Source-Channel Communication in Sensor Networks

Michael Gastpar[1] and Martin Vetterli[12]

[1] University of California, Berkeley CA 94720, USA,
gastpar@eecs.berkeley.edu,
http://www.eecs.berkeley.edu/~gastpar
[2] Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland.

Abstract. Sensors acquire data, and communicate this to an interested party. The arising coding problem is often split into two parts: First, the sensors compress their respective acquired signals, potentially applying the concepts of distributed source coding. Then, they communicate the compressed version to the interested party, the goal being not to make any errors. This coding paradigm is inspired by Shannon's separation theorem for point-to-point communication, but it leads to suboptimal performance in general network topologies. The optimal performance for the general case is not known.

In this paper, we propose an alternative coding paradigm based on joint source-channel coding. This coding paradigm permits to determine the optimal performance for a class of sensor networks, and shows how to achieve it. For sensor networks outside this class, we argue that the goal of the coding system could be to approach our condition for optimal performance as closely as possible. This is supported by examples for which our coding paradigm significantly outperforms the traditional separation-based coding paradigm. In particular, for a Gaussian example considered in this paper, the distortion of the best coding scheme according to the separation paradigm decreases like $1/\log M$, while for our coding paradigm, it decreases like $1/M$, where M is the total number of sensors.

1 Introduction

In a sensor network, the goal is typically to reconstruct the measured physical phenomenon to within some prescribed distortion level, and this at the smallest possible cost on the communication link. What coding strategy should the sensors use? For the case of a single sensor, i.e., for the ergodic point-to-point communication scenario, Shannon proved that separate source and channel code design is an optimal strategy (asymptotically as the delay and the complexity become unconstrained [12]). This fact, known as the separation principle, is both conceptually and practically appealing. Therefore, it is also a tempting coding paradigm in a network context: each sensor compresses its measurements using the best possible distributed coding techniques, see e.g. [1,13,15];

the resulting source codewords are then transmitted across the channel using capacity-achieving codes, see e.g. [4, Ch. 14]. It is well known that in spite of its elegance, this coding paradigm does *not* lead to optimal performance in networks, see e.g. [4, p. 449], [7]. In other words, the concepts of capacity and rate-distortion do not characterize the best achievable performance. Rather, joint source-channel coding techniques can significantly outperform the separation-based coding paradigm in these cases.

In this paper, we propose an alternative coding paradigm based on recent results on the source-channel communication problem in the point-to-point case [5, 6,7]. We investigate a particular sensor network topology, which is defined in detail in Section 2. M separate sensors observe each a different noisy version of a physical phenomenon S. The sensors communicate over a multi-access channel to a central observer who wishes to produce an estimate $\hat{S}$ of the physical phenomenon in such a way as to minimize the distortion. The precise shape of the distortion measure $d(s, \hat{s})$ depends on the problem at hand.

In Section 3, we evaluate the performance achievable by the separation-based coding paradigm in our sensor network topology. Results are given in particular for a Gaussian example, for which we prove that the achievable distortion decays like $1/\log M$, where M is the number of sensors.

In Section 4, we develop a simple joint source-channel coding strategy for the same Gaussian example. We establish two key facts for our strategy: First, we determine that the distortion decays like $1/M$, where M is the number of sensors, thus considerably outperforming the separation-based coding paradigm. Second, we prove that as M tends to infinity, our strategy achieves the smallest possible distortion.

Thereafter, we extend our results beyond the Gaussian example, establishing a general joint source-channel coding paradigm for the considered sensor network topology, using the arguments of [5,6,7]. We show that it sometimes leads to provably optimum performance, but even when it does not, we illustrate that it can considerably outperform the separation-based coding paradigm.

In Section 5, we outline the extension of our basic sensor network topology to include communication *between* the sensors. In particular, we find for a class of sensor networks that this additional degree of freedom does not enhance the asymptotic performance (as the number of sensors M tends to infinity).

2   The Considered Sensor Network

Consider the sensor network shown in Figure 1: The physical phenomenon is characterized by the sequence of random vectors

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

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

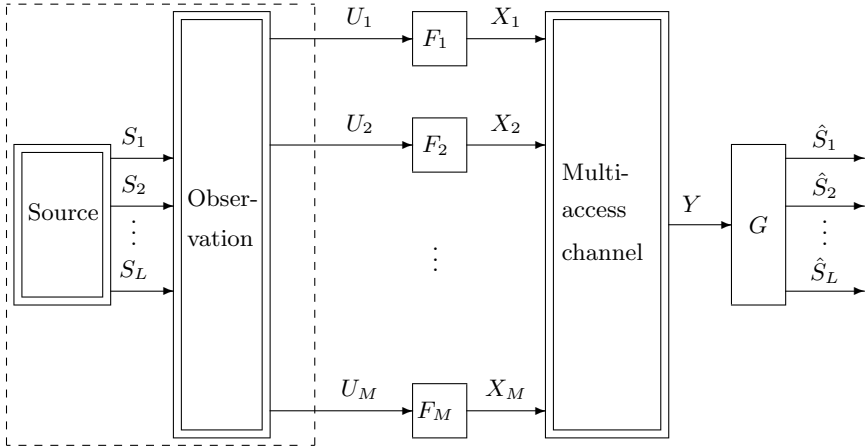$$S^n \overset{def}{=} \{S[n]\}_{n \in Z}. \tag{2}$$

Fig. 1. The sensor network topology considered in this paper.

