Scaling Laws for Homogeneous Sensor Networks*

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

Martin Vetterli†⋆
†School of I&C
Swiss Federal Inst. of Technology (EPFL)
CH-1015 Lausanne, Switzerland
Martin.Vetterli@epfl.ch

Abstract

Sensor networks typically observe noisy versions of the desired data, and their goal is to inform an interested party. In our sensor network scenario, B_s sources are simultaneously observed by M sensors. More precisely, in every time slot, each sensor observes a certain noisy combination of the B_s sources. All these observations have to be communicated in a suitable form to a central data collection point whose goal is to learn the underlying B_s sources at the highest possible fidelity. For the case where all involved distributions are Gaussian, we show that the best possible scaling behavior of the distortion at the data collection point is like $1/M$. Then, we discuss the power-bandwidth trade-offs that permit to achieve this optimal distortion scaling law.

1 Introduction

Many sensor network scenarios involve both a compression and a transmission problem: it is rarely possible to transmit the observed signals *exactly* since, *(i)*, the underlying sensed world is analog, and *(ii)*, the communication links have limited capacities. Instead, the goal must be to communicate compressed versions of the observed signals. In network situations, compression and transmission have to be addressed jointly.[1] While exact information-theoretic performance results are rare for network situations, the perspective of *scaling laws*, spearheaded in [2], has proved more successful. Here, we consider the joint compression/transmission problem for a class of sensor networks from a scaling law perspective. In our (discrete-time) model, there are B_s underlying sources. Sensor m senses a particular combination of these sources, subject to measurement noise. The sensor network is *homogeneous* in the sense that all sensors have the same capabilities. Key observations are discussed for the case of Gaussian statistics and mean-squared error

*This research is supported, in part, by startup funds from the University of California, Berkeley, and by the Swiss National Science Foundation (NCCR-MICS).

[1] In point-to-point communications, Shannon's separation theorem [1, Thm. 21] establishes that compression can be handled separately from transmission (in the information-theoretic sense, i.e., if complexity and delay are left unconstrained). This is not true for networks in general, and it turns out to be a strongly suboptimal coding strategy for the type of sensor networks considered in this paper. More precisely, we are able to show that at fixed total sensor power P_{tot} and fixed target distortion D, the necessary number of sensors for the separation-based approach is *exponentially* higher than for the optimal coding scheme.

target distortion: It is shown that the distortion decreases at best like $1/M$ as a function of the number of sensors M. We then go on to discuss the power-bandwidth trade-offs that permit to achieve this optimum distortion scaling law. For example, it is found that when source and channel bandwidth are equal, a *constant* total power, shared by all M sensors, is necessary and sufficient to achieve the optimal distortion scaling law. Extensions beyond the case of Gaussian statistics, and to more general topologies, are discussed in [3, 4, 5]. An overview of this and further issues of interest to wireless sensor networks can be found in [6] and the references therein.

2 Sensor Network Model and Problem Statement

The sensor network model studied in this paper is shown in Figure 1. There is a physical phenomenon, characterized by B_s 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.[2] 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. Again, we model this observation process in a probabilistic fashion as a conditional distribution of the observations given the state. Based on the respective sensor readings, 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

$$\{S[n]\}_{n\in Z} \quad = \quad \{(S_1[n], S_2[n], \ldots, S_{B_s}[n])\}_{n\in Z}. \tag{1}$$

To simplify the notation in the rest of the paper, we denote sequences as $S^n \stackrel{def}{=} \{S[n]\}_{n\in Z}$. 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 $S[n]$ is not directly observed by the sensors. Rather, sensor m observes a sequence $U_m^n = \{U_m[n]\}_{n\in Z}$ which depends on the physical phenomenon according to a conditional probability distribution, which we denote by $P(u_m|s_1, \ldots, s_{B_s})$. Based on the observations $U_m[n]$, sensor m transmits a signal $X_m^n = F_m(U_m^n)$ on the multi-access channel. The transmitted signals satisfy a power, or more generally, a cost constraint of the form

$$E\rho(X_1^n, X_2^n, \ldots, X_M^n) \quad \leq \quad \Gamma. \tag{2}$$

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 $\rho_i(\cdot)$ and maximum expected cost Γ_i. This is a generalization of the individual power constraints for each sensor.

