Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network

Michael Gastpar
University of California, Berkeley
Wireless Foundations Center, Dept. of EECS, Berkeley, CA 94720-1770, USA
Email: gastpar@berkeley.edu

***Abstract*— One of the simplest sensor network models has one single underlying Gaussian source of interest, observed by many sensors, subject to independent Gaussian observation noise. The sensors communicate over a standard Gaussian multiple-access channel to a fusion center whose goal is to estimate the underlying source with respect to mean-squared error. In this note, a theorem of Witsenhausen is shown to imply that an optimal communication strategy is uncoded transmission, i.e., each sensors' channel input is merely a scaled version of its noisy observation.**

I. Introduction

In this note, we show how a result of Witsenhausen can be used to close a small gap in an argument that we presented earlier concerning a certain simple Gaussian "sensor" network: a single underlying Gaussian "source," i.e., a sequence of independent and identically distributed (i.i.d.) Gaussian random variables, is observed in a memoryless fashion by M sensors, subject to white Gaussian noise. The sensors are connected to a fusion center via a standard Gaussian multiple access channel. The fusion center needs to recover the underlying source sequence to within the smallest mean squared error possible.

The interesting insight concerning this network is that the standard digital communication approach leads to an exponentially suboptimal performance. To put this in context, it is well known that digital communication (in the information-theoretic sense) does not lead to optimal performance in general network problems, see e.g. [1], [2]. One may be tempted to suspect that the performance gap due to digital communication strategies is negligible. However, the above simple sensor network example has shown that the gap may be of *scaling-law relevance*, more precisely, it is *exponential* in the number of users.

The key question, then, is ow to design more general communication strategies, beyond the well studied digital algorithms. Unfortunately, the tools to develop and analyze optimal strategies for such networks are largely non-existent. As a matter of fact, the "coding" scheme used to show that digital communication is suboptimal in a scaling sense is *uncoded transmission*, meaning that each sensor merely transmits a signal *proportional* to its observations, without any further coding. Not only could this scheme be shown to perform exponentially superior to the best digital communication scheme, it could also be shown to perform *optimally* in a scaling-law sense [3], [4], [5].

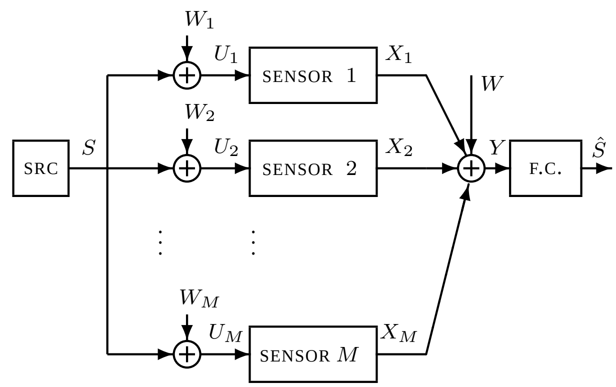

Fig. 1. For this simple Gaussian "sensor" network, uncoded transmission is shown to be exactly optimal: The fusion center (F. C.) needs to reconstruct S, i.e., the underlying source (SRC). Each sensor must satisfy a transmit power constraint of P.

In this sense, the crown jewels are gone, but the previous work left open the question whether uncoded transmission is strictly optimal rather than only in a scaling-law sense. In this note, we show how a result of Witsenhausen can be used to answer this question in the affirmative, at least for some cases.

II. The Sensor Network Model

The "sensor" network considered in this paper is illustrated in Figure 2, and the special case for which we establish the exact optimality is illustrated in Figure 1. The underlying source $\{S[n]\}_{n>0}$ is a sequence of independent and identically distributed (i.i.d.) real-valued Gaussian random variables of mean zero and variance σ_S^2. Sensor m observes a sequence $\{U_m[n]\}_{n>0}$ defined as

$$U_m[n] \quad = \quad \alpha_m S[n] + W_m[n], \qquad (1)$$

where $\alpha_m \geq 0$ are fixed and known constants (for the purpose of this note; more general cases are analyzed in [5]), and $\{W_m[n]\}_{n>0}$ is a sequence of i.i.d. Gaussian random variables of mean zero and variance σ_W^2. Note that therefore, the assumption $\alpha_m \geq 0$ is without loss of generality. Sensor m can apply an *arbitrary* coding function to the observation sequence such as to generate a sequence of channel inputs, $\{X_m[n]\}_{n>0} = f_m(\{U_m[n]\}_{n>0})$. The only constraint is that