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 k observes a sequence $U_k^n = \{U_k[n]\}_{n \in Z}$ which depends on the physical phenomenon according to a conditional probability distribution, which we denote by

$$P(u_k | s_1, \ldots, s_L). \tag{3}$$

Based on the observations $U_k[n]$, sensor k transmits a signal

$$X_k^n = F_k (U_k^n) \tag{4}$$

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

$$E\rho(X_1^n, X_2^n, \ldots, X_M^n) \leq \Gamma. \tag{5}$$

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 $\rho_i(\cdot)$ and maximum expected cost Γ_i. 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

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

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:

$$\Delta = E d \left(S^n, \hat{S}^n \right). \tag{7}$$

For a particular coding scheme $(F_1, F_2, \ldots, F_M, G)$, the performance is determined by the required cost Γ and the incurred distortion Δ. The goal of the analysis is to determine the set of optimal trade-offs (Γ, Δ), where optimal is to be understood in an information-theoretic sense, i.e., irrespective of delay and complexity.

Example 1 (Gaussian case). An important special case of the sensor network topology of Figure 1 is illustrated in Figure 2: In this case, $L = 1$, and
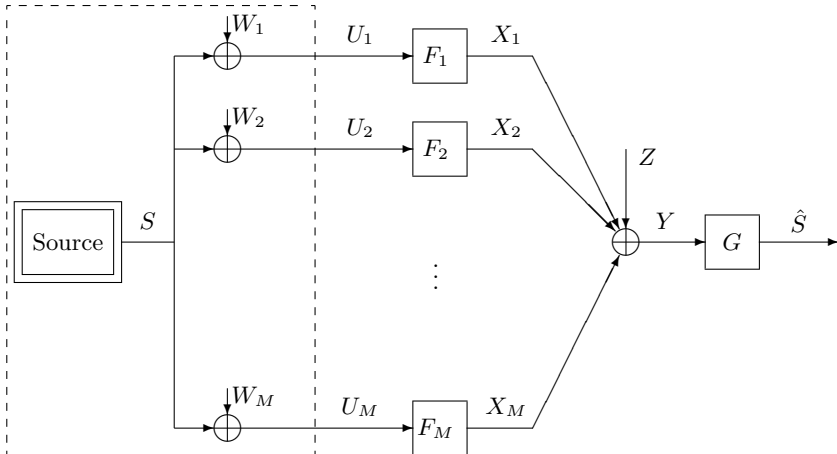

Fig. 2. The Gaussian example.

$$U_k[n] = S[n] + W_k[n], \tag{8}$$

where $\{S[n]\}_n$ is a sequence of independent and identically distributed (iid) Gaussian random variables of variance σ_S^2, and $\{W_k[n]\}_n$ is a sequence (in n, for $n = 1, 2, 3, \ldots$) of iid Gaussian random variables of mean zero and variance σ_W^2. Moreover, for the sake of the example, we also assume that W_k and W_l are independent for all $k \neq l$. The constraint on the signals transmitted by the sensors is a sum power constraint, i.e.,

$$\sum_{k=1}^{M} E|X_k|^2 \leq MP. \tag{9}$$

The final destination receives

$$Y[n] = \sum_{k=1}^{M} X_k[n] + Z[n], \tag{10}$$

where $Z[n]$ is iid Gaussian noise of variance σ_Z^2. The distortion measure in this example is the mean-squared error, i.e.,

$$D = \sum_{j=1}^{n} E|S[n] - \hat{S}[n]|^2. \tag{11}$$

The goal of the analysis is to determine the best power-distortion trade-off. More precisely, we want to determine the minimum distortion for a fixed power MP. The performance of a communication scheme employing source-channel separation is analyzed in two steps: the rate-distortion and the capacity-cost problem.

3   Separate Source and Channel Coding

In extension of the point-to-point case (summarized in Appendix A), a general coding paradigm for sensor networks can again be formulated as the combination of source coding and channel coding. In this section, we outline the performance that can be achieved using a separation-based coding strategy. Then, we provide an explicit answer for a simple Gaussian sensor network with a topology according to Figure 1. Note that it is well-known that this coding paradigm does not lead to optimal network designs in general, see e.g. [4, p. 449].

