

An Upper Bound on the Sum-Rate Distortion Function and its Corresponding Rate Allocation Schemes for the CEO Problem

Jun Chen, *Student Member, IEEE*, Xin Zhang,
Toby Berger, *Fellow, IEEE*, Stephen B. Wicker, *Senior Member, IEEE*

Abstract—We consider a distributed sensor network in which several observations are communicated to the fusion center using limited transmission rate. The observation must be separately encoded so that the target can be estimated with minimum average distortion. We address the problem from an information theoretic perspective and establish the inner and outer bound of the admissible rate-distortion region. We derive an upper bound on the sum-rate distortion function and its corresponding rate allocation schemes by exploiting the contra-polymatroid structure of the achievable rate region. The quadratic Gaussian case is analyzed in detail and the optimal rate allocation schemes in the achievable rate region are characterized. We show that our upper bound on the sum-rate distortion function is tight for the quadratic Gaussian CEO problem in the case of same SNRs at the sensors.

Index Terms—CEO problem, contra-polymatroid, decentralized estimation, Gaussian source, multiterminal source coding, mean-squared error, rate allocation, water-filling.

I. INTRODUCTION

IN this paper, we consider the following distributed sensor network (see Fig.1). $\{X(t)\}_{t=1}^{\infty}$ is the target data sequence

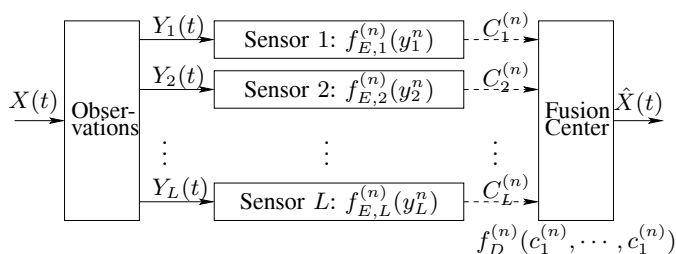


Fig. 1. Model of distributed sensor network

that the fusion center is interested in. This data sequence cannot be observed directly. L sensors are deployed, which observe corrupted versions of $\{X(t)\}_{t=1}^{\infty}$ separately. The data rate at which sensor i ($i = 1, 2, \dots, L$) may communicate information about its observations to the fusion center is limited

to R_i bits per second¹. Due to wide geographical separation of the sensors or other reasons, the sensors are not permitted to communicate with each other, i.e. sensor i has to send data based solely on its own noisy observations $\{Y_i(t)\}_{t=1}^{\infty}$. Finally, the decision $\{\hat{X}(t)\}_{t=1}^{\infty}$ is computed from the combined data at fusion center. Gel'fand and Pinsker [1] considered a model analogous to ours for noiseless reproduction of $\{X(t)\}_{t=1}^{\infty}$. Our Model has been studied by Flynn and Gray [2] in the case of two sensors where they derived an achievable rate-distortion region. A closely related problem, called CEO problem, was introduced in [3] for discrete case, and [4] for quadratic Gaussian case. For CEO problem, we are only interested in the tradeoff between the estimation distortion and the total rate at which the sensors may communicate information about their observations to the fusion center. Oohama [5] derived the sum-rate distortion function for the quadratic Gaussian CEO problem when there are infinite sensors and the SNRs at all the sensors are identical. Viswanath [6] formulated a similar multiterminal Gaussian source coding problem and characterized the sum-rate distortion function for a class of quadratic distortion metrics.

For simplicity, we also call our problem as CEO problem, although our discussion is not restricted to the trade off between sum rate and distortion.

A. A Note about Notation.

- (i) We usually use capital letters (say, U) to indicate a random variable. U^n denotes the random vector $[U(1), U(2), \dots, U(n)]$ and $u^n = [u(1), u(2), \dots, u(n)]$ denotes a realization of U^n .
- (ii) The notion $A \rightarrow B \rightarrow C$ means that A, B, C form a Markov chain.
- (iii) We use calligraphic letters to indicate a set (say, \mathcal{A}) and use $|\mathcal{A}|$ to denote the cardinality of \mathcal{A} .
- (iv) $\mathcal{I}_K \triangleq \{1, 2, \dots, K\}$ for any positive integer K .
- (v) If $\mathcal{B} = \{i_1, i_2, \dots, i_k\}$, then $W_{\mathcal{B}} \triangleq (W_{i_1}, W_{i_2}, \dots, W_{i_k})^T$.

¹The communication channel between sensors and fusion center maybe noisy. If channel coding is allowed, we can convert the noisy channel into an equivalent noiseless channel with certain capacity which may depends on transmitter power, allocated bandwidth or other factors. So the restriction on the transmission rate for sensors may result from the restrictions on bandwidth, power and/or other resources. The restrictions on transmission rates may also be imposed due to the processing limitation of the fusion center.

B. System Model and Problem Formulation

Let $\{X(t), Y_1(t), \dots, Y_L(t)\}_{t=1}^{\infty}$ be a temporally memoryless source with instantaneous joint probability distribution $P_{XY_1 \dots Y_L}$ on $\mathcal{X} \times \mathcal{Y}_1 \times \dots \times \mathcal{Y}_L$, where L is the number of sensors, \mathcal{X} is the common alphabet of the random variables $X(t)$ for $t = 1, 2, \dots$, \mathcal{Y}_i ($i = 1, 2, \dots, L$) is the common alphabet of the random variables $Y_i(t)$ for $t = 1, 2, \dots$. Sensor i encodes a block $y_i^n = [y_i(1), \dots, y_i(n)]$ of length n from its observed data using a source code $c_i^{(n)} = f_{E,i}^{(n)}(y_i^n)$ of rate $R_i^{(n)} \geq \frac{1}{n} \log |\mathcal{C}_i^{(n)}|$. The codewords from the L sensors, $c_1^{(n)}, \dots, c_L^{(n)}$, are sent to the fusion center. The task of fusion center is to recover the target data sequence $x^n = [x(1), \dots, x(n)]$ with minimal expected distortion defined as $d^{(n)} = \frac{1}{n} E \sum_{t=1}^n d(X(t), \hat{X}(t))$, where $d(x, \hat{x})$ is a given distortion measure and \hat{X}^n is the estimate of random target sequence X^n . The fusion center implements a mapping $f_D^{(n)} : \mathcal{C}_1^{(n)} \times \dots \times \mathcal{C}_L^{(n)} \rightarrow \mathcal{X}^n$, i.e. the estimate at fusion center is of the form $\hat{x}^n = f_D^{(n)}(c_1^{(n)}, \dots, c_L^{(n)})$.

The rest of this paper is divided into four sections. In Section II we derive the inner and outer bound of the admissible rate-distortion region. In Section III, we establish an upper bound on the sum-rate distortion function by exploiting the contra-polymatroid structure of the achievable rate region. The rate allocation schemes to achieve this upper bound are characterized. In Section IV, we consider the case of correlated memoryless Gaussian observations and squared distortion measure. The parametric representation of the inner bound is computed. We derive an explicit formula of the upper bound on the sum-rate distortion function and characterize the corresponding rate allocation schemes. We show that the rate allocation schemes that attain the upper bound possess a generalized water-filling interpretation. Moreover, our upper bound is shown to be tight for the quadratic Gaussian CEO problem in the case of same SNRs at the sensors. In Section V we suggest several directions for further research.

For simplicity, in Section II and Section III we assume $\max(|\mathcal{X}|, |\mathcal{Y}_1|, \dots, |\mathcal{Y}_L|) < \infty$ and the distortion measure $d : \mathcal{X} \times \mathcal{X} \rightarrow [0, d_{\max}]$ to be bounded, i.e., $d_{\max} < \infty$.

II. THE INNER AND OUTER BOUNDS FOR RATES-DISTORTION REGION

Definition 1: The rate vector $\mathbf{R} \triangleq (R_1, R_2, \dots, R_L)$ is said to be D -admissible if $\forall \varepsilon > 0, \exists n_0$ such that $\forall n > n_0$ there exist encoders:

$$\begin{aligned} f_{E,1}^{(n)} : \mathcal{Y}_1^n &\rightarrow \mathcal{C}_1^{(n)} & \log |\mathcal{C}_1^{(n)}| &\leq n(R_1 + \varepsilon) \\ f_{E,2}^{(n)} : \mathcal{Y}_2^n &\rightarrow \mathcal{C}_2^{(n)} & \log |\mathcal{C}_2^{(n)}| &\leq n(R_2 + \varepsilon) \\ &\vdots & & \\ f_{E,L}^{(n)} : \mathcal{Y}_L^n &\rightarrow \mathcal{C}_L^{(n)} & \log |\mathcal{C}_L^{(n)}| &\leq n(R_L + \varepsilon) \end{aligned}$$

and a decoder:

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

such that $\frac{1}{n} E \left[\sum_{t=1}^n d(X(t), \hat{X}(t)) \right] \leq D + \varepsilon$.

Let $\mathcal{R}(D)$ denote the set of all D -admissible rate vectors.

A. Inner Bound

Theorem 1: Given the joint distribution of the random variables (X, Y_1, \dots, Y_L) : $P_{XY_1 \dots Y_L}$, for $D \geq 0$, define $\mathcal{W}_{in}(D)$ as the set of random vectors $W_{\mathcal{I}_L} = (W_1, \dots, W_L)^T$ jointly distributed with X and $Y_{\mathcal{I}_L}$ such that the following two properties are satisfied:

- (i) $W_i \rightarrow Y_i \rightarrow (X, Y_{\{i\}^c}, W_{\{i\}^c})$ for all $i \in \mathcal{I}_L$.
- (ii) There exists a function $f : \mathcal{W}_1 \times \dots \times \mathcal{W}_L \rightarrow \mathcal{X}$ such that $Ed(X, \hat{X}) \leq D$ where $\hat{X} = f(W_1, \dots, W_L)$.

