

# Fidelity and Resource Sensitive Data Gathering

Ameesh Pandya, Aman Kansal, Gregory Pottie, and Mani Srivastava

Dept. of Electrical Engineering  
University of California, Los Angeles  
Los Angeles, CA 90024  
{ameesh, kansal, pottie, mbs}@ee.ucla.edu

## Abstract

Sensor networks collect data at multiple distributed nodes and transfer the acquired information to points of interest. The raw data collected by each individual sensor is typically not of interest. Instead, a reduced representation of the measured phenomenon is to be generated. Multiple readings, however, add to the information about the phenomenon by providing its description at multiple points in space for distributed phenomena and multiple perspectives for a localized phenomenon. We also note that sensor readings have noise, and multiple readings can help mitigate the effect of this noise. Thus, while all the sensor readings need not be communicated, enough data must be exchanged to reliably reproduce the phenomenon. Considering the above effects, it becomes important to determine how much data should be transmitted from multiple sensors such that only useful information is exchanged and energy or bandwidth are not wasted on redundant data. We address this question using information theoretic techniques. The effects of sensor noise and correlation in the sensor readings are explicitly modelled.

## I. INTRODUCTION

In this paper we consider the process of data gathering in sensor networks, which is the key functionality of such systems. The common process, which underlies most sensor networking applications, can be viewed as follows. A phenomenon of interest exists in the environment within the sensing range of the deployed system. Multiple sensors collect readings about this phenomenon, which are subject to noise in the sensor transducers. The sensed data is now communicated to points of interest. The sensors may communicate among themselves and with the destinations to transmit this data, or parameters of interest derived from this data, in the most efficient manner. In-network processing may take place as the data travels through the network. Determination of the most efficient communication techniques for this process leads to multi-point network information theoretic problems.

We can however model the problem at a reduced complexity by considering those aspects which occur more commonly in practice. For instance, with the current technology [1], [2], communication cost dominates processing costs and hence, all processing should be performed locally, as close to the source as possible, such that only relevant data needs to be communicated over longer network paths. We develop such a model for the data gathering process and derive the optimal data rates required to communicate an estimate of the phenomenon at the required fidelity. We assume that the network does not communicate the complete set of raw measurements collected by the sensors but just enough information to meet the fidelity requirements. Such an assumption is valid for sensor networks as the data sinks are typically interested only in the measured phenomenon or its location and not the identities or raw measurements from the individual sensor nodes, enabling the network to fuse sensor data close to the source.

This work was supported in part by the National Science Foundation (NSF) Center for Embedded Networked Sensing under Cooperative Agreement #CCR-0121778, ITR-Fundamental Limits in Sensor Network #CCR-0217260, NSF grant #CNS-0306408, and the Office of Naval Research (ONR).

Similar problems have been solved before [3], [4] and are summarized in section II. However, the effect of sensor noise has not been explicitly considered. This is significant since the presence of noise changes the desired communication strategy. For instance, if noise at different sensors is independent, readings from two sensors placed at approximately the same location, can be used to average out the effect of noise leading to a better estimate. Data from such co-located sensors might have been considered redundant in the absence of noise, but in a practical implementation noise cannot be ignored and it becomes relevant to communicate this data.

### A. Key Contributions

We model the data gathering process with multiple noisy sensors as an information theoretic problem and derive the optimal data rate required to communicate an estimate of the phenomenon at the desired distortion level. Since these multiple sensors will share the same wireless channel, the total data rate required, by all the sensors together, is considered. This problem is a variation of the Gaussian CEO problem [5], [6], [7] as discussed later. The sensors communicate compressed data to a local fusion center and then the fused estimates are communicated over the network.

### B. Outline

The next section summarizes the prior work in this field and shows how our work builds upon it. Section III specifies our problem formulation and the abstractions used. Section IV derives a lower bound on the optimal rate-distortion relationship and compares it to a previously known special case. An upper bound is derived in Section V and shown to lie close to the lower bound, establishing that the lower bound is close to the rate-distortion function. Section VI concludes.

## II. RELATED WORK

The data carrying capacity of multi-hop wireless networks was estimated in [8] when each node generated data independently. This clearly does not model the data as generated in sensor networks, which typically comes from a set of common sources and hence is likely to be correlated. The correlated data model was considered in [4] and a rate-distortion relationship was derived to show how data generated from multiple sensors could be used to reduce distortion. The data gathering problem for a correlated source was also considered in [3], with a non band-limited field phenomenon.

We begin with the same model for the phenomenon as used in the above papers, i.e., a multivariate stochastic process with non-zero correlation among its spatial components. However, we also account for sensor noise. The noise itself is modelled as a Gaussian stochastic process.

Slepian and Wolf [9] had calculated the rate required when multiple sources transmit correlated data. That problem does not model sensor noise. Also, they consider lossless reproduction of sources. In practice, especially due to scalability concerns in large distributed networks, we do not need to reproduce the complete raw data. Rather, only a fused version, such as a feature of interest, is required, and hence lossy coding suffices. The case when only one of the sources is reproduced and the others are treated as helpers, was studied in [10], [11].