[2]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.

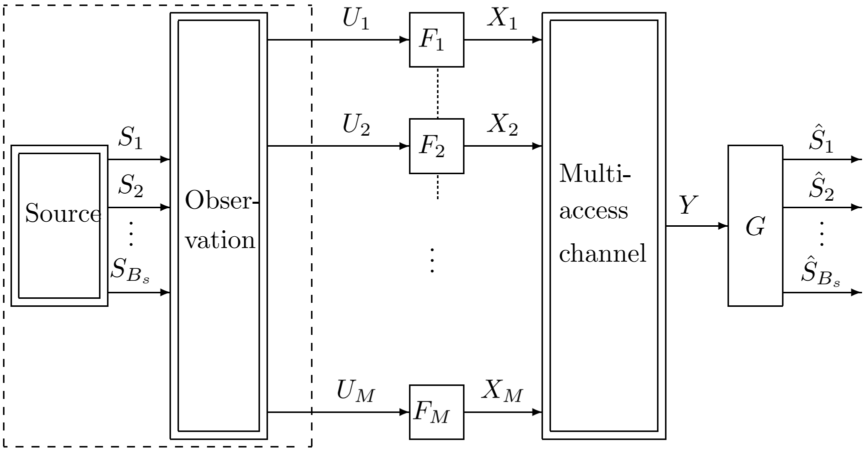

Figure 1: The sensor network topology considered in this paper.

An interesting variation is the case where the sensors also have radio receivers, and hence can send messages to each other. This is hinted at by the dotted lines in Figure 1, and discussed, in part, in [3, 5].

The final destination uses the output of the multi-access channel to construct estimates $\hat{S}^n = (\hat{S}_1^n, \hat{S}_2^n, \ldots, \hat{S}_{B_s}^n)$. In some of the most interesting sensor network applications, the sensor can acquire large amounts of data, and there is no hope for communicating the exact values of all sensor readings to the central data collection point. Under these circumstances, $\hat{S}^n$ cannot be made equal to S^n, and the relevant problem becomes to make the two as close to each other 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:

$$\Delta \quad = \quad Ed\left(S^n, \hat{S}^n\right). \tag{3}$$

The relevant figure of merit is therefore the trade-off between the *cost* Γ of the transmission (Equation (2)), and the achieved *distortion level* Δ (Equation (3)). The problem studied in this paper is that of finding the optimal trade-offs (Γ, Δ), where optimal is to be understood in an information-theoretic sense, i.e., irrespective of delay and complexity.

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

$$f_1(M) \quad \sim \quad f_2(M), \tag{4}$$

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

$$f_1(M) \quad \overset{=}{\sim} \quad f_2(M). \tag{5}$$

3 Single-source Gaussian Sensor Networks

3.1 Network Model

An important special case of the general sensor network topology of Figure 1 is illustrated in Figure 2: In this case, $B_s = 1$, and

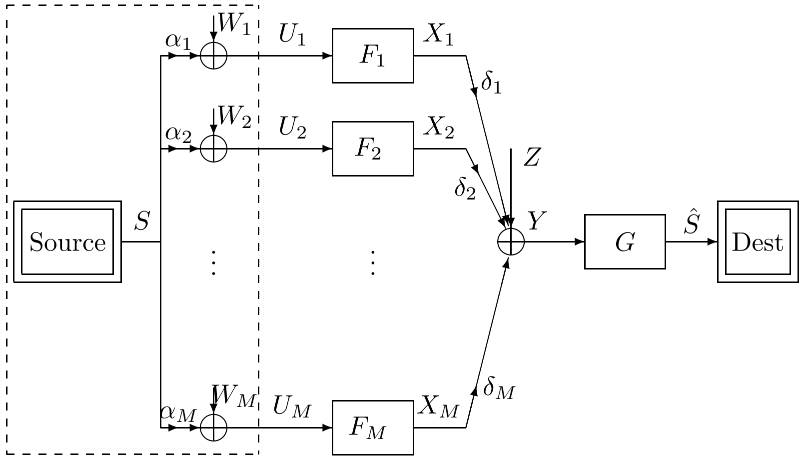

Figure 2: The single-source Gaussian example.

