Power-Bandwidth-Distortion Scaling Laws for Sensor Networks

Michael Gastpar
University of California, Berkeley
Dept. of EECS
Berkeley, CA 94720-1770, USA
gastpar@eecs.berkeley.edu

Martin Vetterli*
School of I&C
Swiss Federal Institute of Technology (EPFL)
CH-1015 Lausanne, Switzerland
martin.vetterli@epfl.ch

ABSTRACT

The goal of a class of sensor networks is to monitor an underlying physical reality at the highest possible fidelity. Sensors acquire noisy measurements and have to communicate them over a power- and possibly bandwidth-constrained interference channel to a set of base stations. The goal of this paper is to analyze, as a function of the number of sensors, the trade-offs between the degrees of freedom of the underlying physical reality, the communication resources (power, temporal and spatial bandwidth), and the resulting distortion at which the physical reality can be estimated by the base stations. The distortion can be expressed as the sum of two fundamentally different terms. The first term reflects the fact that the measurements are noisy. It depends on the number of sensors and on their locations, but it cannot be influenced by the communication resources. The second contribution to the distortion can be controlled by the communication resources, and the key question becomes: What resources are necessary to make it decay at least as fast as the first distortion term, as a function of the number of sensors? This question is answered threefold: First, a lower bound to the power-bandwidth trade-off is derived, showing that at least a constant to linearly increasing total power is required for typical cases (as a function of M). But is this also sufficient? In the second answer, communication strategies are considered where each sensor applies the best possible distributed compression algorithm, followed by capacity-achieving channel codes. For such a *separation* strategy, it is shown for typical cases that the power must increase *exponentially* as a function of the number of sensors, suggesting that the lower bound derived in this paper is far too optimistic. However, in the third answer, it is shown that this is not the case: For some example scenarios, the power requirements of the lower bound are indeed achievable, but *joint* source-channel coding is required. Finally, the problem of sensor synchronization is considered, and it is shown that the scaling laws derived in this paper continue to hold under a Rician fading model.

*Also with Dept. of EECS, UC Berkeley.

Categories and Subject Descriptors

H.1.1 [**Systems and Information Theory**]: Information Theory; E.4 [**Coding and Information Theory**]: Data compaction and compression,Error control codes

General Terms

Theory,Performance

Keywords

sensor networks, OPTA (optimum performance theoretically attainable), joint source-channel coding, separation theorem

1. INTRODUCTION

The class of sensor networks of interest to this study 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 [9], and an extension thereof in [8]. The present paper generalizes this analysis to multiple data sources and multiple base stations. The considered network contains L (discrete-time) sources. The parameter L models both the temporal and spatial bandwidth of the underlying physical process. Much of the paper concerns the case where the sources 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 source 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.

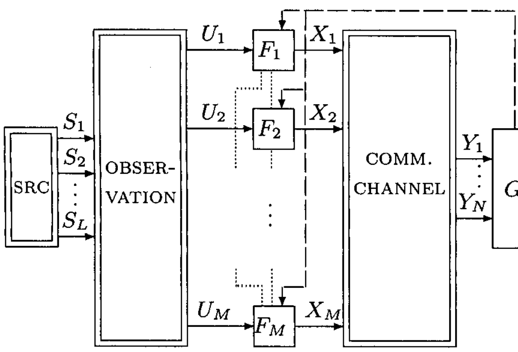

Figure 1: The "monitoring" sensor network topology considered in this paper: L sources are observed in a noisy fashion by M sensors (the boxes labeled F_1 through F_M) that communicate to N base stations over a power- and bandwidth-constrained interference channel. The base stations need to estimate the outputs of the underlying sources.

The remainder of the paper is organized as follows: In Section 2, we define the sensor network topology and model studied in this paper. In particular, in Subsection 2.1, we specify a special case of the general topology, the Gaussian sensor network. Sections 3, 4, 5, and 6 are devoted to this special case. First, in Section 3, we establish a lower bound to the distortion, revealing two fundamentally different terms. The first term reflects the non-ideal observation process, and the second the communication constraints. In Section 4, we discuss in details what communications resources have to be invested to make both distortion terms decay at the same rate. Then, in Section 5, we discuss schemes that achieve the optimum distortion scaling law for particular power-bandwidth trade-offs. Up to this point, we assume that the sensors operate with perfect synchronization. In Section 6, we extend our results to the unsynchronized case. For a model of unsynchronized sensors, we show that the scaling laws obtained earlier continue to hold.

Finally, Section 7 shows how our results can be extended beyond the case of Gaussian statistics, along the lines of *measure-matching* [7].

2. THE SENSOR NETWORK MODEL

The sensor network model studied in this paper is shown in Figure 1. 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 directly be observed. 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 Figure 1, 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 inter-sensor 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 link is again modeled by a conditional distribution. The output of the link is observed by a central data collection unit, whose goal 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

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

To simplify the notation in the rest of the paper, we denote sequences as

$$S^j \stackrel{def}{=} \{S[i]\}_{i=1}^{j}. \qquad (2)$$

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, which we denote by

$$P(u_m|s_1, \ldots, s_L). \qquad (3)$$

Moreover, sensor m may receive information from other sensors as well as from the destination, as illustrated by the dotted lines in Figure 1. For notational purposes, let us summarize this information by V_m^j. Based on this as well as the sensor readings U_m^j, sensor m transmits a signal of length jK (since K channel uses are available for each source sample), given by

$$X_m^{jK} = F_m\left(U_m^j, V_m^j\right), \qquad (4)$$

on the multi-access channel. The transmitted signals satisfy a power, or more generally, a cost constraint of the form

$$E\left[\rho(X_1^{jK}, X_2^{jK}, \ldots, X_M^{jK})\right] \quad \leq \quad \Gamma_j. \qquad (5)$$

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 $\rho^{(1)}(\cdot), \rho^{(2)}(\cdot), \ldots$ and maximum expected costs $\Gamma_j^{(1)}, \Gamma_j^{(2)}, \ldots$, respectively. This is a generalization of the individual power constraints for each sensor.