3.1   Distributed Source Coding

The particular source coding problem corresponding to Figure 1 is known as the CEO problem, proposed and partially solved in [2]. More precisely, the problem for sensor k is to encode its observations into a bit stream of R_k bits per sample. The determination of the set of the smallest $(R_1, R_2, \ldots, R_M)$ that permit the source decoder to reconstruct $S_1, \ldots, S_L$ at a specified fidelity Δ (as in Equation (7)) is an open problem in general. We call this the rate-distortion region, denoted by $\mathcal{R}(\Delta)$.

Example 2 (Gaussian case). For the Gaussian sensor network of Example 1, the particular problem of encoding U_k as in Figure 2 into bits and reconstructing S from these bits, has been studied in the shape of the so-called quadratic Gaussian CEO problem in the literature [11,14]. The distortion D_{CEO} depends on the total rate used by the sensors, $R_{tot} = R_1 + R_2 + \ldots + R_M$ as

$$D_{CEO} = \frac{\sigma_W^2}{2R_{tot}}, \tag{12}$$

when the total rate R_{tot} is large. More precise results for small R_{tot} can be found in [11].

3.2 Multi-access with Dependent Messages

For the multi-access problem in Figure 1, the goal is to determine the set of all achievable rate pairs (R_1, R_2) when the channel inputs satisfy the cost constraint Γ (as in Equation (5)), i.e., the capacity region $\mathcal{C}(\Gamma)$. When the messages of the different users are independent from each other, the capacity region is known, see e.g. [4]. However, in a sensor network situation, the messages of the sensors may typically be correlated since the underlying signals are. The capacity region for the case where the messages are potentially correlated is only partially known [3].

Example 3 (Gaussian case). For the Gaussian sensor network of Example 1, the goal is to determine the capacity region for the additive white Gaussian multiple-access channel with inputs $X_1, X_2, \ldots, X_M$ and output Y. When the messages may be dependent, the maximum sum rate R_{tot} can be upper bounded by

$$R_{tot} \leq \frac{1}{2} \log_2 \left(1 + \frac{M^2 P}{\sigma_Z^2} \right). \tag{13}$$

This bound follows by allowing *arbitrary* dependence between the inputs of the multi-access channel. However, in the Gaussian sensor network of Example 1, since the encoders are separate, the messages cannot be arbitrarily dependent; rather, their dependence follows from the source structure. For this reason, the bound should not be expected to be tight, but rather too optimistic.

3.3 Achievable Cost-Distortion Trade-Offs

For separate source and channel coding, a cost-distortion trade-off (Γ, Δ) is achievable only if

$$\mathcal{R}(\Delta) \cap \mathcal{C}(\Gamma) \neq \emptyset. \tag{14}$$

This follows immediately from the definition of the rate-distortion and the capacity-cost regions. In other words, if the two regions do not intersect, it is not possible to achieve the trade-off (Γ, Δ) by a strategy composed of optimal source compression followed by capacity-approaching channel coding.

Remark 1. Condition (14) is only achievable if each sensor is allowed to observe the entire sequence U_k^n before deciding what X_k^n to transmit.

Example 4 (Gaussian case). For the Gaussian sensor network of Example 1, the optimum power-distortion trade-offs that can be achieved by separate source and channel coding can be determined by combining the results of Examples 2 and 3. For a fixed total sensor power MP, the smallest achievable distortion is bounded by inserting the upper bound to the total rate R_{tot} on the multi-access channel (from Equation (13)) in the minimum distortion for the CEO problem

(Equation (12)). Hence, the distortion achieved by the separation-based scheme behaves at best like

$$D_{sep}(MP) \geq \frac{\sigma_W^2}{\log_2\left(1 + \frac{M^2 P}{\sigma_Z^2}\right)}, \tag{15}$$

when the number of sensors M is large. Here, σ_W^2 is the variance of the observation noise and also the variance of the noise in the multi-access channel, and MP is the total sensor transmit power.

4   Joint Source-Channel Coding

Since it is well-known that the separation-based coding paradigm does not lead to optimal system designs in general, we now develop an alternative coding paradigm for our sensor network topology, illustrated in Figure 1. This is motivated by a particular feature of the Gaussian example discussed above. Therefore, we first study an alternative coding scheme for the Gaussian case that outperforms separation-based strategies considerably. Thereafter, we extend this insight into a general coding paradigm, using an approach reminiscent of [5,6,7].

4.1   The Gaussian Case

An Achievable Distortion. For the point-to-point transmission of an iid Gaussian source across an additive white Gaussian channel, it is well-known that uncoded transmission is optimal [10], see also [5,6,7]. For the Gaussian example illustrated in Figure 2 and defined in Example 1, it is therefore intuitive to study the strategy of uncoded forwarding by the sensors. The following power-distortion trade-off is achieved by this strategy.