Rate-distortion bounds when the sources code and transmit a noisy version of the phenomenon and the receiver fuses information from multiple sources to reproduce the phenomenon with non-zero distortion were first considered in [12]. Rate-distortion bounds for that problem were derived in [5], [6], [7], [13] for Gaussian sources. However, the above solutions assume that all sensors are measuring exactly the same value of the phenomenon.

This does not address the case of distributed phenomenon (such as a temperature field) or the case when the different sensors are measuring multiple perspectives of a single phenomenon. We extend the problem formulation to the distributed case.

### III. PROBLEM DESCRIPTION

The data gathering problem (Figure 1) can be abstracted to the following mathematical formulation. Let  $\{X_1(t), \dots, X_L(t)\}_{t=1}^{\infty}$  represent an  $L$ -dimensional source. The source sequences are assumed temporally memoryless and stationary.

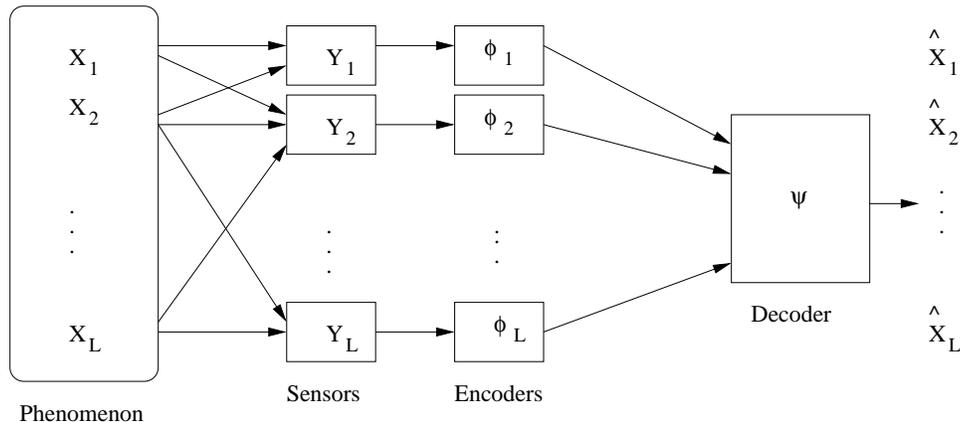


Fig. 1. The distributed Gaussian CEO System.

The vector  $\{X_1(t), \dots, X_L(t)\}$  is modeled as a zero mean Gaussian random variable with a non-singular covariance matrix  $R_{\mathbf{X}}$  for all  $t$ . The sensor readings  $\{Y_1(t), \dots, Y_L(t)\}$  are noisy versions of  $\{X_1(t), \dots, X_L(t)\}$ . Each of  $X_i(t)$  and  $Y_i(t)$  take values on the real lines  $\mathcal{X}_i$  and  $\mathcal{Y}_i$  respectively. We use boldface letters to represent  $L$ -dimensional vectors; for instance  $\mathbf{X} = \{X_1(t), \dots, X_L(t)\}$ . The observed readings  $Y_i(t)$ 's are modeled as:

$$\mathbf{Y} = H\mathbf{X} + \mathbf{N}$$

where  $H \in \mathfrak{R}^{L \times L}$  is a positive definite attenuation matrix and  $\mathbf{N}$  is an additive white zero mean Gaussian noise vector with covariance matrix  $R_{\mathbf{N}}$ .  $N_i$ 's model the noise in the sensor transducers and hence are independent of each other.

Data sequences,  $Y_i^n$ , of block length  $n$ , are separately encoded to  $\{\varphi_i(Y_i^n)\}_{i=1}^L$  and sent to a fusion center. The encoder functions  $\varphi_i$ 's are defined by

$$\varphi_i : \mathcal{Y}_i^n \rightarrow \mathcal{C}_i = \{1, 2, \dots, |\mathcal{C}_i|\}$$

and the transmitted rates are

$$\frac{1}{n} \log |\mathcal{C}_i| \leq r_i, \quad i = 1, \dots, L$$

The sum of the rates, i.e., the total bandwidth used, is denoted by  $r$ ,

$$r = \sum_{i=1}^L r_i$$

Here, the decoder receives the codewords,  $\mathbf{C}$ , to produce the estimate  $\hat{\mathbf{X}}^n$ . The decoder function  $\psi_L$  is given by

$$\psi_L : \mathcal{C}_1 \times \mathcal{C}_2 \times \dots \times \mathcal{C}_L \rightarrow \mathcal{X}^n$$

The distortion in the reproduction is defined as:

$$D^n(\mathbf{X}^n, \hat{\mathbf{X}}^n) = \frac{1}{n} \sum_{t=1}^n \text{tr}(E\|\mathbf{X} - \hat{\mathbf{X}}\|^2) \quad (1)$$

(where  $\text{tr}(A)$  represents the trace of matrix  $A$ ) and is subject to the distortion constraint:

$$D^n(\mathbf{X}^n, \hat{\mathbf{X}}^n) \leq D \quad (2)$$

The distributed Gaussian CEO problem is to find the rate distortion relationship between  $r$  and  $D$ . This will quantify the minimum rate required to achieve a required distortion.

#### IV. DERIVING THE RATE-DISTORTION RELATIONSHIP

We first derive the lower bound on the rate-distortion function. Then, we numerically compare it with an upper bound to evaluate how close our derived bound is to the exact function.