The final destination uses the output of the multi-access channel to construct estimates

$$\hat{S}^j = (\hat{S}_1^j, \hat{S}_2^j, \ldots, \hat{S}_L^j). \qquad (6)$$

The task is to make the estimate $\hat{S}^n$ as close to S^n as possible, in the sense of an appropriately chosen distortion measure $d(s^n, \hat{s}^n)$. 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:

$$\Delta_j \quad = \quad E\left[d\left(S^j, \hat{S}^j\right)\right]. \qquad (7)$$

[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.

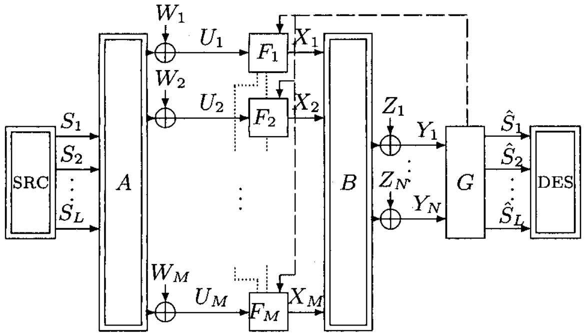

Figure 2: The considered Gaussian sensor network: A and B are appropriate matrices.

The relevant figure of merit is therefore the trade-off between the *cost* Γ_j of the transmission (Equation (5)), and the achieved *distortion level* Δ_j (Equation (7)). The problem studied in this paper is that of finding the optimal trade-offs (Γ_j, Δ_j) in the limit as $j \to \infty$, i.e., the optimum trade-offs *irrespective of coding delay and complexity.* We denote the respective limits by Γ and Δ, respectively.

2.1 Gaussian Sensor Networks

Of particular interest to the arguments of this paper is a special case of the sensor network of Figure 1, namely when all involved statistics are Gaussian. The resulting scenario is illustrated in Figure 2. In particular, there are L physical sources (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.

Source Bandwidth L and Observation Process

The source is characterized by L independent and identically distributed (iid)[2] circularly complex Gaussian random variables with mean zero and variance σ_S^2. The observation process is modeled as

$$U_m[i] \quad = \quad W_m[i] + \sum_{l=1}^{L} a_{m,l} S_l[i], \tag{8}$$

for $m = 1, 2, \ldots, M$, where $W_m[i]$ is iid circularly complex Gaussian with mean zero and variance σ_W^2.[3] The coefficients $a_{m,l}$ model how strongly source S_l influences the measurement made by sensor m. We collect the coefficients $a_{m,l}$ into the matrix $A \in \mathbb{C}^{M\times L}$ defined as

$$A \quad = \quad \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,L} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,L} \\ a_{3,1} & a_{3,2} & \cdots & a_{3,L} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M,1} & a_{M,2} & \cdots & a_{M,L} \end{pmatrix}. \tag{9}$$

[2]Assuming iid sources is without loss of generality since the matrix A in Eqn. (9) below can be chosen arbitrarily.

[3]Note that the assumption that for any $m = 1, 2, \ldots, M$, the measurement noise $W_m[i]$ has the same variance σ_W^2 is without loss of generality since the coefficients $a_{m,l}$ can be chosen arbitrarily.

The matrix A has $\min\{M, L\}$ singular values. Our interest is in the case where $L \leq M$, and we denote the L singular values as

$$\alpha_1, \alpha_2, \ldots, \alpha_L. \tag{10}$$

In this paper, we consider the case where the matrix A has full rank. Hence, L models the product of the spatial and temporal bandwidth of the underlying source.

Spatial Bandwidth N of the Communication Channel

The communication channel is the standard additive white Gaussian multiple access channel, modeled as

$$Y_n \quad = \quad Z_n + \sum_{m=1}^{M} b_{n,m} X_m, \tag{11}$$

where Z_n is iid 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

$$B \quad = \quad \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} & \cdots & b_{1,M} \\ b_{2,1} & b_{2,2} & b_{2,3} & \cdots & b_{2,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{N,1} & b_{N,2} & b_{N,3} & \cdots & b_{N,M} \end{pmatrix}. \tag{12}$$

The matrix B has $\min\{N, M\}$ singular values that we denote by $\beta_1, \beta_2, \ldots$. Our interest is in the case where $N \leq M$, and in this paper, we consider the case where the matrix B has full rank. Hence, N models the spatial bandwidth of the communication channel.

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.

Power on the Communication Channel

The power of sensor m is determined by

$$P_m \quad = \quad \lim_{j\to\infty} \frac{1}{j} \sum_{i=1}^{j} E\left[|X_m[i]|^2\right]. \tag{13}$$

The power on the communication channel is constrained as follows:

$$\sum_{m=1}^{M} P_m \quad \leq \quad \frac{P_{tot}}{K}, \tag{14}$$

i.e., P_{tot} denotes the total power available per source output $(S_1, S_2, \ldots, S_L)$.

Target Distortion

The goal of the sensor network is to minimize the mean-squared error,

$$D \quad = \quad \frac{1}{L} \sum_{l=1}^{L} \lim_{j\to\infty} \frac{1}{j} \sum_{i=1}^{j} E\left[\left|S_l[i] - \hat{S}_l[i]\right|^2\right]. \tag{15}$$

2.2 Simplifying Assumption

For the remainder of this paper, we assume that the spread of the singular values of the matrices A and B is small. The key reason for this is to keep the notation simple: If the singular values have a large spread, our solutions must be modified to include an argument sometimes referred to as "inverse water-filling" (see e.g. [4, p.349]). The full solution will be presented in [10].

