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, *Student Member, IEEE*, Toby Berger, *Fellow, IEEE*, and 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 signal-to-noise ratios 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 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 $(i = 1, 2, \ldots, L)$ may communicate information about its observations to the fusion center is limited to R_i bits per second.[1] 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.

Manuscript received July 15, 2003; revised February 1, 2004. The work of J. Chen and T. Berger was supported in part by the National Science Foundation (NSF) under Grant CCR-0330059, Grant CCR-1980616, and Grant ARO P-40116-PH-MUR, Cornell University, Ithaca, NY. The work of S. B. Wicker was supported in part by the MURI Program and the NSF ITR and Sensor Networking Programs, Cornell University, Ithaca, NY.

The authors are with Cornell University, Ithaca, NY 14853 USA (e-mail: jc353@cornell.edu).

Digital Object Identifier 10.1109/JSAC.2004.830888

[1]The 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.

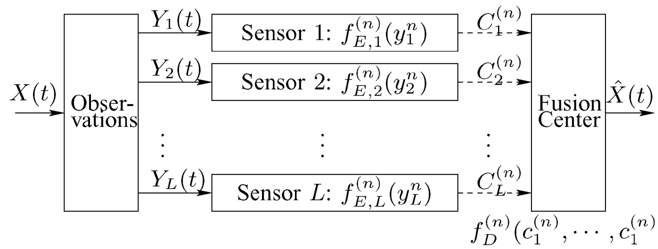


Fig. 1. Model of distributed sensor network.

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 signal-to-noise ratio (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 tradeoff between sum rate and distortion.

A. Note About Notation

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

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 $\mathcal{X} \times \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_L$, where L is the number of sensors, $\mathcal{X}$ is the common alphabet of the random variables $X(t)$ for $t = 1, 2, \dots, \mathcal{Y}_i\, (i = 1, 2, \cdots, L)$ 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 $R_i^{(n)} \geq (1/n)\log|\mathcal{C}_i^{(n)}|$. 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 $d^{(n)} = (1/n)E\sum_{t=1}^{n} d(X(t), \hat{X}(t))$, 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 $f_D^{(n)} : \mathcal{C}_1^{(n)} \times \cdots \times \mathcal{C}_L^{(n)} \to \mathcal{X}^n$, 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 Sections II and III, we assume $\max(|\mathcal{X}|, |\mathcal{Y}_1|, \dots, |\mathcal{Y}_L|) < \infty$ and the distortion measure $d : \mathcal{X} \times \mathcal{X} \to [0, d_{\max}]$ to be bounded, i.e., $d_{\max} < \infty$.

II. 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

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

and a decoder

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

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

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 $\mathcal{W}_{\text{in}}(D)$ 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.

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

Let

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

then $\mathcal{R}(D) \supseteq \mathcal{R}_{\text{in}}(D) \triangleq$ convex hull[2] of $\{\bigcup_{W_{\mathcal{I}_L} \in \mathcal{W}_{\text{in}}(D)} \mathcal{R}(W_{\mathcal{I}_L})\}$. We call $\mathcal{R}_{\text{in}}(D)$ the achievable rate region with respect to distortion D.

The proof of this theorem is standard, which is based on Cover's random binning argument [7] and a Markov lemma [8]–[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 $\{\mathcal{C}^{(n)} = (\mathcal{C}_1^{(n)}, \mathcal{C}_2^{(n)}, \dots, \mathcal{C}_L^{(n)})\}$[3] (where $\mathcal{C}_i^{(n)}$ denotes the codebook of sensor i) as follows.

At sensor i, independently generate $M_i = 2^{n(I(Y_i;W_i)+\varepsilon)}$ codewords of blocklength n, index them $C_i^{(n)}(j), j = 1, 2, \dots, M_i$, and let $\mathcal{C}_i^{(n)} = \{C_i^{(n)}(j)\}_{j=1}^{M_i}$. The codewords are generated by drawing independent identically distributed (i.i.d.) symbols from the marginal distribution P_{W_i}. Randomly assign the codewords to one of 2^{nR_i} bins[4] using a uniform distribution over the indices of the bins.

Encoding Scheme: At sensor i, given observation y_i^n, if it is typical, map it onto the $C_i^{(n)}(j) \in \mathcal{C}_i^{(n)}$ with the smallest index j such that $(y_i^n, C_i^{(n)}(j))$ are jointly typical. Let $C_i^{(n)}(y_i^n)$ denotes the $C_i^{(n)}$ onto which y_i^n is mapped. (Note: By Markov lemma, $P((C_1^{(n)}(Y_1^n), \dots, C_L^{(n)}(Y_L^n))$ are jointly typical$) \to 1$ as $n \to \infty$.) The index of the bin which contains $C_i^{(n)}(y_i^n)$ is sent. Let $b_i(y_i^n)$ denote this bin index. If y_i^n is not typical or there does not exist $C_i^{(n)}(j) \in \mathcal{C}_i^{(n)}$ such that $(y_i^n, C_i^{(n)}(j))$ 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.

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 the codeword $C_i^{(n)}$ is

[2]It follows directly from a time sharing argument.

[3]Here, $\mathcal{C}_i^{(n)}$ 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 is why here we have 2^{nR_i} bins at encoder i, while generally $|\mathcal{C}_i^{(n)}| > 2^{nR_i}$.