##### A. Outer Region

*Theorem 4.1:* For a given distortion  $D$ , the sum of the rates of the coded sensor data streams is bounded as

$$r(D) \geq \frac{1}{2} \log^+ \left[ \frac{|\Theta H' R_{\mathbf{N}}^{-1}|^2 |R_{\mathbf{N}}| (\prod \lambda_i) |R_{\mathbf{X}}|}{|\Theta| \left[ \frac{D}{L} |R_{\mathbf{X}}|^{\frac{1}{L}} - (\prod \lambda_i)^{\frac{1}{L}} |\Theta|^{\frac{1}{L}} \right]^L} \right]$$

where  $i$  ranges over  $1, \dots, L$ ,  $\Theta = (R_{\mathbf{X}}^{-1} + H' R_{\mathbf{N}}^{-1} H)^{-1}$ ,  $\lambda_i$  are the eigenvalues of  $R_{\mathbf{X}}$  and  $\log^+ x = \max\{\log x, 0\}$ .

Note that all logarithms in this paper are to base 2. Theorem 4.1 holds for  $\text{tr}(R_{\mathbf{N}}) \leq D \leq \text{tr}(R_{\mathbf{X}})$ , otherwise  $r(D) = 0$ . Intuitively, if the noise in sensor readings is higher than the acceptable distortion,  $D$ , then there is no need to transmit any data. The result assumes that the required distortion is greater than that achievable using an optimal estimator, i.e.  $D \geq \Theta$ .

*Proof:*

The mutual information between the source  $\mathbf{X}$  and  $\hat{\mathbf{X}}$  is related to the distortion as [14]:

$$\begin{aligned} \frac{1}{n} I(\mathbf{X}^n; \hat{\mathbf{X}}^n) &\geq \sum_{i=1}^L \frac{1}{2} \log \frac{\lambda_i}{D_i} \\ &= \frac{1}{2} \log \left( \prod_{i=1}^L \frac{\lambda_i}{D_i} \right) \end{aligned} \quad (3)$$

where  $\{\lambda_i\}_{i=1}^L$  are the eigenvalues of the covariance matrix  $R_{\mathbf{X}}$  of the phenomenon  $\mathbf{X}$  and

$$D_i = \begin{cases} K & \text{if } K < \lambda_i \\ \lambda_i & \text{otherwise} \end{cases}$$

and  $K$  is such that  $\sum_{i=1}^L D_i = D$ . The  $D_i$ 's can be calculated by reverse water filling [14].

From (3) we have,

$$\frac{2}{nL} I(\mathbf{X}^n; \hat{\mathbf{X}}^n) \geq \log \left( \prod_{i=1}^L \frac{\lambda_i}{D_i} \right)^{1/L} \quad (4)$$

We now need to relate the mutual information  $I(\mathbf{X}^n; \hat{\mathbf{X}}^n)$  with the compressed data rate,  $r$ , at which the codewords are sent. Since  $\hat{\mathbf{X}}$  is estimated from the compressed data  $\mathbf{C}$  it

follows that  $\mathbf{X}^n \rightarrow \mathbf{C}^n \rightarrow \hat{\mathbf{X}}^n$  is a Markov chain and hence  $I(\mathbf{X}^n; \hat{\mathbf{X}}^n) \leq I(\mathbf{X}^n; \mathbf{C}^n)$ . Using this in (4), raising both sides to the power 2 and rearranging terms we get:

$$\left[ \prod_{i=1}^L D_i \right]^{\frac{1}{L}} \geq \left[ \prod_{i=1}^L \lambda_i \right]^{\frac{1}{L}} \exp \left[ \frac{-2}{nL} I(\mathbf{X}^n; \mathbf{C}^n) \right] \quad (5)$$

where  $\exp(z)$  represents  $2^z$ . Next, we express  $I(\mathbf{X}^n; \mathbf{C}^n)$  in terms of  $r$ :

*Lemma 4.2:*

$$\exp \left[ \frac{-2}{nL} I(\mathbf{X}^n; \mathbf{C}^n) \right] \geq \frac{\left( \frac{|\Theta|}{|R_{\mathbf{X}}|} \right)^{\frac{1}{L}}}{1 - \Gamma \left[ \prod_{i=1}^L \frac{\lambda_i}{D_i} \right]^{\frac{1}{L}} \exp \left[ \frac{-2}{L} r \right]}$$

where  $\Gamma = \frac{|\Theta H' R_{\mathbf{N}}^{-1}|^{\frac{2}{L}} |R_{\mathbf{N}}|^{\frac{1}{L}}}{|\Theta|^{\frac{1}{L}}}$  and  $i$  ranges from  $1, \dots, L$ .

(The proof of the above lemma has been moved to Appendix to maintain continuity.)

Also, since the arithmetic mean is greater than or equal to the geometric mean (AM-GM inequality):

$$\left[ \prod_{i=1}^L D_i \right]^{\frac{1}{L}} \leq \frac{1}{L} \sum_{i=1}^L D_i = D \quad (6)$$

Substituting Lemma 4.2 and (6) in (5) we obtain Theorem 4.1. ■

As a verification we consider a previously known special case. Consider  $L = 1$ . Substituting  $L = 1$  and  $H = 1$  in Theorem 4.1,  $R_{\mathbf{X}} = \sigma_X^2$ , and noise variance  $R_{\mathbf{N}} = \sigma_N^2$ , we get:

$$r(D) \geq \frac{1}{2} \log^+ \left( \frac{\sigma_X^4}{D\sigma_X^2 - \sigma_X^2\sigma_N^2 + D\sigma_N^2} \right)$$

which in fact is same as the  $r(D)$  function given in [6], [7]. Thus, the bound yields the exact function for the special case.