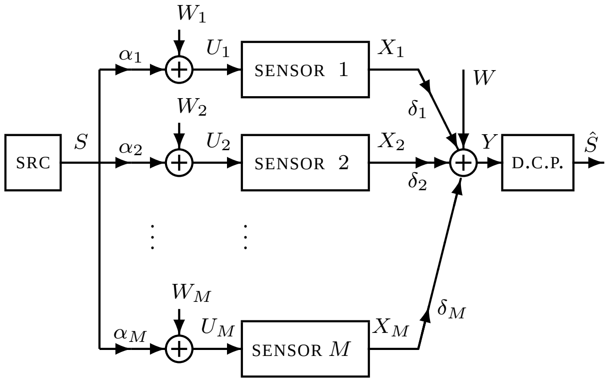

Fig. 2. The considered sensor network: The data collection point (D. C. P.) needs to reconstruct S, i.e., the underlying source (SRC).

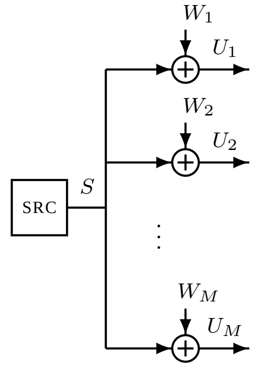

Fig. 3. The CEO source.

the function $f_m(\cdot)$ be chosen to ensure that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} E\left[(X_m[n])^2\right] \quad \leq \quad P_m. \tag{2}$$

The channel output is then given as

$$Y[n] \quad = \quad Z[n] + \sum_{m=1}^{M} \delta_m X_m[n], \tag{3}$$

where $\delta_m \geq 0$ are fixed and known constants (for the purpose of this note; more general cases are analyzed in [5]), and $\{Z[n]\}_{n>0}$ is an i.i.d. sequence of Gaussian random variables of mean zero and variance σ_Z^2. Note that the assumption $\delta_m \geq 0$ is without loss of generality. Upon observing the channel output sequence $\{Y[n]\}_{n>0}$, the decoder (or fusion center) must produce a sequence $\{\hat{S}[n]\}_{n>0} = g(\{Y[n]\}_{n>0})$, and we consider the distortion

$$D \quad = \quad \lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} E\left[(S[n]-\hat{S}[n])^2\right]. \tag{4}$$

This note concerns the determination of the smallest attainable distortion D, for fixed power constraints $P = P_1 = P_2 = \cdots = P_M$, over all possible encoding and decoding functions, $f_m(\cdot)$ and $g(\cdot)$.

III. Uncoded Transmission

In this section, we merely quote the results obtained previously in [3], [4], [5]. Specifically, one simple strategy is for each sensor to transmit

$$X_m[n] \quad = \quad \sqrt{\frac{P_m}{\alpha_m \sigma_S^2 + \sigma_W^2}}\, U_m[n]. \tag{5}$$

It is easily verified that this satisfies the power constraint (Equation (2)). We refer to this communication strategy as *uncoded transmission.* The performance of this simple scheme can be evaluated by straightforward calculations. We summarize the result in the following theorem.

Theorem 1: For the Gaussian "sensor" network with $\alpha_m = \delta_m = 1$ for $m = 1, 2, \ldots, M$, using *uncoded transmission* leads to the following distortion:

$$D = \frac{\sigma_S^2\sigma_W^2}{M\sigma_S^2 + \sigma_W^2}\left(1 + \frac{M(\sigma_S^2\sigma_Z^2/\sigma_W^2)}{\frac{M\sigma_S^2+\sigma_W^2}{\sigma_S^2+\sigma_W^2}P_{tot} + \sigma_Z^2}\right). \tag{6}$$

For a proof, see [3], [4], [5].