$$U_m[n] \quad = \quad \alpha_m S[n] + W_m[n], \tag{6}$$

where $\{S[n]\}_{n \in Z}$ is a sequence of independent and identically distributed (iid) circularly complex Gaussian random variables of variance σ_S^2, and $\{W_m[n]\}_{n \in Z}$ is a sequence of iid circularly complex Gaussian random variables of mean zero and variance σ_W^2. Moreover, for the sake of the example, we also assume that W_m and W_k are independent for all $m \neq k$.

The coefficients α_m may model effects such as the distance of sensor m from the phenomenon, and the coefficients δ_m model different path loss coefficients on the communications section of the network.[3] In this paper, the coefficients α_m and δ_m are assumed fixed and known throughout the network, and that $|\alpha_m| < \infty$, for $m = 1, 2, \ldots, M$.

The constraint on the signals transmitted by the sensors is a sum power constraint,

$$\sum_{m=1}^{M} E|X_m|^2 \quad \leq \quad P_{tot}(M). \tag{7}$$

The final destination receives $Y[n] = \sum_{m=1}^{M} \delta_m X_m[n] + Z[n]$, where $Z[n]$ is iid circularly complex Gaussian noise of variance σ_Z^2.

The distortion measure in this example is the mean-squared error (MSE), i.e., Eqn. (3) becomes $D = \frac{1}{n} \sum_{j=1}^{n} E|S[j] - \hat{S}[j]|^2$. The goal of the analysis is to determine the best trade-offs between the total sensor power $P_{tot}(M)$ and the incurred end-to-end distortion D, in the information-theoretic sense, i.e., at unconstrained processing delay and complexity.

3.2 Scaling Law for Separate Source and Channel Coding

By *separate source and channel coding*, we mean the communication strategy where each sensor (independently) compresses its observation into a source codeword (using the optimum distributed compression technique), and transmits this source codeword across the multi-access channel without making any errors (in the information-theoretic sense, as the length of the codewords tends to infinity). This is discussed in detail in [3].

[3] Note that if the coefficients α_m can be chosen arbitrarily, it is unnecessary to separately consider the case where the additive noises W_m have different variances.

The distortion scaling law for such separate source and channel coding can be lower bounded as follows (based on [7, 8], details see [5]).

Lemma 1. *For any separate source and channel coding scheme, the distortion for the single-source Gaussian sensor network illustrated in Figure 2, in the special case* $\alpha_m = 1$*, for* $m = 1, 2, \ldots, M$*, is lower bounded by*

$$D_{sep}(M, P_{tot}(M)) \;\; \geq \;\; \frac{\sigma_W^2/\sigma_S^2}{\log\left(1 + P_{tot}(M)\sum_{m=1}^{M}|\delta_m|^2/\sigma_Z^2\right)}, \tag{8}$$

where σ_S^2 *is the variance of the underlying source,* σ_W^2 *is the variance of the observation noise,* σ_Z^2 *is the variance of the noise in the multi-access channel, and* $P_{tot}(M)$ *is the total sensor transmit power for the* M *involved sensors.*

Remark. It is easy to show that essentially the same bound holds if the coefficients α_m are not all equal to one (as long as $|\alpha_m| < \infty$, for $m = 1, 2, \ldots, M$).

3.3 Improved Achievable Performance

The analysis of a simple joint compression/transmission coding scheme provides an improved achievable performance, in extension of [3]. The following is proved in [5].