The derived expression is plotted for some values of correlation among the sources in Figure 2. We can see that the data rate required is lower when there is more correlation among the data sources which is intuitive since when the sources are correlated, the total information content is lower.

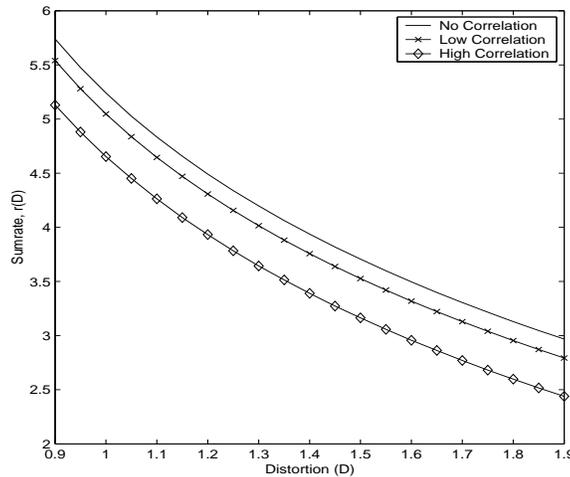


Fig. 2. Evaluating the derived expression from Theorem 4.1.

This theorem provides the minimum amount of non-redundant data that must be transmitted by the set of distributed sensors for reproducing within the distortion constraint. Thus, if a compression algorithm implemented at the sensors transmits only the amount of data mentioned in the above theorem, it will be an optimal scheme. However, the theorem does not claim that the above rate can actually be achieved by a practical compression algorithm. We address this issue next.

## V. ACHIEVABILITY OF THE RATE-DISTORTION BOUND

We now state a theorem which provides an achievable data-rate for reproducing at the given distortion, i.e., an upper bound on the rate required. We will show that this upper bound is close to the minimum bound derived in the previous section, which means that practical compression algorithms can in fact operate close to the minimum bound of Theorem 4.1, and hence it can be used as an approximate measure of the practical data rate required.

*Theorem 5.1:* Consider the random variables  $(X_1, X_2, \dots, X_L, Y_1, Y_2, \dots, Y_L)$  with the joint distribution given by  $p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})$ . Let  $\mathcal{C}_{in}(D)$  be the set of compressed data vectors,  $\mathbf{C} = (C_1, C_2, \dots, C_L)$ , such that:

- 1)  $C_i \rightarrow Y_i \rightarrow (\mathbf{X}, \mathbf{Y} \setminus Y_i, \mathbf{C} \setminus C_i)$  form a Markov chain for  $i = 1, \dots, L$  where  $\mathbf{Y} \setminus Y_i$  refers to  $(Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_L)$  etc.
- 2) There exists a decoding function

$$f : \mathcal{C}_1 \times \mathcal{C}_2 \dots \times \mathcal{C}_L \rightarrow \mathcal{X}_1 \times \mathcal{X}_2 \dots \times \mathcal{X}_L$$

such that  $Ed(\mathbf{X}, \hat{\mathbf{X}}) \leq D$  where  $d(\mathbf{X}, \hat{\mathbf{X}})$  is as defined in equation (1) and  $\hat{\mathbf{X}} = f(\mathbf{C})$ . Let  $\mathcal{R} = \{(r_1, r_2, \dots, r_L) : \sum_{i \in Z} r_i \geq I(C_Z; Y_Z | C_{Z^c}), \forall Z \subseteq (1, \dots, L)\}$  where  $Z^c$  is the complement of  $Z$ . Then,

$$r_{in}(D) \triangleq \text{convex hull of } \left\{ \bigcup_{\mathbf{C} \in \mathcal{C}_{in}(D)} \mathcal{R} \right\}$$

Here  $r_{in}(D)$  is the inner region for the rate-distortion relationship.

The proof of this theorem is an extension of the proof for Theorem 1 in [7], from the case of a scalar source to that of a vector source,  $\mathbf{X}$ . The proof is based on the joint typicality of codewords from different sensors [15], [16], [17], [18]. The encoding and decoding procedures used in the proof in [7] extend directly to the vector case. We omit the details for brevity.

To compare the inner and outer regions, and thus verify that the derived outer region is close to the actual rate-distortion relationship, we now evaluate both the bounds. The expression for the inner region of the rate-distortion relationship in the above theorem can be evaluated numerically for the jointly Gaussian source used in Theorem 4.1 as follows.