IV. Digital Communication

Although this is somewhat tangential to the main result presented in this note, we also briefly quote the best performance attainable via a digital strategy. More precisely, such a strategy enforces a strict separation between source compression and channel coding in the sense that the channel code merely transmits bits without making errors (in the limit of long block length). The maximum number of bits that can be transported by the considered multiple-access channel is upper bounded by

$$R_{tot} \quad \leq \quad \frac{1}{2}\log_2\left(1 + \frac{MP_{tot}}{\sigma_Z^2}\right). \tag{7}$$

Moreover, if we have a total rate of R_{tot} available to the M sensors, then the smallest distortion we can attain follows from the so-called *CEO problem*. This problem was introduced in [6], [7] and the quadratic Gaussian version described above was solved by Oohama [8], with some recent refinements [9]. From this work, the *sum* rate (i.e., the total rate over all M encoders) in order to achieve a certain distortion D is determined as

$$R(D) = \log_2^+\left(\frac{\sigma_S^2}{D}\left(\frac{D\sigma_S^2 M}{D\sigma_S^2 M - \sigma_S^2\sigma_W^2 + D\sigma_W^2}\right)^M\right). \tag{8}$$

For the purpose of this paper, we use Oohama's simpler lower bound, which can be obtained easily from the above, noting that $\sigma_S^2/D \geq 1$,

$$R(D) \geq M\log_2^+\left(\frac{D\sigma_S^2 M}{D\sigma_S^2 M - \sigma_S^2\sigma_W^2 + D\sigma_W^2}\right). \tag{9}$$

Conversely, the smallest achievable distortion satisfies

$$D(R) \geq \frac{\sigma_S^2 \sigma_W^2}{\sigma_S^2 M \left(1 - 2^{-R/M}\right) + \sigma_W^2}. \quad (10)$$

By noting that $1 - 2^{-R/M} \leq R/M$, this implies the lower bound

$$D(R) \geq \frac{\sigma_S^2 \sigma_W^2}{\sigma_S^2 R + \sigma_W^2}. \quad (11)$$

Combining this with Equation (7) leads to the following theorem.

Theorem 2: For the Gaussian "sensor" network with $\alpha_m = \delta_m = 1$ for $m = 1, 2, \ldots, M$, using *digital communication* incurs a distortion of at least

$$D \geq \frac{\sigma_S^2 \sigma_W^2}{\sigma_S^2 \log_2 \left(1 + \frac{M P_{tot}}{\sigma_Z^2}\right) + \sigma_W^2}. \quad (12)$$

The scaling-law insight referred to in the introduction is precisely the difference between Equation (6) and Equation (12): The former gives a distortion that decays like $1/M$ while the latter gives one that decays like $1/\log(M)$.

V. A Simple Cut-set Bound

In this section, we merely quote the simple bound used previously in [3], [4], [5]. This lower bound to the distortion is slightly better than what uncoded transmission (Theorem 1) shows to be achievable, and therefore leaves open the question of whether uncoded transmission performs exactly optimally. We state the result in the following theorem, and then provide a proof outline merely in preparation for the main result of this note.

Theorem 3: For the Gaussian "sensor" network with $\alpha_m = \delta_m = 1$ for $m = 1, 2, \ldots, M$, using *any* scheme, the incurred distortion must satisfy

$$D \geq \frac{\sigma_S^2 \sigma_W^2}{M\sigma_S^2 + \sigma_W^2} \left(1 + \frac{M(\sigma_S^2 \sigma_Z^2 / \sigma_W^2)}{M P_{tot} + \sigma_Z^2}\right). \quad (13)$$

Proof outline: We outline the key steps of the proof in order to explain the contrast to the tighter bound presented below in Theorem 5.

Consider *any* mapping that attains an average distortion D (in the sense of Equation (4)) and uses average powers $P_1, P_2, \ldots, P_M$ (in the sense of Equation (2)). Clearly, we must have (by the data processing inequality)

$$I(U_1^N, U_2^N, \ldots, U_M^N; \hat{S}^N) \leq I(X_1^N, X_2^N, \ldots, X_M^N; Y^N), \quad (14)$$

where we use the notational shorthand $U_m^N = (U_m[1], U_m[2], \ldots, U_m[N])$ (and likewise for $X_m^N, \hat{S}^N$, and Y^N) for sequences of random variables. Since our mapping attains an average distortion D, the converse to the rate-distortion theorem implies that

$$\frac{1}{N} I(U_1^N, U_2^N, \ldots, U_M^N; \hat{S}^N) \geq \min I(U_1, U_2, \ldots, U_M; \hat{S}^N), \quad (15)$$

where the minimum is over all conditional distributions $p(\hat{s}|u_1, \ldots, u_M)$ satisfying $E[(S - \hat{S})^2] \leq D$. For our simple case, the minimum can be evaluated in closed form as

$$\frac{1}{2} \log \frac{\sigma_S^2}{D} + \frac{1}{2} \log \frac{M\sigma_S^2}{M\sigma_S^2 + \sigma_W^2 - \frac{\sigma_S^2}{D}\sigma_W^2}. \quad (16)$$