2.3 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

$$f_1(M) \quad \sim \quad f_2(M), \tag{16}$$

which simply means that

$$\lim_{M\to\infty} f_1(M)/f_2(M) = c, \tag{17}$$

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

$$f_1(M) \quad \overset{=}{\sim} \quad f_2(M). \tag{18}$$

3. LOWER BOUND TO THE OPTIMAL DISTORTION SCALING LAWS

3.1 Lower Bound

In this section, we derive a lower bound to the distortion that can be achieved in the Gaussian sensor network defined in Subsection 2.1. Our lower bound is based on the idealization that the sensors can collaborate perfectly. While this performance should not be expected to be achievable in general, we show in the remainder of the paper that its *scaling behavior* is indeed achievable for a relevant class of sensor networks.

THEOREM 1. *The distortion that can be achieved in the Gaussian sensor network defined in Subsection 2.1 (subject to the simplifying assumption of Subsection 2.2) cannot be smaller than*

$$D_{lower}(M, P_{tot}, L, K, N) = \frac{1}{L}\sum_{l=1}^{L} \frac{\sigma_S^2\sigma_W^2}{\alpha_l^2\sigma_S^2+\sigma_W^2} + \left(\frac{1}{\mu + \frac{P_{tot}}{KN\sigma_Z^2}\sqrt[N]{\prod_{n=1}^{N}\beta_n^2}}\right)^{\frac{KN}{L}} \sqrt[L]{\prod_{l=1}^{L}\frac{\alpha_l^2\sigma_S^4}{\alpha_l^2\sigma_S^2+\sigma_W^2}}, \tag{19}$$

where 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 noise in the multi-access channel, P_{tot} is the total sensor transmit power for the K channel uses, N is the number of destination terminals, and

$$\mu = \left(\frac{1}{N}\sum_{n=1}^{N}\frac{1}{\beta_n^2}\right)\sqrt[N]{\prod_{n=1}^{N}\beta_n^2}. \tag{20}$$

Remark. 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.

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

Discussion

Theorem 1 shows that the distortion is the sum of two characteristic contributions. The first term reflects the non-ideal observation process, and can be influenced by altering the matrix A (e.g., by increasing the number of sensors M, or by changing their locations), and by decreasing the variance of the measurement noise. The second term is due to the communication constraints, i.e., the total sensor power, and the spatial and temporal bandwidth of the channel.

The key question then becomes the following: For fixed matrix (sequence) $A^{(M)}$ and observation noise variance σ_W^2, how much, in terms of communication resources, must be invested to make that second term decay (at least) as fast as the first term, as a function of the number of sensors M? Theorem 1 gives a lower bound to this trade-off. This is discussed in more detail below in Section 4.

3.2 Example: The Circulant Sensor Network

Consider L sources located uniformly spaced on the unit circle. Suppose M sensors are uniformly spaced on a concentric circle of larger radius, and N receiver antennas are spaced regularly on a concentric circle of even larger radius. This geometry is illustrated in Figure 3. For the sake of this discussion, we assume that both N and L divide M, and that M is even.

Under this geometry, the observation coefficients have a symmetry property. For the first source, S_1, we can write (for some function $a(m)$)

$$a_{m,1} = a(m), \text{ for } m = 1, 2, \ldots, M, \tag{21}$$

and the function $a(m)$ satisfies $a(m) = a(M - m + 2)$, for $m = 2, 3, \ldots, M/2$. For the remaining sources $S_l, l = 2, 3, \ldots, L$, the symmetry of the overall situation suggests that

$$a_{m,l} = a\left(\left(m - (l-1)\frac{M}{L}\right) \bmod M\right), \tag{22}$$

for $m = 1, 2, \ldots, M$, where, for ease of notation, we define $a(0) = a(M)$. That is, the coefficients $a_{m,l}$ are the same as the coefficients $a_{m,1}$ up to a cyclic shift of $(l-1)M/L$ positions. In that case, the matrix $A^H A$ is circulant, and its eigenvalues α_l^2 are easily found to be

$$\alpha_l^2 = L\sum_{n=1}^{M/L}\left|\mathcal{F}\{a\}|_{(n-1)L+l}\right|^2, \tag{23}$$

for $l = 1, 2, \ldots, L$, where the discrete Fourier transform is defined in the unitary form as

$$\mathcal{F}\{a\}|_r = \frac{1}{\sqrt{M}}\sum_{m=1}^{M} a(m)e^{-j\frac{2\pi}{M}(r-1)(m-1)}, \tag{24}$$

for $r = 1, 2, \ldots, M$.

To make matters more concrete, consider the case where

$$a(m) = \exp\left(-\frac{\min\{m-1, M-m+1\}}{\rho M}\right). \tag{25}$$

For this choice of $a(m)$, and the case $L = 4$ and $M = 8$ (as

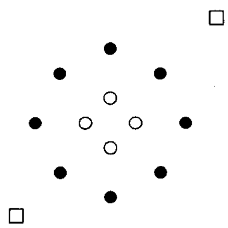

Figure 3: The circulant sensor network with $L = 4$ sources (the empty circles), $M = 8$ sensors (the disks), and $N = 2$ receiving terminals (the squares).