We numerically minimize the value of  $I(C_Z; Y_Z | C_{Z^c})$  for a Gaussian model of the source  $\mathbf{X}$  and noise  $\mathbf{N}$ . From the known expression for mutual information in the Gaussian case:

$$I(C_Z; Y_Z | C_{Z^c}) = \frac{1}{2} \log^+ \frac{|R_{\mathbf{Y}_Z \mathbf{C}_{Z^c}}| |R_{\mathbf{C}}|}{|R_{\mathbf{Y}_Z \mathbf{C}}| |R_{\mathbf{C}_{Z^c}}|}$$

To find the optimal  $\mathbf{C}$ , we begin with  $\mathbf{C} = L\mathbf{Y} + \mathbf{T}$  where  $L$  is a diagonal matrix and  $\mathbf{T}$  is a vector Gaussian random variable. The values of  $L$  and the variances of the components of  $\mathbf{T}$  are found using numerical optimization, thus leading to the optimal value of  $I(C_Z; Y_Z | C_{Z^c})$ .

The evaluated upper and lower bounds are plotted in Figure 3 for  $L = 3$ . The first figure shows that when there is no correlation among the  $L$  sources, then the upper and lower bounds exactly overlap. When the correlation among the sources in  $\mathbf{X}$  is non-zero, even then the upper and lower bounds are close. This shows that the derived lower bound is close to

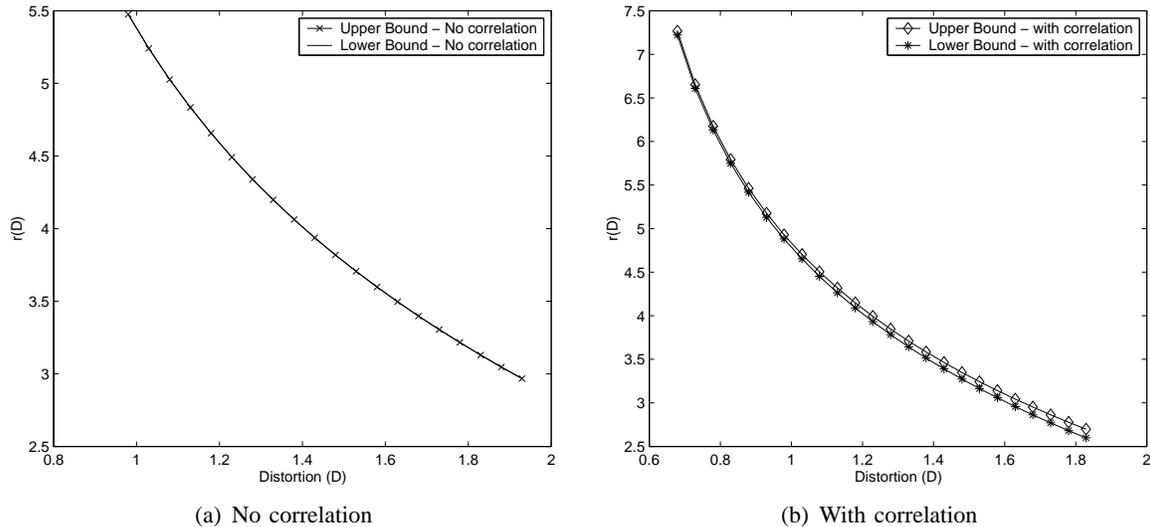


Fig. 3. Evaluating the inner and outer regions (a) when there is no correlation among the source components, and (b) with correlation among the source components

the actual rate-distortion function. Thus, the derived lower bound can be used as an estimate of the minimum data rate required to achieve the desired fidelity. The analytical derivation of the achievable bound, or the exact rate-distortion function, however, is still an open problem.

## VI. CONCLUSIONS

We derived the optimal data rate required to communicate when a set of sensors measure a distributed phenomenon. The effect of sensor noise was explicitly considered. This is useful for determining the minimum amount of resources required for achieving a desired level of fidelity and for evaluating the performance of data compression schemes used in the network.

In this work, we considered the fused estimate to be a reproduction of the phenomenon. However, the data rate will change if other estimates were considered, such as functions of the reproduced phenomenon. We showed that the derived lower bound on the rate-distortion function is close to the exact function, but an analytical derivation of the exact rate-distortion function is still an open problem.

## APPENDIX

From the definition of mutual information we have

$$I(\mathbf{X}^n; \mathbf{C}^n) = h(\mathbf{X}^n) - h(\mathbf{X}^n | \mathbf{C}^n) \quad (7)$$

Substituting the known entropy measure for the multivariate Gaussian vector  $\mathbf{X}$  as  $h(\mathbf{X}) = \frac{1}{2} \log(2\pi e)^L |R_{\mathbf{X}}|$  in (7) and rewriting:

$$\exp \left[ \frac{-2}{nL} I(\mathbf{X}^n; \mathbf{C}^n) \right] = \frac{1}{2\pi e |R_{\mathbf{X}}|^{\frac{1}{2}}} \exp \left[ \frac{2}{nL} h(\mathbf{X}^n | \mathbf{C}^n) \right] \quad (8)$$

Now, we find a bound on  $h(\mathbf{X}^n | \mathbf{C}^n)$  in terms of the required quantity  $r$ .