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

in $\mathcal{B}_i(b_i)$ (where $\mathcal{B}_i(b_i)$ denotes the bin with index b_i at sensor i) and $(C_1^{(n)}, C_2^{(n)}, \ldots, C_L^{(n)})$ are jointly typical, then decode it as $(\hat{C}_1^{(n)}, \hat{C}_2^{(n)}, \ldots, \hat{C}_L^{(n)})$; 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 $\hat{x}(k) = f(\hat{C}_{1,k}^{(n)}, \hat{C}_{2,k}^{(n)}, \ldots, \hat{C}_{L,k}^{(n)})$ for $k = 1, 2, \ldots, n$. Here, $\hat{C}_{i,k}^{(n)}$ is the kth symbol of the codeword $\hat{C}_i^{(n)}$.

B. Outer Bound

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

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

Let

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

then

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

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 $i \in \mathcal{I}_L$

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

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) $V_i \to Y_i \to (X, Y_{\{i\}^c})$.
ii) There exists a function $f_i : \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_{i-1} \times \mathcal{Y}_{i+1} \times \cdots \times \mathcal{Y}_L \times \mathcal{V}_i \to \mathcal{X}$ such that $Ed(X, \hat{X}) \leq D$ where $\hat{X} = f_i(Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_L, V_i)$.

That is to say, the upper bound [reduced from inner bound $\mathcal{R}_{\text{in}}(D)$] 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 $1, \ldots, i-1, i+1, \ldots, L$ are loose enough to guarantee the perfect recovery of $\{Y_1(t), \ldots, Y_{i-1}(t), Y_{i+1}(t), \ldots, Y_L(t)\}_{t=1}^{\infty}$. (Note: By Slepian–Wolf Theorem [14], if

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

then $\{Y_1(t), \ldots, Y_{i-1}(t), Y_{i+1}(t), \ldots, Y_L(t)\}_{t=1}^{\infty}$ 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 $\{Y_1(t), \ldots, Y_{i-1}(t), Y_{i+1}(t), \ldots, Y_L(t)\}_{t=1}^{\infty}$.

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 Schemes in the Achievable Rate Region

In this section, we analyze the minimum sum rate in the achievable rate region with respect to distortion D, i.e.,

$$R_{\sum}(D) \triangleq \inf_{(R_1, R_2, \ldots, R_L) \in \mathcal{R}_{\text{in}}(D)} \sum_{i=1}^{L} R_i. \tag{1}$$

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

$$\vartheta(D) \triangleq \left\{ (R_1, R_2, \ldots, R_L) : (R_1, R_2, \ldots, R_L) \in \mathcal{R}_{\text{in}}(D), \sum_{i=1}^{L} R_i = R_{\sum}(D) \right\} \tag{2}$$

in which every element corresponds to a rate allocation scheme that minimizes the sum rate in the achievable rate region with respect to distortion D.

A major step toward the solutions to (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 [6] and [10]. A similar combinatorial structure called polymatroid has been applied to study the capacity region of multiaccess fading channel in [16].

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

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

is a *contra-polymatroid* if f satisfies

1) $f(\emptyset) = 0$ (normalized).
2) $f(S) \leq f(T)$ if $S \subset T$ (nondecreasing).
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 $W_i \to Y_i \to (Y_{\{i\}^c}, W_{\{i\}^c})$ for all $i \in \mathcal{I}_L$ under the probability distribution $P_{Y_1 \ldots Y_L W_1 \ldots W_L}$, then

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

is a contra-polymatroid.

Proof: Let

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

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

1) By definition.

2)

$$\begin{aligned} f(\mathcal{T}) &= I(Y_{\mathcal{T}}; W_{\mathcal{T}} \,|\, W_{\mathcal{T}^c}) \\ &= I(Y_{\mathcal{T}}; W_{\mathcal{S}} \,|\, W_{\mathcal{S}^c}) + I\left(Y_{\mathcal{T}}; W_{\mathcal{T} \setminus \mathcal{S}} \,|\, W_{\mathcal{T}^c}\right) \\ &\geq I(Y_{\mathcal{T}}; W_{\mathcal{S}} \,|\, W_{\mathcal{S}^c}) \geq I(Y_{\mathcal{S}}; W_{\mathcal{S}} \,|\, W_{\mathcal{S}^c}) \\ &= f(\mathcal{S}), \quad \text{if } \mathcal{S} \subseteq \mathcal{T}. \end{aligned}$$

3) Since

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

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

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

■

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 $\mathbf{v}_{\pi(i)}(\pi) = f(\{\pi(1), \ldots, \pi(i)\}) - f(\{\pi(1), \ldots, \pi(i-1)\})$, $i = 2, \ldots, L$. Then, the points $v(\pi)$ 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 $\mathcal{I}_L, (R_1(\pi), \ldots, R_L(\pi))$ gives a vertex of $\mathcal{R}(W_{\mathcal{I}_L})$ (where $R_{\pi(1)}(\pi) = I(W_{\{\pi(1)\}}; Y_{\{\pi(1)\}} \,|\, W_{\{\pi(1)\}^c})$, $R_{\pi(i)}(\pi) = I(W_{\{\pi(1),\ldots,\pi(i)\}}; Y_{\{\pi(1),\ldots,\pi(i)\}} \,|\, W_{\{\pi(1),\ldots,\pi(i)\}^c}) - I(W_{\{\pi(1),\ldots,\pi(i-1)\}}; Y_{\{\pi(1),\ldots,\pi(i-1)\}} \,|\, W_{\{\pi(1),\ldots,\pi(i-1)\}^c})$, $i = 2, \ldots, L$) 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 $R_{\sum}(W_{\mathcal{I}_L}) \triangleq \inf_{(R_1,\ldots,R_L) \in \mathcal{R}(W_{\mathcal{I}_L})} \sum_{i=1}^{L} R_i$, then

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

and we have

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

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

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

Then, $\tau(D)$ is compact.