Theorem 1. *For the Gaussian sensor network defined in Example 1, with source variance σ_S^2, observation noise variance σ_W^2, and total transmit power at the sensors of MP, the following distortion is achievable:*

$$D_1(MP) = \frac{\sigma_S^2 \sigma_W^2}{\frac{M^2}{M+(\sigma_Z^2/\sigma_W^2)(\sigma_S^2+\sigma_W^2)/P}\sigma_S^2 + \sigma_W^2}. \tag{16}$$

Proof. Suppose the sensors apply uncoded transmission. More precisely, sensor k scales $U_k[n]$ to meet its power constraint P,

$$X_k[n] = \sqrt{\frac{P}{\sigma_S^2 + \sigma_W^2}} U_k[n]. \tag{17}$$

Recalling that $U_k[n] = S[n] + W_k[n]$, the received signal for the uncoded strategy is

$$Y[n] = \sqrt{\frac{P}{\sigma_S^2 + \sigma_W^2}}\left(MS[n] + \sum_{k=1}^{M} W_k[n]\right) + Z[n]. \tag{18}$$

It remains to specify the decoder. Since the encoding operation is memoryless, the optimum decoder (or estimator) G is also memoryless: there is no benefit from considering multiple symbols jointly. The optimum decoder G is then simply the minimum mean-squared error estimator of $S[n]$, given by the standard formula:

$$\hat{S}[n] = \frac{E\left[SY\right]}{E\left[Y^2\right]}Y[n] = \frac{\sqrt{\frac{P}{\sigma_S^2+\sigma_W^2}}M\sigma_S^2}{\frac{P}{\sigma_S^2+\sigma_W^2}(M^2\sigma_S^2 + M\sigma_W^2) + \sigma_Z^2}Y[n]. \tag{19}$$

The resulting distortion is evaluates to

$$D_1(MP) = \sigma_S^2 - \frac{\frac{P}{\sigma_S^2+\sigma_W^2}M^2\sigma_S^4}{\frac{P}{\sigma_S^2+\sigma_W^2}(M^2\sigma_S^2 + M\sigma_W^2) + \sigma_Z^2}$$

$$= \frac{\sigma_S^2\sigma_W^2}{\frac{M^2}{M+(\sigma_Z^2/\sigma_W^2)(\sigma_S^2+\sigma_W^2)/P}\sigma_S^2 + \sigma_W^2}. \tag{20}$$

$\square$

The main result of Theorem 1 follows from the comparison of (16) with (15): The separation-based scheme is clearly suboptimal in our example. More precisely, the decreasing behavior of the distortion as a function of the number of sensors M is fundamentally different for the two schemes: The separation-based scheme achieves at best a decreasing behavior of $1/\log M$, while Theorem 1 establishes an achievable decreasing behavior of $1/M$.

Asymptotic Optimality. In this paragraph, we first derive a lower bound to the minimum achievable distortion $D_{min}(MP)$ at total sensor power MP. We then establish that this lower bound coincides with $D_1(MP)$ in the limit as the number of sensors becomes large. This proves that asymptotically in M, the strategy of Theorem 1 performs optimally.

The lower bound is found by analyzing the system in which the sensors are ideally linked to the destination. This system can be interpreted as a point-to-point multi-antenna system, where the sender has one antenna with output S and the receiver has M antennas with inputs $U_1, U_2, \ldots, U_M$. The minimum distortion achievable in this system cannot be larger than the minimum distortion achievable in our sensor network. The lower bound can stated as follows:

Theorem 2. *For the Gaussian sensor network defined in Example 1, with source variance σ_S^2, observation noise variance σ_W^2, and total transmit power at the sensors of MP, the minimum achievable distortion satisfies $D_{min}(MP) \geq D_{lower}(MP)$, where*

$$D_{lower}(MP) = \frac{\sigma_S^2\sigma_W^2}{M\sigma_S^2 + \sigma_W^2}. \tag{21}$$

Proof. The lower bound is found by idealization: The receiver is ideally linked to the sensors, and we suppose that the physical phenomenon $S[n]$ itself uses optimal coding.

The smallest distortion achievable in this idealized system obviously cannot be larger than the smallest distortion achievable in the original system.

The idealized system is an ergodic point-to-point source-channel communication system; the separation theorem applies. The capacity of the idealized channel is simply the capacity of the multi-antenna channel with one transmit antenna (the source itself) and M receive antennae (the M sensors, now ideally linked to the destination). The capacity of such a system is well-known:

$$C = \frac{1}{2} \log_2 \left(1 + \frac{M\sigma_S^2}{\sigma_W^2} \right). \tag{22}$$

By the separation theorem, the minimum (mean-squared error) distortion that can be achieved for a Gaussian source across this channel is

$$D_{\mathcal{N}}(C) = \frac{\sigma_S^2 \sigma_W^2}{M\sigma_S^2 + \sigma_W^2}, \tag{23}$$

where $D_{\mathcal{N}}(\cdot)$ denotes the distortion-rate function of the iid Gaussian source of variance σ_S^2. This concludes the proof. □

In the limit as $M \to \infty$, the upper and lower bounds of this paper coincide, establishing the following result on the optimal power-distortion trade-off in the Gaussian sensor network of Example 1.