Let  $S = E(\mathbf{X}|\mathbf{Y})$  where  $E(\cdot)$  denotes expectation. Therefore,  $S = \mathbf{A}\mathbf{Y}$  where  $A \in \mathfrak{R}^{L \times L}$  and is given by [19],

$$A = (R_{\mathbf{X}}^{-1} + H' R_{\mathbf{N}}^{-1} H)^{-1} H' R_{\mathbf{N}}^{-1}$$

If  $\tilde{\mathbf{N}}$  be a zero-mean Gaussian with variance  $R_{\tilde{\mathbf{N}}} = (R_{\mathbf{X}}^{-1} + H'R_{\mathbf{N}}^{-1}H)^{-1}$  [19] then we have,

$$\mathbf{X} = \mathbf{S} + \tilde{\mathbf{N}} \quad (9)$$

and  $\tilde{\mathbf{N}}$  is independent of  $\mathbf{Y}$ . Since the sequences are memoryless, this leads to

$$\mathbf{X}^n(\mathbf{c}^n) = \mathbf{S}^n(\mathbf{c}^n) + \tilde{\mathbf{N}}^n \quad (10)$$

where  $\mathbf{X}^n(\mathbf{c}^n)$  is the conditional random variable  $\mathbf{X}^n$  conditioned on  $\mathbf{C}^n = \mathbf{c}^n$ , and  $\mathbf{S}^n(\mathbf{c}^n)$  is similarly defined. Also, note that since  $\tilde{\mathbf{N}}^n$  is independent of  $\mathbf{Y}^n$ , it is also independent of  $\mathbf{C}^n$ .

Using the entropy power inequality [14] in (10),

$$\exp \left[ \frac{2}{nL} h(\mathbf{X}^n(\mathbf{c}^n)) \right] \geq \exp \left[ \frac{2}{nL} h(\mathbf{S}^n(\mathbf{c}^n)) \right] + \exp \left[ \frac{2}{nL} h(\tilde{\mathbf{N}}^n(\mathbf{c}^n)) \right] \quad (11)$$

Substituting the value of  $h(\tilde{\mathbf{N}})$  in (11),

$$\exp \left[ \frac{2}{nL} h(\mathbf{X}^n(\mathbf{c}^n)) \right] \geq \exp \left[ \frac{2}{nL} h(\mathbf{S}^n(\mathbf{c}^n)) \right] + 2\pi e |R_{\tilde{\mathbf{N}}}|^{1/L} \quad (12)$$

Taking the logarithm of the above equation, we get:

$$\frac{1}{nL} h(\mathbf{X}^n(\mathbf{c}^n)) \geq T \left( \frac{1}{nL} h(\mathbf{S}^n(\mathbf{c}^n)) \right) \quad (13)$$

where:

$$T(z) = \frac{1}{2} \log [2^{2z} + 2\pi e |R_{\tilde{\mathbf{N}}}|^{1/L}]$$

Next, we take expectations on both sides of (13) with respect to  $\mathbf{C}^n$ . Note that from the definition of our conditional random variables, it follows that

$$\begin{aligned} E_{\mathbf{C}^n}[h(\mathbf{X}^n(\mathbf{c}^n))] &= h(\mathbf{X}^n|\mathbf{C}^n) \\ E_{\mathbf{C}^n}[h(\mathbf{S}^n(\mathbf{c}^n))] &= h(\mathbf{S}^n|\mathbf{C}^n) \end{aligned}$$

Observe that  $T(z)$  is a convex function of  $z$ . Applying Jensen's inequality, we get

$$\frac{1}{nL} h(\mathbf{X}^n|\mathbf{C}^n) \geq T \left( \frac{1}{nL} h(\mathbf{S}^n|\mathbf{C}^n) \right) \quad (14)$$

Since  $T(z)$  is monotone increasing with respect to  $z$ , the inequality is preserved. From (12), (13) and (14) we obtain,

$$\exp \left\{ \frac{2}{nL} h(\mathbf{X}^n|\mathbf{C}^n) \right\} \geq \exp \left\{ \frac{2}{nL} h(\mathbf{S}^n|\mathbf{C}^n) \right\} + 2\pi e |R_{\tilde{\mathbf{N}}}|^{1/L} \quad (15)$$

Now, we evaluate  $h(\mathbf{S}^n|\mathbf{C}^n)$ . From the definition of mutual information,

$$\frac{1}{n} h(\mathbf{S}^n|\mathbf{C}^n) = \frac{1}{n} h(\mathbf{S}^n|\mathbf{C}^n, \mathbf{X}^n) + \frac{1}{n} I(\mathbf{X}^n; \mathbf{S}^n|\mathbf{C}^n) \quad (16)$$

To evaluate  $h(\mathbf{S}^n|\mathbf{C}^n)$ , we shall first bound  $\frac{1}{n} I(\mathbf{X}^n; \mathbf{S}^n|\mathbf{C}^n)$ .

$$\begin{aligned} \frac{1}{n} I(\mathbf{X}^n; \mathbf{S}^n|\mathbf{C}^n) &= \frac{1}{n} I(\mathbf{X}^n; \mathbf{S}^n, \mathbf{C}^n) - \frac{1}{n} I(\mathbf{X}^n; \mathbf{C}^n) \\ &\geq \frac{1}{n} I(\mathbf{X}^n; \mathbf{S}^n) - \frac{1}{n} I(\mathbf{X}^n; \mathbf{C}^n) \\ &= \frac{1}{2} \log \frac{|R_{\mathbf{X}}|}{|R_{\tilde{\mathbf{N}}}|} - \frac{1}{n} I(\mathbf{X}^n; \mathbf{C}^n) \end{aligned} \quad (17)$$