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

Proof: See Appendix II. ■

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

Theorem 3:

1)

$$\begin{aligned} R_{\sum}(D) &= \min_{W_{\mathcal{I}_L} \in \mathcal{W}_{\text{in}}(D)} I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L}) \\ &= \min_{\tau(D)} I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L}) \end{aligned}$$

2) Let

$$\eta(D) \triangleq \left\{ W^*_{\mathcal{I}_L} : W^*_{\mathcal{I}_L} \in \mathcal{W}_{\text{in}}(D) \quad I\left(Y_{\mathcal{I}_L}; W^*_{\mathcal{I}_L}\right) = \min_{W_{\mathcal{I}_L} \in \mathcal{W}_{\text{in}}(D)} I(Y_{\mathcal{I}_L}; W_{\mathcal{I}_L}) \right\}.$$

We have $\vartheta(D) = $ convex hull[5] of $\{\bigcup_{W_{\mathcal{I}_L}\in\eta(D)}\vartheta(W_{\mathcal{I}_L})\}$.

Proof: $I(Y_{\mathcal{I}_L};W_{\mathcal{I}_L})$ is a continuous function of $P_{XY_1\cdots Y_LW_1\cdots W_L}$ 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 with respect to 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 $\mathcal{R}_{\text{in}}(D)$ 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 Sections II and III to analyze the quadratic Gaussian CEO problem. Although we focus on the finite discrete case with bounded distortion measure in Sections II and 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 $\{X(t),Y_1(t),Y_2(t),\ldots,Y_L(t)\}_{t=1}^\infty$ be i.i.d. Gaussian vectors such that $Y_1(t),Y_2(t),\ldots,Y_L(t)$ are independent conditional on $X(t)$. We let L auxiliary random variables $W_1,W_2,\ldots,W_L$ be joint Gaussian[6] with $X,Y_1,Y_2,\ldots,Y_L$. Since for any $i\in\mathcal{I}_L, Y_i\to X\to Y_{\{i\}^c}, W_i\to Y_i\to \{Y_{\{i\}^c},W_{\{i\}^c}\}$, we can get the following two equations[7]

$$Y_{\mathcal{I}_L} = (1,\ldots,1)^T X + N_{\mathcal{I}_L} \tag{3}$$
$$W_{\mathcal{I}_L} = \mathbf{L}Y_{\mathcal{I}_L} + T_{\mathcal{I}_L} \tag{4}$$

where $N_{\mathcal{I}_L} = (N_1,N_2,\ldots,N_L)^T$ are independent Gaussian noises at the L sensors with variance $\sigma^2_{N_i}\ (i=1,2,\ldots,L)$, respectively; $\mathbf{L}=\text{diag}\,(l_1,l_2,\ldots,l_L)$ is a scalar matrix; $T_{\mathcal{I}_L} = (T_1,T_2,\ldots,T_L)^T$ are mutually independent Gaussian random variables with variance $\sigma^2_{T_i}$. $N_{\mathcal{I}_L}$ 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 $\mathbf{a}^T = R^T_{Y_{\mathcal{I}_L}X}R^{-1}_{Y_{\mathcal{I}_L}Y_{\mathcal{I}_L}}$ and V is a Gaussian r.v. with variance $\sigma_V^2 = \sigma_X^2 - R^T_{Y_{\mathcal{I}_L}X}R^{-1}_{Y_{\mathcal{I}_L}Y_{\mathcal{I}_L}}R_{Y_{\mathcal{I}_L}X}$ and independent of $Y_{\mathcal{I}_L}$. Due to the fact that $W_{\mathcal{I}_L}\to Y_{\mathcal{I}_L}\to X, T_{\mathcal{I}_L}$ and V are independent. For the Gaussian case with squared distortion measure, the optimal estimate of X from $W_{\mathcal{I}_L}$, i.e., $E(X\mid W_{\mathcal{I}_L})$, is linear minimum mean-square error (MMSE) estimate. So we have

$$\hat{X}\left(W_{\mathcal{I}_L}\right) = R^T_{W_{\mathcal{I}_L}X}R^{-1}_{W_{\mathcal{I}_L}W_{\mathcal{I}_L}}W_{\mathcal{I}_L}$$
$$E\left(X-\hat{X}\left(W_{\mathcal{I}_L}\right)\right)^2 = \sigma_X^2 - R^T_{W_{\mathcal{I}_L}X}R^{-1}_{W_{\mathcal{I}_L}W_{\mathcal{I}_L}}R_{W_{\mathcal{I}_L}X}. \tag{5}$$

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

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

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

then, we have the equation, shown at the bottom of the page.