in Figure 3), the matrix A is found to be

$$A = \begin{pmatrix} 1 & e^{-\frac{2}{8\rho}} & e^{-\frac{4}{8\rho}} & e^{-\frac{2}{8\rho}} \\ e^{-\frac{1}{8\rho}} & e^{-\frac{1}{8\rho}} & e^{-\frac{3}{8\rho}} & e^{-\frac{3}{8\rho}} \\ e^{-\frac{2}{8\rho}} & 1 & e^{-\frac{2}{8\rho}} & e^{-\frac{4}{8\rho}} \\ e^{-\frac{3}{8\rho}} & e^{-\frac{1}{8\rho}} & e^{-\frac{1}{8\rho}} & e^{-\frac{3}{8\rho}} \\ e^{-\frac{4}{8\rho}} & e^{-\frac{2}{8\rho}} & 1 & e^{-\frac{2}{8\rho}} \\ e^{-\frac{3}{8\rho}} & e^{-\frac{3}{8\rho}} & e^{-\frac{1}{8\rho}} & e^{-\frac{1}{8\rho}} \\ e^{-\frac{2}{8\rho}} & e^{-\frac{4}{8\rho}} & e^{-\frac{2}{8\rho}} & 1 \\ e^{-\frac{1}{8\rho}} & e^{-\frac{3}{8\rho}} & e^{-\frac{3}{8\rho}} & e^{-\frac{1}{8\rho}} \end{pmatrix}. \quad (26)$$

For this choice of the function $a(m)$, the Fourier transform is known explicitly. In particular, with the help of [3], we find (for even values of M)

$$\mathcal{F}\{a\}|_r = \frac{(1 - e^{-2/(M\rho)})(1 - e^{-1/(2\rho)} \cos(\pi(r-1)))}{\sqrt{M}(1 - 2e^{-1/(M\rho)} \cos(2\pi(r-1)/M) + e^{-2/(M\rho)})}, \quad (27)$$

for $r = 1, 2, \ldots, M$, and the (squares of the) singular values are easily found by evaluating (23).

As illustrated in Figure 3, the geometry is similar on the communications side. For the first destination terminal, Y_1, we can write (for some function $b(m)$)

$$b_{1,m} = b(m), \text{ for } m = 1, 2, \ldots, M, \quad (28)$$

and the function $b(m)$ satisfies $b(m) = b(M - m + 2)$, for $m = 2, 3, \ldots, M/2$. For the remaining terminals $Y_n, n = 2, 3, \ldots, N$, the symmetry of the overall situation suggests

$$b_{n,m} = b\left(\left(m - (n-1)\frac{M}{N}\right) \bmod M\right), \quad (29)$$

for $m = 1, 2, \ldots, M$, where, for ease of notation, we define $b(0) = b(M)$. That is, the coefficients $b_{n,m}$ are the same as $b_{1,m}$ up to a cyclic shift of $(n-1)M/N$ positions. Then, the matrix BB^H is circulant, and its eigenvalues β_n^2 are

$$\beta_n^2 = N \sum_{j=1}^{M/N} \left| \mathcal{F}\{b\}|_{(j-1)N+n} \right|^2. \quad (30)$$

To make matters more concrete, consider the case where

$$b(m) = \exp\left(-\frac{\min\{m-1, M-m+1\}}{\xi M}\right), \quad (31)$$

whose spectrum is given by (27) with ρ replaced by ξ.

Finite number of sources L and base stations N

Let us now investigate the scenario where L and N are kept fixed while M becomes large. Approximating (27) by a Taylor series, it is easy to show that for small r and large M, $\mathcal{F}\{a\}|_r$ increases like $\sqrt{M}$. In particular, since L is kept constant, this is true for $r = 1, 2, \ldots, L$. Using this in (23), we find that for each $l = 1, 2, \ldots, L$, the first term in the sum on the right hand side of (23) increases linearly in M. But since all the terms in the sum are non-negative, this implies that the squares of the singular values α_l^2 increases at least linearly in M, for $l = 1, 2, \ldots, L$.

Could the squares of the singular values α_l^2 increase *faster* than linearly in M? It is easily seen that this is not possible. In particular,

$$\sum_{m=1}^{M} |a(m)|^2 = 2\frac{1 - e^{-1/\rho}}{1 - e^{-2/(M\rho)}} - 1 - e^{-1/\rho} \quad (32)$$

which increases linearly in M. By Parseval's identity, this must be the same as

$$\sum_{m=1}^{M} \left| \mathcal{F}\{a\}|_r \right|^2, \quad (33)$$

implying via (23) that the sum of the squares of the singular values α_l^2 cannot increase faster than linearly in M, and hence, none of the squares of the singular values can.

Combining this, we can express the squares of the singular values as

$$\alpha_l^2 = M\tilde{\alpha}_l^2, \quad (34)$$

for $l = 1, 2, \ldots, L$, where $\tilde{\alpha}_l^2$ converges to a constant as M becomes large. Using the same argument to analyze (30) yields

$$\beta_n^2 = M\tilde{\beta}_n^2. \quad (35)$$

Finally, applying the expressions found in (34) and (35) to Theorem 1, we find a lower bound to the distortion scaling law as

$$D_{lower}(M, P_{tot}, L, K, N) = \frac{c_1}{M} + c_2 \left(\frac{1}{c_3 + \frac{c_4 M P_{tot}}{K N \sigma_Z^2}} \right)^{KN/L}, \quad (36)$$

where c_1, c_2, c_3, and c_4 are defined as

$$c_1 = \frac{1}{L} \sum_{l=1}^{L} \frac{\sigma_S^2 \sigma_W^2}{\tilde{\alpha}_l \sigma_S^2 + \sigma_W^2/M} \quad (37)$$

$$c_2 = \sqrt[L]{\prod_{l=1}^{L} \frac{\tilde{\alpha}_l \sigma_S^4}{\tilde{\alpha}_l \sigma_S^2 + \sigma_W^2/M}} \quad (38)$$

$$c_3 = \left(\frac{1}{N} \sum_{n=1}^{N} \frac{1}{\tilde{\beta}_n^2} \right) \sqrt[N]{\prod_{n=1}^{N} \tilde{\beta}_n^2} \quad (39)$$

$$c_4 = \sqrt[N]{\prod_{n=1}^{N} \tilde{\beta}_n^2}. \quad (40)$$

Note that strictly speaking, for small M, the values of c_1, c_2, c_3, and c_4 are also functions of M. However, they rapidly converge to a constant value as M increases.