Theorem 3. *For the Gaussian sensor network defined in Example 1, with source variance σ_S^2, observation noise variance σ_W^2, and a total transmit power at the sensors of MP,*

$$\lim_{M \to \infty} \frac{D_{lower}(MP)}{D_1(MP)} = 1, \tag{24}$$

hence,

$$D_{min}(MP) = \frac{\sigma_S^2 \sigma_W^2}{M\sigma_S^2 + \sigma_W^2}, \tag{25}$$

and the minimum is achieved when the sensors use a simple scaling, $X_k[n] = \gamma_k U_k[n]$, and the final destination uses $\hat{X}[n] = \gamma Y[n]$.

Proof. The theorem follows directly by combining Theorems 1 and 2. □

Remark 2. In the limiting case as $M \to \infty$, the minimum distortion $D(MP)$ does not depend on P and σ_Z^2. Note however that the result *does* depend on the fact that the total available power at the sensor increases linearly with the number of sensors M. This can be extended to cases where the total sensor power behaves according to a different law along the lines of the analysis in [8, 9].

Remark 3 (real-time processing). There is no causality or real-time constraint on the encoding and decoding functions in the setup of Example 1; in fact, the scheme discussed in Section 3 does not satisfy any such constraint. In contrast to this, Theorem 3 shows that the globally optimum trade-off (in the limit $M \to \infty$) *can be achieved* by a causal and real-time system in the sense that $X_k[n]$ can be generated without waiting for future source outputs $U_k[n+j]$, $j = 1, 2, \ldots$.

The Gaussian example can be extended using concepts similar to [5,6,7]. This is the issue of the next section.

4.2   Generalization of the Gaussian Case

The Gaussian case discussed above can be summarized by two key insights:

1. Uncoded transmission (at the sensors) considerably outperforms any approach based on the separation paradigm (even for a relatively small number of sensors M).
2. In the limit as $M \to \infty$, uncoded transmission performs optimally.

In this section, both of these features are extended beyond the Gaussian case. We first establish a general sufficient condition for a given sensor coding system $(F_1, F_2, \ldots, F_M, G)$ to perform optimally, thus extending the second feature of the Gaussian case. Thereafter, we extend the first feature of the Gaussian case by suggesting a general coding paradigm as an alternative to the separation-based coding paradigm. We show that for a class of sensor networks that includes the Gaussian case, our coding paradigm, while not necessarily optimal, considerably outperforms the separation-based paradigm.

Optimal Performance. We now establish a general sufficient condition for the optimality of a sensor network with a topology according to Figure 1. It can be stated as follows.

Theorem 4. *If in the sensor network of Figure 1, the code $(F_1, F_2, \ldots, F_M, G)$ satisfies the cost constraint $E\rho(X_1, X_2, \ldots, X_M) \leq \Gamma$, and*

$$d(s, \hat{s}) = -c_2 \log_2 P(s|\hat{s}) + d_0(s) \tag{26}$$

$$I(S; U_1, U_2, \ldots, U_M) = I(S; \hat{S}), \tag{27}$$

then it performs optimally.

Proof. The proof works by idealizing the sensor network of Figure 1 by a point-to-point communication system. In particular, we consider the communication system where the final destination has direct access to the observations $U_1, \ldots, U_M$. If the code $(F_1, F_2, \ldots, F_M, G)$ achieves optimal performance in this point-to-point communication system, then it *must* achieve optimal performance in the original sensor network of Figure 1.

The conditions for optimality for the idealized point-to-point source-channel communication system can be stated as follows (see Appendix A and [5,6,7]): The point-to-point communication system is optimal if

$$\rho(s) = c_1 D(p_{U_1,U_2,\ldots,U_M|s}||p_{U_1,U_2,\ldots,U_M}) + \rho_0 \tag{28}$$

$$d(s,\hat{s}) = -c_2 \log_2 P(s|\hat{s}) + d_0(s) \tag{29}$$

$$I(S;\hat{S}) = I(S;U_1, U_2, \ldots, U_M), \tag{30}$$

where $D(\cdot||\cdot)$ denotes the Kullback-Leibler distance, see e.g. [4]. If the sensor network satisfies these conditions, it must perform optimally. However, in our problem, there is no cost constraint on the source signal S. In other words, $\rho(s)$ can always be chosen appropriately. Hence, the condition on $\rho(s)$ can be removed, which completes the proof.   □

The conditions of Theorem 4 are sufficient for optimality, but they are only achievable for a particular class of sensor networks; in the general case of Figure 1, they cannot be met. The goal of the following discussion is to illustrate the special class for which the conditions of Theorem 4 can be satisfied. We first illustrate these two issues for the Gaussian example studied above.

