sigma_S^2+\sigma_W^2}\right] + 2^{E_A\left[\log\sqrt[L]{\prod_{\ell=1}^{L}\frac{\alpha_\ell^2\sigma_S^4}{\alpha_\ell^2\sigma_S^2+\sigma_W^2}}\right]}2^{-\frac{R}{L}}. \tag{49}$$

Note that if $\alpha_{\min}^2$ increases without bound as $M \to \infty$ (as is the case in the examples of Section III-C), then the RHS of (47) tends to zero, and (48) becomes valid for *any* rate $R \ge 0$. Hence, (44) can be simplified using formula (48).

The capacity $C(P_{\text{tot}})$ needed to complete the proof is simply the capacity of the Gaussian vector channel characterized by the matrix B, in parametric form given by

$$C(P_{\text{tot}}) = \sum_{n=1}^{\tilde{N}}\left(\log_2\left(\beta_n^2\gamma\right)\right)^+ \tag{50}$$

where γ is chosen such that

$$\sum_{n=1}^{\tilde{N}}\left(\gamma - \frac{1}{\beta_n^2}\right)^+ = P_{\text{tot}}. \tag{51}$$

In order to obtain a more compact description, let us assume a small spread of the singular values β_n, for $n = 1,2,\ldots,\tilde{N}$, or alternatively a sufficiently large power P_{tot}, the capacity of K uses of the multiple-input–multiple-output channel characterized by the matrix B and by additive white Gaussian noises of variance σ_Z^2 is given by

$$C(P_{\text{tot}}) = K\tilde{N}\log_2\left(\frac{\mathcal{G}(\beta)}{\mathcal{H}(\beta)} + \mathcal{G}(\beta)\frac{P_{\text{tot}}}{K\tilde{N}\sigma_Z^2}\right) \tag{52}$$

where P_{tot} is the total power available for K channel uses. Notice that this is true whether or not feedback is available since the capacity of a memoryless channel is not increased by feedback. Using this in (43) yields the claimed bound. ■

APPENDIX II

Proof: (Outline of the Proof of Theorem 4): By assumption, since the matrices A and B are matched, there exists a diagonal matrix Λ_F that satisfies (36). Denote the diagonal entries of Λ_F by $\lambda_m = \{\Lambda_F\}_{mm}$. Let sensor m transmit $X_m = c\lambda_m U_m$, where c must be chosen to meet the power constraint. The data collection point then simply uses the minimum mean-squared error estimator of $(S_1,S_2,\ldots,S_L)$ based on the observations $(Y_1,Y_2,\ldots,Y_N)$. By standard arguments, the resulting mean-squared error is found to be

$$D = \sigma_S^2\left(1 - \frac{1}{L}\sum_{\ell=1}^{L}\frac{c^2|\gamma_\ell|^2\alpha_\ell^2\beta_\ell^2}{c^2|\gamma_\ell|^2\beta_\ell^2\left(\alpha_\ell^2+\frac{\sigma_W^2}{\sigma_S^2}\right)+\frac{\sigma_Z^2}{\sigma_S^2}}\right). \tag{53}$$

To determine c, we calculate the total power

$$\sum_{m=1}^{M}E\left[|X_m|^2\right] = \sum_{m=1}^{M}c^2|\lambda_m|^2E\left[|U_m|^2\right] = \sum_{m=1}^{M}c^2|\lambda_m|^2\left(\sigma_W^2+\sum_{\ell=1}^{L}|a_{m,\ell}|^2\right). \tag{54}$$

Hence, if c is such that (54) is at most P_{tot}, the power constraint is satisfied. Introducing this in (53) and using simple elementary manipulation yields the claimed bound. ■

APPENDIX III

Proof: (Outline of the Proof of Theorem 6): Equation (40) is established by adapting (52) to the case of fading with the co-

efficients known at the receiver, see [14] and [19]. Specifically, this can be expressed for the case $K = N = 1$ as

$$C(P_{\text{tot}}) = E_{\beta_1}\left[\log_2\left(1 + \frac{\beta_1^2 P_{\text{tot}}}{\sigma_Z^2}\right)\right] \tag{55}$$

where E_{β_1} is the expectation over β_1, the singular value of the matrix B. By Jensen's inequality, this is upper bounded

$$C(P_{\text{tot}}) \leq \log_2\left(1 + \frac{P_{\text{tot}}}{\sigma_Z^2} E_{\beta_1}\left[\beta_1^2\right]\right) \tag{56}$$

Using this in (43) yields the claimed bound, noting that $\beta_1^2 = \sum_{m=1}^{M} |b_m|^2$. —The lower bound (41) is simply the performance for the scheme where each sensor with $E[b_m] \geq \delta > 0$, in each time step, transmits an appropriately scaled version of its observation U_m, and the scaling law follows immediately by combining (40) with (41). ■

REFERENCES

[1] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.

[2] D. Marco, E. J. Duarte-Melo, M. Liu, and D. L. Neuhoff, "On the many-to-one transport capacity of a dense wireless sensor network and compressibility of its data," in *Lecture Notes in Computer Science*, L. J. Guibas and F. Zhao, Eds. New York: Springer, Apr. 2003, Proc. 2nd Int. Workshop Inf. Sensor Netw. (IPSN'03), pp. 1–16.