Theorem 2. *For the single-source Gaussian sensor network discussed in this section and illustrated in Figure 2, the following distortion is achievable:*

$$D_1(M, P_{tot}(M)) \;\; = \;\; \frac{\sigma_S^2\sigma_W^2}{\sigma_W^2 + \sum_{m=1}^{M}|\alpha_m|^2\sigma_S^2}\left(1 + \frac{(\sigma_S^2\sigma_Z^2/\sigma_W^2)\sum_{m=1}^{M}|\alpha_m|^2}{\sigma_Z^2 + P_{tot}(M)b(M)}\right), \tag{9}$$

where

$$b(M) \;\; = \;\; \frac{\left(\sigma_W^2 + \sum_{m=1}^{M}|\alpha_m|^2\sigma_S^2\right)\sum_{m=1}^{M}|\alpha_m|^2}{\sum_{m=1}^{M}(|\alpha_m|^2\sigma_S^2 + \sigma_W^2)|\alpha_m|^2/|\delta_m|^2}, \tag{10}$$

and σ_S^2 *is the variance of the underlying source,* σ_W^2 *is the variance of the observation noise,* σ_Z^2 *is the variance of the noise in the multi-access channel, and* $P_{tot}(M)$ *is the total sensor transmit power for the* M *involved sensors.*

3.4 Lower Bound

A simple lower bound to the distortion for any scheme is proved in [5] and can be stated as follows.

Theorem 3. *For the single-source Gaussian sensor network discussed in this section and illustrated in Figure 2, the achievable distortion is lower bounded by*

$$D_{lower}(M, P_{tot}(M)) \;\; = \;\; \frac{\sigma_S^2\sigma_W^2}{\sigma_W^2 + \sum_{m=1}^{M}|\alpha_m|^2\sigma_S^2}\left(1 + \frac{(\sigma_S^2\sigma_Z^2/\sigma_W^2)\sum_{m=1}^{M}|\alpha_m|^2}{\sigma_Z^2 + P_{tot}(M)\sum_{m=1}^{M}|\delta_m|^2}\right), \tag{11}$$

where σ_S^2 *is the variance of the underlying source,* σ_W^2 *is the variance of the observation noise,* σ_Z^2 *is the variance of the noise in the multi-access channel, and* $P_{tot}(M)$ *is the total sensor transmit power for the* M *involved sensors.*

3.5 Optimal Scaling Law

The comparison of Thms. 2 and 3 yields the following optimal scaling law (see [5]).

Theorem 4 (optimal Gaussian/MSE scaling law). *The optimal scaling law for the single-source Gaussian sensor network discussed in this section, and illustrated in Figure 2, is given by*

$$D(M, P_{tot}(M)) \sim \frac{\sigma_S^2 \sigma_W^2}{\sigma_W^2 + \sum_{m=1}^M |\alpha_m|^2 \sigma_S^2} \left(1 + \frac{(\sigma_S^2 \sigma_Z^2 / \sigma_W^2) \sum_{m=1}^M |\alpha_m|^2}{\sigma_Z^2 + P_{tot}(M) \sum_{m=1}^M |\delta_m|^2}\right), \quad (12)$$

provided that $b(M)$ and $\sum_{m=1}^M |\delta_m|^2$ have the same dependence on M in the sense that

$$\lim_{M\to\infty} \frac{b(M)}{\sum_{m=1}^M |\delta_m|^2} = c > 0. \quad (13)$$

where σ_S^2 is the variance of the underlying source, σ_W^2 is the variance of the observation noise, and M is the number of involved sensors. If moreover $P_{tot}(M)$ is an unboundedly increasing function of M, then (12) becomes a strong scaling law.