The goal is now to upper bound the right hand side. By the fact that the channel is memoryless, we find along standard arguments that

$$I(X_1^N, X_2^N, \ldots, X_M^N; Y^N) \leq \sum_{n=1}^{N} I(X_1[n], X_2[n], \ldots, X_M[n]; Y[n]). \quad (17)$$

Since the considered mapping uses average powers $P_1, P_2, \ldots, P_M$, we can further bound

$$\sum_{n=1}^{N} I(X_1[n], X_2[n], \ldots, X_M[n]; Y[n]) \leq \max \sum_{n=1}^{N} I(X_1[n], X_2[n], \ldots, X_M[n]; Y[n]), \quad (18)$$

where the maximum is over all joint distributions such that $\frac{1}{n}\sum_{n=1}^{N} P_{m,n} \leq P_m$, where we define

$$P_{m,n} \stackrel{\text{def}}{=} E[(X_m[n])^2]. \quad (19)$$

Now, we can relax this and maximize over all distributions for which

$$\sum_{m=1}^{M} \frac{1}{n} \sum_{n=1}^{N} P_{m,n} \leq \sum_{m=1}^{M} P_m \stackrel{\text{def}}{=} P_{tot}. \quad (20)$$

The latter maximum is easily found to be

$$\sum_{n=1}^{N} I(X_1[n], \ldots, X_M[n]; Y[n]) \leq \frac{1}{2} \log_2 \left(1 + \frac{M P_{tot}}{\sigma_Z^2}\right),$$

which implies the bound.

VI. Witsenhausen's Bound

Prior art, as summarized in Theorems 1 and 3, establishes upper and lower bounds that match in a scaling-law sense, but leave open a small gap. The contribution of this note is to close this gap by tightening the lower bound to the distortion, i.e., Theorem 3. In other words, this note establishes that uncoded transmission is *exactly* optimal for the considered simplistic Gaussian sensor network.

The key result is the following lemma, due to Witsenhausen [10].

Lemma 4: Consider two sequences $U_1[n]$ and $U_2[n]$ sampled i.i.d. from the joint distribution $p(u_1, u_2)$, and two arbitrary real-valued functions $f_1(\cdot)$ and $g_1(\cdot)$ satisfying

$$E[f_1(U_1)] = E[g_1(U_2)] = 0, \quad (21)$$

$$E[f_1^2(U_1)] = E[g_1^2(U_2)] = 1. \quad (22)$$

Define

$$\rho^* = \sup_{f_1, g_1} E[f_1(U_1) g_1(U_2)]. \quad (23)$$

Then, for any real-valued functions $f_N(\cdot)$ and $g_N(\cdot)$ satisfying

$$E[f_N(U_1^N)] = E[g_N(U_2^N)] = 0, \quad (24)$$
$$E[f_N^2(U_1^N)] = E[g_N^2(U_2^N)] = 1, \quad (25)$$

we have that

$$\sup_{f_N, g_N} E[f_N(U_1^N) g_N(U_2^N)] \leq \rho^*. \quad (26)$$

Remark 1: The quantity ρ^* is also known as the "maximum correlation coefficient" of the random variables U_1 and U_2 in the literature.

This lemma can be used to establish the converse bound given in the following theorem.

Theorem 5: For the Gaussian "sensor" network with $\alpha_m = \delta_m = 1$ and $P_m = P_{tot}/M$ for $m = 1, 2, \ldots, M$, using *any* scheme, the incurred distortion must satisfy

$$D \geq \frac{\sigma_S^2 \sigma_W^2}{M\sigma_S^2 + \sigma_W^2} \left(1 + \frac{M(\sigma_S^2 \sigma_Z^2 / \sigma_W^2)}{\frac{M\sigma_S^2 + \sigma_W^2}{\sigma_S^2 + \sigma_W^2} P_{tot} + \sigma_Z^2} \right). \quad (27)$$

The proof of this theorem is outlined in the appendix.

Remark 2: Hence, the main conclusion of this note is that for the simple Gaussian "sensor" network shown in Figure 1, "uncoded transmission" by the sensors, as characterized in Theorem 1, is *exactly* optimal.