Let

$$\mathcal{R}(W_{\mathcal{I}_L}) = \{(R_1, \dots, R_L) : \sum_{i \in \mathcal{A}} R_i \geq I(W_{\mathcal{A}}; Y_{\mathcal{A}} | W_{\mathcal{A}^c}) \forall \mathcal{A} \subseteq \mathcal{I}_L\}$$

then

$$\mathcal{R}(D) \supseteq \mathcal{R}_{in}(D) \triangleq \text{convex hull}^2 \text{ of } \left\{ \bigcup_{W_{\mathcal{I}_L} \in \mathcal{W}_{in}(D)} \mathcal{R}(W_{\mathcal{I}_L}) \right\}.$$

We call $\mathcal{R}_{in}(D)$ the achievable rate region w.r.t. distortion D .

The proof of this theorem is standard, which is based on Cover's random binning argument [7] and a Markov lemma [8] [9] [10] that ensures the joint typicality of the codewords from different sensors (It will be made precise in the description of the encoding scheme). In the interest of conservation and simplicity, we just describe the encoding and decoding procedures and omit the details of the proof.

Let (W_1, W_2, \dots, W_L) and function f satisfy the conditions given in Theorem 1. Construct the random codebooks $\{\mathcal{C}_i^{(n)} = (\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}, \dots, \mathcal{C}_L^{(n)})\}^3$ (where $\mathcal{C}_i^{(n)}$ denotes the codebook of Sensor i) as follows:

At Sensor i , generate $M_i = 2^{n(I(Y_i; W_i) + \varepsilon)}$ i.i.d. codewords $\mathcal{C}_i^{(n)}$ according to the product distribution $\prod_{l=1}^n P_{W_i}$ and index them $\mathcal{C}_i^{(n)}(j), j = 1, 2, \dots, M_i$. Let $\mathcal{C}_i^{(n)} = \{\mathcal{C}_i^{(n)}(j)\}_{j=1}^{M_i}$. Randomly assign the indices of the codewords to one of 2^{nR_i} bins⁴ using a uniform distribution over the indices of the bins.

Encoding Scheme:

At Sensor i , given observation y_i^n , if it's typical, map it onto the $\mathcal{C}_i^{(n)}(j) \in \mathcal{C}_i^{(n)}$ with the smallest index j such that $(y_i^n, \mathcal{C}_i^{(n)}(j))$ are jointly typical. Let $\mathcal{C}_i^{(n)}(y_i^n)$ denotes the $\mathcal{C}_i^{(n)}$ onto which y_i^n is mapped. (Note: By Markov lemma, $P((\mathcal{C}_1^{(n)}(Y_1^n), \dots, \mathcal{C}_L^{(n)}(Y_L^n)) \text{ are jointly typical}) \rightarrow 1$ as $n \rightarrow \infty$.) The index of the bin which contains $\mathcal{C}_i^{(n)}(y_i^n)$ is sent. Let $b_i(y_i^n)$ denote this bin index. If y_i^n is not typical or there does not exist $\mathcal{C}_i^{(n)}(j) \in \mathcal{C}_i^{(n)}$ such that $(y_i^n, \mathcal{C}_i^{(n)}(j))$ are jointly typical, then a special error symbol is sent. This special error symbol does not increase the rate R_i in the limit of large n , so we may safely ignore it.

²It follows directly from a time sharing argument.

³Here $\mathcal{C}_i^{(n)}$ actually is not the $\mathcal{C}_i^{(n)}$ stated in the Definition 1. As we will see, we will not send the codewords in $\mathcal{C}_i^{(n)}$ directly. Instead, we will send the index of bin. That's why here we have 2^{nR_i} bins at encoder i , while generally $|\mathcal{C}_i^{(n)}| > 2^{nR_i}$.

⁴We suppose 2^{nR_i} is an integer. When n is large enough, this assumption causes no essential loss.

Decoding Scheme:

Given (b_1, b_2, \dots, b_L) , if there exists a unique $(C_1^{(n)}, C_2^{(n)}, \dots, C_L^{(n)})$ such that $C_i^{(n)} \in \mathcal{B}_i(b_i)$ (where $\mathcal{B}_i(b_i)$ denotes the bin with index b_i at Sensor i) and $(C_1^{(n)}, C_2^{(n)}, \dots, C_L^{(n)})$ are jointly typical, then decode it as $(\hat{C}_1^{(n)}, \hat{C}_2^{(n)}, \dots, \hat{C}_L^{(n)})$; otherwise declare an error and incur the maximum distortion d_{max} . If the received vector contains special error symbol, also declare an error and incur the maximum distortion d_{max} . Assuming no error, produce the estimate $\hat{x}(k) = f(\hat{C}_1^{(n)}(k), \hat{C}_2^{(n)}(k), \dots, \hat{C}_L^{(n)}(k))$ for $k = 1, 2, \dots, n$. Here $C_i^{(n)}(k)$ is the k^{th} symbol of the codeword $C_i^{(n)}$.

B. Outer Bound

Theorem 2: Given the joint distribution of the random variables (X, Y_1, \dots, Y_L) : $P_{XY_1 \dots Y_L}$, for $D \geq 0$, define $\mathcal{Z}_{out}(D)$ as the set of random vectors $Z_{\mathcal{I}_L} = (Z_1, \dots, Z_L)^T$ jointly distributed with X and $Y_{\mathcal{I}_L}$ such that the following two properties are satisfied:

- (i) $Z_i \rightarrow Y_i \rightarrow (X, Y_{\{i\}^c})$ for all $i \in \mathcal{I}_L$.
- (ii) There exists function $g : \mathcal{Z}_1 \times \dots \times \mathcal{Z}_L \rightarrow \mathcal{X}$ such that $Ed(X, \hat{X}) \leq D$ where $\hat{X} = g(Z_1, \dots, Z_L)$.

Let

$$\mathcal{R}(Z_{\mathcal{I}_L}) = \{(R_1, \dots, R_L) : \sum_{i \in \mathcal{A}} R_i \geq I(Z_{\mathcal{A}}; Y_{\mathcal{I}_L} | Z_{\mathcal{A}^c}), \forall \mathcal{A} \subseteq \mathcal{I}_L\},$$

then

$$\mathcal{R}(D) \subseteq \mathcal{R}_{out}(D) \triangleq \bigcup_{Z_{\mathcal{I}_L} \in \mathcal{Z}_{out}(D)} \mathcal{R}(Z_{\mathcal{I}_L}).$$

Proof: See Appendix I. ■

C. Discussion

- 1) Our inner bound can be specialized to the results of Wyner and Ziv [11], Draper and Wornell [12] and Berger *et al.* [13].

Specifically, it is easy to show that for any $i \in \mathcal{I}_L$,

$$\begin{aligned} & \inf\{R_i : (R_1, \dots, R_L) \in \mathcal{R}(D)\} \\ &= \inf\{R_i : (R_1, \dots, R_L) \in \mathcal{R}_{in}(D)\} \\ &= \inf_{V_i \in \mathcal{V}_i(D)} I(Y_i; V_i | Y_{\{i\}^c}) \end{aligned}$$

where $\mathcal{V}_i(D)$ is the set of random variables V_i jointly distributed with X and $Y_{\mathcal{I}_L}$ such that the following two properties are satisfied:

- (i) $V_i \rightarrow Y_i \rightarrow (X, Y_{\{i\}^c})$;
- (ii) There exists a function

$$\begin{aligned} f_i : \mathcal{Y}_1 \times \dots \times \mathcal{Y}_{i-1} \times \mathcal{Y}_{i+1} \times \dots \times \mathcal{Y}_L \times \mathcal{V}_i &\rightarrow \mathcal{X} \\ \text{such that } Ed(X, \hat{X}) &\leq D \text{ where} \\ \hat{X} &= f_i(Y_1, \dots, Y_{i-1}, Y_{i+1}, \dots, Y_L, V_i). \end{aligned}$$

That is to say, the upper bound (reduced from inner bound $\mathcal{R}_{in}(D)$) on the minimum rate required by Sensor i is tight. Note: In order to minimize the rate

required by Sensor i , a sufficient condition is that the rate constraints on Sensor $1, \dots, i-1, i+1, \dots, L$ are loose enough to guarantee the perfect recovery of $\{Y_1(t), \dots, Y_{i-1}(t), Y_{i+1}(t), \dots, Y_L(t)\}_{t=1}^{\infty}$ (Note: By Slepian-Wolf Theorem [14], if

$$\sum_{j \in \mathcal{S}} R_j \geq H(Y_{\mathcal{S}} | Y_{\mathcal{S}^c \setminus \{i\}}), \forall \mathcal{S} \subseteq \mathcal{I}_L \setminus \{i\},$$

then $\{Y_1(t), \dots, Y_{i-1}(t), Y_{i+1}(t), \dots, Y_L(t)\}_{t=1}^{\infty}$ can be recovered losslessly). But this condition is not necessary because what we need is just a sufficient statistic for $\{X(t)\}_{t=1}^{\infty}$, not the raw data $\{Y_1(t), \dots, Y_{i-1}(t), Y_{i+1}(t), \dots, Y_L(t)\}_{t=1}^{\infty}$.

- 2) The inner and outer bounds generally do not meet, so the complete characterization of the admissible rate-distortion region is still an open problem. Furthermore, even there exists a complete characterization of the admissible region, it is still a formidable, if not impossible, task to compute the explicit expression of the admissible region for a specific case since the associated optimization problem is very complicated in general.
- 3) Our inner bound and outer bound differ from those of [9] [8] only in the distortion constraint. But [9] and [8] focus mainly on the case $L = 2$. As pointed out in [15], CEO problem can be converted into Berger-Tung problem and thus is a special case of the latter.

III. OPTIMAL RATE ALLOCATION IN THE ACHIEVABLE REGION

In this section, we analyze the minimum sum rate in the achievable rate region w.r.t. distortion D , i.e.,

$$R_{\Sigma}(D) \triangleq \inf_{(R_1, R_2, \dots, R_L) \in \mathcal{R}_{in}(D)} \sum_{i=1}^L R_i. \quad (1)$$

We show that “inf” in equation (1) can be replaced by “min”. Furthermore, we characterize the following set

$$\begin{aligned} \vartheta(D) &\triangleq \{(R_1, R_2, \dots, R_L) : \\ &(R_1, R_2, \dots, R_L) \in \mathcal{R}_{in}(D), \sum_{i=1}^L R_i = R_{\Sigma}(D)\} \end{aligned} \quad (2)$$

in which every element corresponds to an optimal rate allocation scheme that minimizes the sum rate in the achievable rate region w.r.t. distortion D .