Example 5 (Gaussian case). Let us study conditions (26) and (27) for the Gaussian example. Consider first condition (26). Since $S[n]$ and $\hat{S}[n]$ are jointly Gaussian, we can write $S[n]$ in terms of $\hat{S}[n]$ as follows:

$$S[n] = \frac{E\left[S\hat{S}\right]}{E\left[\hat{S}^2\right]}\hat{S}[n] + W'[n], \tag{31}$$

where $W'[n]$ is additive white Gaussian noise. Since $\hat{S}$ is the minimum mean-squared error estimate of S, we find that $E[S\hat{S}]/E[\hat{S}^2] = 1$, and hence $S[n] = \hat{S}[n] + W'[n]$. This immediately reveals that $P(s|\hat{s})$ is given by the distribution of $W'[n]$. Since it is Gaussian, we find

$$-\log_2 P(s|\hat{s}) = d_1(s-\hat{s})^2 + d_0(s), \tag{32}$$

i.e., mean-squared error distortion. The variance of the noise $W'[n]$ is irrelevant for this argument; it only influences the constant d_1. This means that even for finite M, condition (26) is satisfied by the Gaussian example.

For condition (27) in our Gaussian example, equality is only achieved asymptotically as the number of sensors tends to infinity. We now analyze this in detail. From (19), $\hat{S}$ can be rewritten as

$$\hat{S}[n] = \frac{\frac{P}{\sigma_S^2+\sigma_W^2}M\sigma_S^2}{\frac{P}{\sigma_S^2+\sigma_W^2}(M^2\sigma_S^2 + M\sigma_W^2) + \sigma_Z^2}\left(MS[n] + \sum_{i=1}^{M}W_k[n] + \sqrt{\frac{\sigma_S^2+\sigma_W^2}{P}}Z[n]\right).$$

It follows immediately that

$$I(S;U_1, U_2, \ldots, U_M) = I(S;\hat{S}, U_1, U_2, \ldots, U_M)$$

$$= I(S;\hat{S}) + \sum_{k=1}^{M}I(S;U_k|\hat{S}, U_1, \ldots, U_{k-1}). \tag{33}$$

The next goal is to determine the first term in the sum, i.e., $I(S; U_1|\hat{S})$. This term is easily calculated by first replacing $\hat{S}$ by a scaled version $\hat{S}' = MS[n] + \sum_{k=1}^{M} W_k[n] + \sqrt{(\sigma_S^2 + \sigma_W^2)/P} Z[n]$. Using the shorthand

$$\beta = \frac{\sigma_S^2 + \sigma_W^2}{P} \frac{\sigma_Z^2}{\sigma_W^2}, \tag{34}$$

we can evaluate

$$I(S; U_1|\hat{S}) = I(S; U_1|\hat{S}')$$
$$= \frac{1}{2} \log \frac{(M + \beta)\sigma_S^2\sigma_W^2 \left((\sigma_S^2 + \sigma_W^2)(M^2\sigma_S^2 + (M + \beta)\sigma_W^2) - (M\sigma_S^2 + \sigma_W^2)^2\right)}{(M + \beta - 1)\sigma_S^2\sigma_W^4(M^2\sigma_S^2 + (M + \beta)\sigma_W^2)}$$
$$= \frac{1}{2} \log \frac{(M + \beta)\sigma_S^2\sigma_W^2((M^2 - M + \beta)\sigma_S^2\sigma_W^2 + M\sigma_W^4)}{(M + \beta - 1)\sigma_S^2\sigma_W^4(M^2\sigma_S^2 + (M + \beta)\sigma_W^2)}$$
$$= \frac{1}{2} \log \frac{M^3\sigma_S^2 + M^2((\beta - 1)\sigma_S^2 + \sigma_W^2) + M((\beta - 1)\sigma_S^2 + \beta\sigma_W^2) + \beta^2\sigma_S^2}{M^3\sigma_S^2 + M^2((\beta - 1)\sigma_S^2 + \sigma_W^2) + M(2\beta - 1)\sigma_W^2 + \beta(\beta - 1)\sigma_W^2}, \tag{35}$$

and hence, as $M \to \infty$,

$$I(S; U_1|\hat{S}) \to 0. \tag{36}$$

Note that this convergence is very rapid: the coefficients of both M^3 and M^2 are the same in the numerator and the denominator. A similar argument establishes that

$$I(S; U_k|\hat{S}, U_1, U_2, \ldots, U_{k-1}) \to 0, \tag{37}$$

as $M \to \infty$, hence

$$I(S; U_1, U_2, \ldots, U_M) \to I(S; \hat{S}), \tag{38}$$

hence condition (27) is satisfied in the limit as $M \to \infty$.

This short argument immediately reveals a class of extensions of the Gaussian example for which the conditions of Theorem 4 are also achievable, as follows:

Example 6 (Simple extension of Gaussian case). Suppose that the source S in Figure 2 is still Gaussian, that the observation noises W_k satisfy

$$\sum_{k=1}^{M} W_k \sim \mathcal{N}(0, M\sigma_W^2), \tag{39}$$

i.e., the sum of the observation noises is Gaussian and its variance grows linearly in the number of sensors M, and finally that the variance of W (see Figure 2) vanishes in comparison to the signal. Then, the same asymptotic behavior is observed.