VII. Extensions

To keep things as simple and insightful as possible, we have concentrated on the simplest possible configuration, shown in Figure 1. The obvious next question concerns the scenario illustrated in Figure 2: Some sensor measurements are more valuable than others, and some sensors have a stronger channel to the fusion center than others. The simple cut-set approach discussed above is particularly weak in this setting due to its inability to take into account cases where good measurements are not aligned with the strong channels.

The following distortion can be shown to be achievable:

Theorem 6: For the single-source Gaussian sensor network discussed in this section and illustrated in Figure 2, the following distortion is achievable:

$$D = \frac{\sigma_S^2 \sigma_W^2}{\sigma_W^2 + \sum_{m=1}^M |\alpha_m|^2 \sigma_S^2} \cdot \left(1 + \frac{(\sigma_S^2 \sigma_Z^2 / \sigma_W^2) \sum_{m=1}^M |\alpha_m|^2}{\sigma_Z^2 + P_{tot}(M) b(M)} \right), \quad (28)$$

where

$$b(M) = \frac{\left(\sigma_W^2 + \sum_{m=1}^M |\alpha_m|^2 \sigma_S^2 \right) \sum_{m=1}^M |\alpha_m|^2}{\sum_{m=1}^M (|\alpha_m|^2 \sigma_S^2 + \sigma_W^2) |\alpha_m|^2 / |\delta_m|^2}, \quad (29)$$

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.

A proof of this theorem is given in [5].

By contrast, the simple cut-set approach derived in Section V above yields the following upper bound on performance (and thus, lower bound to the distortion).

Theorem 7: For the single-source Gaussian sensor network discussed in this section and illustrated in Figure 2, the achievable distortion is lower bounded by

$$D \geq \frac{\sigma_S^2 \sigma_W^2}{\sigma_W^2 + \sum_{m=1}^M |\alpha_m|^2 \sigma_S^2} \cdot \left(1 + \frac{(\sigma_S^2 \sigma_Z^2 / \sigma_W^2) \sum_{m=1}^M |\alpha_m|^2}{\sigma_Z^2 + P_{tot}(M) \sum_{m=1}^M |\delta_m|^2} \right), \quad (30)$$

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.

A proof of this theorem is given in [5].

Again, we note that the difference between the two is rather small. The Witsenhausen bound can be used to tighten the gap somewhat, and a general bound is given in Equation (49) in the appendix. Note, however, that this bound is not sufficient to prove that uncoded transmission is "exactly" optimal in the case of general α_m, δ_m and P_m, for $m = 1, 2, \ldots, M$.

VIII. Received-power Constraints

Let us reconsider the model of Figure 2, as defined in Section II. We are now removing the transmit power constraint, i.e., Condition (2). Instead, the only constraint is that the functions $f_m(\cdot)$, for $m = 1, 2, \ldots, M$, be chosen to ensure that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^N E\left[(Y[n])^2 \right] \leq Q. \quad (31)$$

This is a received-power constraint, as considered in [11]. For this setting, uncoded transmission can be shown to be exactly optimal for more general choices of parameters. Specifically, the following theorem is proved in [11].

Theorem 8: For the Gaussian "sensor" network with $\alpha_m = 1$ for $m = 1, 2, \ldots, M$, and received-power constraint (as in Equation (31)) using *any* scheme, the smallest distortion is attained by uncoded transmission and is given by

$$D = \frac{\sigma_S^2 \sigma_W^2}{M\sigma_S^2 + \sigma_W^2} \left(1 + \frac{M\sigma_S^2}{\sigma_W^2 \left(1 + \frac{Q}{\sigma_Z^2} \right)} \right). \quad (32)$$

It is also interesting to note that this theorem continues to hold even if arbitrary causal feedback is available at some or all of the transmitters, and under arbitrary models of transmitter cooperation (i.e., arbitrary "free" additional links between sensors). This is explained in more detail in [11].

IX. Some Concluding Remarks

In this note, we have closed a small gap in our previous arguments, using an insight due to Witsenhausen. This concerns a joint source-channel coding problem in a Gaussian "sensor" network. The improved upper bound shows that uncoded transmission is not only scaling-law optimal; it is exactly optimal.