A major step toward the solutions to equation (1) and (2) is to exploit the contra-polymatroid structure of the achievable rate region. The contra-polymatroid theory has been used to study distributed source coding for years, see [10] [6]. A similar combinatorial structure called polymatroid has been applied to study the capacity region of multiaccess fading channel in [16].

Definition 2: Let $f : 2^{\mathcal{I}_L} \rightarrow \mathcal{R}_+$ be a set function. The polyhedron

$$\mathcal{G}(f) \equiv \{(x_1, \dots, x_L) : \sum_{i \in \mathcal{S}} x_i \geq f(\mathcal{S}), \forall \mathcal{S} \subseteq \mathcal{I}_L\}$$

is a *contra-polymatroid* if f satisfies

- 1) $f(\emptyset) = 0$ (normalized).
- 2) $f(S) \leq f(T)$ if $S \subset T$ (non-decreasing).
- 3) $f(S) + f(T) \leq f(S \cup T) + f(S \cap T)$ (supermodular).

If f satisfies the three properties, f is called a *rank function*.

Lemma 1: If $W_i \rightarrow Y_i \rightarrow (Y_{\{i\}^c}, W_{\{i\}^c})$ for all $i \in \mathcal{I}_L$ under the probability distribution $P_{Y_1 \dots Y_L W_1 \dots W_L}$, then

$$\mathcal{R}(W_{\mathcal{I}_L}) \triangleq \{(R_1, \dots, R_L) : \sum_{i \in \mathcal{A}} R_i \geq I(W_{\mathcal{A}}; Y_{\mathcal{A}} | W_{\mathcal{A}^c}), \forall \mathcal{A} \subseteq \mathcal{I}_L\}$$

is a contra-polymatroid.

Proof: Let

$$f(S) \triangleq I(W_S; Y_S | W_{S^c}), \forall S \subseteq \mathcal{I}_L.$$

We only need to show that f satisfies the three properties of contra-polymatroid.

- 1) By definition.
- 2)

$$\begin{aligned} f(T) &= I(Y_T; W_T | W_{T^c}) \\ &= I(Y_T; W_S | W_{S^c}) + I(Y_T; W_{T \setminus S} | W_{T^c}) \\ &\geq I(Y_T; W_S | W_{S^c}) \geq I(Y_S; W_S | W_{S^c}) \\ &= f(S), \text{ if } S \subseteq T. \end{aligned}$$

- 3) Since

$$\begin{aligned} f(S) &= I(Y_S; W_S | W_{S^c}) \\ &= H(W_S | W_{S^c}) - H(W_S | Y_S, W_{S^c}) \\ &\stackrel{(a)}{=} H(W_S | W_{S^c}) - H(W_S | Y_S) \\ &\stackrel{(b)}{=} H(W_{\mathcal{I}_L}) - H(W_{S^c}) - \sum_{i \in S} H(W_i | Y_i) \end{aligned}$$

(where both (a) and (b) follow from the fact that $W_i \rightarrow Y_i \rightarrow \{Y_{\{i\}^c}, W_{\{i\}^c}\}$ for all $i \in \mathcal{I}_L$), it follows that

$$\begin{aligned} &f(S) + f(T) - f(S \cup T) - f(S \cap T) \\ &= (H(W_{(S \cap T)^c}) - H(W_{T^c})) \\ &\quad - (H(W_{S^c}) - H(W_{(S \cup T)^c})) \\ &= H(W_{S^c \cap T} | W_{T^c}) - H(W_{S^c \cap T} | W_{(S \cup T)^c}) \\ &\leq 0. \end{aligned}$$

An important property on the characterization of the vertices of contra-polymatroid is given in [16]. For completeness, we restate it here.

Lemma 2 ([16], Lemma 3.3): Let $\mathcal{G}(f)$ be a contra-polymatroid. If π is a permutation on the set \mathcal{I}_L , define the vector $\mathbf{v}(\pi) \in \mathcal{R}^L$ by $\mathbf{v}_{\pi(1)}(\pi) = f(\{\pi(1)\})$ and for $\mathbf{v}_{\pi(i)}(\pi) = f(\{\pi(1), \dots, \pi(i)\}) - f(\{\pi(1), \dots, \pi(i-1)\})$, $i = 2, \dots, L$. Then the points $\mathbf{v}(\pi)$ are precisely the vertices of $\mathcal{G}(f)$.

Since it has been shown in Lemma 1 that $\mathcal{R}(W_{\mathcal{I}_L})$ is a contra-polymatroid, we can conclude that for each permutation

π on the set \mathcal{I}_L , $(R_1(\pi), \dots, R_L(\pi))$ gives a vertex of $\mathcal{R}(W_{\mathcal{I}_L})$ (where

$$\begin{aligned} R_{\pi(1)}(\pi) &= I(W_{\{\pi(1)\}}; Y_{\{\pi(1)\}} | W_{\{\pi(1)\}^c}), \\ R_{\pi(i)}(\pi) &= I(W_{\{\pi(1), \dots, \pi(i)\}}; Y_{\{\pi(1), \dots, \pi(i)\}} | W_{\{\pi(1), \dots, \pi(i)\}^c}) \\ &\quad - I(W_{\{\pi(1), \dots, \pi(i-1)\}}; Y_{\{\pi(1), \dots, \pi(i-1)\}} | W_{\{\pi(1), \dots, \pi(i-1)\}^c}), \end{aligned}$$

$i = 2, \dots, L$), and thus the contra-polymatroid $\mathcal{R}(W_{\mathcal{I}_L})$ has totally $L!$ vertices (Note: these vertices may not be distinct). It is easy to check that for each of these $L!$ vertices, the sum rate $\sum_{i=1}^L R_{\pi(i)}(\pi) = I(W_{\mathcal{I}_L}; Y_{\mathcal{I}_L})$. So the sum rate constraint $I(W_{\mathcal{I}_L}; Y_{\mathcal{I}_L})$ is attainable. Furthermore, let $\vartheta(W_L)$ be the convex hull of these $L!$ vertices. It is obvious that we have $\vartheta(W_L) \subseteq \mathcal{R}(W_{\mathcal{I}_L})$ and every point in $\vartheta(W_L)$ attains the sum rate constraint $I(W_{\mathcal{I}_L}; Y_{\mathcal{I}_L})$.

We summarize the above result in the following lemma.

Lemma 3: Let $R_{\Sigma}(W_{\mathcal{I}_L}) \triangleq \inf_{(R_1, \dots, R_L) \in \mathcal{R}(W_{\mathcal{I}_L})} \sum_{i=1}^L R_i$, then

$$R_{\Sigma}(W_{\mathcal{I}_L}) = \min_{(R_1, \dots, R_L) \in \mathcal{R}(W_{\mathcal{I}_L})} \sum_{i=1}^L R_i = I(W_{\mathcal{I}_L}; Y_{\mathcal{I}_L})$$

and we have

$$\begin{aligned} R_{\Sigma}(D) &= \inf_{W_{\mathcal{I}_L} \in \mathcal{W}_{in}(D)} R_{\Sigma}(W_{\mathcal{I}_L}) \\ &= \inf_{W_{\mathcal{I}_L} \in \mathcal{W}_{in}(D)} I(W_{\mathcal{I}_L}; Y_{\mathcal{I}_L}). \end{aligned}$$

Now we proceed to show that ‘‘inf’’ in Lemma 3 can be replaced by ‘‘min’’.

Lemma 4: There is no loss of generality to assume that $|\mathcal{W}_i| \leq |\mathcal{Y}_i| + 2^L - 1$ for all $i \in \mathcal{I}_L$ in Theorem 1.

Proof: By invoking the support lemma [17, pp.310], W_i must have $|\mathcal{Y}_i| - 1$ letters to preserve the probability distribution P_{Y_i} and 2^L more to preserve $I(Y_{\mathcal{A}}; W_{\mathcal{A}} | W_{\mathcal{A}^c})$ (for any nonempty set $\mathcal{A} \subseteq \mathcal{I}_L$) and D . ■

Lemma 5: Given the joint distribution of the discrete random variables $(X, Y_1, \dots, Y_L) : P_{XY_1 \dots Y_L}$, $x \in \mathcal{X}$, $y_i \in \mathcal{Y}_i$ ($i \in \mathcal{I}_L$), let $\tau(D)$ be the set of joint distribution $P_{XY_1 \dots Y_L W_1 \dots W_L}$ of $(X, Y_1, \dots, Y_L, W_1, \dots, W_L)$ ($w_i \in \mathcal{W}_i$, $|\mathcal{W}_i| \leq |\mathcal{Y}_i| + 2^L - 1 \forall i \in \mathcal{I}_L$) with the following properties satisfied:

- (i) $\sum_{w_1 \dots w_L} P_{XY_1 \dots Y_L W_1 \dots W_L}(x, y_1, \dots, y_L, w_1, \dots, w_L) = P_{XY_1 \dots Y_L}(x, y_1, \dots, y_L), \forall x \in \mathcal{X}, \forall y_i \in \mathcal{Y}_i$ ($i \in \mathcal{I}_L$).
- (ii) $W_i \rightarrow Y_i \rightarrow (X, Y_{\{i\}^c}, W_{\{i\}^c})$ for all $i \in \mathcal{I}_L$
- (iii) There exists a function $f : \mathcal{W}_1 \times \dots \times \mathcal{W}_L \rightarrow \mathcal{X}$ such that $Ed(X, f(W_1, \dots, W_L)) \leq D$.

Then $\tau(D)$ is compact.

Remark: Here we view $P_{XY_1 \dots Y_L W_1 \dots W_L}$ as a point in the $|\mathcal{X}| \times |\mathcal{Y}_1| \times \dots \times |\mathcal{Y}_L| \times |\mathcal{W}_1| \times \dots \times |\mathcal{W}_L|$ -dimensional Euclidean space.

Proof: See Appendix II. ■

By Lemma 3, 4 and 5, it is straightforward to get the following theorem.