Increasing number of sources L and base stations N

When L increases with M, the analysis of (27) becomes more involved. Clearly, when L increases very slowly with M, the analysis of the previous paragraph still applies. The case where L increases significantly with M is currently under investigation.

3.3 Extensions of the circulant network example

The above example can be readily extended to a "multi-circulant' network where there are multiple circles of sensors around the sources, both outside and inside of the source circle. Moreover, the circulant example can also be extended to a *Toeplitz sensor network*, and essentially the same asymptotic analysis applies, along the lines of [11].

4. THE NECESSARY POWER AND BANDWIDTH

The key insight of Section 3 is that there is a term in the distortion expression that does not depend on the communication capabilities, namely, the first summand in (19). 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 $A^{(M)}$, 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 example, for the considered circulant sensor network, the behavior is like $1/M$.

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 (19) decays at the same rate as the first? This question can be answered based on Theorem 1. We summarize this insight in the following corollary:

COROLLARY 2. *Suppose that the slowest-growing of the singular values α_l, $l = 1, 2, \ldots, L$, of the observation matrix sequence $A^{(M)}$ grows like $M^{\delta/2}$. Then, the optimal distortion scaling law is $1/M^\delta$, and the power, P_{tot}, and the bandwidth, K and N, required to sustain this distortion scaling law must satisfy*

$$\left(\frac{1}{\mu + \frac{P_{tot}}{KN\sigma_Z^2}\sqrt[N]{\prod_{n=1}^{N}\beta_n^2}}\right)^{KN/L} \sim \frac{1}{M^\delta} \qquad (41)$$

Example (Circulant sensor network, cont'd). For the circulant sensor network analyzed in Section 3.2, we find the necessary total power to sustain the optimal distortion scaling law of $1/M$ must scale at least as follows:

$$P_{tot} \sim KNM^{\frac{L}{KN}-1}. \qquad (42)$$

A key conclusion from this formula is that in order to prevent the power from increasing exponentially with the number of sensors, the total channel bandwidth KN must be at least equal to the number of underlying sources L.

5. CODING SCHEMES

Sections 3 and 4 were concerned with lower bounds only, but leave open whether the corresponding performance is actually achievable. In this section, we show how to achieve this lower bound, at least in some cases. More precisely, we discuss schemes that achieve optimal power-bandwidth-distortion tradeoffs. The analysis in this section is still restricted to the Gaussian sensor networks as defined in Section 2.1.

First, in Subsection 5.1, 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 the sensors each want to communicate a 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 that lower bound is too optimistic.

In Subsection 5.2, we show that this is not the case: we present a class of networks for which the scaling behavior as predicted by Theorem 1 can indeed be achieved. In fact, in previous work [9], we established the achievability of the lower bound of Theorem 1 for the simple case of a single source and destination (i.e., $L = K = N = 1$). This was established by a simple coding scheme where the sensors only scale their sensor readings, and exploit the channel structure to do the rest. In [8], we showed how these arguments can be extended to the case of multiple sources, but only one base station (i.e., $L = K > 1$, but $N = 1$).

The present paper addresses the case where $L = N > 1$, but $K = 1$. Using this result in combination with the result of [8], we outline how to address the more general case where $L = KN$.

5.1 Separate Source and Channel Coding

It is well known that separate source and channel coding does not lead to optimal performance in general networks. In extension of this, it has been shown in [9, 8] that it is not only suboptimal, it may even lead to an entirely different scaling behavior. The present paper extends this analysis to the more complex network topologies considered here. 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 Figure 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. Therefore, we instead give a *lower bound* to the corresponding distortion that can be achieved with a scheme that separates source from channel coding, combining the rate-distortion results of [2, 12, 15] for the so-called CEO problem with 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 [14]. To compare the resulting expression to the lower bound established in Theorem 1, consider the scenario of the circulant example presented in Section 3.2.

Example (Circulant sensor network, cont'd). For the circulant sensor network analyzed in Section 3.2, and suppose for simplicity that the columns of the matrix A are orthonormal with respect to each other, and that the rows of the matrix B are also orthonormal with respect to each other.

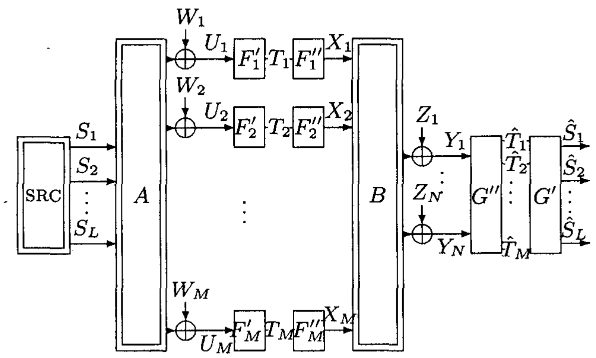

Figure 4: Schematic rendering of separate source and channel coding for the single-source Gaussian sensor network. The bit sequence that sensor m needs to communicate to the base stations is denoted by T_m. F'_m and F''_m denote source and channel encoding at sensor m, respectively, and G'' and G' denote channel and source decoding at the base stations, respectively.

Then, the desired lower bound can be expressed as

$$D(M, P_{tot}, L, K, N) = \frac{\sigma_S^2\sigma_W^2}{M\sigma_S^2+\sigma_W^2} + \frac{\sigma_S^4}{\sigma_S^2+\sigma_W^2/M}\frac{1}{1+\frac{KN\sigma_S^2}{L\sigma_W^2}\log_2\left(1+\frac{MP_{tot}}{KN\sigma_Z^2}\right)} \quad (43)$$

In order to compare this to the lower bound established in Theorem 1, consider for example the total power P_{tot}. In order to obtain an overall scaling behavior of $1/M$, the lower bound requires at least the total power given in Equation (42). However, according to (43), a separation-based strategy requires a total power of at least

$$P_{tot} \sim e^M, \quad (44)$$

that is, the total power must grow *exponentially* as more and more sensors are added to the network.