Clearly, it will be interesting to investigate the significance of the Witsenhausen bound for more general source-channel communication networks. A partial account of this was given by Kang and Ulukus in [12]. As shown in the proof of Theorem 5, one way of using the bound is in the shape of an "enhanced" cut-set bound: To bound the mutual information across the channel, in the regular cut-set bound, *all* joint distributions of the channel input signals must be considered. The Witsenhausen argument permits to limit the class of joint distributions in a non-trivial way: The maximum amount of correlation between the channel inputs can be bounded in terms of the "maximum correlation coefficient" of the underlying source signals. In the Gaussian case considered in this paper, this is a simple matter as the maximum correlation coefficient is merely the standard correlation coefficient. More generally, finding the maximum correlation coefficient is a non-trivial issue for which algorithms have been devised (see e.g. the ACE algorithm in [13]).

It is perhaps interesting to point out a connection to another method of understanding the ("exact") optimality of uncoded transmission, as proposed in [14], [15] and sometimes referred to as "measure-matching." In this perspective, we *fix* the encoder to be uncoded transmission (as in Equation (5)), and the decoder to be the corresponding minimum mean-squared error estimate. The measure-matching approach then determines the channel input cost function as well as the distortion measure for which the considered encoder and decoder are information-theoretically optimal. Straightforward evaluation of the measure-matching conditions as given in [14], [15] leads, not surprisingly, to the mean-squared error distortion measure. Slightly less trivially, the cost function is found to be the received-power constraint discussed in Section VIII, i.e., Equation (31).

Acknowledgment

The author is indebted to Prof. Amos Lapidoth (ETH Zürich, Switzerland) for making him aware of the result by Witsenhausen [10] via the analysis of a related problem [16].

Appendix

Proof: [Proof of Theorem 5] For the purpose of this proof, we will first consider the case of general values of α_m, δ_m and P_m, and then specialize our results to the claimed scenario. We consider *any* mapping of length N that attains an average distortion D (in the sense of Equation (4)) and uses average powers $P_1, P_2, \ldots, P_M$ (in the sense of Equation (2)). Clearly, we must have (by the data processing inequality)

$$I(U_1^N, U_2^N, \ldots, U_M^N; \hat{S}^N) \leq I(X_1^N, X_2^N, \ldots, X_M^N; Y^N). \quad (33)$$

As shown in the proof of Theorem 3, the left hand side can be lower bounded by the remote rate-distortion function (see e.g. [17])

$$\begin{aligned} & I(U_1^N, U_2^N, \ldots, U_M^N; \hat{S}^N) \\ & \quad \geq \quad \frac{1}{2}\log\frac{\sigma_S^2}{D} + \frac{1}{2}\log\frac{A\sigma_S^2}{A\sigma_S^2 + \sigma_W^2 - \frac{\sigma_S^2}{D}\sigma_W^2}, \end{aligned} \quad (34)$$

where $A = \sum_{m=1}^M \alpha_m^2$. The new argument is used to bound the right hand side of Equation (33) more tightly. We start as in the proof of Theorem 3 by arguing that

$$\begin{aligned} & I(X_1^N, X_2^N, \ldots, X_M^N; Y^N) \\ & \quad \leq \quad \max \sum_{n=1}^N I(X_1[n], X_2[n], \ldots, X_M[n]; Y[n]), \end{aligned} \quad (35)$$

where the maximum is over all joint distributions such that $\frac{1}{n}\sum_{n=1}^N P_{m,n} \leq P_m$, where $P_{m,n}$ is as defined in Equation (19). The maximizing distribution is Gaussian, therefore,

$$\begin{aligned} & \sum_{n=1}^N I(X_1[n], X_2[n], \ldots, X_M[n]; Y[n]) \\ & \quad = \quad \sum_{n=1}^N \frac{1}{2}\log(1 + \frac{1}{\sigma_Z^2}\delta^T\Sigma_n\delta), \end{aligned} \quad (36)$$

where $\delta = (\delta_1, \ldots, \delta_M)^T$ and Σ_n is a matrix whose entry in row m and column m' is given by

$$\{\Sigma_n\}_{m,m'} \quad = \quad E\left[X_m[n]X_{m'}[n]\right]. \quad (37)$$

Now, we note that

$$X_m[n] \quad = \quad f_{m,n}(U_m^N) \quad (38)$$

for any function $f_{m,n}(\cdot)$ satisfying

$$E[f_{m,n}(U_m^N)] = 0 \quad \text{and} \quad E[f_{m,n}^2(U_m^N)] = P_{m,n},$$

where the sufficiency of mappings satisfying the first condition can be shown by standard arguments, and the second condition is merely the definition of $P_{m,n}$ as in Equation (19). By Lemma 4, we have that

$$E[f_{m,n}(U_m^N)f_{m',n}(U_{m'}^N)] \quad \leq \quad \rho_{m,m'}^*\sqrt{P_{m,n}P_{m',n}}, \quad (39)$$

where

$$\rho_{m,m'}^* \quad \stackrel{\text{def}}{=} \quad \sup_{f_1, g_1} E[f_1(U_m)g_1(U_{m'})]. \quad (40)$$