Theorem 3:

- 1) $R_{\Sigma}(D) = \min_{W_{\mathcal{I}_L} \in \mathcal{W}_{in}(D)} I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L})$
 $= \min_{\tau(D)} I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L})$;
- 2) Let $\eta(D) \triangleq \{W_{\mathcal{I}_L}^* : W_{\mathcal{I}_L}^* \in \mathcal{W}_{in}(D),$
 $I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L}^*) = \min_{W_{\mathcal{I}_L} \in \mathcal{W}_{in}(D)} I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L})\}$

We have

$$\vartheta(D) = \text{convex hull}^5 \text{ of } \left\{ \bigcup_{W_{\mathcal{I}_L} \in \eta(D)} \vartheta(W_{\mathcal{I}_L}) \right\}.$$

Proof: $I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L})$ is a continuous function of $P_{XY_1 \dots Y_L W_1 \dots W_L}$ and thus is able to attain its minimum value over the compact set $\tau(D)$. Here the Euclidean metric is assumed implicitly in both the domain and the range of $I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L})$. ■

Remark: $R_{\Sigma}(D)$ is the minimum sum rate in the achievable region w.r.t. distortion D and thus is an upper bound of the sum-rate distortion function of the CEO problem. $\vartheta(D)$ is the collection of rate vectors in $\mathcal{R}_{in}(D)$ that attain the sum rate bound $R_{\Sigma}(D)$. We call $\vartheta(D)$ as the optimal rate allocation region in the achievable rate region. Every point in region $\vartheta(D)$ corresponds to a rate allocation scheme (Note: the corresponding coding scheme is guaranteed by Theorem 1) that achieves the sum rate bound $R_{\Sigma}(D)$.

In the next section, we apply the results obtained in Section II and Section III to analyze the quadratic Gaussian CEO problem. Although we focus on the finite discrete case with bounded distortion measure in Section II and Section III, many of our results hold for more general cases. For example, Theorem 1 can be extended to the Gaussian case with squared distortion measure by standard techniques [18], [19]. Specifically, the Markov lemma which is fundamental in the proof of Theorem 1 has been generalized by Oohama [19] to the Gaussian case. Lemma 3 in Section III also holds for the quadratic Gaussian case.

IV. QUADRATIC GAUSSIAN CEO PROBLEM

In this section, we will evaluate the achievable rate region defined in Theorem 1 for the Gaussian case with squared distortion measure.

Let $\{X(t), Y_1(t), Y_2(t), \dots, Y_L(t)\}_{t=1}^{\infty}$ be i.i.d Gaussian vectors such that $Y_1(t), Y_2(t), \dots, Y_L(t)$ are independent conditional on $X(t)$. We let L auxiliary random variables W_1, W_2, \dots, W_L be joint Gaussian⁶ with X, Y_1, Y_2, \dots, Y_L . Since for any $i \in \mathcal{I}_L$, $Y_i \rightarrow X \rightarrow Y_{\{i\}^c}$, $W_i \rightarrow Y_i \rightarrow \{Y_{\{i\}^c}, W_{\{i\}^c}\}$, we can get the following two equations⁷

$$Y_{\mathcal{I}_L} = \mathbf{1}X + N_{\mathcal{I}_L} \quad (3)$$

$$W_{\mathcal{I}_L} = \mathbf{L}Y_{\mathcal{I}_L} + T_{\mathcal{I}_L} \quad (4)$$

⁵It follows from a time sharing argument. Note that although $\vartheta(W_{\mathcal{I}_L})$ is convex, $\left\{ \bigcup_{W_{\mathcal{I}_L} \in \eta(D)} \vartheta(W_{\mathcal{I}_L}) \right\}$ may not be convex.

⁶It is not clear whether such a restriction will cause any loss of generality. But it greatly simplifies the computation.

⁷We can also let $Y_{\mathcal{I}_L} = \mathbf{k}X + N_{\mathcal{I}_L}$, where $\mathbf{k} = (k_1, k_2, \dots, k_L)^T$. But since all are zero-mean, (3) can always be got by scaling. In the case when $k_i = 0$, we can let $\sigma_{N_i} = \infty$ in (3).

where $N_{\mathcal{I}_L} = (N_1, N_2, \dots, N_L)^T$ are independent Gaussian noises at the L sensors with variance $\sigma_{N_i}^2$ ($i = 1, 2, \dots, L$) respectively; $\mathbf{L} = \text{diag}(l_1, l_2, \dots, l_L)$ is a scalar matrix; $T_{\mathcal{I}_L} = (T_1, T_2, \dots, T_L)^T$ are mutually independent Gaussian random variables with variance $\sigma_{T_i}^2$. $N_{\mathcal{I}_L}$ are independent of X and $T_{\mathcal{I}_L}$ are independent of $Y_{\mathcal{I}_L}$.

A. Distortion

We rewrite (3) in the form

$$X = \mathbf{a}^T Y_{\mathcal{I}_L} + V$$

where $\mathbf{a}^T = R_{Y_{\mathcal{I}_L} X}^T R_{Y_{\mathcal{I}_L} Y_{\mathcal{I}_L}}^{-1}$ and V is a Gaussian r.v. with variance $\sigma_V^2 = \sigma_X^2 - R_{Y_{\mathcal{I}_L} X}^T R_{Y_{\mathcal{I}_L} Y_{\mathcal{I}_L}}^{-1} R_{Y_{\mathcal{I}_L} X}$ and independent of $Y_{\mathcal{I}_L}$. Due to the fact that $W_{\mathcal{I}_L} \rightarrow Y_{\mathcal{I}_L} \rightarrow X$, $T_{\mathcal{I}_L}$ and V are independent. For the Gaussian case with squared distortion measure, the optimal estimate of X from $W_{\mathcal{I}_L}$, i.e. $E(X|W_{\mathcal{I}_L})$, is linear MMSE estimate. So we have

$$\hat{X}(W_{\mathcal{I}_L}) = R_{W_{\mathcal{I}_L} X}^T R_{W_{\mathcal{I}_L} W_{\mathcal{I}_L}}^{-1} W_{\mathcal{I}_L}$$

$$E(X - \hat{X}(W_{\mathcal{I}_L}))^2 = \sigma_X^2 - R_{W_{\mathcal{I}_L} X}^T R_{W_{\mathcal{I}_L} W_{\mathcal{I}_L}}^{-1} R_{W_{\mathcal{I}_L} X} \quad (5)$$

To get the expression for the covariances, we introduce the following lemma first.

Lemma 6: If matrix $R(\mathbf{l}; \mathbf{c})$ (where $\mathbf{l} = (l_1, \dots, l_L)^T$, $\mathbf{c} = (c_1, \dots, c_L)^T$, and $l_i, c_i \neq 0$) has the form

$$R(\mathbf{l}; \mathbf{c}) = \mathbf{1} \mathbf{1}^T + \text{diag}(c_1, \dots, c_L)$$

then we have

$$r(\mathbf{l}; \mathbf{c}) = \det R(\mathbf{l}; \mathbf{c}) = c_1 \dots c_L \left(1 + \frac{l_1^2}{c_1} + \dots + \frac{l_L^2}{c_L} \right)$$

$$R^{-1}(\mathbf{l}; \mathbf{c}) = \frac{1}{r(\mathbf{l}; \mathbf{c})}$$

$$\begin{pmatrix} r(l_2, \dots, l_L; c_2, \dots, c_L) & \dots & -l_1 l_L c_2 c_3 \dots c_{L-1} \\ \vdots & \ddots & \vdots \\ -l_1 l_L c_2 c_3 \dots c_{L-1} & \dots & r(l_1, \dots, l_{L-1}; c_1, \dots, c_{L-1}) \end{pmatrix}$$

and,

$$\mathbf{1}^T R^{-1}(\mathbf{l}; \mathbf{c}) = \frac{\left(\frac{l_1}{c_1}, \dots, \frac{l_L}{c_L} \right)}{1 + \frac{l_1^2}{c_1} + \dots + \frac{l_L^2}{c_L}}$$

$$\mathbf{1}^T R^{-1}(\mathbf{l}; \mathbf{c}) \mathbf{1} = \frac{\frac{l_1^2}{c_1} + \dots + \frac{l_L^2}{c_L}}{1 + \frac{l_1^2}{c_1} + \dots + \frac{l_L^2}{c_L}}.$$

Proof: See Appendix III. ■

Now we start to evaluate (5).

$$\begin{aligned} \mathbf{a}^T &= R_{Y_{\mathcal{I}_L} X}^T R_{Y_{\mathcal{I}_L} Y_{\mathcal{I}_L}}^{-1} = \sigma_X^2 (1, \dots, 1) R_{Y_{\mathcal{I}_L} Y_{\mathcal{I}_L}}^{-1} \\ R_{W_{\mathcal{I}_L} X} &= E[\mathbf{L}Y_{\mathcal{I}_L} + T_{\mathcal{I}_L}][Y_{\mathcal{I}_L}^T \mathbf{a} + V] = E[\mathbf{L}Y_{\mathcal{I}_L} Y_{\mathcal{I}_L}^T \mathbf{a}] \\ &= \mathbf{L} R_{Y_{\mathcal{I}_L} Y_{\mathcal{I}_L}} R_{Y_{\mathcal{I}_L} Y_{\mathcal{I}_L}}^{-1} \mathbf{1} \sigma_X^2 = \sigma_X^2 \mathbf{1} \end{aligned}$$

where $\mathbf{l} = (l_1, l_2, \dots, l_L)^T$.

Using Lemma 6, we directly get

$$\begin{aligned} E(X - \hat{X}(W_{\mathcal{I}_L}))^2 &= \sigma_X^2 - R_{W_{\mathcal{I}_L} X}^T R_{W_{\mathcal{I}_L} W_{\mathcal{I}_L}}^{-1} R_{W_{\mathcal{I}_L} X} \\ &= \frac{\sigma_X^2}{1 + \frac{1}{\sigma_{N_1}^2 + \frac{1}{\mu_1^2}} + \dots + \frac{1}{\sigma_{N_L}^2 + \frac{1}{\mu_L^2}}} \end{aligned}$$

where $\mu_i = \frac{l_i}{\sigma_{T_i}^2}$ for $i = 1, 2, \dots, L$.