3.6 Discussion

In order to gain insight into Thm. 4, consider the simple case where $\alpha_m = \delta_m = 1$, for $m = 1, 2, \ldots, M$. Thms. 2 and 3 bound the minimum achievable distortion D as

$$\frac{\sigma_S^2 \sigma_W^2}{M\sigma_S^2 + \sigma_W^2}\left(1 + \frac{M(\sigma_S^2\sigma_Z^2/\sigma_W^2)}{MP_{tot}(M) + \sigma_Z^2}\right) \le D \le \frac{\sigma_S^2 \sigma_W^2}{M\sigma_S^2 + \sigma_W^2}\left(1 + \frac{M(\sigma_S^2\sigma_Z^2/\sigma_W^2)}{\frac{M\sigma_S^2+\sigma_W^2}{\sigma_S^2+\sigma_W^2} P_{tot}(M) + \sigma_Z^2}\right).$$

Clearly, no matter how high a total power is chosen, the scaling behavior is at best like $1/M$. More precisely, the following observations can be made:[4]

1. *Unbounded total power.* When $P_{tot}(M)$ is an *unboundedly increasing* function of M, the asymptotic behavior is exactly as if the destination knew $U_1, U_2, \ldots, U_M$ precisely, i.e., $D(M) = \sigma_S^2\sigma_W^2/(M\sigma_S^2 + \sigma_W^2)$, irrespective of the precise shape of the function $P_{tot}(M)$.

2. *Bounded total power.* If $P_{tot}(M)$ is a bounded, non-decreasing function of M, and hence, more and more sensors have to share a bounded common power budget, we obtain a weak scaling law $D(M, P_{tot}) = \sigma_S^2\sigma_W^2/(M\sigma_S^2+\sigma_W^2)(1+\sigma_S^2\sigma_Z^2/(P_{tot}\sigma_W^2))$, i.e., the scaling behavior of distortion is still $1/M$, but there is a small loss with respect to the ideal case where the destination knows the precise values of $U_1, U_2, \ldots, U_M$.

3. *Scaling-law suboptimality of separate source and channel coding.* For the sensor network considered in this section, separate source and channel coding is not only suboptimal, it is *suboptimal in a scaling sense.* To illustrate this point, fix the target distortion D, and the number of sensors M. Suppose that the lower bound of Thm. 3 requires a total power of $P_{tot}(M)$. Then, the strategy of Lemma 1 requires a total power $P_{tot,sep}(M)$ of at least

$$P_{tot,sep}(M) \ge \frac{\sigma_Z^2}{M}\left(2^{M\frac{\sigma_S^2+\sigma_W^2}{\sigma_S^4}\frac{MP_{tot}(M)+\sigma_Z^2}{MP_{tot}(M)+\sigma_Z^2+M\sigma_S^2\sigma_Z^2/\sigma_W^2}} - 1\right). \quad (14)$$

[4] Similar asymptotic statements hold for many cases, including the case where $0 < |\alpha_m| < \infty$ and $0 < |\delta_m| < \infty$, for $m = 1, 2, \ldots, M$, see [5].

4 Multi-source Gaussian Sensor Network

4.1 Network Model

In extension of the single-source Gaussian sensor network model presented in Section 3, we consider in this section the Gaussian sensor network with *multiple* underlying sources. The model of sensing is as follows. B_s physical phenomena (or: B_s physical degrees of freedom) are modeled by a discrete-time random process each, $\{S_l\}_{l \in Z}$, for $l = 1, \ldots, B_s$. The quantity B_s can equivalently be interpreted as the *bandwidth* of the underlying physical process. There are two interesting cases. In the first case, B_s is a fixed, finite number, and the goal is to determine the system performance as the number of sensors M becomes large. This models a *dense* sensor network: A finite number B_s of degrees of freedom is monitored by more and more sensors. In the second case, the number of sources B_s increases with the number of sensors, hence modeling an *expanding* sensor network situation. Even though most of the results presented in this section apply to both cases, our focus will be on the dense situation. The expanding case is treated in more detail in [5]. For the purpose of this section, we assume that the B_s sources are independent of each other, and that each process is specified by an iid sequence of zero-mean Gaussian random variables of variance σ_S^2.[5] The B_s underlying phenomena are simultaneously sensed by M sensors. Sensor m senses the following signal:

$$U_m[n] \;=\; \sum_{l=1}^{B_s} \alpha_{ml} S_l[n] + W_m[n], \tag{15}$$

where the observation noise processes W_m are exactly as defined in the single-source Gaussian model in Section 3. For notational simplicity, we define u_m to be the sensed power at sensor m, i.e.,

$$u_m \;=\; \sum_{l=1}^{B_s} |\alpha_{m,l}|^2 \sigma_S^2 + \sigma_W^2 \tag{16}$$

and we collect the coefficients $\alpha_{m,l}$, $m = 1, 2, \ldots, M$, $l = 1, 2, \ldots, B_s$, into an $M \times B_s$ matrix A, i.e.,

$$A \;=\; \begin{pmatrix} \alpha_{1,1} & \alpha_{1,2} & \cdots & \alpha_{1,B_s} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{M,1} & \alpha_{M,2} & \cdots & \alpha_{M,B_s} \end{pmatrix} \tag{17}$$

On the communication side, we consider the same model as in Section 3, the only difference being that we now allow B_c *channel uses for each source sample* $U_m[n]$. The total power available *for all the* B_c *channel uses together* is called $P_{tot}(M)$. Note that instead of B_c channel uses, it is equivalent to consider a bandwidth increase by a factor of B_c. Hence, this model allows to trade-off power and bandwidth. We discuss this issue in detail below.

Finally, the distortion measure is again the MSE. In particular, we will express our scaling law in terms of the average MSE, averaged over all B_s sources, i.e., $D = \frac{1}{nB_s} \sum_{l=1}^{B_s} E||S_l^n - \hat{S}_l^n||^2$.

[5] Assuming iid sources is without loss of generality in the sense that the matrix A in Eqn. (17) below can be chosen arbitrarily.

4.2 Scaling Law for Separate Source and Channel Coding

By analogy to Lemma 1, the distortion for separate source and channel coding can be lower bounded as follows (see [5] for details).

Lemma 5. *For the multi-source Gaussian sensor network described in this section, suppose that the observation coefficients satisfy $|\alpha_{m,l}| \leq \alpha_{max} < \infty$, for $m = 1, \ldots, M$ and $l = 1, \ldots, B_s$. Then, a lower bound to the distortion for separate source and channel coding is*

$$D_{sep}(M, P_{tot}(M)) \;\geq\; \frac{\sigma_W^2/\sigma_S^2}{B_c \alpha_{max}^2 \log\left(1 + P_{tot}(M) \sum_{m=1}^{M} |\delta_m|^2/(B_c \sigma_Z^2)\right)}, \tag{18}$$

where σ_S^2 is the variance of the underlying source, σ_W^2 is the variance of the observation noise, σ_Z^2 is the variance of the noise in the multi-access channel, and $P_{tot}(M)$ is the total sensor transmit power for the B_c channel uses.

4.3 Improved Achievable Performance for $B_c = B_s$

The scheme leading to Thm. 2 can be extended to the case of multiple sources whenever $B_c = B_s$ (see [5] for details).

Theorem 6. *The following distortion is achievable in the multi-source Gaussian sensor network described in this section, with $\delta_m = 1$, for $m = 1, 2, \ldots, M$:*[6]

$$D_1(M, P_{tot}(M)) \;=\; \frac{1}{B_s} \sum_{l=1}^{B_s} \frac{\sigma_S^2 \sigma_W^2}{\nu_l^2 \sigma_S^2 + \sigma_W^2} + \frac{1}{P_{tot}(M)} \sum_{l=1}^{B_s} \frac{\nu_l^2 u_{[l]} \sigma_Z^2}{(\nu_l^2 + \sigma_W^2/\sigma_S^2)^2} \tag{19}$$

where $\nu_1, \nu_2, \ldots, \nu_{B_s}$ are the singular values of the matrix A, σ_S^2 is the variance of the underlying source, σ_W^2 is the variance of the observation noise, σ_Z^2 is the variance of the noise in the multi-access channel, $P_{tot}(M)$ is the total sensor transmit power for the $B_c = B_s$ channel uses, and $u_{[l]}$ denotes the l-th largest of the sensed powers u_m defined in Eqn. (16).