Proof: See Appendix III. ■

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

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

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

$$r(\mathbf{l};\mathbf{c}) = \det R(\mathbf{l};\mathbf{c}) = c_1\ldots c_L\left(1+\frac{l_1^2}{c_1}+\cdots+\frac{l_L^2}{c_L}\right)$$
$$R^{-1}(\mathbf{l};\mathbf{c}) = \frac{1}{r(\mathbf{l};\mathbf{c})}\times\begin{pmatrix} r(l_2,\ldots,l_L;c_2,\ldots,c_L) & \cdots & -l_1l_Lc_2c_3\ldots c_{L-1}\\ \vdots & \ddots & \vdots \\ -l_1l_Lc_2c_3\ldots c_{L-1} & \cdots & r(l_1,\ldots,l_{L-1};c_1,\ldots,c_{L-1})\end{pmatrix}$$

and

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

Now, we start to evaluate (5)

$$\mathbf{a}^T = R_{Y_{\mathcal{I}_L}X}^T R_{Y_{\mathcal{I}_L}Y_{\mathcal{I}_L}}^{-1} = \sigma_X^2(1,\ldots,1)R_{Y_{\mathcal{I}_L}Y_{\mathcal{I}_L}}^{-1}$$
$$R_{W_{\mathcal{I}_L}X} = E[\mathbf{L}Y_{\mathcal{I}_L}+T_{\mathcal{I}_L}]\left[Y_{\mathcal{I}_L}^T\mathbf{a}+V\right] = E\left[\mathbf{L}Y_{\mathcal{I}_L}Y_{\mathcal{I}_L}^T\mathbf{a}\right] = \mathbf{L}R_{Y_{\mathcal{I}_L}Y_{\mathcal{I}_L}}R_{Y_{\mathcal{I}_L}Y_{\mathcal{I}_L}}^{-1}(1,\ldots,1)^T\sigma_X^2 = \sigma_X^2\mathbf{l}$$

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

Using Lemma 6, we directly get

$$E\left(X-\hat{X}\left(W_{\mathcal{I}_L}\right)\right)^2 = \sigma_X^2 - R_{W_{\mathcal{I}_L}X}^T R_{W_{\mathcal{I}_L}W_{\mathcal{I}_L}}^{-1} R_{W_{\mathcal{I}_L}X} = \frac{1}{\frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_1}^2+\frac{1}{\mu_1^2}}+\cdots+\frac{1}{\sigma_{N_L}^2+\frac{1}{\mu_L^2}}}$$

where $\mu_i^2 = (l_i^2/\sigma_{T_i}^2)$ for $i = 1, 2, \ldots, L$.

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

$$\frac{1}{D} \le \frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_1}^2+\frac{1}{\mu_1^2}}+\cdots+\frac{1}{\sigma_{N_L}^2+\frac{1}{\mu_L^2}}. \tag{6}$$

Clearly, a nontrivial D should be in the range of $D_0 \le D \le \sigma_X^2$, where $D_0 = (1/\sigma_X^2 + 1/\sigma_{N_1}^2 + \cdots + 1/\sigma_{N_L}^2)^{-1}$, 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

$$I(\mathbf{X};\mathbf{Y}) = \frac{1}{2}\log^+\frac{\det R_{\mathbf{X}}\det R_{\mathbf{Y}}}{\det R_{\mathbf{XY}}}$$
$$I(\mathbf{X};\mathbf{Y}\,|\,\mathbf{Z}) = \frac{1}{2}\log^+\frac{\det R_{\mathbf{XZ}}\det R_{\mathbf{YZ}}}{\det R_{\mathbf{XYZ}}\det R_{\mathbf{Z}}}$$

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

So

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

By Lemma 6

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

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

$$\sum_{i\in\mathcal{A}}R_i \ge I\left(Y_{\mathcal{A}};W_{\mathcal{A}}\,|\,W_{\mathcal{A}^c}\right) = \frac{1}{2}\log^+\left\{\left[\prod_{i\in\mathcal{A}}\left(\mu_i^2\sigma_{N_i}^2+1\right)\right]\frac{\frac{1}{\sigma_X^2}+\sum_{i\in\mathcal{I}_L}\frac{\mu_i^2}{\mu_i^2\sigma_{N_i}^2+1}}{\frac{1}{\sigma_X^2}+\sum_{i\in\mathcal{A}^c}\frac{\mu_i^2}{\mu_i^2\sigma_{N_i}^2+1}}\right\},\quad \forall\mathcal{A}\subseteq\mathcal{I}_L. \tag{7}$$

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

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

where

$$\Lambda(D) \triangleq \left\{\boldsymbol{\mu} = (\mu_1,\ldots,\mu_L) : \frac{1}{D} \le \frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_1}^2+\frac{1}{\mu_1^2}}+\cdots+\frac{1}{\sigma_{N_L}^2+\frac{1}{\mu_L^2}}\right\}$$
$$\mathcal{R}_{\text{in}}(\boldsymbol{\mu}) \triangleq \left\{(R_1,\ldots,R_L) : \sum_{i\in\mathcal{A}}R_i \ge \frac{1}{2}\log^+\left\{\left[\prod_{i\in\mathcal{A}}\left(\mu_i^2\sigma_{N_i}^2+1\right)\right]\times\frac{\frac{1}{\sigma_X^2}+\sum_{i\in\mathcal{I}_L}\frac{\mu_i^2}{\mu_i^2\sigma_{N_i}^2+1}}{\frac{1}{\sigma_X^2}+\sum_{i\in\mathcal{A}^c}\frac{\mu_i^2}{\mu_i^2\sigma_{N_i}^2+1}}\right\},\quad\forall\mathcal{A}\subseteq\mathcal{I}_L\right\}.$$

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 $\mathcal{R}_{\text{in}}(D)$.