It can be shown that when U_m and $U_{m'}$ are jointly Gaussian, the supremum in the above expression is attained by choosing

$$f_1(U_m) \quad = \quad \frac{U_m}{\sqrt{\alpha_m\sigma_S^2 + \sigma_W^2}} \quad \text{and} \quad (41)$$

$$g_1(U_{m'}) \quad = \quad \frac{U_{m'}}{\sqrt{\alpha_{m'}\sigma_S^2 + \sigma_W^2}}. \quad (42)$$

A proof of this well-known fact can be found, e.g., in [18]. It is perhaps important to point out that beyond the Gaussian case, the optimal functions $f_1(\cdot)$ and $g_1(\cdot)$ are not always as simple. This has been studied in much detail, see e.g. the ACE algorithm in [13]. For our scenario, we now find

$$\begin{aligned} \rho^*_{m,m'} &= \frac{E[U_m U_{m'}]}{\sqrt{E[U_m^2]E[U_{m'}^2]}} \quad (43) \\ &= \frac{E[(\alpha_m S + W_m)(\alpha_{m'} S + W_{m'})]}{\sqrt{E[(\alpha_m S + W_m)^2]E[(\alpha_{m'} S + W_{m'})^2]}} \\ &= \frac{\alpha_m \alpha_{m'} \sigma_S^2}{\sqrt{(\alpha_m \sigma_S^2 + \sigma_W^2)(\alpha_{m'} \sigma_S^2 + \sigma_W^2)}}. \quad (44) \end{aligned}$$

To continue, we define $\overline{\Sigma} = \frac{1}{N}\sum_{n=1}^N \Sigma_n$, and note that $\overline{\Sigma}$ also has its correlation coefficients bounded by $\rho^*_{m,m'}$. The concavity of the logarithm implies that

$$\begin{aligned} &\sum_{n=1}^N I(X_1[n], X_2[n], \ldots, X_M[n]; Y[n]) \\ &\leq \frac{1}{2}\log(1 + \frac{1}{\sigma_Z^2}\delta^T \overline{\Sigma}\delta) \leq \max \frac{1}{2}\log(1 + \frac{1}{\sigma_Z^2}\delta^T \overline{\Sigma}\delta), \end{aligned} \quad (45)$$

where the maximum is over all covariance matrices $\overline{\Sigma}$ satisfying

$$\{\overline{\Sigma}\}_{m,m} \leq \frac{1}{N}\sum_{n=1}^N P_{m,n} \leq P_m \quad (46)$$

$$\begin{aligned} \{\overline{\Sigma}\}_{m,m'} &\leq \rho^*_{m,m'} \frac{1}{N}\sum_{n=1}^N \sqrt{P_{m,n}P_{m',n}} \\ &\leq \rho^*_{m,m'}\sqrt{P_m P_{m'}}, \text{ for } m \neq m'. \end{aligned} \quad (47)$$

Since the expression inside the logarithm in Equation (45) is merely a weighted sum (with non-negative coefficients) of the entries of the matrix $\overline{\Sigma}$, it is clear that the maximizing covariance matrix is simply the one that attains these two constraints with equality. Hence, we find the following rate upper bound:

$$\begin{aligned} &I(X_1^N, X_2^N, \ldots, X_M^N; Y^N) \\ &\leq \frac{1}{2}\log\left(1 + \frac{1}{\sigma_Z^2}\left(\sum_{m=1}^M \delta_m^2 P_m \right.\right. \\ &\left.\left. + 2\sum_{m=1}^M \sum_{m'=m+1}^M \rho^*_{m,m'}\delta_m\delta_{m'}\sqrt{P_m P_{m'}}\right)\right) \end{aligned} \quad (48)$$

To conclude the proof, let us denote $R = I(U_1^N, U_2^N, \ldots, U_M^N; \hat{S}^N)$ and invert the relationship (34) to obtain

$$D \geq \frac{\sigma_S^2}{1 + \frac{A\sigma_S^2}{\sigma_W^2}} + \frac{\sigma_S^2 2^{-2R}}{1 + \frac{\sigma_W^2}{A\sigma_S^2}}. \quad (49)$$

However, R is upper bounded by Equation (48). Combining this with Equation (49) gives a general lower bound on D.