So the distortion constraint $E(X - \hat{X}(W_{\mathcal{I}_L}))^2 \leq D$ becomes

$$\frac{1}{D} \leq \frac{1}{\sigma_X^2} + \frac{1}{\sigma_{N_1}^2 + \frac{1}{\mu_1^2}} + \dots + \frac{1}{\sigma_{N_L}^2 + \frac{1}{\mu_L^2}}. \quad (6)$$

Clearly, a nontrivial D should be in the range of $D_0 \leq D \leq \sigma_X^2$, where $D_0 = \frac{1}{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_{N_1}^2} + \dots + \frac{1}{\sigma_{N_L}^2}}$, which is the MMSE of X given $Y_{\mathcal{I}_L}$.

B. Rates

For joint Gaussian random vectors $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$, we have:

$$\begin{aligned} I(\mathbf{X}; \mathbf{Y}) &= \frac{1}{2} \log^+ \frac{\det R_{\mathbf{X}} \det R_{\mathbf{Y}}}{\det R_{\mathbf{X}\mathbf{Y}}} \\ I(\mathbf{X}; \mathbf{Y}|\mathbf{Z}) &= \frac{1}{2} \log^+ \frac{\det R_{\mathbf{X}\mathbf{Z}} \det R_{\mathbf{Y}\mathbf{Z}}}{\det R_{\mathbf{X}\mathbf{Y}\mathbf{Z}} \det R_{\mathbf{Z}}} \end{aligned}$$

where $\log^+ x = \max(\log x, 0)$.

So,

$$I(Y_{\mathcal{A}}; W_{\mathcal{A}}|W_{\mathcal{A}^c}) = \frac{1}{2} \log^+ \frac{\det R_{Y_{\mathcal{A}}W_{\mathcal{A}^c}} \det R_{W_{\mathcal{I}_L}}}{\det R_{Y_{\mathcal{A}}W} \det R_{W_{\mathcal{A}^c}}}.$$

By Lemma 6,

$$\det R_{W_{\mathcal{I}_L}} = (1 + \sigma_X^2 \sum_{i \in \mathcal{I}_L} \frac{\mu_i^2}{\mu_i^2 \sigma_{N_i}^2 + 1}) \prod_{i \in \mathcal{I}_L} \sigma_{T_i}^2 (\mu_i^2 \sigma_{N_i}^2 + 1)$$

$$\det R_{W_{\mathcal{A}^c}} = (1 + \sigma_X^2 \sum_{i \in \mathcal{A}^c} \frac{\mu_i^2}{\mu_i^2 \sigma_{N_i}^2 + 1}) \prod_{i \in \mathcal{A}^c} \sigma_{T_i}^2 (\mu_i^2 \sigma_{N_i}^2 + 1).$$

Also, notice that

$$\det R_{Y_{\mathcal{A}}W_{\mathcal{I}_L}} = \det R_{Y_{\mathcal{A}}W_{\mathcal{A}^c}} \det R_{T_{\mathcal{A}}},$$

we get the rate constraints

$$\begin{aligned} \sum_{i \in \mathcal{A}} R_i &\geq I(Y_{\mathcal{A}}; W_{\mathcal{A}}|W_{\mathcal{A}^c}) \\ &= \frac{1}{2} \log^+ \left\{ \left[\prod_{i \in \mathcal{A}} (\mu_i^2 \sigma_{N_i}^2 + 1) \right] \frac{\frac{1}{\sigma_X^2} + \sum_{i \in \mathcal{I}_L} \frac{\mu_i^2}{\mu_i^2 \sigma_{N_i}^2 + 1}}{\frac{1}{\sigma_X^2} + \sum_{i \in \mathcal{A}^c} \frac{\mu_i^2}{\mu_i^2 \sigma_{N_i}^2 + 1}} \right\}, \quad (7) \end{aligned}$$

$\forall \mathcal{A} \subseteq \mathcal{I}_L$.

By equation (6) and (7), we get the achievable rate region

$$\mathcal{R}_{in}(D) = \bigcup_{\mu \in \Lambda(D)} \mathcal{R}_{in}(\mu)$$

where

$$\Lambda(D) \triangleq \{ \mu = (\mu_1, \dots, \mu_L) : \frac{1}{D} \leq \frac{1}{\sigma_X^2} + \frac{1}{\sigma_{N_1}^2 + \frac{1}{\mu_1^2}} + \dots + \frac{1}{\sigma_{N_L}^2 + \frac{1}{\mu_L^2}} \},$$

$$\mathcal{R}_{in}(\mu) \triangleq \{ (R_1, \dots, R_L) : \sum_{i \in \mathcal{A}} R_i \geq \frac{1}{2} \log^+ \left\{ \left[\prod_{i \in \mathcal{A}} (\mu_i^2 \sigma_{N_i}^2 + 1) \right] \frac{\frac{1}{\sigma_X^2} + \sum_{i \in \mathcal{I}_L} \frac{\mu_i^2}{\mu_i^2 \sigma_{N_i}^2 + 1}}{\frac{1}{\sigma_X^2} + \sum_{i \in \mathcal{A}^c} \frac{\mu_i^2}{\mu_i^2 \sigma_{N_i}^2 + 1}} \right\}, \forall \mathcal{A} \subseteq \mathcal{I}_L \}.$$

Oohama [20] derived an expression different than ours and claimed it to be the rate-distortion region. But his result seems not very correct. Since one can let $r_i = 0$ for all i and show that the resulting rate region contains unachievable points.

The reason is probably the lack of proper constraints between r_i ($i \in \mathcal{I}_L$) and distortion D in his result.

Now we proceed to derive the minimum sum rate distortion function $R_{\Sigma}(D)$ and optimal rate allocation region $\vartheta(D)$ in the achievable rate region $\mathcal{R}_{in}(D)$.

By Lemma 3 in Section III, the sum rate constraint $I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L})$ is attainable. We have

$$\begin{aligned} R_{\Sigma}(D) &= \inf_{\mu \in \Lambda(D)} I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L}) \\ &= \inf_{\mu \in \Lambda(D)} \frac{1}{2} \log^+ \frac{\sigma_X^2}{D} \prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1). \end{aligned}$$

Note that both $\frac{1}{\sigma_X^2} + \frac{1}{\sigma_{N_1}^2 + \frac{1}{\mu_1^2}} + \dots + \frac{1}{\sigma_{N_L}^2 + \frac{1}{\mu_L^2}}$ and

$\frac{1}{2} \log^+ \frac{\sigma_X^2}{D} \prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1)$ are monotone increasing functions of μ_i^2 ($\forall i \in \mathcal{I}_L$). So in order to minimize $\frac{1}{2} \log^+ \frac{\sigma_X^2}{D} \prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1)$, the distortion constraint should be tight, i.e.,

$$\begin{aligned} R_{\Sigma}(D) &= \inf_{\mu \in \Lambda(D)} \frac{1}{2} \log^+ \frac{\sigma_X^2}{D} \prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1) \\ &= \frac{1}{D} = \frac{1}{\sigma_X^2} + \frac{1}{\sigma_{N_1}^2 + \frac{1}{\mu_1^2}} + \dots + \frac{1}{\sigma_{N_L}^2 + \frac{1}{\mu_L^2}} \frac{1}{2} \log^+ \frac{\sigma_X^2}{D} \prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1). \end{aligned}$$

So we can apply Lagrange multiplier to find the optimal $\hat{\mu} = (\hat{\mu}_1, \dots, \hat{\mu}_L)$ that minimizes $\frac{1}{2} \log^+ \frac{\sigma_X^2}{D} \prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1)$. Moreover, it is easy to see that minimizing $\frac{1}{2} \log^+ \frac{\sigma_X^2}{D} \prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1)$ is equivalent to minimizing $\prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1)$. Define

$$G(\mu) = \prod_{i=1}^L (\mu_i^2 \sigma_{N_i}^2 + 1) + \lambda \left(1 + \sigma_X^2 \sum_{i=1}^L \frac{\mu_i^2}{\mu_i^2 \sigma_{N_i}^2 + 1} - \frac{\sigma_X^2}{D} \right).$$

Without loss of generality, assume $\sigma_{N_1} \leq \dots \leq \sigma_{N_L}$. Find the largest \tilde{L} such that $\frac{\tilde{L}}{\sigma_{N_{\tilde{L}}}^2} - (\frac{1}{D_0(\tilde{L})} - \frac{1}{D}) \geq 0$, where $D_0(\tilde{L}) = \frac{1}{\frac{1}{\sigma_X^2} + \frac{1}{\sigma_{N_1}^2} + \dots + \frac{1}{\sigma_{N_{\tilde{L}}}^2}}$. Then we get

$$\begin{aligned} \hat{\mu}_i^2 &= \frac{\frac{\tilde{L}}{\sigma_{N_i}^2} - (\frac{1}{D_0(\tilde{L})} - \frac{1}{D})}{\sigma_{N_i}^2 (\frac{1}{D_0(\tilde{L})} - \frac{1}{D})}, \quad i = 1, \dots, \tilde{L}; \\ \hat{\mu}_i^2 &= 0, \quad i = \tilde{L} + 1, \dots, L. \end{aligned} \quad (8)$$

and,

$$R_{\Sigma}(D) = \frac{1}{2} \log^+ \left\{ \frac{\sigma_X^2}{D} \prod_{i=1}^{\tilde{L}} \left[\frac{\tilde{L}}{\sigma_{N_i}^2 (\frac{1}{D_0(\tilde{L})} - \frac{1}{D})} \right] \right\}. \quad (9)$$

The optimal rate allocation region $\vartheta(D)$ in $\mathcal{R}_{in}(D)$ is the convex hull of the $\tilde{L}!$ vertices $\{(R_1(\pi), \dots, R_L(\pi))\}_{\pi}$, where π is a permutation on the set $\mathcal{I}_{\tilde{L}}$. The coordinates of vertices