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

$$R_\Sigma(D) = \inf_{\boldsymbol{\mu}\in\Lambda(D)} I\left(Y_{\mathcal{I}_L};W_{\mathcal{I}_L}\right) = \inf_{\boldsymbol{\mu}\in\Lambda(D)}\frac{1}{2}\log^+\left[\left(1+\sum_{i=1}^{L}\frac{\mu_i^2\sigma_X^2}{\mu_i^2\sigma_{N_i}^2+1}\right)\prod_{i=1}^{L}\left(\mu_i^2\sigma_{N_i}^2+1\right)\right].$$

Note that both $1/\sigma_X^2 + \sum_{i=1}^L(1/(\sigma_{N_1}^2 + 1/\mu_1^2))$ and $(1 + \sum_{i=1}^L(\mu_i^2\sigma_X^2/(\mu_i^2\sigma_{N_i}^2+1)))\prod_{i=1}^L(\mu_i^2\sigma_{N_i}^2+1)$ are monotone increasing functions of $\mu_i^2 (\forall i \in \mathcal{I}_L)$. So in order to minimize $(1 + \sum_{i=1}^L(\mu_i^2\sigma_X^2/(\mu_i^2\sigma_{N_i}^2+1)))\prod_{i=1}^L(\mu_i^2\sigma_{N_i}^2+1)$, the distortion constraint should be tight

$$R_\Sigma(D) = \inf_{\boldsymbol{\mu}\in\Lambda(D)}\frac{1}{2}\log^+\left[\left(1+\sum_{i=1}^{L}\frac{\mu_i^2\sigma_X^2}{\mu_i^2\sigma_{N_i}^2+1}\right)\prod_{i=1}^{L}\left(\mu_i^2\sigma_{N_i}^2+1\right)\right] = \inf_{\frac{1}{D}=\frac{1}{\sigma_X^2}+\frac{1}{\sigma_{N_1}^2+\frac{1}{\mu_1^2}}+\cdots+\frac{1}{\sigma_{N_L}^2+\frac{1}{\mu_L^2}}}\frac{1}{2}\log^+\frac{\sigma_X^2}{D}\prod_{i=1}^{L}\left(\mu_i^2\sigma_{N_i}^2+1\right).$$

So we can apply Lagrange multiplier to find the optimal $\hat{\mu} = (\hat{\mu}_1,\ldots,\hat{\mu}_L)$ that minimizes $(1/2)\log^+(\sigma_X^2/D)$

$\prod_{i=1}^{L}(\mu_i^2\sigma_{N_i}^2+1)$. Moreover, it is easy to see that minimizing $(1/2)\log^{+}(\sigma_X^2/D)\prod_{i=1}^{L}(\mu_i^2\sigma_{N_i}^2+1)$ is equivalent to minimizing $\prod_{i=1}^{L}(\mu_i^2\sigma_{N_i}^2+1)$. Define

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

Without loss of generality, assume $\sigma_{N_1}\le\cdots\le\sigma_{N_L}$. Find the largest $\tilde{L}$ such that

$$\frac{\tilde{L}}{\sigma_{N_{\tilde{i}}}^2}-\left(\frac{1}{D_0(\tilde{L})}-\frac{1}{D}\right)\ge 0$$

where $D_0(\tilde{L})=(1/\sigma_X^2+1/\sigma_{N_1}^2+\cdots+1/\sigma_{N_{\tilde{L}}}^2)^{-1}$. Then, we get

$$\hat{\mu}_i^2=\frac{\frac{\tilde{L}}{\sigma_{N_i}^2}-\left(\frac{1}{D_0(\tilde{L})}-\frac{1}{D}\right)}{\sigma_{N_i}^2\left(\frac{1}{D_0(\tilde{L})}-\frac{1}{D}\right)},\quad i=1,\ldots,\tilde{L}$$
$$\hat{\mu}_i^2=0,\quad i=\tilde{L}+1,\ldots,L \tag{8}$$

and

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

The optimal rate allocation region $\vartheta(D)$ in $\mathcal{R}_{\text{in}}(D)$ is the convex hull of the $\tilde{L}!$ vertices $\{(R_1(\pi),\ldots,R_L(\pi))\}_{\pi}$, where π is a permutation on the set $\mathcal{I}_{\tilde{L}}$. The coordinates of vertices $(R_1(\pi),\ldots,R_L(\pi))$ are determined, as shown in the equation at the bottom of the page.

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.