Next, we derive a lower bound on  $h(\mathbf{S}^n | \mathbf{C}^n, \mathbf{X}^n)$ . For this let  $\mathbf{S}^n(\mathbf{x}^n, \mathbf{c}^n)$  be a conditional random variable  $\mathbf{S}^n$  conditioned by  $(\mathbf{X}^n, \mathbf{C}^n) = (\mathbf{x}^n, \mathbf{c}^n)$ . We have a similar definition for  $\mathbf{Y}^n(\mathbf{x}^n, \mathbf{c}^n)$ . Also observe that  $Y_i^n \rightarrow \mathbf{X}^n \rightarrow \mathbf{C}^n \setminus C_i^n$  forms a Markov chain. The conditional independence of  $(Y_i^n, C_i^n)$ ,  $i = 1, 2, \dots, L$  given  $\mathbf{X}^n$  also extends to  $\mathbf{S}^n(\mathbf{x}^n, \mathbf{c}^n)$ . Hence, from the definition of  $\mathbf{S}^n$  we have:

$$\mathbf{S}^n(\mathbf{x}^n, \mathbf{c}^n) = A\mathbf{Y}^n(\mathbf{x}^n, \mathbf{c}^n)$$

The above equation implies:

$$\begin{aligned} h(\mathbf{S}^n(\mathbf{x}^n, \mathbf{c}^n)) &= h(A\mathbf{Y}^n(\mathbf{x}^n, \mathbf{c}^n)) \\ &= \log |A| + h(\mathbf{Y}^n(\mathbf{x}^n, \mathbf{c}^n)) \end{aligned}$$

Taking expectation on both the sides of above equation with respect to  $(\mathbf{X}^n, \mathbf{C}^n)$  we get,

$$\frac{1}{n}h(\mathbf{S}^n | \mathbf{X}^n, \mathbf{C}^n) = \log |A| + \frac{1}{n}h(\mathbf{Y}^n | \mathbf{X}^n, \mathbf{C}^n) \quad (18)$$

From the definition of sum rate  $r$ , we have

$$\begin{aligned} r &\geq \sum_{i=1}^L \log |C_i^n| \geq \sum_{i=1}^L \frac{1}{n}h(C_i^n) \geq \frac{1}{n}h(\mathbf{C}^n) \\ &\geq \frac{1}{n}h(\mathbf{C}^n | \mathbf{X}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \\ &= \frac{1}{n}I(\mathbf{Y}^n; \mathbf{C}^n | \mathbf{X}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \\ &= \frac{1}{n}h(\mathbf{Y}^n | \mathbf{X}^n) - \frac{1}{n}h(\mathbf{Y}^n | \mathbf{X}^n, \mathbf{C}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \\ &= \frac{1}{2} \log(2\pi e)^L |R_{\mathbf{N}}| - \frac{1}{n}h(\mathbf{Y}^n | \mathbf{X}^n, \mathbf{C}^n) + \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \end{aligned}$$

From the above equation we have:

$$\frac{1}{n}h(\mathbf{Y}^n | \mathbf{X}^n, \mathbf{C}^n) \geq \frac{1}{2} \log(2\pi e)^L |R_{\mathbf{N}}| - \left[ r - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \right] \quad (19)$$

From (18) and (19), we have:

$$\frac{1}{n}h(\mathbf{S}^n | \mathbf{X}^n, \mathbf{C}^n) \geq \log |A| + \frac{1}{2} \log(2\pi e)^L |R_{\mathbf{N}}| - \left[ r - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \right] \quad (20)$$

Substituting (17) and (20) in (16), and dividing both the sides by  $L$  we have:

$$\begin{aligned} \frac{1}{nL}h(\mathbf{S}^n | \mathbf{C}^n) &\geq \frac{1}{L} \log |A| + \frac{1}{2L} \log(2\pi e)^L |R_{\mathbf{N}}| - \frac{1}{L} \left[ r - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \right] \\ &\quad + \frac{1}{2L} \log \frac{|R_{\mathbf{X}}|}{|R_{\tilde{\mathbf{N}}}|} - \frac{1}{nL}I(\mathbf{X}^n; \mathbf{C}^n) \end{aligned} \quad (21)$$

Substituting (21) in (15) we obtain :

$$\begin{aligned} \exp \left[ \frac{2}{nL}h(\mathbf{X}^n | \mathbf{C}^n) \right] &\geq \\ \frac{|A|^{\frac{2}{L}} 2\pi e |R_{\mathbf{N}}|^{\frac{1}{L}} |R_{\mathbf{X}}|^{\frac{1}{L}}}{|R_{\tilde{\mathbf{N}}}|^{\frac{1}{L}}} \exp \left\{ \frac{-2}{L} \left[ r - \frac{1}{n}I(\mathbf{X}^n; \mathbf{C}^n) \right] \right\} &\exp \left[ \frac{-2}{nL}I(\mathbf{X}^n; \mathbf{C}^n) \right] + 2\pi e |R_{\tilde{\mathbf{N}}}|^{\frac{1}{L}} \end{aligned}$$

where,  $A = \Theta H' R_{\mathbf{N}}^{-1}$  and  $R_{\tilde{\mathbf{N}}} = \Theta$ .