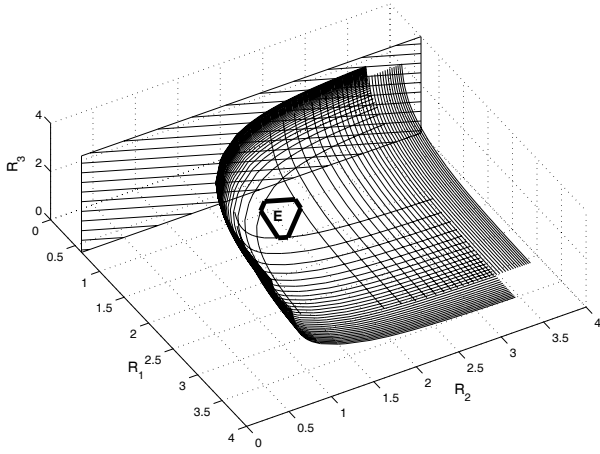


Fig. 2. Achievable rate region for the quadratic Gaussian CEO problem

$(R_1(\pi), \dots, R_L(\pi))$ are determined as follows

$$\begin{aligned}
 R_{\pi_1}(\pi) &= I(Y_{\{\pi(1)\}}; W_{\{\pi(1)\}} | W_{\{\pi(1)\}^c}) |_{\mu=\hat{\mu}} \\
 &= \frac{1}{2} \log^+ \left\{ \frac{\frac{\frac{\tilde{L}}{\sigma_{N\pi_1}^2}}{D_0(L) - \frac{1}{D}}}{1 - D \left[\frac{1}{\sigma_{N\pi_1}^2} - \frac{1}{L} \left(\frac{1}{D_0(L)} - \frac{1}{D} \right) \right]}} \right\}; \\
 R_{\pi_i}(\pi) &= I(Y_{\{\pi_1, \dots, \pi_i\}}; W_{\{\pi_1, \dots, \pi_i\}} | W_{\{\pi_1, \dots, \pi_i\}^c}) |_{\mu=\hat{\mu}} \\
 &\quad - I(Y_{\{\pi_1, \dots, \pi_{i-1}\}}; W_{\{\pi_1, \dots, \pi_{i-1}\}} | W_{\{\pi_1, \dots, \pi_{i-1}\}^c}) |_{\mu=\hat{\mu}} \\
 &= \frac{1}{2} \log^+ \left\{ \frac{\frac{\frac{\tilde{L}}{\sigma_{N\pi_i}^2}}{D_0(L) - \frac{1}{D}} \left(1 - D \sum_{j=1}^{i-1} \left[\frac{1}{\sigma_{N\pi_j}^2} - \frac{1}{L} \left(\frac{1}{D_0(L)} - \frac{1}{D} \right) \right] \right)}{1 - D \sum_{j=1}^i \left[\frac{1}{\sigma_{N\pi_j}^2} - \frac{1}{L} \left(\frac{1}{D_0(L)} - \frac{1}{D} \right) \right]} \right\}; \\
 &\quad i = 2, \dots, \tilde{L} \\
 R_i(\pi) &= 0; \quad i = \tilde{L} + 1, \dots, L.
 \end{aligned}$$

The 3-D case is illustrated in Fig. 2, where E corresponds to $\vartheta(D)$.

From the above analysis, it is clear that the number of sensors we use depends on the amount of available sum-rate and we always choose the sensors with small noise variances first. We call this phenomenon as "Generalized Waterfilling". We can also see that generally the optimal rate allocation scheme in the achievable rate region is not unique unless some further constraints are imposed.

For the quadratic Gaussian CEO problem studied in [4], [5], the SNRs at different sensors are identical, i.e. $\sigma_{N_1} = \dots = \sigma_{N_L} = \sigma_N$. In this case, we have

$$\hat{\mu}^2 = \hat{\mu}_1^2 = \dots = \hat{\mu}_L^2 = \frac{\frac{L}{\sigma_N^2} - \left(\frac{1}{D_0} - \frac{1}{D} \right)}{\sigma_N^2 \left(\frac{1}{D_0} - \frac{1}{D} \right)}$$

and

$$\begin{aligned}
 R_{\Sigma}(D) &= \frac{1}{2} \log^+ \left\{ \frac{\sigma_X^2}{D} \prod_{i=1}^L \left[\frac{L}{\sigma_{N_i}^2 \left(\frac{1}{D_0} - \frac{1}{D} \right)} \right] \right\} \\
 &= \frac{1}{2} \log^+ \left\{ \frac{\sigma_X^2}{D} \left(\frac{D \sigma_X^2 L}{D \sigma_X^2 L - \sigma_X^2 \sigma_N^2 + D \sigma_N^2} \right)^L \right\}. \quad (10)
 \end{aligned}$$

Actually equation (10) is exactly the sum-rate distortion function, not just an upper bound. The direct coding theorem is provided by Theorem 1. For the converse we apply Oohama's bounding technique [5, Section III] with the single difference that we do not let $L \rightarrow \infty$ at the end of his proof.

It is easy to check that if we let $L \rightarrow \infty$ in equation (10),

$$R_{\Sigma}(D) \rightarrow \frac{1}{2} \log^+ \left(\frac{\sigma_X^2}{D} \right) + \frac{\sigma_N^2}{2\sigma_X^2} \left[\frac{\sigma_X^2}{D} - 1 \right]^+,$$

which coincides with the result in [5]. Here

$$\left[\frac{\sigma_X^2}{D} - 1 \right]^+ = \max \left(\left[\frac{\sigma_X^2}{D} - 1 \right]^+, 0 \right).$$

As to the corresponding optimal rate allocation region, it is convex hull of the $L!$ vertices $\{(R_1(\pi), \dots, R_L(\pi))\}_{\pi}$ with

$$\begin{aligned}
 R_{\pi_1} &= \frac{1}{2} \log^+ \left\{ \frac{\frac{\frac{L}{\sigma_N^2}}{\frac{1}{D_0} - \frac{1}{D}}}{1 - D \left[\frac{1}{\sigma_N^2} - \frac{1}{L} \left(\frac{1}{D_0} - \frac{1}{D} \right) \right]} \right\} \\
 R_{\pi_i} &= \frac{1}{2} \log^+ \left\{ \frac{\frac{\frac{L}{\sigma_N^2}}{\frac{1}{D_0} - \frac{1}{D}} \left(1 - (i-1) D \left[\frac{1}{\sigma_N^2} - \frac{1}{L} \left(\frac{1}{D_0} - \frac{1}{D} \right) \right] \right)}{1 - i D \left[\frac{1}{\sigma_N^2} - \frac{1}{L} \left(\frac{1}{D_0} - \frac{1}{D} \right) \right]} \right\}, \\
 &\quad i = 2, \dots, L.
 \end{aligned}$$

where π is a permutation on the set \mathcal{I}_L .

A surprising consequence of the above rate allocation result is that:

Even when the SNRs are identical at different sensors, there exists a (actually uncountably infinite) rate allocation scheme that assigns different rates to different sensors but is still able to minimize the sum rate.

This is fundamentally different from the classic water filling results. The main reason is that the observed processes at different sensors are correlated. So it is possible to compensate the performance loss due to the decreasing of the rate allocated at one sensor by an increasing of same amount of rate at another sensor. Furthermore, the classic waterfilling method tries to equalize the marginal utility of different components. For our model, different components are not independent, so marginal utility is not a correct measure. Instead, we shall consider "conditional marginal utility".

V. CONCLUSION

In this paper, we studied the rate distortion region for the CEO problem with emphasis on the sum rate distortion function and the optimal rate allocation schemes in the achievable rate region. It will be extremely nice if one can find a complete characterization of the rate distortion region for the CEO problem or even just for the quadratic Gaussian case. But it seems impossible to overcome the difficulties with the

existing techniques. It's very clear that CEO problem is close related with many other distributed source coding problems, say Berger-Tung problem. One can expect that a complete solution to one of them will automatically lead to complete solutions to all the others.

For real applications, simple, robust and universal distributed coding schemes are preferred, but the existing literature [21] [22] in this direction is very limited. Much more work should be done in the near future. The final goal is to obtain a comprehensive understanding of distributed source coding systems, which involves the fundamental tradeoffs among system complexity, compression efficiency and performance robustness. A parallel tradeoff between diversity (i.e., performance robustness) and multiplexing (i.e., transmission efficiency) in MIMO systems has already been established by Zheng and Tse in [23].

APPENDIX I PROOF OF THEOREM 2

Proof:

$$\begin{aligned}
& n \sum_{i \in \mathcal{A}} R_i \\
&= \sum_{i \in \mathcal{A}} \log |\mathcal{C}_i^{(n)}| \geq \sum_{i \in \mathcal{A}} H(C_i^{(n)}) \\
&\geq H(C_{\mathcal{A}}^{(n)}) \geq H(C_{\mathcal{A}}^{(n)} | C_{\mathcal{A}^c}^{(n)}) \\
&= I(Y_{\mathcal{I}_L}^n; C_{\mathcal{A}}^{(n)} | C_{\mathcal{A}^c}^{(n)}) \\
&= H(Y_{\mathcal{I}_L}^n | C_{\mathcal{A}^c}^{(n)}) - H(Y_{\mathcal{I}_L}^n | C_{\mathcal{I}_L}^{(n)}) \\
&= \sum_{t=1}^n [H(Y_{\mathcal{I}_L}(t) | C_{\mathcal{A}^c}^{(n)} Y_{\mathcal{I}_L}^{t-1}) - H(Y_{\mathcal{I}_L}(t) | C_{\mathcal{I}_L}^{(n)} Y_{\mathcal{I}_L}^{t-1})] \\
&= \sum_{t=1}^n [H(Y_{\mathcal{I}_L}(t) | Z_{\mathcal{A}^c}(t)) - H(Y_{\mathcal{I}_L}(t) | Z_{\mathcal{I}_L}(t))] \\
&= \sum_{t=1}^n I(Y_{\mathcal{I}_L}(t); Z_{\mathcal{A}}(t) | Z_{\mathcal{A}^c}(t)),
\end{aligned}$$

where $Z_i(t) \triangleq (C_i^{(n)}, Y_{\mathcal{I}_L}^{t-1})$.