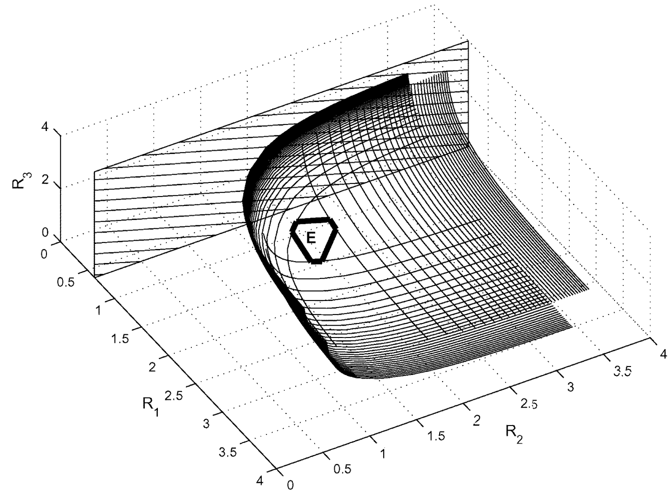


Fig. 2. Achievable rate region for the quadratic Gaussian CEO problem.

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

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

and

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

Actually, equation (10) is exactly the sum-rate distortion function, not just an upper bound. The direct coding theorem is

$$R_{\pi_1}(\pi)=I\left(Y_{\{\pi(1)\}};W_{\{\pi(1)\}}\,|\,W_{\{\pi(1)\}^c}\right)|_{\mu=\hat{\mu}}$$
$$=\frac{1}{2}\log^{+}\left\{\frac{\frac{\frac{\tilde{L}}{\sigma_{N_{\pi(1)}}^2}}{\frac{1}{D_0(\tilde{L})}-\frac{1}{D}}}{1-D\left[\frac{1}{\sigma_{N_{\pi(1)}}^2}-\frac{1}{\tilde{L}}\left(\frac{1}{D_0(\tilde{L})}-\frac{1}{D}\right)\right]}\right\}$$
$$R_{\pi(i)}(\pi)=I\left(Y_{\{\pi(1),\ldots,\pi(i)\}};W_{\{\pi(1),\ldots,\pi(i)\}}\,|\,W_{\{\pi(1),\ldots,\pi(i)\}^c}\right)|_{\mu=\hat{\mu}}-I\left(Y_{\{\pi(1),\ldots,\pi(i-1)\}};W_{\{\pi(1),\ldots,\pi(i-1)\}}\,|\,W_{\{\pi(1),\ldots,\pi(i-1)\}^c}\right)|_{\mu=\hat{\mu}}$$
$$=\frac{1}{2}\log^{+}\left\{\frac{\frac{\frac{\tilde{L}}{\sigma_{N_{\pi(i)}}^2}}{\frac{1}{D_0(\tilde{L})}-\frac{1}{D}}\left(1-D\sum_{j=1}^{i-1}\left[\frac{1}{\sigma_{N_{\pi(j)}}^2}-\frac{1}{\tilde{L}}\left(\frac{1}{D_0(\tilde{L})}-\frac{1}{D}\right)\right]\right)}{1-D\sum_{j=1}^{i}\left[\frac{1}{\sigma_{N_{\pi(j)}}^2}-\frac{1}{\tilde{L}}\left(\frac{1}{D_0(\tilde{L})}-\frac{1}{D}\right)\right]}\right\},i=2,\ldots,\tilde{L}$$
$$R_i(\pi)=0,\quad i=\tilde{L}+1,\ldots,L$$

provided by Theorem 1. For the converse we apply Oohama's bounding technique [5, Sec. 3] with the single difference that we do not let $L \to \infty$ at the end of this proof.

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

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

which coincides with the result in [5]. Here

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

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

$$R_{\pi(1)} = \frac{1}{2}\log^+\left\{\frac{\frac{\frac{L}{\sigma_N^2}}{\frac{1}{D_0}-\frac{1}{D}}}{1 - D\left[\frac{1}{\sigma_N^2} - \frac{1}{L}\left(\frac{1}{D_0} - \frac{1}{D}\right)\right]}\right\}$$

$$R_{\pi(i)} = \frac{1}{2}\log^+ \times \left\{\frac{\frac{\frac{L}{\sigma_N^2}}{\frac{1}{D_0}-\frac{1}{D}}\left(1 - (i-1)D\left[\frac{1}{\sigma_N^2} - \frac{1}{L}\left(\frac{1}{D_0} - \frac{1}{D}\right)\right]\right)}{1 - iD\left[\frac{1}{\sigma_N^2} - \frac{1}{L}\left(\frac{1}{D_0} - \frac{1}{D}\right)\right]}\right\}$$

$$i = 2, \dots, L.$$

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 water-filling 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. It is very clear that the 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 multiple-input–multiple-output (MIMO) systems has already been established by Zheng and Tse in [23].

Appendix I
Proof of Theorem 2

Proof:

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

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

1) Since

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

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 tth 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 $(Z_1(t), \dots, Z_L(t)) \in \mathcal{Z}_{\text{out}}(D_t)$.

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

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

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

Let γ be a random variable such that

$$P\{\gamma = t\} = \lambda_t$$

with γ independent of $(X(t), Y_1(t),\ldots,Y_L(t), Z_1(t),\ldots, Z_L(t)), t = 1,\ldots,n$. Let $X = X(t), Y_i = Y_i(t), Z_i = (Z_i(t), t)$ if $\gamma = t$, for $i = 1,\ldots,L$. 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 is easy to check the following.