From the above equation and (8), we obtain the required relation between the rate  $r$  and  $I(\mathbf{X}^n; \mathbf{C}^n)$ :

$$\exp \left[ \frac{-2}{nL} I(\mathbf{X}^n; \mathbf{C}^n) \right] = \frac{|A|^{2/L} |R_{\mathbf{N}}|^{1/L}}{|R_{\tilde{\mathbf{N}}}|^{1/L}} \exp \left\{ \frac{-2}{L} \left[ r - \frac{1}{n} I(\mathbf{X}^n; \mathbf{C}^n) \right] \right\} \exp \left\{ \frac{-2}{nL} I(\mathbf{X}^n; \mathbf{C}^n) \right\} + \frac{|R_{\tilde{\mathbf{N}}}|^{1/L}}{|R_{\mathbf{X}}|^{1/L}}$$

The above equation can be rearranged to obtain:

$$\exp \left[ \frac{-2}{nL} I(\mathbf{X}^n; \mathbf{C}^n) \right] \left[ 1 - \frac{|A|^{2/L} |R_{\mathbf{N}}|^{1/L}}{|R_{\tilde{\mathbf{N}}}|^{1/L}} \eta \right] \geq \frac{|R_{\tilde{\mathbf{N}}}|^{1/L}}{|R_{\mathbf{X}}|^{1/L}}$$

where,  $\eta = \exp \left\{ \frac{-2}{L} \left[ r - \frac{1}{n} I(\mathbf{X}^n; \mathbf{C}^n) \right] \right\}$ .

Noting that  $R_{\tilde{\mathbf{N}}} = \Theta$  and  $I(\mathbf{X}^n; \mathbf{C}^n) \geq \frac{1}{n} I(\mathbf{X}^n; \hat{\mathbf{X}}^n) \geq \frac{1}{2} \log \left( \prod_{i=1}^L \frac{\lambda_i}{D_i} \right)$  (Section IV-A) in the above equation we get Lemma 4.2.

## REFERENCES

- [1] Mica wireless measurement system. Datasheet, Crossbow Technology Inc. [http://www.xbow.com/Products/Product\\_pdf\\_files/Wireless\\_pdf/MICA.pdf](http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICA.pdf).
- [2] Stargate: Xscale processor platform. [http://www.xbow.com/Products/Product\\_pdf\\_files/Wireless\\_pdf/6020-0049-01\\_C-Stargate.pdf](http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/6020-0049-01_C-Stargate.pdf). Data Sheet.
- [3] D Marco, E Duarte-Melo, M Liu, and D L Neuhoff. On the many-to-one transport capacity of a dense wireless sensor network and the compressibility of its data. In *Proceedings of the International Workshop on Information Processing in Sensor Networks (IPSN)*, April 2003.
- [4] A Scaglione and S D Servetto. On the interdependence of routing and data compression in multi-hop sensor networks. In *Proceedings of the 8th annual international conference on Mobile computing and networking*, pages 140–147. ACM Press, 2002.
- [5] H Viswanathan and T Berger. The quadratic gaussian CEO problem. *IEEE Transactions on Information Theory*, 43(5):1549–1559, September 1997.
- [6] Y Oohama. The rate-distortion function for the quadratic gaussian CEO problem. *IEEE Transactions on Information Theory*, 44(3):1057–1070, May 1998.
- [7] 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 (Special Issue on Fundamental Performance limits of wireless Sensor networks)*. to appear.
- [8] P Gupta and P R Kumar. The capacity of wireless networks. *IEEE Transactions on Information Theory*, IT-46(2):388–404, March 2000.
- [9] D Slepian and J K Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, IT-19:471–480, July 1973.
- [10] Y Oohama. Gaussian multiterminal source coding. *IEEE Transactions on Information Theory*, 43(6):1912–1923, November 1997.
- [11] A Pandya, A Kansal, G Pottie, and M Srivastava. Lossy source coding of multiple gaussian sources: m-helper problem. In *Proceedings of IEEE Information Theory Workshop*, San Antonio, TX, October 2004.
- [12] T Berger, Z Zhang, and H Vishwanathan. The CEO problem. *IEEE Transactions on Information Theory*, 42(3):887–902, May 1996.
- [13] V Prabhakaran, DNC Tse, and K Ramchandran. Rate region of the quadratic gaussian CEO problem. In *IEEE International Symposium on Information Theory*, June '2004.
- [14] T Cover and J Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc, 1991.
- [15] T Cover. A proof of the data compression theorem of slepian and wolf for ergodic sources. *IEEE Transactions on Information Theory*, 21:226–228, March 1975.
- [16] T Berger. *The Information Theory Approach to Communications*, chapter Multiterminal source coding, pages 171–231. CISM Courses and Lectures, no. 229. Vienna/New York:Springer-Verlag, 1978.
- [17] T S Han and K Kobayashi. A unified achievable rate region for a general class of multi-terminal source coding systems. *IEEE Transactions on Information Theory*, IT-26:277288, May 1980.
- [18] S Tung. *Multiterminal source coding*. PhD thesis, Cornell University, Ithaca, NY, May 1978.
- [19] A Sayed. *Fundamentals of Adaptive Filtering*. Wiley-IEEE Computer Society Press, 2003.