5.2 Joint Source and 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 (43), and the lower bound of (36).[4] In this subsection, we show that for certain cases of interest, our lower bound is indeed achievable.

Single base station

The first special case we consider is the scenario when $N = 1$. 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 δ_m, at the expense of extra notation.

[4]The circulant example clearly illustrated this gap, simply by comparing the required powers, Equations (42) and (Eq-sep-P).

THEOREM 3. *For the Gaussian sensor network defined in Section 2.1 with $L = K$ and $N = 1$, and with $b_m = 1$, for $m = 1, 2, \ldots, M$, the following distortion can be achieved:*

$$D_1(M, P_{tot}(M)) = \frac{1}{L}\sum_{l=1}^{L}\frac{\sigma_S^2\sigma_W^2}{\alpha_l^2\sigma_S^2+\sigma_W^2} + \frac{1}{P_{tot}(M)}\sum_{l=1}^{L}\frac{\alpha_l^2 u_{[l]}\sigma_Z^2}{(\alpha_l^2+\sigma_W^2/\sigma_S^2)^2} \quad (45)$$

where $\alpha_1, \alpha_2, \ldots, \alpha_L$ denote the singular values of A, 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 noise in the multi-access channel, P_{tot} is the total sensor transmit power for the K channel uses, and $u_{[l]}$ denotes the l-th largest of the sensed powers $u_m = \sum_{l=1}^{L}|\alpha_{m,l}|^2\sigma_S^2 + \sigma_W^2$.

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 $U_1, U_2, \ldots, U_M$. Clearly, Equation (45) describes the same distortion scaling law as Theorem 1 whenever the singular values of the matrix sequence $A^{(M)}$ increase at the same rate as the singular values of the matrix sequence $B^{(M)}$, as was the case for example in the circulant scenario of Section 3.2.

Multiple Base Stations

When there are many ($N > 1$) base stations, a more efficient strategy can be implemented: each channel use enables to estimate up to N sources, 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. *Denote the singular value decomposition of the matrix $A \in \mathbb{C}^{M\times L}$ by $A = U_aD_aV_a^H$, and of the matrix $B \in \mathbb{C}^{N\times M}$ by $B = U_bD_bV_b^H$. 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*

$$V_b^H\Lambda_F U_a = Q\,diag(\gamma_1, \gamma_2, \ldots, \gamma_L)Q^H. \quad (46)$$

for 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.

Examples. The following are examples of matched matrices according to Definition 1.

1. When A and B are simply vectors ($L = N = 1$) with enough non-zero entries, they are matched. This is a special case of the analysis of Subsection 5.2.
2. Circulant case: When both A^HA 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:*

$$D_1(M, P_{tot}, L = N, K = 1)$$
$$= \frac{1}{L}\sum_{l=1}^{L} \frac{\sigma_S^2\sigma_W^2}{|\gamma_l|^2\alpha_l^2\sigma_S^2 + \sigma_W^2}$$
$$+\frac{\nu LM}{P_{tot}/\sigma_Z^2} \cdot \frac{1}{L}\sum_{l=1}^{L} \frac{|\gamma_l|^2\alpha_l^2\sigma_S^2}{\beta_l^2(|\gamma_l|^2\alpha_l^2\sigma_S^2 + \sigma_W^2)^2}, \quad (47)$$

where $\nu = \frac{1}{LM}\left(\|\Lambda_F A\|_F^2\sigma_S^2 + \|\Lambda_F\|_F^2\sigma_W^2\right)$ *and* $\|\cdot\|_F$ *denotes the Frobenius norm.*

Remark. If the matrices A and B are almost matched, i.e., there exists a Λ_F such that (46) is close to diagonal, the achieved performance is close to (47), but the precise formula involves the eigenvectors of the matrices A and B, and is therefore considerably less compact.

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

THEOREM 5. *Suppose that the following conditions are satisfied:*
1. the matrices $A^{(M)}$ *and* $B^{(M)}$ *are matched for every* M,
2. the corresponding values of $|\gamma_l|$, $l = 1, 2, \ldots, L$, *are strictly larger than zero,*
3. the expression $\frac{1}{LM}\left(\|\Lambda_F A\|_F^2\sigma_S^2 + \|\Lambda_F\|_F^2\sigma_W^2\right)$ *is bounded,*
4. the slowest-growing of the singular values $\alpha_l^{(M)}$ *increases at least like* $\sqrt{M}$, *and*
5. $\beta_n^{(M)} = \tilde{\beta}_n f_\beta(M)$, *for* $n = 1, 2, \ldots, N$. *Then, the optimum distortion scaling law is given by*

$$D(M, P_{tot}, L = N, K = 1)$$
$$\sim \frac{1}{L}\sum_{l=1}^{L} \frac{\sigma_S^2\sigma_W^2}{\alpha_l^2\sigma_S^2 + \sigma_W^2} + \frac{L}{P_{tot}f_\beta^2(M)}. \quad (48)$$

To illustrate this theorem, consider again the following example:
Example (Circulant sensor network, cont'd). For the circulant sensor network analyzed in Section 3.2, for the case where $L = N$ and $K = 1$, the conditions of Theorem 5 are satisfied, and we find that the optimum distortion scaling law is

$$D(M, P_{tot}, L = N, K = 1) \sim \frac{c_1}{M} + \frac{c_2}{c_4} \cdot \frac{L\sigma_Z^2}{MP_{tot}}, \quad (49)$$

where c_1, c_2, and c_4 are defined in (37)-(40). Hence, the necessary and sufficient total power to sustain the optimal distortion scaling law of $1/M$ scales as

$$P_{tot} \sim L, \quad (50)$$

i.e., the total power scales linearly in the number of underlying sources. This is the minimum power, as a comparison with (42) reveals (recall that $L = N$).