1) $Z_i \to Y_i \to (X, Y_{\{i\}^c})$ for all $i \in \mathcal{I}_L$ since

$$I\left(Z_i; XY_{\{i\}^c} \,|\, Y_i\right) = \sum_{t=1}^{n} \lambda_i I\left(Z_i(t); X(t)Y_{\{i\}^c}(t) \,|\, Y_i(t)\right) = 0.$$

2) There exists function $g: \mathcal{Z}_1 \times \cdots \times \mathcal{Z}_L \to \mathcal{X}$ such that $Ed(X,\hat{X}) \le D$ where $\hat{X} = g(Z_1,\ldots,Z_L)$. This is because the decoder can use the function g_t corresponding to $(Z_1(t),\ldots,Z_L(t)), t = 1,\ldots,n$.

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

We can check the following.

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

Now, Lemma 7 follows by setting $\lambda_i = (1/n)$. ■

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 $\{P^n_{XY_1\ldots Y_L W_1\ldots W_L}\}_{n=1}^{\infty}$ in $\tau(D)$. Let $P^*_{XY_1\ldots Y_L W_1\ldots W_L}$ be the limiting distribution. It is obvious that $P^*_{XY_1\ldots Y_L W_1\ldots W_L}$ satisfies Property (i).

Without loss of generality, suppose $P_{Y_1}(y_1) > 0\ \forall y_1 \in \mathcal{Y}_1$. Note that $P^n_{XY_1\ldots Y_L W_1\ldots W_L} \to P^*_{XY_1\ldots Y_L W_1\ldots W_L}$ implies

$$P^n_{XY_2\ldots Y_L W_1\ldots W_L \,|\, Y_1} \to P^*_{XY_2\ldots Y_L W_1\ldots W_L \,|\, Y_1},$$
$$P^n_{XY_2\ldots Y_L W_2\ldots W_L \,|\, Y_1} \to P^*_{XY_2\ldots Y_L W_2\ldots W_L \,|\, Y_1},$$

and

$$P^n_{W_1 \,|\, Y_1} \to P^*_{W_1 \,|\, Y_1}.$$

Since

$$\begin{aligned} P^n_{XY_2\cdots Y_L W_1\ldots W_L \,|\, Y_1}&(x,y_2,\ldots,y_L,w_1,\ldots,w_L \,|\, y_1) \\ &= P^n_{XY_2\ldots Y_L W_2\ldots W_L \,|\, Y_1}(x,y_2,\ldots,y_L,w_2,\ldots,w_L \,|\, y_1) \\ &\quad \times P_{W_1 \,|\, Y_1}(w_1 \,|\, y_1) \end{aligned}$$

we have

$$\begin{aligned} P^*_{XY_2\ldots Y_L W_1\ldots W_L \,|\, Y_1}&(x,y_2,\ldots,y_L,w_1,\ldots,w_L \,|\, y_1) \\ &= \lim_{n\to\infty} P^n_{XY_2\cdots Y_L W_1\cdots W_L \,|\, Y_1}(x,y_2,\ldots,y_L,w_1,\ldots,w_L \,|\, y_1) \\ &= \lim_{n\to\infty} P^n_{XY_2\cdots Y_L W_2\ldots W_L \,|\, Y_1}(x,y_2,\ldots,y_L,w_2,\ldots, \\ &\quad w_L \,|\, y_1) \\ P^n_{W_1 \,|\, Y_1}&(w_1 \,|\, y_1) \\ &= \lim_{n\to\infty} P^n_{XY_2\cdots Y_L W_2\ldots W_L \,|\, Y_1}(x,y_2,\ldots,y_L,w_2,\ldots,w_L \,|\, y_1) \\ &\quad \lim_{n\to\infty} P^n_{W_1 \,|\, Y_1}(w_1 \,|\, y_1) \\ &= P^*_{XY_2\ldots Y_L W_2\ldots W_L \,|\, Y_1}(x,y_2,\ldots,y_L,w_2,\ldots,w_L \,|\, y_1) \\ &\quad P^*_{W_1 \,|\, Y_1}(w_1 \,|\, y_1), \quad \forall x,y_1,\ldots,y_L,w_1,\ldots,w_L. \end{aligned}$$

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

Let f_n be the function associated with $P^n_{XY_1\ldots Y_L W_1\ldots W_L}$ such that $E_{P^n_{XY_1\ldots Y_L W_1\ldots W_L}} d(X, f_n(W_1,\ldots,W_L)) \le D$. Since $\max(|\mathcal{X}|, |\mathcal{W}_1|,\ldots,|\mathcal{W}_L|) < \infty$, there are only finite number of functions from $\mathcal{W}_1 \times \cdots \times \mathcal{W}_L$ 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

$$\begin{aligned} E_{P^*_{XY_1\ldots Y_L W_1\ldots W_L}} & d(X, f^*(W_1,\ldots,W_L)) \\ &= \lim_{k\to\infty} E_{P^{n_k}_{XY_1\ldots Y_L W_1\ldots W_L}} d(X, f^*(W_1,\ldots,W_L)) \\ &\le D. \end{aligned}$$