4.4 Lower Bound to the Distortion

The arguments leading to Thm. 3 are extended in [5] to yield the following lower bound.

Theorem 7. *The distortion that can be achieved in the multi-source Gaussian sensor network described in this section, under the additional assumption that the spread of the singular values $\nu_1, \nu_2, \ldots, \nu_{B_s}$ of the matrix A is small,*[7] *cannot be smaller than*

$$D_{lower}(M, P_{tot}(M))$$
$$= \frac{1}{B_s} \sum_{l=1}^{B_s} \frac{\sigma_S^2 \sigma_W^2}{\nu_l^2 \sigma_S^2 + \sigma_W^2} + \left(\frac{1}{1 + \frac{P_{tot}(M)}{B_c \sigma_Z^2} \sum_{m=1}^{M} |\delta_m|^2} \right)^{B_c/B_s} \sqrt[B_s]{\prod_{l=1}^{B_s} \frac{\nu_l^2 \sigma_S^4}{\nu_l^2 \sigma_S^2 + \sigma_W^2}}, \tag{20}$$

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, and $P_{tot}(M)$ is the total sensor transmit power for the B_c channel uses.

---

[6] It is straightforward to extend the result to the case of general δ_m, at the expense of extra notation [5].

[7] The full solution involves "inverse water-filling" and is given in [5].

4.5 Optimal Scaling Laws for $B_c = B_s$

It is clear that the expressions given in Thms. 6 and 7 do not coincide in general. However, under certain simple assumptions on the singular values, both expressions describe the same scaling behavior, i.e., the same dependence on the number of sensors M. The goal of this section is to characterize this fact more precisely.

To this end, consider the scaled matrix sequence $\tilde{A}^{(M)} = A^{(M)}/\sqrt{M}$ with singular values $\mu_1^{(M)}, \mu_2^{(M)}, \ldots, \mu_{B_s}^{(M)}$. At least for dense sensor networks (where B_s is a constant, but M increases without bound), it is easy to see that many interesting scenarios have (scaled) singular values $\mu_l^{(M)}$ that behave nicely in the sense that, *(i)*, $0 < \mu_l^{(M)} < \infty$, for $l = 1, 2, \ldots, B_s$, for all M, and *(ii)*, they all converge to finite non-zero values $\mu_1, \mu_2, \ldots, \mu_{B_s}$ as $M \to \infty$. If these assumptions are satisfied, we obtain the following scaling law (see [5] for details).

Theorem 8. *The optimal strong scaling law for the multi-source Gaussian sensor network described in this section (with $\delta_m = 1$, for $m = 1, 2, \ldots, M$),*[6] *under the additional assumption that the spread of the (scaled) singular values $\mu_1, \mu_2, \ldots, \mu_{B_s}$ of the matrix A is small,*[7] *and that the total power $P_{tot}(M)$ is a non-decreasing function of M, is given by*

$$D(M, P_{tot}(M)) \sim \frac{1}{M}\left(\frac{1}{B}\sum_{l=1}^{B}\frac{\sigma_S^2\sigma_W^2}{\mu_l^2\sigma_S^2 + \sigma_W^2/M}\right) + \frac{1}{M\frac{P_{tot}(M)}{B\sigma_Z^2} + 1}\sqrt[B]{\prod_{l=1}^{B}\frac{\mu_l^2\sigma_S^4}{\mu_l^2\sigma_S^2 + \sigma_W^2/M}}$$

where $B_s = B$ is the number of sources and σ_S^2 is the variance of each source, σ_W^2 is the variance of the observation noise at each of the M sensors, $P_{tot}(M)$ is the total power used by the sensors for the $B_c = B$ channel uses, and σ_Z^2 is the additive noise on the multi-access channel.