Discussion

This theorem shows that at least in some paradigmatic cases, the performance lower bound derived in Sections 3 and 4 can be achieved. However, to achieve this performance, *joint* source-channel codes are required. A modular, separate design of the compression and communication stages leads to substantial sub-optimality.

Moreover, since the lower bound of Sections 3 and 4 includes both feedback and arbitrary collaboration between the sensors, the above theorem also shows that for those cases, neither feedback nor collaboration has a scaling-law relevant role: The optimum power-distortion scaling law can already be achieved by a simple joint source-channel coding scheme.

5.3 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, and hence, source and channel bandwidth are again equal. Then, the resulting situation may still be favorably matched. The special case $N = 1$ was analyzed in [8], and a more general case can be addressed by combining the arguments outlined in Sections 5.2 and 5.2.

When $L = N$, but the matrices A and B are not matched, 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 considered in [6].

6. SYNCHRONIZATION

The lower bounds derived in Sections 3 and 4 apply whether or not perfect synchronization is available to the sensors. However, the "uncoded forwarding" coding scheme presented in [9, 8] (and extended to more general cases in Section 5 above) 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 $m = 1, 2, \ldots, M$, iid 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 wide-band) 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 non-zero 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

$$Y = Z + \sum_{m=1}^{M} b_m X_m, \quad (51)$$

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:

$$\begin{aligned} D_{lower}(M, P_{tot}) &= \frac{\sigma_S^2 \sigma_W^2}{\sigma_W^2 + \sigma_S^2 \sum_{m=1}^M |a_m|^2} \\ &+ \frac{1}{1 + \frac{P_{tot}}{\sigma_Z^2} \sum_{m=1}^M E\left[|b_m|^2\right]} \frac{\sigma_S^4 \sum_{m=1}^M |a_m|^2}{\sigma_W^2 + \sigma_S^2 \sum_{m=1}^M |a_m|^2}. \end{aligned} \quad (52)$$

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 non-zero mean. Then, the distortion achieved by a simple forwarding scheme is found to be at most

$$\begin{aligned} D_1(M, P_{tot}) &= \frac{\sigma_W^2}{\left(\sum_{m=1}^M |a_m|^2\right)^2} \sum_{m=1}^M \frac{|a_m|^2 E\left[|b_m|^2\right]}{|E\left[b_m\right]|^2} \\ &+ \frac{\sigma_S^2}{\left(\sum_{m=1}^M |a_m|^2\right)^2} \sum_{m=1}^M \frac{|a_m|^4 Var(b_m)}{|E\left[b_m\right]|^2} \\ &+ \frac{\sigma_Z^2}{P_{tot}\left(\sum_{m=1}^M |a_m|^2\right)^2} \sum_{m=1}^M \frac{|a_m|^2(|a_m|^2\sigma_S^2 + \sigma_W^2)}{|E\left[b_m\right]|^2} \end{aligned} \quad (53)$$

Clearly, this distortion does not generally coincide with the lower bound (52). However, in interesting cases, (52) and (53) describe the same scaling behavior. One of these cases is described by the following result:

THEOREM 6. *Suppose that* $0 < |a_m| < a_{max}$, *for* $m = 1, 2, \ldots, M$. *If at least a fraction* ϵM *of the* b_m, $m = 1, 2, \ldots, M$, *have* $|E[b_m]| > \delta$, *and the total power* P_{tot} *is a non-decreasing function of* M, *then the optimum distortion scaling law is*

$$D(M, P_{tot}) \sim \frac{d_1}{M} + \frac{d_2}{MP_{tot}}, \quad (54)$$

where d_1 *and* d_2 *are constants.*

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

Another distribution of the coefficients b_m for which a simple strategy achieves scaling-law optimal performance is when enough of the coefficients are sufficiently dependent.

7. BEYOND GAUSSIAN STATISTICS

Sections 3-6 were concerned exclusively with the Gaussian sensor networks defined in Section 2.1. Here, we briefly outline extensions.

Non-Gaussian Sources

Reconsider the Gaussian sensor network of Section 2.1, but suppose that the underlying sources have an arbitrary distribution. It is intuitive that under mild conditions on that distribution, a lower bound of the type of Theorem 1 holds. The respective conditions will be explored in [10].

Sufficient Condition for Optimal Performance

For general sensor networks of the topology defined in Section 2, the situation is more involved. A sufficient characterization of optimal codes can be given along the lines of [7] as follows.

THEOREM 7. *In the sensor network problem defined in Section 2, if the coding functions* $F_1, F_2, \ldots, F_M, G$ *are chosen such that*

$$\rho(x_1, x_2, \ldots, x_M) = c_1 D(p_{Y|x_1,x_2,\ldots,x_M} || p_Y) + \rho_0 \quad (55)$$

$$E_{S|u_1,u_2,\ldots,u_M}\left[d(S, \hat{s})\right] = -c_2 \log_2 P(u_1, u_2, \ldots, u_M|\hat{s}) \quad (56)$$

$$I(U_1U_2\ldots U_M; \hat{S}) = I(X_1X_2\ldots X_M; Y), \quad (57)$$

then they perform optimally.

This will be discussed in more detail and from a scaling-law perspective in [10].

8. CONCLUSIONS

This paper analyzes the power-bandwidth-distortion tradeoffs in a certain class of sensor networks where the sensors observe noisy combinations of the underlying data, and a central data collector would like to estimate the underlying data at the highest possible fidelity. The first step of the analysis consists in showing that the distortion generally is the sum of two terms. The first term is due to the fact that the sensor observations are noisy. It is only the second term that one can control with the power and bandwidth of the communication link. The second step of our analysis is then to discuss the power-bandwidth pairs that are needed in order to make the second term decrease at the same rate (as a function of M) as the first. This provides general lower bounds on the power-bandwidth-distortion tradeoffs. The third part of the paper shows scenarios for which the lower bounds are achievable, including a case where the sensors are not synchronized. Extensions of this work include the situation where the underlying number of sources L increases rapidly, more general models of unsynchronized sensors, the analysis of the situation when the involved statistics are not Gaussian, and strategies involving feedback and collaboration between the sensors [6, 10].