For reasons of space and insight, we will not write out the general bound here. Instead, let us now specialize to the case $\alpha_m = \delta_m = 1$ and $P_m = P = P_{tot}/M$. Then, we find

$$\rho^*_{m,m'} = \rho^* \stackrel{\text{def}}{=} \frac{\sigma_S^2}{\sigma_S^2 + \sigma_W^2}, \quad (50)$$

and thus, Equation (48) evaluates to

$$\begin{aligned} &I(X_1^N, X_2^N, \ldots, X_M^N; Y^N) \\ &\leq \frac{1}{2}\log\left(1 + \frac{P_{tot}\frac{M\sigma_S^2 + \sigma_W^2}{\sigma_S^2 + \sigma_W^2}}{\sigma_Z^2}\right). \end{aligned} \quad (51)$$

Substituting this expression for R in Equation (49), noting that $A = M$, and rearranging terms gives the claimed lower bound. ■

REFERENCES

[1] T. M. Cover, A. A. E. Gamal, and M. Salehi, “Multiple access channels with arbitrarily correlated sources,” *IEEE Transactions on Information Theory*, vol. 26, pp. 648–657, November 1980.

[2] R. Ahlswede and T. S. Han, “On source coding with side information via a multiple-access channel and related problems in multi-user information theory,” *IEEE Transactions on Information Theory*, vol. IT–29, pp. 396–412, May 1983.

[3] M. Gastpar and M. Vetterli, “On the capacity of wireless networks: The relay case,” in *Proc IEEE Infocom 2002*, vol. 3, (New York, NY), pp. 1577 – 1586, June 2002.

[4] 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.

[5] M. Gastpar and M. Vetterli, “Power, spatio-temporal bandwidth, and distortion in large sensor networks,” *IEEE Journal on Selected Areas in Communications (Special Issue on Self-Organizing Distributive Collaborative Sensor Networks)*, vol. 23, pp. 745–754, April 2005.

[6] T. Berger, Z. Zhang, and H. Viswanathan, “The CEO problem,” *IEEE Transactions on Information Theory*, vol. IT–42, pp. 887–902, May 1996.

[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] J. Chen, X. Zhang, T. Berger, and S. Wicker, “An upper bound on the sum-rate distortion function and its corresponding rate allocation schemes for the CEO problem,” *IEEE Journal on Selected Areas in Communications*, vol. 22, pp. 977–987, Aug 2004.

[10] H. S. Witsenhausen, “On sequences of pairs of dependent random variables,” *SIAM J. Appl. Math.*, vol. 28, pp. 100–113, January 1975.

[11] M. Gastpar, “On capacity under receive and spatial spectrum-sharing constraints,” *IEEE Transactions on Information Theory*, vol. 53, February 2007. To appear.

[12] W. Kang and S. Ulukus, “An outer bound for multiple access channels with correlated sources,” in *40th Conf. on Information Sciences and Systems (CISS)*, (Princeton, NJ), March 2006.

[13] L. Breiman and J. Friedman, “Estimating optimal transformations for multiple regression and correlation,” *Journal of the American Statistical Association*, vol. 80, pp. 580–598, September 1985.

[14] 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.

[15] M. Gastpar, *To Code Or Not To Code*. PhD thesis, Ecole Polytechnique Fédérale (EPFL), Lausanne, Switzerland, 2002.

[16] A. Lapidoth and S. Tinguely, “Sending a bi-variate Gaussian source over a Gaussian MAC,” in *Proc IEEE Int Symp Info Theory*, (Seattle, WA), July 9-14 2006.

[17] M. Gastpar, “A lower bound to the AWGN remote rate-distortion function,” in *Proc IEEE Statistical Signal Processing Workshop*, (Bordeaux, France), pp. 1176–1181, July 2005.

[18] H. O. Lancaster, “Some properties of the bivariate normal distribution considered in the form of a contingency table,” *Biometrika*, vol. 44, pp. 289 – 292, 1957.