Remark 4. This example is particularly interesting because only the sum of the observation noises has to be Gaussian: by the central limit theorem, this condition is satisfied in many practical cases as the number of sensors becomes large.

General Coding Paradigm. In the Gaussian case, the performance of the scheme of Theorem 1 cannot be argued to be optimal. Nevertheless, it is considerably superior to the performance of separate source and channel coding: In the latter, the distortion (as a function of the number of sensors) only decays like $1/\log M$, while in the former, it decays like $1/M$. Hence, even at relatively small M, the joint source-channel coding approach of Theorem 1 outperforms the separate source and channel coding.

This behavior can also be observed in terms of the conditions of Theorem 4: For the Gaussian case, condition (26), i.e.,

$$-\log_2 P(s|\hat{s}) = d_1(s - \hat{s})^2 + d_0(s), \tag{40}$$

was shown to be satisfied (for any M), while condition (27) was evaluated in (35) to be

$$
\begin{aligned}
&I(S; U_1, U_2, \ldots, U_M) \\
&= I(S; \hat{S}) + \frac{1}{2}\log\frac{M^3\sigma_S^2 + M^2\ldots}{M^3\sigma_S^2 + M^2\ldots} + \ldots.
\end{aligned}
\tag{41}
$$

This converges rapidly as M tends to infinity, and condition (27) is asymptotically satisfied.

These observations propose an alternative coding paradigm for sensor networks with a topology according to Figure 1, namely to code in such a way as to approach the conditions of Theorem 4 as closely as possible:

Coding Paradigm. *The goal of the coding scheme for sensor networks with a topology according to Figure 1 is to approach*

$$d(s, \hat{s}) = -c_2\log_2 P(s|\hat{s}) + d_0(s) \tag{42}$$
$$I(S; U_1, U_2, \ldots, U_M) = I(S; \hat{S}), \tag{43}$$

as closely as possible.

Remark 5. Note that *neither* the above coding paradigm *nor* the separation-based coding paradigm can be shown to lead to optimal performance in general sensor networks with a topology according to Figure 1. Recall that the optimal performance for the general case of Figure 1 is *not known* to date.

In our coding paradigm, the precise meaning of approaching the formulae of Theorem 4 "as closely as possible" is currently under investigation. For the Gaussian case studied in this paper, one such approaching behavior is achieved by the strategy of Theorem 1, as shown in Equation (41).

5   Communication between the Sensors

In the sensor network topology of Figure 1, the sensors can only communicate to the destination; they cannot communicate with each other. An interesting variation on the consideration of this paper is to allow the sensors to communicate

with each other. Our arguments can be used to obtain directly the following statement:

Theorem 5 (communication between the sensors). *Consider the sensor network of Figure 1, but allow now for communication between the sensors. If in this revised sensor network, the code $(F_1, F_2, \ldots, F_M, G)$ satisfies the cost constraint $E\rho(X_1, X_2, \ldots, X_M) \leq \Gamma$, and*

$$d(s, \hat{s}) = -c_2 \log_2 P(s|\hat{s}) + d_0(s) \tag{44}$$

$$I(S; U_1, U_2, \ldots, U_M) = I(S; \hat{S}), \tag{45}$$

then it performs optimally.

Proof. This follows again by the idealization used to prove Theorem 4: This idealization does include communication between the sensors.   □

Remark 6. For the general case involving communication between the sensors, we cannot compare to the separation-based code design: Its performance it unknown to date. However, it must be expected to perform suboptimally, in line with the arguments discussed above.

Remark 7. While in general, the possibility of the sensors to communicate with each other may be expected to enhance the performance, Theorem 5 establishes that for all sensor networks that satisfy Theorem 4, communication between the sensors *does not* improve the performance. This includes in particular our Gaussian example (Example 1). To emphasize the point, suppose that in the Gaussian example, each sensor is linked to every other sensor by an ideal cable. Then, not only can the sensors apply a much more efficient compression, but they can also act like a multiple-antenna transmitter, thus harvesting gains in capacity. Do we get a better performance than without the ideal cables between the sensors? Theorem 5 establishes that the answer is negative (asymptotically as the number of sensors $M \to \infty$): The uncoded transmission scheme of Theorem 1 achieves just the same performance. In other words, in this case, there is *no penalty* for the fact that the sensors are distributed, rather than joint.

6 Conclusions

In this paper, we analyzed a particular sensor network topology. We first derived the performance of a coding scheme designed according to the source-channel separation principle. For the considered Gaussian example, for instance, it was shown that the distortion decays like $1/\log M$, where M is the total number of sensors. Thereafter, we considered *joint* source-channel coding. The optimal performance and coding scheme is not known in general. We proposed an alternative coding paradigm and derived a class of sensor networks for which codes designed according to our paradigm achieve optimal performance. For the considered Gaussian example, it was shown that the distortion for a code according to our paradigm decays like $1/M$, i.e., considerably better than the separation-based scheme.