APPENDIX

A. OUTLINE OF PROOFS

PROOF (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 [13, Thm. 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 [1, p. 78] (see also [5, p. 136]). The only modification is to replace the standard rate-distortion function by the "remote" rate-distortion function. The minimum distortion D_{min} that can be achieved by arbitrarily collaborating sensors, for a given total power P_{tot}, can be characterized as

$$D_{min} = D_{remote}(C(P_{tot})), \quad (58)$$

where $D_{remote}(R)$ is the remote distortion-rate function, given by

$$D_{remote}(R) = \min \frac{1}{L}\sum_{l=1}^{L} E|S_l - \hat{S}_l|^2, \qquad (59)$$

where the minimization is over all distributions

$$p(\hat{s}_1, \ldots, \hat{s}_L | u_1, \ldots, u_M) \qquad (60)$$

that satisfy $I(U_1U_2 \ldots U_M; \hat{S}_1\hat{S}_2 \ldots \hat{S}_L) \leq R$. Under the assumption that the spread of the singular values α_l of the matrix A is small, this evaluates to

$$D_{remote}(R) = \frac{1}{L}\sum_{l=1}^{L} \frac{\sigma_S^2 \sigma_W^2}{\alpha_l^2 \sigma_S^2 + \sigma_W^2} + \sqrt[L]{\prod_{l=1}^{L} \frac{\alpha_l^2 \sigma_S^4}{\alpha_l^2 \sigma_S^2 + \sigma_W^2}}\, 2^{-\frac{R}{L}}. \qquad (61)$$

The capacity $C(P_{tot})$ needed to complete the proof is simply the capacity of the Gaussian vector channel characterized by the matrix B. Assuming a small spread of the singular values β_n, for $n = 1, 2, \ldots, N$, 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

$$C(P_{tot}) = KN \log_2 \left(\left(\frac{1}{N}\sum_{n=1}^{N} \frac{1}{\beta_n^2} \right) \sqrt[N]{\prod_{n=1}^{N} \beta_n^2} + \frac{P_{tot}}{KN\sigma_Z^2} \sqrt[N]{\prod_{n=1}^{N} \beta_n^2} \right), \qquad (62)$$

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 (61) yields the claimed bound. □

PROOF (THEOREM 4). By assumption, since the matrices A and B are matched, there exists a diagonal matrix Λ_F that satisfies (46). 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 chose 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

$$D = \sigma_S^2 \left(1 - \frac{1}{L}\sum_{l=1}^{L} \frac{c^2|\gamma_l|^2 \alpha_l^2 \beta_l^2}{c^2|\gamma_l|^2 \beta_l^2 (\alpha_l^2 + \sigma_W^2/\sigma_S^2) + \sigma_Z^2/\sigma_S^2} \right). \qquad (63)$$

To determine c, we calculate the total power,

$$\sum_{m=1}^{M} E|X_m|^2 = \sum_{m=1}^{M} c^2 |\lambda_m|^2 E|U_m|^2 = \sum_{m=1}^{M} c^2 |\lambda_m|^2 \left(\sigma_W^2 + \sum_{l=1}^{L} |a_{m,l}|^2 \right). \qquad (64)$$

Hence, if c is such that (64) is at most P_{tot}, the power constraint is satisfied. Introducing this in (63) and simple elementary manipulation yields the claimed bound. □

B. REFERENCES

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

[2] T. Berger, Z. Zhang, and H. Viswanathan. The CEO problem. *IEEE Transactions on Information Theory,* IT–42:887–902, May 1996.

[3] W. L. Briggs and V. E. Henson. *The DFT: An owner's manual for the Discrete Fourier Transform.* SIAM, Philadelphia, 1995.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

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

[6] M. Gastpar. On the role of feedback in large sensor networks. In *Proc International Zurich Seminar,* Zurich, Switzerland, February 2004. Invited.

[7] M. Gastpar, B. Rimoldi, and M. Vetterli. To code, or not to code: Lossy source-channel communication revisited. *IEEE Transactions on Information Theory,* 49(5):1147–1158, May 2003.

[8] M. Gastpar and M. Vetterli. Scaling laws for homogeneous sensor networks. In *Proc 2003 Allerton Conference,* October 2003. Invited paper.

[9] M. Gastpar and M. Vetterli. Source-channel communication in sensor networks. In Leonidas J. Guibas and Feng Zhao, editors, *2nd International Workshop on Information Processing in Sensor Networks (IPSN'03),* pages 162–177. Lecture Notes in Computer Science, Springer, New York, NY, April 2003.

[10] M. Gastpar and M. Vetterli. Scaling laws for sensor networks. *to be submitted to IEEE Transactions on Information Theory,* 2004.

[11] R. M. Gray. On the asymptotic eigenvalue distribution of Toeplitz matrices. *IEEE Transactions on Information Theory,* IT–18:725–730, November 1972.

[12] Y. Oohama. The rate-distortion function for the quadratic Gaussian CEO problem. *IEEE Transactions on Information Theory,* IT–44(3):1057–1070, May 1998.

[13] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal,* 27:379–423, 623–656, 1948.

[14] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *Bell Labs Technical Memorandum,* June 1995. Also published in *European Transactions on Telecommunications,* 10(6):585-596, Nov.-Dec. 1999.

[15] H. Viswanathan and T. Berger. The quadratic Gaussian CEO problem. *IEEE Transactions on Information Theory,* IT–43(5):1549–1559, September 1997.