If $P_{tot}(M)$ increases unboundedly with M, then the above is a strong scaling law.

4.6 The Bandwidth vs Power Trade-Off

For the special class of sensor networks studied in Section 4.5, the best scaling law for the distortion behaves like $1/M$, *irrespective* of power and bandwidth. Hence, a relevant question is: Which power-bandwidth pairs permit to achieve the $1/M$ scaling law? The lower bound given in Thm. 7 permits to analyze this trade-off to some extent, as the following corollary shows:

Corollary 9. *In the multi-source Gaussian sensor network described in this section (with $\delta_m = 1$, for $m = 1, 2, \ldots, M$),*[6] *under the additional assumption that the spread of the singular values $\nu_1, \nu_2, \ldots, \nu_{B_s}$ of the matrix A is small,*[7] *the total power $P_{tot}(M)$ required to sustain a (weak or strong) scaling law of*

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

satisfies $P_{tot}(M) \geq P_{tot,lower}(M)$, where

$$P_{tot,lower}(M) \sim B_c M^{\frac{B_s}{B_c}-1}, \tag{22}$$

where B_s is the number of sources (or, equivalently, the source bandwidth), and B_c is the number of channel uses per source sample (or, equivalently, the channel bandwidth).

Remark. We observe the following:
1. If $B_c < B_s$, the total power *must* increase with M.
2. If $B_c = B_s$, then a constant total power is sufficient to achieve the optimum (weak) distortion scaling law. This is established in Thm. 8.
3. If $B_c > B_s$, a total power that decreases with M may be sufficient. The achievability of such an operating point is currently under investigation. In particular, simple (joint source-channel) feedback schemes are able to harvest the gains offered by the additional bandwidth (see e.g. [9] and the references therein).

5 Conclusions and Extensions

The key characteristic of the sensor networks studied in this paper is that the data of interest cannot be observed directly by the sensors. Rather, only noisy measurements are available. For the case of Gaussian statistics under MSE distortion, the paper determines the information-theoretic scaling behavior in terms of the power-bandwidth-distortion trade-offs. It is found that the distortion decreases at best like $1/M$, where M is the number of sensors. Thereafter, the paper studies power-bandwidth trade-offs that are necessary and sufficient to actually achieve this optimal distortion scaling,

Future extensions of this work include the following. *1. Beyond Gaussian statistics.* Along the lines of [10], the presented results can be generalized to other statistics and distortion measures. Initial results can be found in [3]. *2. Further sensor network topologies.* Topologies beyond the one shown in Fig. 1 are of interest, including the case of multiple data collection points, and of feedback links. Initial work appears in [4].

References

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

[2] P. Gupta and P. R. Kumar, "The capacity of wireless networks," *IEEE Transactions on Information Theory*, vol. IT–46, pp. 388–404, March 2000.

[3] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *2nd International Workshop on Information Processing in Sensor Networks (IPSN'03)* (L. J. Guibas and F. Zhao, eds.), pp. 162–177, New York, NY: Lecture Notes in Computer Science, Springer, April 2003.

[4] M. Gastpar, "On source-channel communication in networks," in *DIMACS Workshop on Network Information Theory* (P. Gupta, G. Kramer, and A. van Wijngarden, eds.), AMS DIMACS, March 2003.

[5] M. Gastpar and M. Vetterli, "Scaling laws for sensor networks," *to be submitted to IEEE Transactions on Information Theory*, November 2003.

[6] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Commun. ACM*, vol. 43, pp. 51–58, May 2000.

[7] H. Viswanathan and T. Berger, "The quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. IT–43, pp. 1549–1559, September 1997.

[8] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Transactions on Information Theory*, vol. IT–44, pp. 1057–1070, May 1998.

[9] M. Gastpar and B. Rimoldi, "Source-channel communication with feedback," in *Proc 2002 IEEE Information Theory Workshop*, (Paris, France), April 2003. Invited paper.

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