References

1. T. Berger. Multiterminal source coding. *Lectures presented at CISM Summer School on the Information Theory Approach to Communications*, July 1977.
2. T. Berger, Z. Zhang, and H. Viswanathan. The CEO problem. *IEEE Transactions on Information Theory*, IT–42:887–902, May 1996.
3. T. M. Cover, A. A. El Gamal, and M. Salehi. Multiple access channels with arbitrarily correlated sources. *IEEE Transactions on Information Theory*, 26(6):648–657, November 1980.
4. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, New York, 1991.
5. M. Gastpar. *To Code Or Not To Code*. PhD thesis, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland, 2002.
6. M. Gastpar, B. Rimoldi, and M. Vetterli. To code or not to code. In *Proc IEEE Int Symp Info Theory*, page 236, Sorrento, Italy, June 2000.
7. M. Gastpar, B. Rimoldi, and M. Vetterli. To code, or not to code: Lossy source-channel communication revisited. *submitted to IEEE Transactions on Information Theory*, May 2001. Revised July 2002.
8. M. Gastpar and M. Vetterli. On the capacity of large Gaussian relay networks. *submitted to IEEE Transactions on Information Theory*, September 2002.
9. M. Gastpar and M. Vetterli. On the capacity of wireless networks: The relay case. In *Proc IEEE Infocom 2002*, New York, NY, June 2002.
10. T. J. Goblick. Theoretical limitations on the transmission of data from analog sources. *IEEE Transactions on Information Theory*, IT–11(4):558–567, October 1965.
11. Y. Oohama. The rate-distortion function for the quadratic Gaussian CEO problem. *IEEE Transactions on Information Theory*, IT–44(3):1057–1070, May 1998.
12. C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. Journal*, 27:379–423, 623–656, 1948.
13. D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT–19:471–480, 1973.
14. H. Viswanathan and T. Berger. The quadratic Gaussian CEO problem. *IEEE Transactions on Information Theory*, IT–43(5):1549–1559, September 1997.
15. A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the receiver. *IEEE Transactions on Information Theory*, IT–22:1–11, January 1976.

A   Review: Point-to-Point Source-Channel Communication

In this section, we provide a brief review of the information-theoretic results for the point-to-point source-channel communication system, illustrated in Figure 3. The source is defined by a source distribution $P_S(s)$ and a distortion measure $d(s, \hat{s})$. The channel is defined by a conditional distribution $P_{Y|X}(y|x)$ and an input cost function $\rho(x)$. For the purpose of this brief review, we suppose that the encoder F maps a sequence of n source symbols onto a sequence of n channel input symbols. We also suppose that the decoder is synchronized with the encoder, and maps a sequence of n channel output symbols onto a sequence of

Fig. 3. The general point-to-point source-channel communication problem.

n source reconstruction symbols. The goal of the code (F, G) is to produce a minimum distortion,

$$\Delta = Ed\left(S^n, \hat{S}^n\right),\tag{46}$$

using, simultaneously, a minimum power (or more generally, cost) on the channel,

$$\Gamma = E\rho(X^n).\tag{47}$$

The key problem of source-channel communication is to determine the optimal cost-distortion pairs (Γ, Δ). We consider this problem in the information-theoretic sense, i.e., we are interested in the optimum *irrespective of the coding complexity and delay.*

Shannon's separation theorem determines the optimal trade-off between cost and distortion by the condition

$$R(\Delta) = C(\Gamma).\tag{48}$$

For a more detailed treatment, see e.g. [5, Ch. 1]. By the operational meaning of the rate-distortion and capacity-cost function, this simultaneously furnishes a coding paradigm, i.e., a way to implement the optimal coding (F, G). Hence, the communication system is optimal if it satisfies a *rate-matching* condition: the minimum rate for the source compression (the rate-distortion function) must be equal to the maximum rate for the channel code (the capacity-cost function).

Recently, an alternative perspective has been presented [5,6,7]. The optimal trade-off satisfies

$$\rho(x^n) = c_1 D(P_{Y^n|x^n}||P_{Y^n}) + \rho_0\tag{49}$$
$$d(s^n, \hat{s}^n) = -c_2 \log_2 P_{S^n|\hat{S}^n} + d_0(s^n)\tag{50}$$
$$I(S^n; \hat{S}^n) = I(X^n; Y^n),\tag{51}$$

where $c_1 \geq 0, c_2 \geq 0$ and ρ_0 are constants, $d_0(s)$ is an arbitrary function of s, and $D(\cdot||\cdot)$ denotes the Kullback-Leibler distance, see e.g. [4]. Hence, the communication system is optimal if it satisfies a *measure-matching* condition: the probability measures of the source and the channel and the cost and distortion measure must be matched in the right way by the coding system.

In this paper, we extend both these perspectives to the case of the considered sensor network. Previously, we have also applied our measure-matching perspective to obtain capacity results for relay networks [9,8].