[3] P. Ishwar, A. Kumar, and K. Ramchandran, "Distributed sampling for dense sensor networks: a bit-conservation principle," in *Lecture Notes in Computer Science, Springer*, J. Guibas and F. Zhao, Eds. New York: Springer, Apr. 2003, Proc. 2nd Int. Workshop Inf. Sensor Netw. (IPSN'03), pp. 17–31.

[4] J. Barros and S. Servetto, "Reachback capacity with noninterfering nodes," presented at the IEEE Int. Symp. Inf. Theory, Yokohama, Japan, 2003.

[5] M. Gastpar and M. Vetterli, "On the capacity of large Gaussian relay networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 765–779, Mar. 2004.

[6] ——, "Source-channel communication in sensor networks," in *Lecture Notes in Computer Science*, J. Guibas and F. Zhao, Eds. New York: Springer, Apr. 2003, Proc. 2nd Int. Workshop Inf. Sensor Netw. (IPSN'03), pp. 162–177.

[7] ——, "Power-bandwidth-distortion scaling laws for sensor networks," in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw. (IPSN'04)*, Berkeley, CA, Apr. 2004, pp. 320–329.

[8] V. C. Marcenko and A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Math, USSR-Sb*, vol. 1, pp. 457–483, 1967.

[9] W. Silverstein and D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *J. Multivariate Anal.*, vol. 54, pp. 175–192, 1995.

[10] M. Gastpar and M. Vetterli, "Scaling laws for homogeneous sensor networks," presented at the 41st Annu. Allerton Conf. Commun., Control, Comput., Monticello, IL, Oct. 2003.

[11] T. Berger, Z. Zhang, and H. Viswanathan, "The CEO problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[12] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 43, no. 5, pp. 1549–1559, Sep. 1997.

[13] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1057–1070, May 1998.

[14] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," Bell Labs Tech. Memo., Jun. 1995.

[15] M. Gastpar, "On the role of feedback in large sensor networks," in *Proc. Int. Zurich Seminar*, Zurich, Switzerland, Feb. 2004, pp. 188–191.

[16] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379-423–623-656, 1948.

[17] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.

[18] I. Csiszár and J. Körner, *Information Theory: Coding Theory for Discrete Memoryless Systems*. New York: Academic, 1981.

[19] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–596, Nov.–Dec. 1999.

Michael Gastpar (S'99–A'02–M'04) received the Dipl. El.-Ing. degree from the Swiss Federal Institute of Technology, Zürich (ETHZ), Switzerland, in 1997, the M.S. degree from the University of Illinois at Urbana–Champaign, Urbana, in 1999, and the Doctorat ès Science degree from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 2002, all in electrical engineering. He was a student in engineering and philosophy at the University of Edinburgh, Edinburgh, U.K., and the University of Lausanne, Lausanne, Switzerland.

He was a Summer Researcher in the Mathematics of Communications Department, Bell Laboratories, Lucent Technologies, Murray Hill, NJ. He is now an Assistant Professor at the University of California, Berkeley. His research interests are in network information theory and related coding and signal processing techniques, with applications to sensor networks and neuroscience.

Dr. Gastpar won the 2002 EPFL Best Thesis Award and a National Science Foundation (NSF) CAREER Award in 2004.

Martin Vetterli (S'86–M'86–SM'90–F'95) received the Dipl. El.-Ing. degree from Swiss Federal Institute of Technology, Zürich (ETHZ), Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, CA, in 1982, and the Doctorat ès Science degree from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1986.

He was a Research Assistant at Stanford University and EPFL. He has worked for Siemens and AT&T Bell Laboratories. In 1986, he joined Columbia University, New York, where he was an Associate Professor of Electrical Engineering and Co-Director of the Image and Advanced Television Laboratory. In 1993, he joined the University of California, Berkeley, where he was a Professor in the Department of Electrical Engineering and Computer Sciences until 1997, and holds now an Adjunct Professor position. Since 1995, he has been a Professor of Communication Systems at EPFL, where he Chaired the Communications Systems Division (1996-1997), and heads the Audio-Visual Communications Laboratory. He held visiting positions at ETHZ (1990) and Stanford University (1998). Since 2001, he directs the National Competence Center in research on mobile information and communication systems (www.mics.org). He is the coauthor of *Wavelets and Subband Coding* (Englewood Cliffs, NJ: Prentice-Hall, 1995) with J. Kovacevic. He has published about 100 journal papers on a variety of topics in signal/image processing and communications and holds seven patents. His research interests include wavelets, signal processing for communications, computational complexity, image/video processing and compression and distributed signal processing and communications.

Dr. Vetterli is a member of the Society for Industrial and Applied Mathematics (SIAM) and the Swiss Council on Science and Technology. He received the Best Paper Award of EURASIP in 1984 for his paper on multidimensional subband coding, the Research Prize of the Brown Bovery Corporation, Switzerland, in 1986 for his doctoral dissertation, and the IEEE Signal Processing Society's Senior Award in 1991 and 1996 (for papers with D. LeGall and K. Ramchandran, respectively). He received the Swiss National Latsis Prize in 1996, the SPIE Presidential Award in 1999, and the IEEE Signal Processing Technical Achievement Award in 2001. He was a Plenary Speaker at various conferences (e.g., 1992 IEEE ICASSP). He was the Area Editor for Speech, Image, Video, and Signal Processing of the IEEE TRANSACTIONS ON COMMUNICATIONS. He is also on the Editorial Boards of *Annals of Telecommunications*, *Applied and Computational Harmonic Analysis*, and the *Journal of Fourier Analysis and Applications*.