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

APPENDIX III
PROOF OF LEMMA 6

Proof: First, we consider the following special case. If matrix $R(c_1,c_2,\ldots,c_L)$ of the form

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

we get

$$\begin{aligned} r(c_1,c_2,\ldots,c_L) &= \det R(c_1,c_2,\ldots,c_L) \\ &= \det \begin{pmatrix} 1+c_1 & 1 & \cdots & 1 \\ -c_1 & c_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -c_1 & 0 & \cdots & c_L \end{pmatrix} \\ &= (1+c_1)c_2\ldots c_L + c_1c_3\ldots c_L + \cdots + c_1c_2\ldots c_{L-1} \\ &= c_1\ldots c_L\left(1 + \frac{1}{c_1} + \cdots + \frac{1}{c_L}\right). \end{aligned}$$

Since

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

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

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

Back to $R(\mathbf{l}; \mathbf{c})$. Since $R(\mathbf{l}; \mathbf{c}) = \mathbf{1}\,\mathbf{1}^T + \mathrm{diag}(c_1, \dots, c_L)$, $\det R(\mathbf{l}; \mathbf{c}) = l_1^2 \dots l_L^2 \det R(c_1/l_1^2, \dots, c_L/l_L^2)$. 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

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

■

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," *Probl. Inform. Transmission*, vol. 15, no. 2, pp. 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] Y. Oohama, "The rate-distortion function for the quadratic Gaussian CEO problem," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1057–1070, May 1998.

[6] P. Viswanath, Sum rate of multiterminal gaussian source coding, in *IEEE Trans. Inform. Theory*, ser. DIMACS series in discrete mathematics and theoretical computer science, submitted for publication.

[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 Elec. Eng., 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, Austria: 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, 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. Inform. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[12] S. C. Draper and G. W. Wornell, "Successively structured CEO problems," in *Proc. ISIT 2002*, Lausanne, Switzerland, June 30–July 5 2002, p. 65.

[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. Inform. Theory*, vol. IT-19, pp. 471–480, July 1973.

[15] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 106–117, Jan. 1999.

[16] D. N. C. Tse and S. V. Hanly, "Multiaccess fading channels—Part I: Polymatroid structure, optimal resource allocation, and throughput capacities," *IEEE Trans. Inform. Theory*, vol. 44, 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, July 1978.

[19] Y. Oohama, "Gaussian multiterminal source coding," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1912–1923, Nov. 1997.

[20] ——, "Multiterminal source coding for correlated memoryless Gaussian sources with several side informations at the decoder," in *Proc. IEEE Information Theory and Communications Workshop*, Kruger National Park, South Africa, 1999, p. 100.

[21] P. Ishwar, R. Puri, S. S. Pradhan, and K. Ramchandran, "On compression for robust estimation in sensor networks," in *Proc. ISIT 2003*, Yokohama, Japan, June 29–July 4 2003, p. 193.

[22] J. Chen and T. Berger, "Robust coding schemes for distributed sensor networks with unreliable sensors", submitted for publication.

[23] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

Jun Chen (S'03) 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, Ithaca, NY, in 2004. He is working towards the Ph.D. degree at Cornell University.

His research interests are in information theory, communication, and networking.

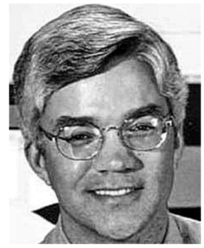
Xin Zhang (S'04) received the B.S. degree in electronic engineering from Tsinghua University, Beijing, China, in 2001. She is working toward the M.S. and Ph.D. degrees in electrical engineering at Cornell University, Ithaca, NY.

Her research interests include wireless communication and sensor network.

Toby Berger (S'60–M'66–SM'74–F'78) was born in New York 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* (Englewood Cliffs, NJ: Prentice-Hall, 1971) and a coauthor of *Digital Compression for Multimedia: Principles and Standards* (San Francisco, CA: Morgan Kaufmann, 1998) and *Information Measures for Discrete Random Fields* (Beijing, New York: Science Press, 1998).

Dr. Berger 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. He is a Life Member of Tau Beta Pi. He has served as the Editor-in-Chief of the IEEE TRANSACTIONS ON INFORMATION THEORY and as President of the IEEE Information Theory Group.



Stephen B. Wicker (S'83–M'83–SM'97) received the B.S.E.E. degree (High Honors) from the University of Virginia, Charlottesville, in 1982, the M.S.E.E. degree from Purdue University, West Lafayette, IN, in 1983, and the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, in 1987.

He is a Professor of Electrical and Computer Engineering at Cornell University. He teaches and conducts research in wireless information networks, digital systems, self-configuring networks, and game theory. 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. He is the author of *Codes, Graphs, and Iterative Decoding* (Norwell, MA: Kluwer, 2002), *Turbo Coding* (Norwell, MA: Kluwer, 1999), *Error Control Systems for Digital Communication and Storage* (Englewood Cliffs, NJ: Prentice-Hall, 1995), and *Reed–Solomon Codes and Their Applications* (Piscataway, NJ: IEEE Press, 1994). His current research focuses on the use of game theory in the development of highly distributed, adaptive networks.

Professor Wicker 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 an 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.