1) Since

$$\begin{aligned}
& I(Z_i(t); X(t)Y_{\{i\}^c}(t) | Y_i(t)) \\
&= I(X(t)Y_{\{i\}^c}(t); Z_i(t)Y_i(t)) - I(X(t)Y_{\{i\}^c}(t); Y_i(t)) \\
&= I(X(t)Y_{\{i\}^c}(t); C_i^{(n)}Y_{\mathcal{I}_L}^{t-1}Y_i(t)) \\
&\quad - I(X(t)Y_{\{i\}^c}(t); Y_i(t)) \\
&\leq I(X(t)Y_{\{i\}^c}(t); Y_i^n Y_{\{i\}^c}^{t-1}Y_i(t)) \\
&\quad - I(X(t)Y_{\{i\}^c}(t); Y_i(t)) \\
&= I(X(t)Y_{\{i\}^c}(t); Y_i(t)) - I(X(t)Y_{\{i\}^c}(t); Y_i(t)) \\
&= 0,
\end{aligned}$$

it implies that $Z_i(t)$ and $(X(t), Y_{\{i\}^c}(t))$ are conditionally independent given $Y_i(t)$.

2) $\hat{X}(t)$ is the t^{th} coordinate of $f_D(C_1^{(n)}, \dots, C_L^{(n)})$ so that we can write $\hat{X}(t)$ as a deterministic function of $(Z_1(t), \dots, Z_L(t))$. Let $\hat{X}(t) = g_t(Z_1(t), \dots, Z_L(t))$ and $D_t = Ed(X(t), \hat{X}(t))$.

From 1) and 2), we can see that $(Z_1(t), \dots, Z_L(t)) \in \mathcal{Z}_{out}(D_t)$.

Let $D = \frac{1}{n} \sum_{t=1}^n D_t$. Theorem 2 follows from Lemma 7 which we shall prove below:

Lemma 7: Let $(Z_1(t), \dots, Z_L(t)) \in \mathcal{Z}_{out}(D_t)$ and $D = \frac{1}{n} \sum_{t=1}^n D_t$, then there exists $(Z_1, \dots, Z_L) \in \mathcal{Z}_{out}(D)$ such that $I(Y_{\mathcal{I}_L}; Z_{\mathcal{A}} | Z_{\mathcal{A}^c}) = \frac{1}{n} \sum_{t=1}^n I(Y_{\mathcal{I}_L}(t); Z_{\mathcal{A}}(t) | Z_{\mathcal{A}^c}(t))$, for all $\mathcal{A} \subseteq \mathcal{I}_L$.

Proof: Let $\sum_{t=1}^n \lambda_t = 1$, $\lambda_t > 0$, $t = 1, \dots, n$.

Let γ be a random variable such that $P\{\gamma = t\} = \lambda_t$ with γ independent of $(X(t), Y_1(t), \dots, Y_L(t), Z_1(t), \dots, Z_L(t))$, $t = 1, \dots, n$.

Let $X = X(t), Y_i = Y_i(t), Z_i = (Z_i(t), t)$ if $\gamma = t$, for $i = 1, \dots, L$. We should note that $\forall i \in \mathcal{I}_L$, if Z_i is given, then γ is determined; while X and Y_i are independent of γ .

It's easy to check that

- 1) $Z_i \rightarrow Y_i \rightarrow (X, Y_{\{i\}^c})$ form a Markov chain for all $i \in \mathcal{I}_L$ since $I(Z_i; XY_{\{i\}^c} | Y_i) = \sum_{t=1}^n \lambda_t I(Z_i(t); X(t)Y_{\{i\}^c}(t) | Y_i(t)) = 0$.
- 2) There exists function $g : \mathcal{Z}_1 \times \dots \times \mathcal{Z}_L \rightarrow \mathcal{X}$ such that $Ed(X, \hat{X}) \leq D$ where $\hat{X} = g(Z_1, \dots, Z_L)$. This is because the decoder can use the function g_t corresponding to $(Z_1(t), \dots, Z_L(t))$, $t = 1, \dots, n$.

So $(Z_1, \dots, Z_L) \in \mathcal{Z}_{out}(D)$.

We can check that

$$I(Y_{\mathcal{I}_L}; Z_{\mathcal{A}} | Z_{\mathcal{A}^c}) = \sum_{t=1}^n \lambda_t I(Y_{\mathcal{I}_L}(t); Z_{\mathcal{A}}(t) | Z_{\mathcal{A}^c}(t)).$$

Now Lemma 7 follows by setting $\lambda_i = \frac{1}{n}$. ■

APPENDIX II PROOF OF LEMMA 5

Proof: The boundedness of $\tau(D)$ is obvious. So we only need to show that $\tau(D)$ is closed.

Consider a Cauchy sequence $\{P_{XY_1 \dots Y_L W_1 \dots W_L}^n\}_{n=1}^{\infty}$ in $\tau(D)$. Let $P_{XY_1 \dots Y_L W_1 \dots W_L}^*$ be the limiting distribution. It is obvious that $P_{XY_1 \dots Y_L W_1 \dots W_L}^*$ satisfies Property (i).

Without loss of generality, suppose $P_{Y_1}(y_1) > 0 \forall y_1 \in \mathcal{Y}_1$. Note that $P_{XY_1 \dots Y_L W_1 \dots W_L}^n \rightarrow P_{XY_1 \dots Y_L W_1 \dots W_L}^*$ implies $P_{XY_2 \dots Y_L W_1 \dots W_L | Y_1}^n \rightarrow P_{XY_2 \dots Y_L W_1 \dots W_L | Y_1}^*$, $P_{XY_2 \dots Y_L W_2 \dots W_L | Y_1}^n \rightarrow P_{XY_2 \dots Y_L W_2 \dots W_L | Y_1}^*$ and $P_{W_1 | Y_1}^n \rightarrow P_{W_1 | Y_1}^*$.

Since

$$\begin{aligned}
& P_{XY_2 \dots Y_L W_1 \dots W_L | Y_1}^n(x, y_2, \dots, y_L, w_1, \dots, w_L | y_1) \\
&= P_{XY_2 \dots Y_L W_2 \dots W_L | Y_1}^n(x, y_2, \dots, y_L, w_2, \dots, w_L | y_1) \\
& P_{W_1 | Y_1}^n(w_1 | y_1),
\end{aligned}$$

we have

$$\begin{aligned}
& P_{XY_2 \dots Y_L W_1 \dots W_L | Y_1}^*(x, y_2, \dots, y_L, w_1, \dots, w_L | y_1) \\
&= \lim_{n \rightarrow \infty} P_{XY_2 \dots Y_L W_1 \dots W_L | Y_1}^n(x, y_2, \dots, y_L, w_1, \dots, w_L | y_1) \\
&= \lim_{n \rightarrow \infty} P_{XY_2 \dots Y_L W_2 \dots W_L | Y_1}^n(x, y_2, \dots, y_L, w_2, \dots, w_L | y_1) \\
&\quad P_{W_1 | Y_1}^n(w_1 | y_1) \\
&= \lim_{n \rightarrow \infty} P_{XY_2 \dots Y_L W_2 \dots W_L | Y_1}^n(x, y_2, \dots, y_L, w_2, \dots, w_L | y_1) \\
&\quad \lim_{n \rightarrow \infty} P_{W_1 | Y_1}^n(w_1 | y_1) \\
&= P_{XY_2 \dots Y_L W_2 \dots W_L | Y_1}^*(x, y_2, \dots, y_L, w_2, \dots, w_L | y_1) \\
&\quad P_{W_1 | Y_1}^*(w_1 | y_1), \forall x, y_1, \dots, y_L, w_1, \dots, w_L.
\end{aligned}$$

Hence under $P_{XY_1 \dots Y_L W_1 \dots W_L}^*$, we have $W_1 \rightarrow Y_1 \rightarrow (X, Y_{\{1\}^c}, W_{\{1\}^c})$. Similarly, we can show that under $P_{XY_1 \dots Y_L W_1 \dots W_L}^*$, $W_i \rightarrow Y_i \rightarrow (X, Y_{\{i\}^c}, W_{\{i\}^c})$, $i = 2, \dots, L$. So $P_{XY_1 \dots Y_L W_1 \dots W_L}^*$ satisfies Property (ii).

Let f_n be the function associated with $P_{XY_1 \dots Y_L W_1 \dots W_L}^n$ such that $E_{P_{XY_1 \dots Y_L W_1 \dots W_L}^n} d(X, f_n(W_1, \dots, W_L)) \leq D$. Since $\max(|\mathcal{X}|, |\mathcal{W}_1|, \dots, |\mathcal{W}_L|) < \infty$, there are only finite number of functions from $\mathcal{W}_1 \times \dots \times \mathcal{W}_L$ to \mathcal{X} . So there exists a function f^* that appears infinite times in $\{f_n\}_{n=1}^\infty$. Let $\{n_k\}_{k=1}^\infty$ be a subsequence of $\{n\}_{n=1}^\infty$ such that $f_{n_k} = f^*$. We have

$$\begin{aligned}
& E_{P_{XY_1 \dots Y_L W_1 \dots W_L}^*} d(X, f^*(W_1, \dots, W_L)) \\
&= \lim_{k \rightarrow \infty} E_{P_{XY_1 \dots Y_L W_1 \dots W_L}^{n_k}} d(X, f^*(W_1, \dots, W_L)) \\
&\leq D.
\end{aligned}$$

So $P_{XY_1 \dots Y_L W_1 \dots W_L}^*$ also satisfies Property (iii). Hence $P_{XY_1 \dots Y_L W_1 \dots W_L}^* \in \tau(D)$ and we can conclude that $\tau(D)$ is compact. ■

APPENDIX III PROOF OF LEMMA 6

Proof: Firstly, we consider the following special case:

If matrix $R(c_1, c_2, \dots, c_L)$ of the form

$$R(c_1, c_2, \dots, c_L) = \begin{pmatrix} 1 + c_1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 + c_L \end{pmatrix}$$

We get

$$\begin{aligned}
r(c_1, c_2, \dots, c_L) &= \det R(c_1, c_2, \dots, c_L) \\
&= \det \begin{pmatrix} 1 + c_1 & 1 & \cdots & 1 \\ -c_1 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -c_1 & 0 & \cdots & c_L \end{pmatrix} \\
&= (1 + c_1)c_2 \cdots c_L + c_1 c_3 \cdots c_L + \cdots + c_1 c_2 \cdots c_{L-1} \\
&= c_1 \cdots c_L \left(1 + \frac{1}{c_1} + \cdots + \frac{1}{c_L}\right).
\end{aligned}$$

Since

$$R^{-1}(c_1, c_2, \dots, c_L) = \frac{1}{r(c_1, c_2, \dots, c_L)} \begin{pmatrix} M_{11} & -M_{21} & \cdots & (-1)^{1+L} M_{L1} \\ -M_{12} & M_{22} & \cdots & (-1)^{2+L} M_{L2} \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{1+L} M_{1L} & (-1)^{2+L} M_{2L} & \cdots & M_{LL} \end{pmatrix},$$

where M_{ij} is a minor of $R(c_1, c_2, \dots, c_L)$, obtained by taking the determinant of remainder of $R(c_1, c_2, \dots, c_L)$ with row i and column j ‘‘crossed out’’; and

$$M_{ii} = r(c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_L);$$

$$M_{ij} = (-1)^{i+j-1} \frac{c_1 \cdots c_L}{c_i c_j}$$

So,

$$\mathbf{1}^T R^{-1}(c_1, c_2, \dots, c_L) = \frac{\left(\frac{1}{c_1}, \dots, \frac{1}{c_L}\right)}{1 + \frac{1}{c_1} + \cdots + \frac{1}{c_L}};$$

$$\mathbf{1}^T R^{-1}(c_1, c_2, \dots, c_L) \mathbf{1} = \frac{\frac{1}{c_1} + \cdots + \frac{1}{c_L}}{1 + \frac{1}{c_1} + \cdots + \frac{1}{c_L}}.$$

Back to $R(\mathbf{l}; \mathbf{c})$. Since $R(\mathbf{l}; \mathbf{c}) = \mathbf{1} \mathbf{1}^T + \text{diag}(c_1, \dots, c_L)$, $\det R(\mathbf{l}; \mathbf{c}) = l_1^2 \cdots l_L^2 \det R\left(\frac{c_1}{l_1^2}, \dots, \frac{c_L}{l_L^2}\right)$. Apply the above results,

$$r(\mathbf{l}; \mathbf{c}) = \det R(\mathbf{l}; \mathbf{c}) = c_1 \cdots c_L \left(1 + \frac{l_1^2}{c_1} + \cdots + \frac{l_L^2}{c_L}\right)$$

Similarly

$$M_{ii} = r(l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_L; c_1, \dots, c_{i-1}, c_{i+1}, \dots, c_L);$$

$$M_{ij} = (-1)^{i+j-1} l_i l_j \frac{c_1 \cdots c_L}{c_i c_j}$$

and,

$$\mathbf{1}^T R^{-1}(\mathbf{l}; \mathbf{c}) = \frac{\left(\frac{l_1}{c_1}, \dots, \frac{l_L}{c_L}\right)}{1 + \frac{l_1^2}{c_1} + \cdots + \frac{l_L^2}{c_L}};$$

$$\mathbf{1}^T R^{-1}(\mathbf{l}; \mathbf{c}) \mathbf{1} = \frac{\frac{l_1^2}{c_1} + \cdots + \frac{l_L^2}{c_L}}{1 + \frac{l_1^2}{c_1} + \cdots + \frac{l_L^2}{c_L}}.$$

ACKNOWLEDGMENT

The authors wish to thank two anonymous reviewers for their valuable comments and suggestions.

REFERENCES

- [1] S. I. Gel'fand and M. S. Pinsker, ‘‘Coding of sources on the basis of observations with incomplete information’’. *Problems of Information Transmission*, 15(2):115-125, 1979.
- [2] T. J. Flynn and R. M. Gray, ‘‘Encoding of correlated observations.’’ *IEEE Trans. Inform. Theory*, vol. 33, pp. 773-787, Nov. 1987.
- [3] T. Berger, Z. Zhang, and H. Viswanathan, ‘‘The CEO problem.’’ *IEEE Trans. Inform. Theory*, vol. 42, pp. 887-902, May 1996.
- [4] H. Viswanathan and T. Berger, ‘‘The quadratic Gaussian CEO problem.’’ *IEEE Trans. Inform. Theory*, vol. 43, pp. 1549-1559, Sept. 1997.
- [5] Yasutada Oohama, ‘‘The rate-distortion function for the quadratic Gaussian CEO problem.’’ *IEEE Trans. on Inform. Theory*, vol. 44, no. 3, pp. 1057-1070, May 1998.

- [6] P. Viswanath, "Sum rate of multiterminal gaussian source coding," submitted to *IEEE Trans. on Information Theory*, in DIMACS Series in Discrete Mathematics and Theoretical Computer Science.
- [7] T. M. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources," *IEEE Trans. Inform. Theory*, vol. 21, pp 226-228, Mar. 1975.
- [8] S. Y. Tung, "Multiterminal source coding," Ph.D. dissertation, School of Electrical Engineering, Cornell Univ., Ithaca, NY, May 1978.
- [9] T. Berger, "Multiterminal source coding," in *The Information Theory Approach to Communications (CISM Courses and Lectures, no. 229)*, G. Longo, Ed. Vienna/New York: Springer-Verlag, 1978, pp. 171-231.
- [10] T. S. Han and K. Kobayashi, "A unified achievable rate region for a general class of multiterminal source coding systems," *IEEE Trans. Inform. Theory*, vol. 26, no. 3, pp. 277-288, May 1980.
- [11] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Info. Theory*, vol. 22, no. 1, pp. 1-10, Jan. 1976.
- [12] Stark C. Draper, Gregory W. Wornell, "Successively structured CEO problems," *ISIT 2002*, pp. 65, Lausanne, Switzerland, June 30-July 5, 2002
- [13] T. Berger, K. Housewright, J. Omura, S. Y. Tung, and J. Wolfowitz, "An upper bound on the rate-distortion function for source coding with partial side information at the decoder," *IEEE Trans. Inform. Theory*, vol. IT-25, pp.664-666, Nov., 1979.
- [14] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Info. Theory*, vol.IT-19, pp. 471-480, Jul. 1973.
- [15] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inform. Theory*, Vol45(1), Jan. 1999, pp. 106-117.
- [16] D. N. C. Tse, S. V. Hanly, "Multiaccess fading channels—Part I: poly-matroid structure, optimal resource allocation and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, No. 7, pp 2796-2815, Nov. 1998.
- [17] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
- [18] A. D. Wyner, "The rate-distortion function for source coding with side information at the decoder-II: General sources," *Inform. Contr.*, vol. 38, pp. 60-80, Jul. 1978.
- [19] Yasutada Oohama, "Gaussian multiterminal source coding," *IEEE Trans. on Inform. Theory*, vol. 43, no. 6, pp. 1912-1923, Nov. 1997.
- [20] Y. Oohama, "Multiterminal source coding for correlated memoryless Gaussian sources with several side informations at the decoder," in *IEEE Information Theory and Communications Workshop*, Kruger National Park, South Africa, page 100, 1999.
- [21] P. Ishwar, R. Puri, S. S. Pradhan and K. Ramchandran, "On compression for robust estimation in sensor networks," *ISIT 2003*, pp. 193, Yokohama, Japan, June 29-July 4, 2003.
- [22] J. Chen and T. Berger, "Robust coding schemes for distributed sensor networks with unreliable sensors," submitted to *ISIT 2004*.
- [23] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp.1073-1096, May 2003.



Jun Chen received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001 and the M.S. degree in electrical engineering from Cornell University, NY, USA in 2004. He is currently pursuing his Ph.D. degree in Cornell University. His research interests are in information theory, communication and networking.



Xin Zhang received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2001. Now she is an MS/Ph.D. student in Electrical Engineering, Cornell University. Her research interests include wireless communication and sensor network.



Toby Berger was born in New York, NY on September 4, 1940. He received the B.E. degree in electrical engineering from Yale University, New Haven, CT in 1962, and the M.S. and Ph.D. degrees in applied mathematics from Harvard University, Cambridge, MA in 1964 and 1966, respectively.

From 1962 to 1968 he was a Senior Scientist at Raytheon Company, Wayland, MA, specializing in communication theory, information theory, and coherent signal processing. In 1968 he joined the faculty of Cornell University, Ithaca, NY where

he is presently the Irwin and Joan Jacobs Professor of Engineering. His research interests include information theory, random fields, communication networks, wireless communications, video compression, voice and signature compression and verification, neuroinformation theory, quantum information theory, and coherent signal processing. He is the author of *Rate Distortion Theory: A Mathematical Basis for Data Compression* and a co-author of *Digital Compression for Multimedia: Principles and Standards*, and *Information Measures for Discrete Random Fields*.

Berger has served as editor-in-chief of the *IEEE Transactions on Information Theory* and as president of the *IEEE Information Theory Group*. He has been a Fellow of the Guggenheim Foundation, the Japan Society for Promotion of Science, the Ministry of Education of the People's Republic of China and the Fulbright Foundation. In 1982 he received the Frederick E. Terman Award of the American Society for Engineering Education, and in 2002 he received the Shannon Award from the *IEEE Information Theory Society*. Berger is a Fellow of the IEEE, a life member of Tau Beta Pi, and a blues harmonica player.



Stephen B. Wicker received the B.S.E.E. with High Honors from the University of Virginia in 1982. He received the M.S.E.E. from Purdue University in 1983 and a Ph.D. in Electrical Engineering from the University of Southern California in 1987. He is a Professor of Electrical and Computer Engineering at Cornell University. He is the author of *Codes, Graphs, and Iterative Decoding* (Kluwer, 2002), *Turbo Coding* (Kluwer, 1999), *Error Control Systems for Digital Communication and Storage* (Prentice Hall, 1995) and *Reed-Solomon Codes and Their*

Applications (IEEE Press, 1994). He has served as Associate Editor for *Coding Theory and Techniques* for the *IEEE Transactions on Communications*, and recently concluded his second term as a member of the Board of Governors of the *IEEE Information Theory Society*. Professor Wicker teaches and conducts research in wireless information networks, digital systems, self-configuring networks, and game theory. His current research focuses on the use of game theory in the development of highly distributed, adaptive networks. He was awarded the 1988 Cornell College of Engineering Michael Tien Teaching Award and the 2000 Cornell School of Electrical and Computer Engineering Teaching Award. He has served as a consultant in wireless telecommunications systems, error control coding, and cryptography for various companies and governments in North America, Europe, and Asia.